



HAL
open science

Du diagnostic au pronostic de pannes des entraînements électriques

Abdenour Soualhi

► **To cite this version:**

Abdenour Soualhi. Du diagnostic au pronostic de pannes des entraînements électriques. Energie électrique. Université Claude Bernard - Lyon I, 2013. Français. NNT : 2013LYO10146 . tel-01258196

HAL Id: tel-01258196

<https://theses.hal.science/tel-01258196>

Submitted on 18 Jan 2016

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.

THÈSE DE L'UNIVERSITÉ DE LYON

Délivrée par

L'UNIVERSITÉ CLAUDE BERNARD LYON 1

École Doctorale EEA

Électronique, Électrotechnique, Automatique

DIPLÔME DE DOCTORAT

Spécialité: Génie Électrique

(arrêté du 7 août 2006)

soutenue publiquement le 24 septembre 2013

Par

Abdenour SOUALHI

Du diagnostic au pronostic de pannes des entraînements électriques

Thèse dirigée par M. Hubert RAZIK et M. Guy CLERC
Laboratoire de recherche : AMPÈRE – UMR 5005

JURY

Farid Meibody-Tabar	Professeur des Universités (GREEN-Nancy)	Président
M.E.H. BENBOUZID	Professeur des Universités (LBMS-Brest)	Rapporteur
Pascal MAUSSION	Professeur des Universités (LAPLACE-Toulouse)	Rapporteur
Hubert RAZIK	Professeur des Universités (AMPERE-Lyon)	Directeur
Guy CLERC	Professeur des Universités (AMPERE-Lyon)	Encadrant
Francklin RIVAS	Doctor, Universidad de Los Andes (LabSIULA-Venezuela)	Examineur
LE Pham-Hiep	SAGEM défense sécurité	Examineur

Remerciements

Les travaux que je présente dans ce manuscrit n'auraient pas été possibles sans l'aide de nombreuses personnes dont :

Les membres du jury qui ont eu la volonté d'assister et d'intervenir lors de la soutenance.

Je remercie donc avec sincérité le Professeur Farid Meibody-Tabar pour avoir présidé ce jury, ainsi que les rapporteurs Professeur Mohamed El Hachemi BENBOUZID et Professeur Pascal MAUSSION pour leurs efforts d'attention et de réflexion au cours de leur lecture.

Je remercie à son tour Mr Francklin RIVAS et Mr LE Pham-Hiep, qui ont accepté d'être membres examinateurs de ce travail de thèse.

L'ensemble du laboratoire AMPÈRE, où j'ai effectué l'intégralité de ma thèse, mérite de profonds remerciements. Merci tout d'abord à mon encadrant le professeur Guy CLERC pour ces conseils, son aide et son expérience qu'il a bien voulu me faire partager.

Un grand merci également à mon directeur de thèse le professeur Hubert RAZIK, qui m'a accueilli en thèse au laboratoire AMPÈRE, pour nos longues discussions, les déjeuners et les bons moments que nous avons passé ensemble.

J'adresse aussi à toutes les valeureuses personnes présentes au sein du laboratoire qui ont croisé ma route, de sincères remerciements. J'ai apprécié, et espère avoir contribué, à l'excellente cohésion présente entre les équipes de recherche. J'aimerais remercier mes collègues Oualid MESSAL, Menad BOURKEB, Ronan GERMAN et Maawad MAKDESSI qui ont rendu ces trois années très agréables.

D'un point de vue plus personnel, merci à ma famille qui m'a toujours encouragé et sans qui je ne serais pas arrivé à ce stade.

Je voudrais terminer ces remerciements par la personne qui compte le plus pour moi, qui m'a supporté et soutenu dans les moments difficiles durant toute la durée de ma thèse, je veux bien évidemment parler de l'amour de ma vie, Wafia. Merci Wafia pour ton sourire, ton aide et ton soutien : tout le monde n'a pas la chance de tenir un soleil dans ses bras.

Résumé

Le diagnostic et le pronostic de pannes des systèmes d'entraînement électriques est un enjeu majeur pour assurer la fiabilité et la sûreté de fonctionnement des outils de production notamment dans les secteurs sensibles (militaire, l'aéronautique, l'aérospatiale et nucléaire, etc.). Le travail de recherche présenté dans cette thèse vise à introduire de nouvelles méthodes de diagnostic et de pronostic des défauts d'une machine asynchrone ainsi que des roulements à rouleaux. Ces méthodes, orientées données, utilisent les données de mesure recueillies à partir de capteurs placés sur un système (machine asynchrone, roulement à rouleaux) afin de construire un vecteur de paramètres indicateur de défaut. Les méthodes de classification développées (supervisée, non supervisée) permettent de classer les observations, décrites par le vecteur de paramètres, par rapport aux différents modes de fonctionnement, connus ou inconnus du système, avec ou sans défauts. Des défauts ont été créés au rotor et aux roulements de la machine asynchrone, alimentée par le biais d'un onduleur de tension. La classification non supervisée, basée sur l'algorithme des fourmis artificielles, permet d'analyser les observations inconnues et inexplorées afin de mettre en évidence les classes regroupant les observations similaires. Cela permet d'améliorer la classification et de détecter l'apparition de nouveaux modes de fonctionnement. La classification supervisée, basée sur les modèles de Markov cachés, permet d'associer un degré d'appartenance (sous forme d'une probabilité) lors de l'affectation d'une observation à une ou plusieurs classes. Cela permet de définir un indice de fiabilité à l'affectation réalisée mais aussi de détecter l'apparition de nouveaux modes de fonctionnement. Ces méthodes ne se limitent pas qu'à diagnostiquer les défauts, elles peuvent aussi contribuer au pronostic des défauts. En effet, le pronostic peut être défini comme une extension du problème de diagnostic. La prédiction d'un défaut est réalisée par trois méthodes basées sur les modèles de Markov cachés pour la détection de l'imminence d'un défaut ainsi que par deux méthodes basées sur le système neuro-flou (ANFIS pour Adaptive Neuro Fuzzy Inference System et le neurone neo-flou) pour estimer le temps restant avant son apparition. Des données de vieillissement d'un ensemble de roulements à rouleaux ont été utilisées afin de tester les méthodes proposées. Les résultats obtenus montrent l'efficacité de ces méthodes pour le diagnostic et le pronostic des défauts dans les entraînements électriques.

Abstract

Faults diagnosis and prognosis of electrical drives play a key role in the reliability and safety of production tools especially in key sectors (military, aviation, aerospace and nuclear, etc.). The research presented in this thesis aims to introduce new methods for faults diagnosis and prognosis of an induction motors and roller bearings. These methods use measured data collected from sensors placed on the system (induction motor, roller) in order to construct a feature vector which indicates the state of the system. Supervised and unsupervised classification methods are developed to classify measurements (observations) described by the feature vector compared to known or unknown operating modes, with or without failures. Defects were created in the rotor and the bearing of the induction motor, fed by a voltage inverter. The unsupervised classification technique, based on artificial ant-clustering, allows analyzing the unknown and unexplored observations to highlight classes with similar observations. This allows improving the classification and the detection of new operating modes. The supervised classification, based on hidden Markov models, allows associating a degree of similarity when we affect an observation to one or more classes. This defines a reliability index which allows the detection of new operating modes. These methods are not limited to diagnose faults; they can also contribute to the prognosis of faults. Indeed, the prognosis can be defined as an extension of the problem of diagnosis. The prognosis of faults is carried out by three methods based on hidden Markov models for the detection of an impending failure and by two methods based on the neuro-fuzzy system (ANFIS for Adaptive Neuro fuzzy Inference System and the neo-fuzzy neuron) to estimate the remaining time before its appearance. A set of historical data collected on roller bearings is used to validate the proposed methods. The obtained results show the effectiveness of the proposed methods for faults diagnosis and prognosis of electrical drives.

SOMMAIRE

Introduction générale	1
------------------------------	----------

Chapitre 1 : Le diagnostic et le pronostic industriel

1.1. Introduction	5
1.2. Stratégies de maintenance	5
1.3. Les principaux aspects de la mise en œuvre d'une maintenance préventive conditionnelle	7
1.4. Traitement du signal	10
1.4.1. Analyse temporelle	10
1.4.2. Analyse fréquentielle	12
1.4.3. Analyse temps- fréquence	12
1.5. Le diagnostic	12
1.5.1. Diagnostic basé sur un modèle physique	13
1.5.1.1. Observateurs	13
1.5.1.2. Espace de parité	14
1.5.1.3. Estimation paramétrique	14
1.5.2. Diagnostic orienté données	15
1.5.2.1. Système expert	15
1.5.2.2. Méthodes statistiques multivariées	16
1.5.2.3. Réseau de neurones	16
1.5.2.4. Classification	16
1.6. Le pronostic	19
1.6.1. Pronostic basé sur un modèle physique	20
1.6.2. Pronostic orienté données	21
1.6.3. Pronostic basé sur l'expérience	22
1.7. Conclusion	23

Chapitre 2 : Diagnostic par classification supervisée et non supervisée

2.1. Introduction	25
2.2. Détermination de l'espace de représentation	27
2.2.1. Critère de sélection (J)	29
2.3. Classification	31
2.3.1. La classification non supervisée	31
2.3.1.1. Mesure de similarité	31

2.3.1.2.	Classification hiérarchique	33
2.3.1.2.1.	Critère de liaison	34
2.3.1.2.1.1.	Critère de liaison par saut minimal	34
2.3.1.2.1.2.	Critère de liaison par saut maximal	34
2.3.1.2.1.3.	Critère de liaison par distance moyenne	35
2.3.1.2.1.4.	Critère de Ward	35
2.3.1.3.	Classification par partition	35
2.3.1.4.	Approche heuristique pour l'optimisation de la classification non supervisée	37
2.3.1.4.1.	Comportement des fourmis	37
2.3.1.4.2.	Un modèle mathématique sur le comportement des fourmis	38
2.3.1.4.3.	L'algorithme Fourmis Artificielles (FA)	40
2.3.2.	La classification supervisée	50
2.3.2.1.	Classification par modèles de Markov cachés	52
2.3.2.2.	Proposition	52
2.3.2.3.	Modèle de Markov caché	53
2.3.2.3.1.	Paramètres d'un Modèle de Markov caché	53
2.3.2.3.2.	Apprentissage d'une classe par le modèle de Markov caché	55
2.3.2.4.	Classification d'une observation par un modèle de Markov caché	58
2.4.	Evaluation des méthodes proposées	60
2.4.1.	Algorithme fourmis artificielles	61
2.4.2.	Modèles de Markov cachés	65
2.4.3.	Comparaison des méthodes de classification	68
2.5.	Conclusion	69

Chapitre 3 : Du diagnostic au pronostic

3.1.	Introduction	72
3.2.	Prédiction de l'imminence d'un défaut par les modèles de Markov cachés	72
3.2.1.	Méthode basée sur l'identification du processus de dégradation du système	73
3.2.2.	Méthodes basées sur l'analyse des séquences d'observations	80
3.3.	Estimation du temps de vie restant	86
3.3.1.	Prédiction de séries temporelles par ANFIS	89
3.3.2.	Prédiction de séries temporelles par le neurone neo-flou (NNF)	92
3.3.2.1.	Apprentissage d'un NNF (neurone neo-fou)	93
3.3.3.	Mesure de la performance des modèles de prédiction	94
3.3.4.	ANFIS et NNF pour la prédiction de séries temporelles chaotiques	95
3.4.	Conclusion	98

Chapitre 4 : Application au diagnostic de la machine asynchrone et au pronostic des roulements

4.1.	Introduction	100
4.2.	Présentation de la machine asynchrones à cage d'écureuil	100
4.3.	Diagnostic des défauts d'une machine asynchrone	103
4.3.1.	Présentation du banc de test diagnostic	103
4.3.2.	Modes de fonctionnement étudiés	104
4.3.3.	Construction du vecteur de paramètre par analyse des courants statoriques	106

4.3.3.1.	Analyse fréquentielle	106
4.3.3.1.1.	Prétraitement des courants statoriques	106
4.3.3.1.2.	Ruptures de barres	106
4.3.3.1.3.	Excentricité	107
4.3.3.1.4.	Défauts de roulements	108
4.3.3.1.5.	Défauts au niveau du stator	108
4.3.3.1.6.	Autres paramètres	109
4.3.3.2.	Analyse temporelle	110
4.3.3.2.1.	Prétraitement des courants statoriques	110
4.3.3.2.2.	Ruptures de barres	110
4.3.3.2.3.	Défauts au niveau du stator	112
4.3.3.2.4.	Excentricité	113
4.3.4.	Diagnostic d'une machine alimentée par onduleur	116
4.3.4.1.	Diagnostic avec apparition d'un nouveau mode de fonctionnement	116
4.3.4.1.1.	Détermination de l'espace de représentation (sélection des q-paramètres)	117
4.3.4.1.2.	Validation des méthodes de classification	119
4.3.4.1.2.1.	Algorithme fourmis artificielles (FA)	119
4.3.4.1.2.2.	Classification par les modèles de Markov cachés	123
4.3.4.2.	Diagnostic avec une connaissance limitée sur les modes de fonctionnement	126
4.3.4.2.1.	Détermination de l'espace de représentation (sélection des q-paramètres)	128
4.3.4.2.2.	Validation des méthodes de classification	130
4.3.4.2.2.1.	Algorithme fourmis artificielles	131
4.3.4.2.2.2.	Classification par les modèles de Markov cachés	134
4.4.	Pronostic des roulements	137
4.4.1.	Présentation du banc de test pronostic	137
4.4.2.	Validation des méthodes de pronostic	139
4.4.2.1.	Construction du vecteur de paramètres	139
4.4.2.2.	Identification des différents stades de dégradation	143
4.4.2.3.	Imminence des stades de dégradation du roulement 4	145
4.4.2.3.1.	Méthode (2) : modèle de Markov évolutif	146
4.4.2.3.2.	Méthode (3) : Modèles de Markov cachés	149
4.4.2.4.	Prédiction des défauts du roulement 4	153
4.4.2.5.	Estimation du temps restant avant le prochain stade de dégradation du roulement 4	154
4.5.	Conclusion	165
	Conclusion générale	167
	Bibliographie	170

Liste des figures

Figure 1. 1. Les conséquences économiques d'une journée d'arrêt dans les secteurs industriels [Helle_06]	5
Figure 1. 2. Classification de la maintenance _____	6
Figure 1. 3. Architecture d'une maintenance préventive conditionnelle avec sept modules [Lebold_03]	8
Figure 1. 4. Acquisition de données en fonction des disciplines _____	8
Figure 1. 5. Architecture générale d'un système de pronostic tel que présenté par [Vacht_01]	9
Figure 1. 6. Les éléments de base du diagnostic et pronostic pour la MPC tel que présenté par [Chinnam_04]	10
Figure 1. 7. Techniques de traitement du signal _____	10
Figure 1. 8. Structure d'un système de diagnostic à base d'un modèle _____	13
Figure 1. 9. Générateur de résidus à base d'un observateur _____	14
Figure 1. 10. Générateur de résidus dans l'espace de parité _____	14
Figure 1. 11. Description schématique d'une estimation paramétrique _____	15
Figure 1. 12. Structure d'un système de diagnostic orienté données _____	15
Figure 1. 13. Architecture d'un système expert _____	16
Figure 1. 14. Association d'une observation x_i à l'une des classes $(\Omega_1, \Omega_2, \dots, \Omega_M)$ _____	17
Figure 1. 15. Pronostic Vs diagnostic _____	19
Figure 1. 16. Classification des méthodes de pronostic [Vacht_06]	20
Figure 1. 17. Structure d'un système de pronostic à base d'un modèle _____	21
Figure 1. 18. Structure d'un système de pronostic orienté données _____	21
Figure 2. 1. Représentation d'une observation dans un espace à q-dimensions _____	25
Figure 2. 2. Représentation d'un ensemble de classes dans un espace à q-dimensions _____	26
Figure 2. 3. Eléments d'un système de reconnaissance des formes _____	26
Figure 2. 4. Structure d'une classification supervisée et non-supervisée _____	27
Figure 2. 5. Contours des métriques : Euclidienne, Manhattan et Mahalanobis dans un espace 2D _____	32
Figure 2. 6. Frontières déclinées par les distances de Mahalanobis et Euclidienne _____	32
Figure 2. 7. Exemple d'un dendrogramme _____	34
Figure 2. 8. Effet de chaîne _____	34
Figure 2. 9. Algorithme K-means pour le partitionnement de deux classes _____	36
Figure 2. 10. Comportement d'une colonie de fourmis pour trouver le chemin le plus court vers une source de nourriture _____	38
Figure 2. 11. Détermination du chemin le plus court entre N-villes par les fourmis _____	39
Figure 2. 12. Colonie de fourmis à la recherche de nourriture [Nid_11] _____	40
Figure 2. 13. Comportement des fourmis lors de la phase d'exploration _____	41
Figure 2. 14. Exemple de Règle de comportement d'une fourmi Ant_j lorsqu'elle se situe dans le nid _____	44
Figure 2. 15. Exemple de comportement de deux groupes de fourmis $Ants_i, Ants_j$ dans le cas où elles empruntent deux chemins similaires _____	46
Figure 2. 16. Représentation d'une rotation orthogonale _____	47
Figure 2. 17. Exemple de comportement de deux groupes de fourmis $Ants_i, Ants_j$ dans le cas où elles empruntent deux chemins différents _____	49
Figure 2. 18. Représentation des modes de fonctionnement d'un système sous forme de classes _____	52
Figure 2. 19. Extraction des séquences d'observations _____	55
Figure 2. 20. Organigramme d'une classification supervisée par MMCs _____	59

Figure 2. 21. Répartition des observations d'iris _____	60
Figure 2. 22. Evolution du critère J en fonction du seuil de similarité _____	61
Figure 2. 23. Détermination des paramètres e et f à partir du critère J _____	62
Figure 2. 24. Répartition des observations d'iris avant l'ACP _____	63
Figure 2. 25. Répartition des observations d'iris après une ACP _____	63
Figure 2. 26. Processus d'apprentissage des classes par ré-estimation de Baum-Welsh _____	66
Figure 2. 27. Classification des observations d'iris par les modèles de Markov cachés _____	67
Figure 3. 1. Visualisation des différents stades de dégradation d'un système par des classes _____	72
Figure 3. 2. Organigramme de la première méthode de pronostic basée sur l'indentification du processus de dégradation _____	75
Figure 3. 3. Évolution de la dégradation en fonction du temps (mois) _____	76
Figure 3. 4. Représentation des stades de dégradation des moteurs _____	76
Figure 3. 5. Évolution de la dégradation (q_t) du moteur n°1 _____	77
Figure 3. 6. Évolution de la dégradation (q_t) du moteur n°11 _____	80
Figure 3. 7. Construction d'un modèle de Markov évolutif _____	81
Figure 3. 8. Organigramme de la deuxième méthode de pronostic basée sur un modèle de Markov évolutif _____	82
Figure 3. 9. Organigramme de la troisième méthode de pronostic _____	86
Figure 3. 10. Détermination du temps de vie restant par le suivi de la tendance d'un paramètre de surveillance _____	87
Figure 3. 11. Prédiction à un pas dans le futur _____	88
Figure 3. 12. Prédiction à plusieurs pas dans le futur (multi-pas) _____	88
Figure 3. 13. Vers les systèmes neuro-flous [Gouriveau_07] _____	88
Figure 3. 14. Architecture du prédicteur ANFIS avec quatre entrées et deux fonctions d'appartenance associées à chaque entrée _____	90
Figure 3. 15. Architecture du prédicteur NNF _____	92
Figure 3. 16. Fonctions d'appartenance triangulaires _____	93
Figure 3. 17. Série temporelle de Mackey-Glass _____	95
Figure 3. 18. Fonctions d'appartenance initiales _____	96
Figure 3. 19. Fonctions d'appartenance après la phase d'apprentissage _____	96
Figure 3. 20. Résultats de prédiction pour le NNF _____	97
Figure 3. 21. Résultats de prédiction pour ANFIS _____	98
Figure 4. 1. Constitution générale d'une machine asynchrone à cage d'écuréuil _____	101
Figure 4. 2. Les pourcentages de défauts communs dans les machines asynchrones [Albrecht_86,Repo_85]. _____	101
Figure 4. 3. Diagramme des défaillances sur la machine asynchrone à cage d'écuréuil _____	102
Figure 4. 4. Banc de test diagnostique sur une machine asynchrone de 5.5 Kw _____	103
Figure 4. 5. Acquisition des signaux _____	104
Figure 4. 6. Schéma de principe du processus de diagnostique _____	105
Figure 4. 7. Analyse fréquentielle pour une machine avec 1 barre cassée à pleine charge. FFT sur $Re[I_d]$ _____	107
Figure 4. 8. Variation de l'amplitude correspondant aux fréquences recherchées. Fonctionnement à vide _____	110
Figure 4. 9. Variation de l'amplitude correspondant aux fréquences recherchées. Fonctionnement en charge _____	110
Figure 4. 10. Caractéristique $I_\beta (I_\alpha)$ normalisée par rapport à $I_{\alpha,\beta \text{ eff}}$. Fonctionnement en charge _____	111
Figure 4. 11. Evolution des composantes I_α et I_β des courants stators pour un rotor sain (a) et pour un rotor ayant 4 barres cassées (b) _____	112

Figure 4. 12. Caractéristique normalisée $Q'(P')$ obtenue pour la machine saine, pour une charge de 3 kW en alimentation réseau _____	113
Figure 4. 13. Barre cassée au niveau du rotor de la machine asynchrone à cage d'écureuil _____	116
Figure 4. 14. Variation du critère J en fonction de chaque paramètre _____	117
Figure 4. 15. Sélection de l'ensemble q-paramètres par l'algorithme OSBS _____	118
Figure 4. 16. Choix de la dimension (q) du vecteur de paramètres _____	118
Figure 4. 17. Position des classes de l'ensemble d'apprentissage dans le sous- espace de dimension 3. Les paramètres sont sélectionnés avec l'algorithme OSBS _____	119
Figure 4. 18. Evolution du critère J en fonction du seuil de similarité _____	120
Figure 4. 19. Détermination des paramètres e et f à partir du critère J _____	121
Figure 4. 20. Position des classes dans le sous – espace de dimension 3. Machine saine (à vide) «□», machine saine (en charge) «□», 1 barre cassée (à vide) «□», 1 barre cassée (en charge) «□» _____	121
Figure 4. 21. Détection de l'apparition d'une nouvelle classe _____	122
Figure 4. 22. Processus d'apprentissage des classes par ré-estimation de Baum-Welsh _____	124
Figure 4. 23. Classification de l'ensemble des observations test par modèles de Markov cachés _____	125
Figure 4. 24. Défauts générés au niveau du rotor et des roulements de la machine asynchrone à cage d'écureuil _____	127
Figure 4. 25. Variation du critère J en fonction de chaque paramètre _____	128
Figure 4. 26. Sélection des ensembles de q-paramètres par la méthode OSBS _____	129
Figure 4. 27. Choix de la dimension (q) du vecteur de paramètres _____	129
Figure 4. 28. Position des classes de l'ensemble d'apprentissage dans le sous espace de dimension 3. Les paramètres sont sélectionnés avec l'algorithme OSBS _____	130
Figure 4. 29. Evolution du critère J en fonction du seuil de similarité _____	131
Figure 4. 30. Détermination des paramètres e et f à partir du critère J _____	132
Figure 4. 31. Position des classes obtenues par l'algorithme FA dans le sous – espace de dimension 3 avant l'ACP _____	132
Figure 4. 32. Position des classes obtenues par l'algorithme FA dans le sous – espace de dimension 3 après l'ACP _____	133
Figure 4. 33. Détection de l'apparition d'une nouvelle classe _____	133
Figure 4. 34. Classification de l'ensemble des observations test par modèles de Markov cachés _____	135
Figure 4. 35. Banc de vieillissement des roulements à rouleaux [Qiu_06] _____	137
Figure 4. 36. Schéma de principe du processus de diagnostic et de pronostic _____	138
Figure 4. 37. Evolution de la dégradation des roulements (1, 2, 4 et 4) en fonction des paramètres statistiques _____	142
Figure 4. 38. Evolution du critère J en fonction du seuil de similarité _____	143
Figure 4. 39. Détermination des paramètres e et f à partir du critère J _____	144
Figure 4. 40. Position des classes obtenues par l'algorithme FA dans le sous – espace de dimension 3 après l'ACP _____	144
Figure 4. 41. Défaut de bague interne dans le roulement 3 _____	145
Figure 4. 42. Imminence du premier stade de dégradation du roulement 4 _____	146
Figure 4. 43. Probabilité qu'une séquence d'observations arrive à un fonctionnement (S_j) à $t = 7707$ minutes _____	146
Figure 4. 44. Imminence du deuxième stade de dégradation _____	147
Figure 4. 45. Probabilité qu'une séquence d'observations arrive à un fonctionnement (S_j) à $t = 42766$ minutes _____	147
Figure 4. 46. Imminence du troisième stade de dégradation _____	148
Figure 4. 47. Probabilité qu'une séquence d'observations arrive à un fonctionnement (S_j) à $t = 50110$ minutes _____	148
Figure 4. 48. Imminence du premier stade de dégradation _____	150
Figure 4. 49. Probabilité de générer une séquence d'observations étant donné les MMCs ($\bar{\lambda}_i$) à $t = 5553$ minutes _____	150

Figure 4. 50. Imminence du deuxième stade de dégradation _____	151
Figure 4. 51. Probabilité de générer une séquence d'observations étant donné les MMCs ($\bar{\lambda}_i$) à t = 42667 minutes _____	151
Figure 4. 52. Imminence du troisième stade de dégradation _____	152
Figure 4. 53. Probabilité de générer une séquence d'observations étant donné les MMCs ($\bar{\lambda}_i$) à t = 49910 minutes _____	152
Figure 4. 54. FFT du signal vibratoire du roulement 4 à t=1000, 42667, 49910 minutes _____	153
Figure 4. 55. Défaut de bague externe dans le roulement 4 _____	153
Figure 4. 56. Défaut de rouleau dans le roulement 4 _____	154
Figure 4. 57. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps _____	156
Figure 4. 58. Prédiction de la puissance du signal vibratoire par ANFIS et le neurone neo-flou _____	156
Figure 4. 59. Evolution de la RMSE de la valeur efficace du signal vibratoire en fonction du temps _____	157
Figure 4. 60. Prédiction de la valeur efficace du signal vibratoire par ANFIS et le neurone neo-flou _____	157
Figure 4. 61. Evolution de la RMSE de la valeur moyenne du signal vibratoire en fonction du temps _____	158
Figure 4. 62. Prédiction de la valeur moyenne du signal vibratoire par ANFIS et le neurone neo-flou _____	158
Figure 4. 63. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps _____	159
Figure 4. 64. Prédiction de la puissance du signal vibratoire par ANFIS et le neurone neo-flou _____	159
Figure 4. 65. Evolution de la RMSE de la valeur efficace du signal vibratoire en fonction du temps _____	160
Figure 4. 66. Prédiction de la valeur efficace du signal vibratoire par ANFIS et le neurone neo-flou _____	160
Figure 4. 67. Evolution de la RMSE de la valeur moyenne du signal vibratoire en fonction du temps _____	160
Figure 4. 68. Prédiction de la valeur moyenne du signal vibratoire par ANFIS et le neurone neo-flou _____	161
Figure 4. 69. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps _____	161
Figure 4. 70. Prédiction de la puissance du signal vibratoire par ANFIS et le neurone neo-flou _____	162
Figure 4. 71. Evolution de la RMSE de la valeur efficace du signal vibratoire en fonction du temps _____	162
Figure 4. 72. Prédiction de la valeur efficace du signal vibratoire par ANFIS et le neurone neo-flou _____	162
Figure 4. 73. Evolution de la RMSE de la valeur moyenne du signal vibratoire en fonction du temps _____	163
Figure 4. 74. Prédiction de la valeur moyenne du signal vibratoire par ANFIS et le neurone neo-flou _____	163
Figure 4. 75. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps _____	165

Liste des tableaux

Tableau 1. 1. Points forts et points faibles des stratégies de maintenance	7
Tableau 1. 2. Comparaison des différentes méthodes utilisées pour le diagnostic des défauts	18
Tableau 1. 3. Approches de pronostic et type d'information nécessaire	22
Tableau 2. 1. Résultats de classification	69
Tableau 3. 1. Matrice d'observation B de la séquence d'observations X^1	78
Tableau 3. 2. Matrice d'observation B de la séquence d'observations X^{11}	79
Tableau 3. 3. Variables Forward normalisées obtenues de la séquence d'observations X^{11} et cela jusqu'à l'instant $t=6$	79
Tableau 3. 4. Variables Forward normalisées obtenues de la séquence d'observations X^{11}	83
Tableau 3. 5. Avantage et inconvénient des méthodes de pronostic proposées	86
Tableau 3. 6. Erreur de prédiction pour les 500 échantillons test avec un pas de prédiction $p=6$	97
Tableau 4. 1. Caractéristiques de la machine asynchrone.....	103
Tableau 4. 2. Modes de fonctionnement servant à constituer l'ensemble d'apprentissage et l'ensemble test	104
Tableau 4. 3. Fréquences contenues dans le vecteur de paramètres et induites par l'alimentation onduleur	109
Tableau 4. 4. Paramètres calculés constituant le vecteur initial à q'-paramètres	115
Tableau 4. 5. Modes de fonctionnement étudiés	116
Tableau 4. 6. Composition de l'ensemble d'apprentissage	116
Tableau 4. 7. Composition de l'ensemble test.....	119
Tableau 4. 8. Modes de fonctionnement étudiés	127
Tableau 4. 9. Composition de l'ensemble d'apprentissage	127
Tableau 4. 10. Composition de l'ensemble test	130
Tableau 4. 11. Caractéristiques des roulements.....	138
Tableau 4. 12. Paramètres statistiques caractéristiques d'un signal vibratoire	139
Tableau 4. 13. Configuration d'ANFIS et du neurone neo-flou.....	154
Tableau 4. 14. RMSE de γ_3 à plusieurs horizons de prédiction (p).....	164
Tableau 4. 15. RMSE de γ_3 à plusieurs horizons de prédiction (p).....	164
Tableau 4. 16. RMSE de γ_3 à plusieurs horizons de prédiction (p).....	165

Liste des algorithmes

Algorithme 2. 1. Sequential Backward Selection (SBS)	28
Algorithme 2. 2. Optimized Sequential Backward Selection (OSBS)	29
Algorithme 2. 3. Etape 1 : algorithme simulant le comportement d'une fourmi Ant_j en déplacement lorsqu'elle quitte le nid	43
Algorithme 2. 4. Etape 2 : algorithme simulant le comportement de deux groupes de fourmis $Ants_i$, $Ants_j$ en déplacement vers la source de nourriture	45
Algorithme 2. 5. Etape 3 : algorithme simulant le comportement d'un groupe de fourmis $Ants_i$ lorsqu'il décide de se diviser pour arriver à la source de nourriture	48
Algorithme 2. 6. Etape 4 : détection de nouvelles classes	50
Algorithme 2. 7. Algorithme d'apprentissage de Baum-Welch	58

Notations

\bar{x}	Valeur moyenne d'un signal sur une fenêtre d'échantillons de données
σ^2	Variance
x_i	Observation définie dans un espace à q -dimensions $x_i = [x_{i1}, x_{i2}, \dots, x_{iq}]$
Ω	Classe
J	Critère de sélection
Σ_w	Matrice de variance intra-classe
Σ_B	Matrice de dispersion inter-classes
g	Centre de gravité de l'ensemble des observations
M	Nombre de classes
g_i	Centre de gravité de la classe Ω_i
x_{ij}	$j^{\text{ème}}$ observation de la classe Ω_i
n_i	Nombre d'observations de la classe Ω_i
Cov	Matrice de covariance
Sim	Similarité
e	Paramètre qui contrôle l'importance de la quantité de phéromone déposée
f	Paramètre qui contrôle l'importance de la distance
ξ	Constante liée à la quantité de phéromone déposée par les fourmis
τ	Quantité de phéromone déposée par les fourmis
A	Matrice de probabilité de transition
B	Matrice de probabilité d'observation
π	Vecteur de probabilités initiales
λ	Modèle de Markov caché
$(.)'$	Matrice transposée
X	Séquence d'observations
$\alpha_t^{(k)}(i)$	Variable forward
$\beta_t^{(k)}(i)$	Variable Backward
t	temps
p	Horizon de prédiction
N'	Vitesse de rotation de la machine en (tr/min)
N_s	Vitesse de synchronisme (tr/min)
N_r	Nombre d'encoches (ou de barres) au rotor
n_p	Nombre de paires de pôles
g	Glissement $g = \frac{N_s - N'}{N_s}$
ω_s	Pulsation des courants statoriques $\omega_s = 2\pi f_s$
f_s	Fréquence des courants statoriques

f_r	Fréquence de rotation du rotor $f_r = \frac{(1-g)f_s}{n_p}$
i_a, i_b, i_c	Courants statoriques
V_a, V_b, V_c	Tensions statoriques
[C]	Matrice de Concordia (repère diphasé de Park fixe lié au stator)
$I_{\alpha,\beta}$	Courant statorique suivant l'axe α (ou β)
$V_{\alpha,\beta}$	Tension statorique suivant l'axe α (ou β)

Introduction générale

Les systèmes d'entraînement électriques à base de machine asynchrone sont largement utilisés en milieu industriel en raison de leur faible coût, de leur robustesse et de leur rapport poids/puissance. En réponse à cette demande industrielle, un nouveau challenge a émergé : l'amélioration des aspects fiabilité, sécurité, coût et maintenance (GdR SDSE 1997-2001, GdR ME2MS 2002-2006). En effet, en raison de la complexité accrue des systèmes d'entraînement électriques, leurs coûts d'exploitation, essentiellement liés à la fiabilité, la sécurité et à la maintenance, se sont vus augmentés ces dernières années. Dès lors, le diagnostic et le pronostic de pannes des systèmes d'entraînement électriques sont devenus incontournables.

Les méthodes utilisées pour le diagnostic de pannes ou de défauts ont largement évolué ces dernières années. Elles ont d'abord été utilisées dans la maintenance corrective afin de diagnostiquer des défauts apparents, puis dans la maintenance systématique pour -au final- être utilisées dans la maintenance préventive conditionnelle. La maintenance systématique est définie comme la réalisation d'une maintenance périodique visant à ce qu'un système ne tombe pas en panne. La maintenance préventive conditionnelle, quant-à-elle, est définie comme une action de maintenance prise en se basant sur des informations acquises à partir de mesures effectuées sur le système. L'efficacité de cette dernière est jugée à partir des méthodes de diagnostic utilisées et qui sont en constante amélioration grâce aux recherches effectuées dans ce domaine. Ces recherches, initialement dédiées au diagnostic des défauts, se sont étendus ces dernières années à la prédiction des défauts (pronostic). Le diagnostic et le pronostic des défauts d'un système se révèlent être une stratégie clé et prometteuse qui suscite auprès des industriels énormément d'intérêt car elle permet d'augmenter la fiabilité et la sécurité des systèmes d'entraînement électriques mais aussi de réduire les coûts liés à leur maintenance.

Il existe deux grandes familles d'approches pour le diagnostic et le pronostic des défauts : l'approche basée sur des modèles analytiques et l'approche basée sur des méthodes orientées données. Le choix d'une approche est basé sur la complexité du système, mais aussi sur la disponibilité de la connaissance que l'on souhaite acquérir sur lui. En effet, tous les systèmes sont susceptibles d'évoluer vers différents modes de fonctionnement. Ces modes ne sont pas obligatoirement tous connus car certains d'entre eux peuvent n'avoir été jamais observés sur le système. Pour résoudre ces problèmes, il convient de développer des méthodes de diagnostic disposant d'une certaine faculté d'adaptation à des situations nouvelles (classification non supervisée), et permettant une reconnaissance fiable (classification supervisée) du mode de fonctionnement dans lequel le système évolue. Il convient aussi de développer des méthodes de pronostic capables d'estimer la durée de vie restante avant l'apparition d'un défaut et de détecter l'imminence d'un ou plusieurs défauts.

C'est dans ce contexte que les travaux présentés dans cette thèse portent sur le développement de nouvelles méthodes de diagnostic et de pronostic orientées données par la reconnaissance

des formes RdF. La RdF est basée sur l'extraction d'informations par le biais du traitement des signaux mesurés. Les signaux mesurables (courants, tensions, vitesse, les vibrations...etc.) peuvent fournir des informations significatives sur l'état de santé du système (présence d'un ou plusieurs défauts, niveau de sévérité...). À partir de ces grandeurs caractéristiques du fonctionnement du système, la mise en œuvre de méthodes de diagnostic (classification non supervisée par l'algorithme des fourmis artificielles, classification supervisée par les modèles de Markov cachés) et de pronostic (systèmes neuro-flous ANFIS « Adaptive neuro fuzzy inference system », neurone neo-flou et les modèles de Markov cachés) permet de concevoir des outils de surveillance efficaces au service de la maintenance préventive conditionnelle.

Ce manuscrit comporte quatre chapitres. Le **premier chapitre** présente un état de l'art sur les méthodes existantes en matière de diagnostic et de pronostic utilisées dans la maintenance préventive conditionnelle. Les méthodes de diagnostic et de pronostic, mises au point ces dernières années, sont examinées et comparées. Il en ressort de cette comparaison que les méthodes de diagnostic et de pronostic orientées données permettent de modéliser le comportement d'un système (processus de dégradation, défaillance) uniquement à partir des données de mesure. Le diagnostic utilise la reconnaissance des formes pour détecter des changements dans les données de mesure via des méthodes de classification supervisée et non supervisée.

Dans le **deuxième chapitre**, une introduction à la reconnaissance des formes (RdF) pour le diagnostic des défauts est présentée. L'objectif de la RdF est de classer une observation mesurée dans une forme en la comparant à des formes types. Une forme est constituée d'un ensemble d'observations mesurées à partir d'un système à un instant donné. Une telle forme, est représentée par plusieurs points dans un espace multidimensionnel à q -paramètres. Les formes types sont en réalité des ensembles de points occupant les zones géométriques de l'espace de représentation appelées classes, où sont regroupées des formes semblables. Les classes correspondent aux modes de fonctionnement connus et inconnus du système. Le fait de détecter une nouvelle classe revient à détecter un nouveau mode de fonctionnement et le fait de classer une nouvelle observation revient à identifier un de ces modes. Deux étapes sont nécessaires pour mettre au point nos méthodes de diagnostic :

- Définition de l'espace de représentation : c'est une étape très importante puisqu'elle influence directement la qualité du système de diagnostic. Elle consiste à construire un vecteur de paramètres pertinents vis-à-vis des modes de fonctionnement connus à partir d'une base d'apprentissage. Cette étape s'effectue généralement par une phase d'extraction ou de sélection de paramètres pertinents. L'algorithme OSBS (Optimized Sequential Backward Selection) est proposé pour la sélection des paramètres pertinents. Cet algorithme a l'avantage de prendre en compte toutes les combinaisons possibles entre les paramètres afin d'améliorer la sélection.
- Classification : elle consiste à identifier les régions de l'espace de représentation correspondant à chaque mode de fonctionnement. Cette étape fait appel à la classification supervisée et non supervisée. La classification non supervisée que nous proposons est basée sur l'algorithme des fourmis artificielles. Cet algorithme analyse les observations inconnues et inexplorées afin de mettre en évidence les classes regroupant les observations similaires. Cela permet d'améliorer la classification et de détecter l'apparition d'un ou plusieurs modes de fonctionnement qui n'étaient pas présents dans la base d'apprentissage. La classification supervisée par les modèles de Markov cachés permet d'associer un degré d'appartenance à l'observation réalisée (sous forme d'une probabilité) par rapport aux classes connues de la base

d'apprentissage. Ainsi, il est possible de définir un indice de fiabilité à l'affectation réalisée mais aussi de détecter l'apparition de nouveaux modes de fonctionnement. Ces méthodes ne se limitent pas qu'au diagnostic des défauts, elles peuvent aussi contribuer au pronostic des défauts.

Dans le **troisième chapitre**, La pronostic est défini comme une extension du diagnostic. L'objectif du pronostic est de détecter l'imminence d'un défaut et d'estimer le temps restant avant son apparition. Pour répondre à ces exigences, trois méthodes basées sur les modèles de Markov cachés sont proposées pour détecter l'imminence d'un défaut et deux méthodes basées le système neuro-flou ANFIS et le neurone neo-flou sont proposées pour estimer le temps restant avant l'apparition d'un défaut.

Le **quatrième chapitre** est consacré à l'application de nos méthodes de diagnostic et de pronostic aux machines asynchrones et aux roulements à rouleaux. Au niveau de l'application, la première partie est consacrée au diagnostic de défauts créés au rotor et aux roulements d'une machine asynchrone de 5,5 KW. Une description du banc de test ainsi que des différents modes de fonctionnement étudiés (sain, défaillants), est présentée. Les défauts correspondent à des cassures de barres et un vieillissement des roulements, et ceci, pour différents niveaux de charge avec une alimentation par onduleur de tension. Deux cas de figure sont étudiés. Le premier cas de figure concerne le diagnostic d'une machine asynchrone avec apparition d'un nouveau mode de fonctionnement ; Le second concerne le diagnostic d'une machine asynchrone avec une connaissance limitée sur chacun de ses modes de fonctionnement. L'algorithme de sélection de paramètres présenté au chapitre 2 est mis en application dans les deux cas de figure afin de déterminer l'espace de représentation. Les méthodes de diagnostic proposées sont alors utilisées sur des essais en présence ou non de défauts et les observations sont analysées pour déterminer une éventuelle apparition d'un nouveau mode de fonctionnement. La seconde partie du chapitre est consacrée au pronostic des roulements à rouleaux. Un espace de représentation est défini à partir d'un vecteur de paramètres indiquant la dégradation des roulements. Les méthodes de pronostic proposées sont alors utilisées pour détecter l'imminence de chaque stade de dégradation et estimer le temps restant avant leur apparition.

Dans une **conclusion générale**, nous repositionnons l'ensemble de nos méthodes de diagnostic et de pronostic au regard de l'objectif initial de cette thèse « Du diagnostic au pronostic de pannes des entraînements électriques ». Nous résumerons les principaux résultats de notre travail et nous aborderons les perspectives de travail qui découlent de cette thèse.

Chapitre 1

Le diagnostic et le pronostic industriel

Sommaire

1.1. Introduction	5
1.2. Stratégies de maintenance	5
1.3. Les principaux aspects de la mise en œuvre d'une maintenance préventive conditionnelle	7
1.4. Traitement du signal	10
1.4.1. Analyse temporelle	10
1.4.2. Analyse fréquentielle	12
1.4.3. Analyse temps- fréquence	12
1.5. Le diagnostic	12
1.5.1. Diagnostic basé sur un modèle physique	13
1.5.2. Diagnostic orienté données	15
1.6. Le pronostic	19
1.6.1. Pronostic basé sur un modèle physique	20
1.6.2. Pronostic orienté données	21
1.6.3. Pronostic basé sur l'expérience	22
1.7. Conclusion	23

1.1. Introduction

La fiabilité et la sûreté de fonctionnement d'un système industriel, à moindre coût, sont les principaux objectifs des entreprises industrielles pour rester compétitif dans un marché qui est sans cesse croissant. Cet intérêt, comme le montre la figure 1.1, est alimenté par le fait qu'un arrêt imprévu peut avoir des conséquences économiques très lourdes dans des secteurs clés.

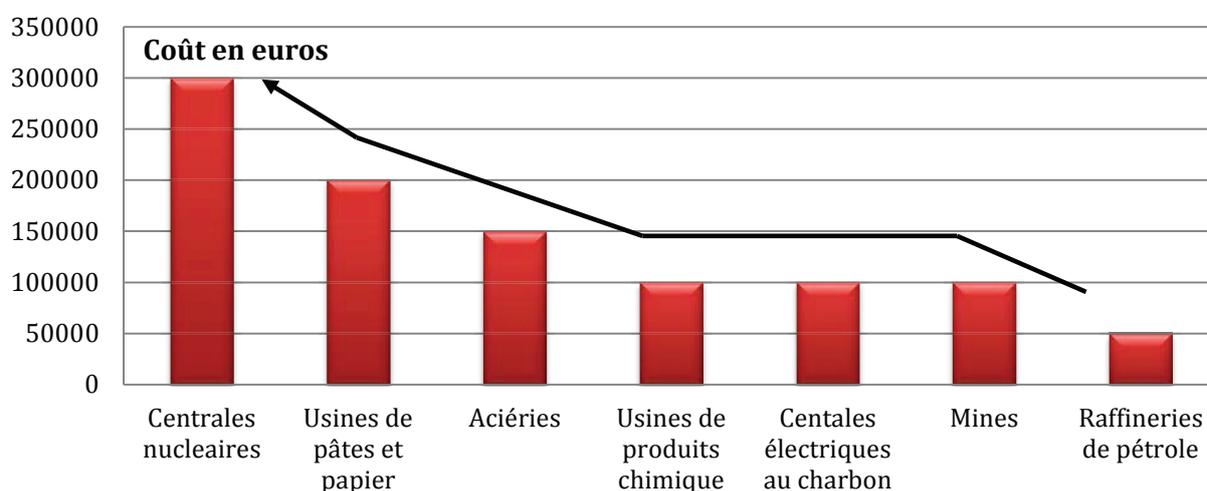


Figure 1. 1. Les conséquences économiques d'une journée d'arrêt dans les secteurs industriels [Helle_06]

Les coûts indiqués ci-dessus soulignent l'importance croissante de la surveillance de l'état de santé des systèmes dans l'industrie par le biais de la maintenance. Les objectifs cités précédemment peuvent être atteints grâce à la mise en place d'une stratégie de maintenance adéquate. Par conséquent, il existe un besoin urgent de développer et d'améliorer continuellement des stratégies de maintenance intelligentes afin d'identifier les besoins en services, optimiser les actions de maintenance et éviter les arrêts imprévus [Pusey_07].

Ce chapitre vise à montrer l'émergence du concept « diagnostic et pronostic » dans l'industrie, à décrire en quoi ils viennent compléter les différentes stratégies de maintenance, et à décrire les bénéfices pouvant être attendus par la mise en œuvre d'une approche de diagnostic et de pronostic au service de la maintenance préventive conditionnelle. De manière plus ciblée, ce chapitre donne un état de l'art sur les méthodes existantes en matière de diagnostic et de pronostic. L'ensemble permet de montrer la diversité des approches envisageables et de choisir parmi elles celle qui définira un cadre de travail pour notre thèse.

1.2. Stratégies de maintenance

[Murthy_03] et [Wang_02] ont effectué ces dernières années des recherches relatives au développement et à l'exécution de plans de maintenance dans le but de garantir la fiabilité et la sûreté de fonctionnement des systèmes. Leurs recherches définissent la maintenance comme étant une action corrective pour réparer une panne déjà existante ou une action préventive pour empêcher l'apparition d'une panne. Au terme de ces recherches, deux stratégies de maintenance ont été identifiées : la maintenance corrective et la maintenance préventive. La figure 1.2 montre la structure hiérarchique de ces stratégies de maintenance [Murthy_03].

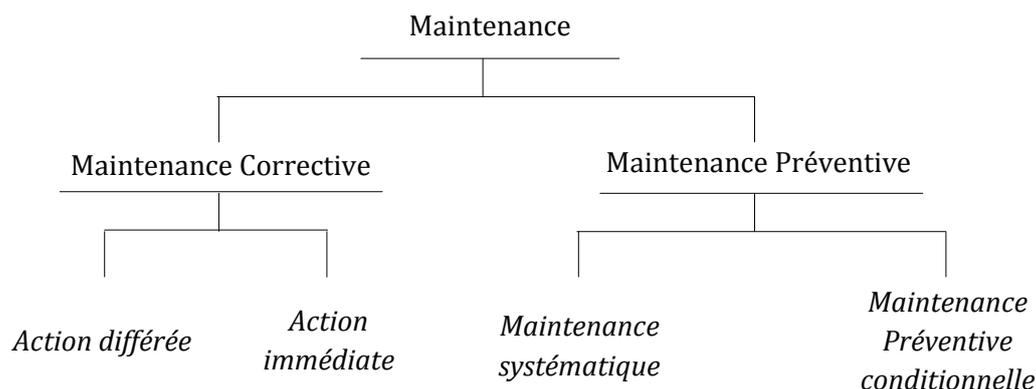


Figure 1. 2. Classification de la maintenance

La maintenance corrective est semblable à des travaux de réparation, elle consiste à faire fonctionner un système jusqu'à ce qu'il tombe en panne puis de prendre des mesures correctives afin de le remettre en état. Ce type de maintenance est souvent nommé réparation ou restauration et comprend la réparation et le remplacement des composants défectueux. L'un de ses principaux avantages est qu'elle ne nécessite aucun investissement dans les activités de maintenance avant l'apparition de la panne. Cependant, ce type de maintenance peut entraîner un ralentissement, voir l'arrêt de la production. Pour les travaux de réparation, certaines approches sont disponibles. Avec une approche « réparations minimales », le composant défaillant est simplement remis à son état de fonctionnement, il continue son travail comme si rien ne s'était passé. La probabilité d'un défaut (taux de défaillance) dans ce cas reste le même que celui avant la panne. L'utilisation d'une telle approche signifie que l'élément remis en état, est seulement rétabli à un état dit « aussi mauvais que vieux » [Hoyland_94]. La réparation minimale peut être exécutée pour diverses raisons, comme le manque de temps, les pièces de rechange, la compétence...etc. Si le composant défaillant est remplacé par un nouveau composant du même type, ou s'il est restauré à un état « aussi bon que neuf », le taux de défaillance diminue au niveau de celui d'un composant tout juste mis en service. Pour des pannes critiques engendrant un arrêt du système, la maintenance corrective doit être effectuée immédiatement. Toutefois, pour des pannes qui n'ont pas ou peu de conséquence sur la production, l'entretien peut être différé à un moment mieux adapté ou plus opportun.

La maintenance préventif est définie comme une « Maintenance effectuée à des intervalles prédéterminés ou selon des critères prescrits et destinés à réduire le taux de défaillance ou la dégradation des équipements » (SS-EN 13306, 2001, p.14). La maintenance préventive inclut deux types de plans de maintenance : la maintenance préventive systématique et la maintenance préventive conditionnelle (SS-EN 13306, 2001). La maintenance systématique est prévue et planifiée sans l'apparition d'aucune activité de surveillance. Elle peut être basée sur le nombre d'heures d'utilisation, le nombre de fois qu'un système est utilisé, le nombre de kilomètres d'utilisation, en fonction de dates prescrites...etc. Ce type de maintenance est le mieux adapté pour un composant qui a des signes d'usure bien visibles et où les tâches de maintenance peuvent être effectuées à un moment qui permettra d'éviter une panne dans le système [Starr_97]. Selon [Yam_01] et [Starr_97], la maintenance systématique est parfois appelée « Maintenance basée sur le temps » et « maintenance préventive planifiée ».

La maintenance préventive conditionnelle (MPC), à l'inverse de la maintenance systématique, ne repose pas sur des actions prévues et planifiées. Elle est effectuée en fonction des besoins

indiqués par l'état de santé du système [Yam_01]. Elle repose sur des paramètres (signatures) pouvant détecter l'état de santé actuel du système et utilisés pour prédire une éventuelle défaillance avant qu'elle ne se produise réellement. Le suivi de ces signatures peut fournir une indication sur l'imminence d'une défaillance ainsi que sur les défauts naissants qui peuvent mener le système ou ses composants à s'écarter d'un niveau de performance acceptable ou dans le pire des cas, provoquer son arrêt. Parmi les avantages de la MPC on trouve sa capacité à détecter un défaut imminent et sa précision dans la prédiction de la défaillance [Sou_13]. Elle contribue également au diagnostic des défauts car il est relativement facile d'associer un défaut spécifique dans le système aux paramètres surveillés. Comme le montre le tableau 1.1, la MPC peut réduire considérablement les coûts d'intervention et accroître la sûreté de fonctionnement des systèmes. Elle optimise le compromis entre les coûts de la maintenance et les performances en augmentant la disponibilité et la fiabilité tout en éliminant les activités de maintenance inutiles.

Tableau 1. 1. Points forts et points faibles des stratégies de maintenance

Stratégie de maintenance	Avantage	Inconvénient
Maintenance corrective	<ul style="list-style-type: none"> + Equipements pas tout le temps entretenus + Aucun coût lié à la maintenance 	<ul style="list-style-type: none"> - Risque élevé de défaillances - Arrêt de production élevé - Coût élevé de pièces de rechange - Travail supplémentaire - Risque élevé au niveau de la sécurité
Maintenance systématique	<ul style="list-style-type: none"> + Moins de pannes catastrophiques + Gestion maîtrisée des coûts de maintenance + Taux de défaillances réduits 	<ul style="list-style-type: none"> - Equipements entretenus même en absence de pannes - Risque de dégradation des équipements durant la maintenance - Risque élevé de pannes imprévues
Maintenance préventive conditionnelle	<ul style="list-style-type: none"> + Réduction de pannes imprévues + Pièces de rechange commandées en cas de besoin + Maintenance effectuée au moment opportun + Durée de vie prolongée 	<ul style="list-style-type: none"> - Coûts d'investissement élevés

1.3. Les principaux aspects de la mise en œuvre d'une maintenance préventive conditionnelle

La maintenance préventive conditionnelle se compose d'un certain nombre de modules fonctionnels: acquisition de données, traitement du signal, surveillance, diagnostic, pronostic et l'aide à la décision. En outre, une interface homme-machine peut être ajoutée afin de fournir un moyen d'afficher des informations vitales sur l'état de santé du système.

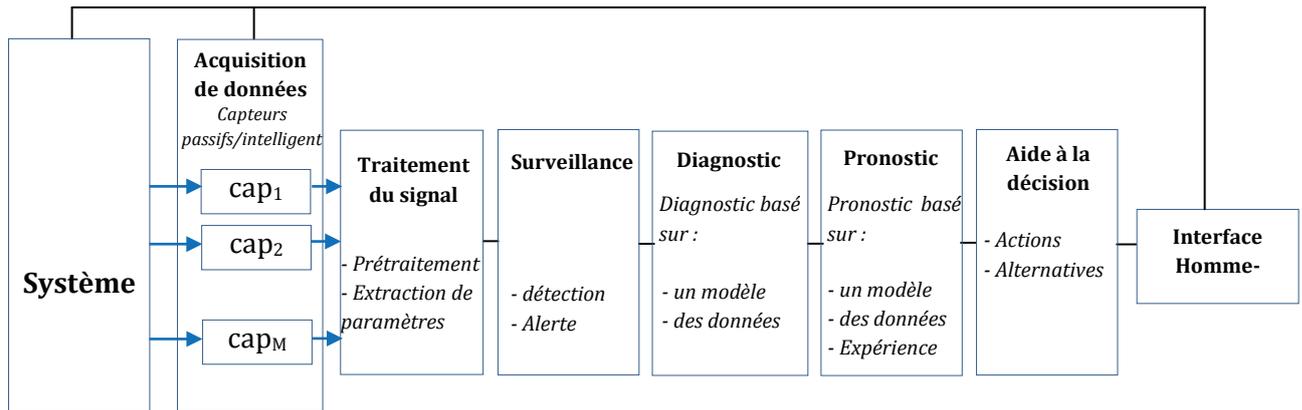


Figure 1. 3. Architecture d'une maintenance préventive conditionnelle avec sept modules [Lebold_03]

Le (MIMOSA) « Machinery Information Management Open Standard Alliance » a proposé une architecture standardisée pour la MPC qui est décrite en termes de couches fonctionnelles : de l'acquisition de données jusqu'à l'interface homme-machine (voir figure 1.3) [Thurston_01].

+ *Couche 1 - Acquisition de données* : le module d'acquisition de données permet d'accéder aux signaux et aux données numériques des capteurs. Comme le montre la figure 1.4, l'acquisition de données couvre différentes disciplines, telles que les mesures mécaniques [Wang_96, Holroyd_05, Roemer_00, Hountalas_00], les mesures électriques [Tsoumas_05, Kar_06], la tribologie [Walter_04] et les mesures non destructifs [Mba_06]. Les mesures mécaniques comprennent les vibrations, l'émission acoustique, la pression, le débit, la température et le stress, tandis que les mesures électriques sont le courant, la tension, la phase et le flux. La tribologie s'intéresse surtout à la lubrification des machines et à l'analyse des débris d'huile. Le contrôle non destructif utilise l'inspection visuelle ou la mesure sans contact.

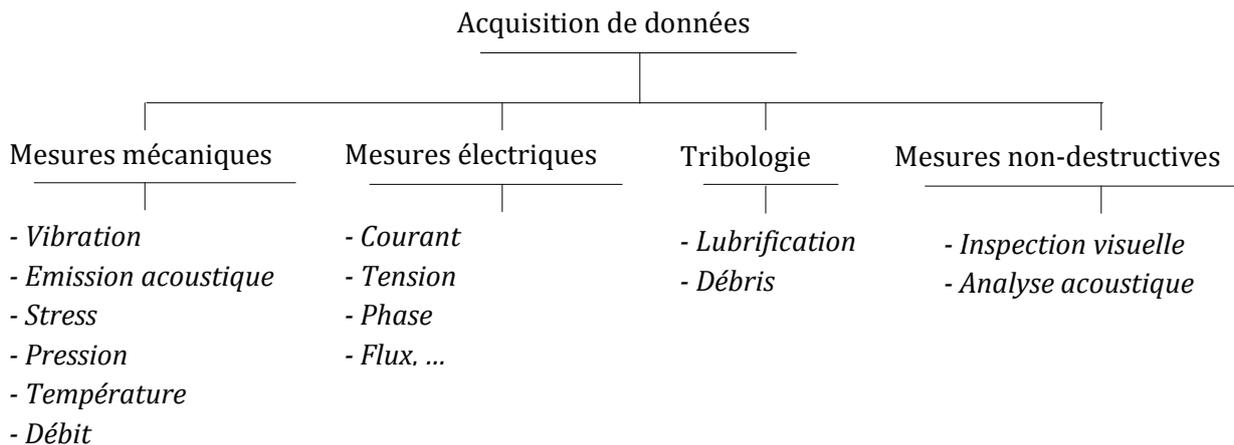


Figure 1. 4. Acquisition de données en fonction des disciplines

+ *Couche 2 - Traitement du signal* : le traitement de signal reçoit des signaux et des données du module d'acquisition de données ou d'autres modules de traitement du signal. La sortie de ce module comprend des données filtrées, des données spectrales, des signaux virtuels et d'autres signatures.

+ *Couche 3 – Surveillance* : le module de surveillance reçoit des données des modules de traitement du signal et d'autres modules de surveillance. Son principal objectif est de comparer les données avec des valeurs de référence. Le module de surveillance doit également être en mesure de générer des alertes basées sur les limites de fonctionnement prédéfinies (par exemple, faible niveau de dégradation, niveau normal de dégradation, stade avancé de dégradation).

+ *Couche 4 – Diagnostic* : le module de diagnostic détermine si l'état du système surveillé, ou composant est dégradé ou non et identifie le défaut responsable de cette dégradation.

+ *Couche 5 – Pronostic* : le module de pronostic a la possibilité de prendre en compte les informations provenant de toutes les couches antérieures. L'objectif principal de ce module est de projeter l'état de santé actuel du système dans le futur.

+ *Couche 6 – Aide à la décision* : le module d'aide à la décision reçoit des données provenant du module de diagnostic et du module de pronostic. Son objectif principal est de recommander des actions de maintenance et des alternatives liées à la gestion du système.

+ *Couche 7 – Interface homme-machine* : l'état de santé du système ainsi que les alertes sont affichés dans ce module avec la possibilité d'accéder aux couches antérieures en fonction des besoins en information de l'utilisateur.

Une autre architecture pour la MPC, semblable à la première, a été présentée par [Vacht_01]. Cette architecture met l'accent sur le module de diagnostic et de pronostic avec comme sortie, un accès direct sur l'état futur du système. Le diagnostiqueur évalue l'état réel du système à partir des mesures en ligne fournies par les capteurs. Son but est de détecter et d'identifier une éventuelle anomalie qui pourrait apparaître durant le fonctionnement du système. Le pronostiqueur, qui reçoit des informations du module de diagnostic, décide du maintien en service du système en se basant sur la prédiction de l'état futur du système.

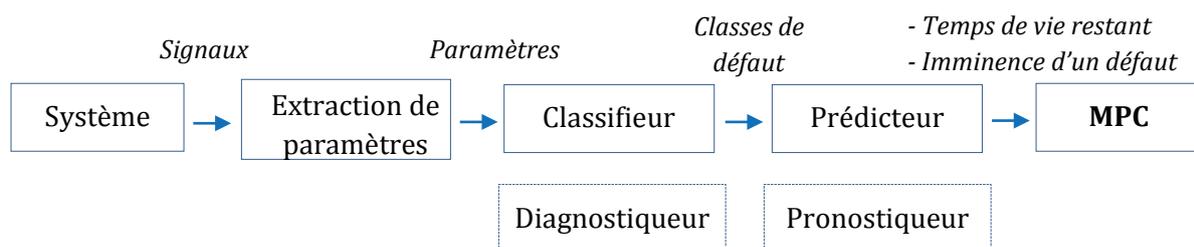


Figure 1. 5. Architecture générale d'un système de pronostic tel que présenté par [Vacht_01]

Une troisième architecture pour la MPC a été présentée par [Chinnam_04]. Cette architecture, en plus du diagnostic et du pronostic, met l'accent sur les capteurs et le traitement du signal. Des capteurs sont montés sur les composants à surveiller afin de capter des signaux de dégradation et qui par des interprétations ultérieures peuvent conduire au développement d'un plan de maintenance sur mesure.

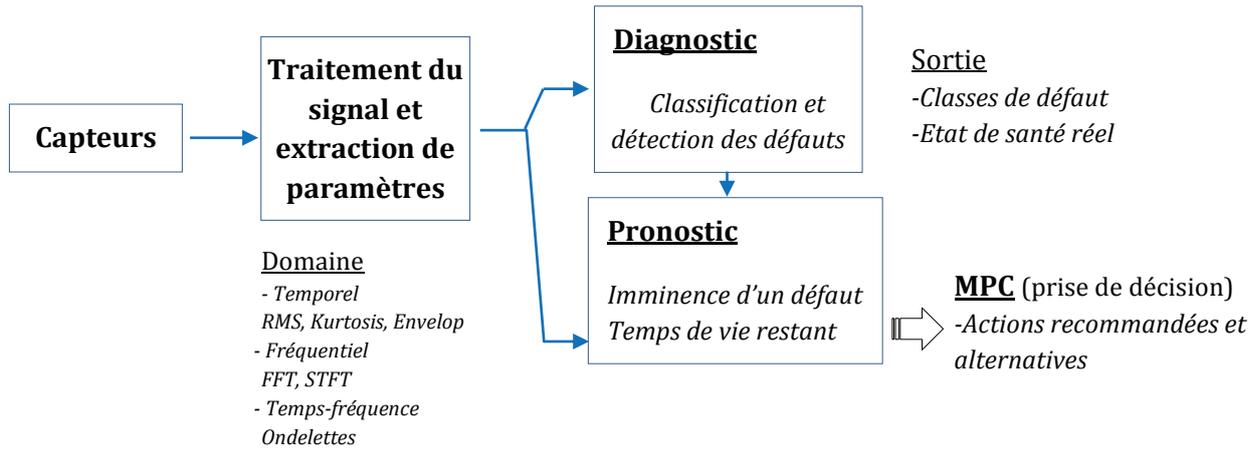


Figure 1. 6. Les éléments de base du diagnostic et pronostic pour la MPC tel que présenté par [Chinnam_04]

Dans les sections suivantes, les techniques de traitement du signal, de diagnostic et de pronostic sont décrites et examinées car ce sont des éléments clés pour assurer la fiabilité et la sûreté de fonctionnement des systèmes.

1.4. Traitement du signal

Le traitement du signal regroupe un ensemble de techniques permettant de créer, d’analyser et de transformer les signaux d’entrée en vue d’extraire des paramètres indicateurs de défauts [Serya_10, Chen_12, Prieto_13]. Comme le montre la figure 1.7, il est possible de classer les techniques de traitement du signal en : Analyse temporelle, Analyse fréquentielle et Analyse temps-fréquence.

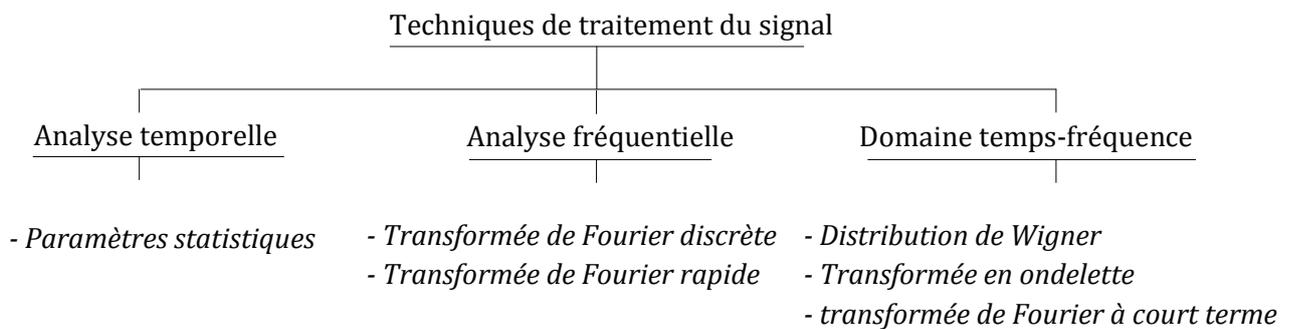


Figure 1. 7. Techniques de traitement du signal

1.4.1. Analyse temporelle

L'analyse temporelle permet d’extraire des paramètres indicateurs de défauts à partir de données brutes du capteur. Les paramètres décrits ici sont appelés « paramètres statistiques » car ils sont basés sur une analyse des caractéristiques temporelles du signal enregistré. On peut citer par exemple :

- La valeur moyenne : la valeur moyenne notée \bar{x} d’un signal sur une fenêtre d’échantillons de données est un paramètre significatif pour presque chaque type de capteur. Elle est définie par :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.1)$$

où x_i est la $i^{\text{ème}}$ donnée et N est le nombre de points échantillon choisi pour représenter le signal.

- La variance : la variance notée σ^2 est une mesure servant à caractériser la dispersion d'une distribution ou d'un échantillon. Elle est définie par :

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} \quad (1.2)$$

Ce paramètre est souvent utilisé comme base métrique pour les classifieurs comme les réseaux bayésiens dynamiques [Wang_07] et les réseaux de neurones [Laerhoven_01].

- La valeur efficace : la valeur efficace (dite aussi valeur RMS, de l'anglais *Root Mean Square*, ou moyenne quadratique) d'un signal, est la racine carré du moment d'ordre deux (ou variance) du signal :

$$RMS = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (1.3)$$

C'est l'un des paramètres le plus utilisé dans l'analyse temporelle. Une augmentation de la valeur du RMS indiquera une dégradation de l'état de santé du système.

- Le Kurtosis : le Kurtosis noté S_{kurt} représente le moment statique d'ordre 4. Il mesure le degré d'écrasement de la distribution du signal vibratoire enregistré et est défini comme étant le rapport entre le moment d'ordre quatre centré et le carré de la variance.

$$S_{kurt} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{(\sigma^2)^2} \quad (1.4)$$

Un système en bon état engendre un signal vibratoire avec un Kurtosis voisin de 3. Pour un système dégradé, l'amplitude du signal est modifiée et le Kurtosis devient supérieur ou égal à 4.

- Le facteur de crête : le facteur de crête noté F_c est une mesure caractéristique d'un signal vibratoire. C'est le rapport entre l'amplitude du pic du signal et la valeur efficace du signal. Il est défini par :

$$F_c = \frac{|x_{pic}|}{RMS} \quad (1.5)$$

Un système en bon état génère un signal vibratoire de faible amplitude, aussi bien en valeur crête qu'en valeur efficace. Le facteur de crête reste donc faible (entre 2 et 6). Un défaut localisé génère une vibration d'amplitude crête forte et d'amplitude efficace faible, donc un facteur de crête important (> 6). Le défaut majeur de ce paramètre et du Kurtosis est qu'ils indiquent environ les mêmes valeurs à l'état neuf et en fin de vie du système [Medjaher_12].

- *Le Skewness* : le Skewnes noté S_{kew} représente le moment statique d'ordre 3 centré sur le cube de l'écart-type. Il mesure la symétrie de la distribution, ou plus précisément le manque de symétrie. Une distribution est symétrique si elle présente la même allure de part et d'autre du signal. Il est défini comme suit :

$$S_{kew} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{(\sigma)^3} \quad (1.6)$$

Si l'écart-type σ est égal à 0, la distribution est symétrique. Si σ est plus petit que 0, la distribution est asymétrique vers la gauche. Si σ est plus grand que 0, la distribution est asymétrique à droite.

Les paramètres décrits ci-dessus ont tous le même inconvénient, ils indiquent la dégradation du système mais n'arrivent pas à identifier le défaut responsable de cette dégradation [Tandon_99].

1.4.2. Analyse fréquentielle

L'analyse du spectre fréquentiel d'un signal est la technique la plus couramment utilisée pour identifier les défauts dans un système. Cette technique est basée sur le fait qu'un défaut localisé génère un signal périodique avec une fréquence caractéristique unique [Li_99]. À l'inverse de l'analyse temporelle, l'analyse fréquentielle permet d'identifier le défaut présent dans le système en identifiant sa fréquence caractéristique. Cette technique est généralement appliquée pendant le régime permanent du système [Did_04]. Un classique parmi les techniques utilisées dans le domaine fréquentiel est l'analyse spectrale par transformée de Fourier rapide FFT (*Fast Fourier Transform*). La technique de spectrogramme permet d'effectuer des analyses fréquentielles des signaux en régime dynamique de fonctionnement du système. Cette technique consiste à effectuer un calcul répétitif de la FFT sur une fenêtre temporelle glissante, ce qui rend cette technique sensible à la longueur de la fenêtre, au type de fenêtrage, à la durée totale de supervision et au pas de glissement de la fenêtre [Fla_93]. Bien que cette technique permet d'analyser des signaux en régime dynamique, la rapidité du régime dynamique, qui est dans le cas des machines asynchrones de l'ordre de 150ms, réduit significativement l'efficacité de cette technique.

1.4.3. Analyse temps- fréquence

L'analyse en temps-fréquence des signaux traite à la fois le domaine temporel et le domaine fréquentiel. Les signaux non stationnaires sont mieux représentés par une distribution temps-fréquence, qui vise à montrer la répartition de l'énergie du signal sur l'espace à deux dimensions temps-fréquence [Burg_95]. Les techniques les plus utilisées pour l'analyse temps-fréquence sont la transformée de Fourier à court terme (STFT) pour *Short-Time Fourier Transform*, la distribution de Wigner-Ville et la transformée en ondelettes [Wang_96].

1.5. Le diagnostic

En raison d'une demande croissante pour de meilleures performances ainsi que pour assurer la fiabilité et la sûreté de fonctionnement des systèmes, le diagnostic, qui est défini comme étant un processus de détection et de localisation de défauts, est devenu une étape incontournable dans le domaine de la surveillance. Le diagnostic d'un défaut, alors que le système est en cours

de fonctionnement, permet d'éviter des arrêts imprévus et par la même occasion, réduit les coûts liés à la maintenance. Par conséquent, le diagnostic des défauts est devenu un sujet de recherche qui suscite auprès des industriels un intérêt grandissant qui a débouché sur de nombreux travaux de recherche afin de développer des méthodes de diagnostic encore plus efficaces. On peut citer par exemple le diagnostic basée sur des méthodes analytiques, statistiques ou sur l'intelligence artificielle [Li_00, Stack_04, Frosini_10, Immo_09, Zhang_11, Loparo_00, Ibrahim_08, Immo_13, Yazidi_11].

Les méthodes de diagnostic étudiées dans [Ven(I)_03], [Ven(II)_03] et [Ven(III)_03] peuvent être classées en deux grandes catégories : les méthodes basées sur un modèle physique et les méthodes orientées données.

1.5.1. Diagnostic basé sur un modèle physique

La plupart des méthodes utilisant un modèle physique sont basées sur les résidus. Ces méthodes nécessitent deux étapes. La première étape génère des irrégularités entre le comportement réel du système et celui estimé par le modèle. De telles irrégularités, aussi appelées résidus, reflètent les défauts potentiels du système. Chaque résidu représente une signature unique d'un défaut et donc les éléments défectueux peuvent être facilement identifiés. La deuxième étape choisit une règle de décision pour le diagnostic. La figure 1.8 donne le schéma conceptuel du diagnostic basé sur un modèle.

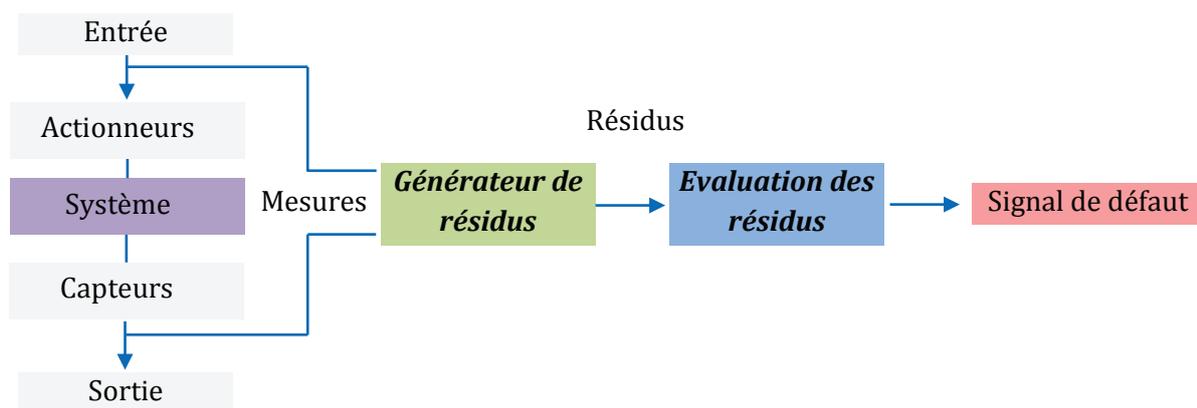


Figure 1. 8. Structure d'un système de diagnostic à base d'un modèle

Il existe une grande variété de méthodes basées sur les résidus, par exemple les méthodes à base d'observateurs, espace de parité, et les méthodes basées sur l'estimation paramétrique.

1.5.1.1. Observateurs

Les méthodes à base d'observateurs consistent à reconstruire à partir d'un modèle analytique et d'un ensemble d'observations partielles du système (entrées/sorties) les variables d'états ou sortie du système considéré [Zolghadri_96], [Said_00]. Les différences entre les variables mesurées du système et les valeurs estimées par le modèle sont utilisées comme des résidus. Un résidu proche ou égal à zéro signifie que le système fonctionne normalement. Par contre, un résidu éloigné de zéro indiquera une divergence entre la sortie du système et la sortie estimée par le modèle. Cette différence permet la détection d'un défaut dans le système. La figure 1.9 donne le schéma conceptuel d'un générateur de résidus à base d'un observateur.

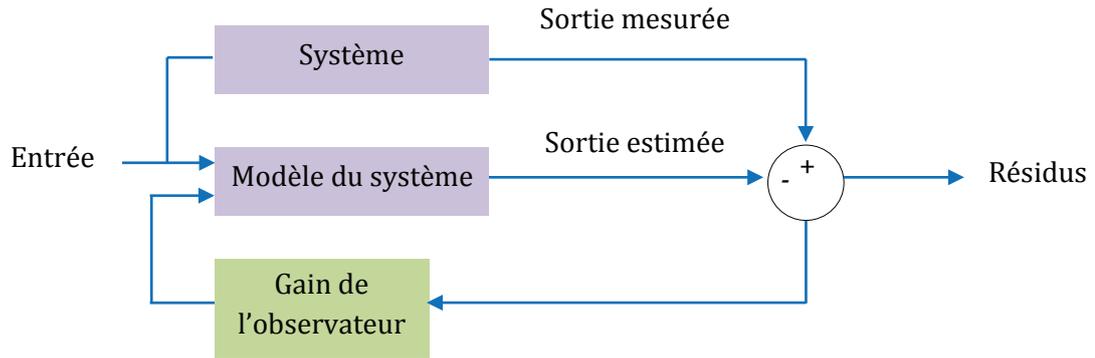


Figure 1. 9. Générateur de résidus à base d'un observateur

1.5.1.2. Espace de parité

Les méthodes de type espace de parité reposent sur la vérification des relations de redondance analytique (statique ou dynamique) parmi les entrées et les sorties du système. Les relations de redondance analytique sont obtenues en éliminant les variables d'état non observables du modèle. Les relations de redondance analytique sont ensuite utilisées pour construire des indicateurs de défaut pour le diagnostic.

Les méthodes de type espace de parité et celles à base d'observateurs sont similaires, comme l'indique les recherches effectuées par [Frank_94a, Wun_90, Patton_94, Chow_84]. Les travaux présentés dans [Ding_08] montrent le développement mathématique pour calculer le vecteur de parité correspondant à un générateur de résidus à base d'observateurs et vice versa. Ainsi, nous pouvons concevoir un générateur de résidus dans l'espace de parité et ensuite transformer le vecteur de parité afin de générer des résidus. La figure 1.10 donne le schéma conceptuel d'une approche de type espace de parité pour un générateur de résidus [Frank_96].

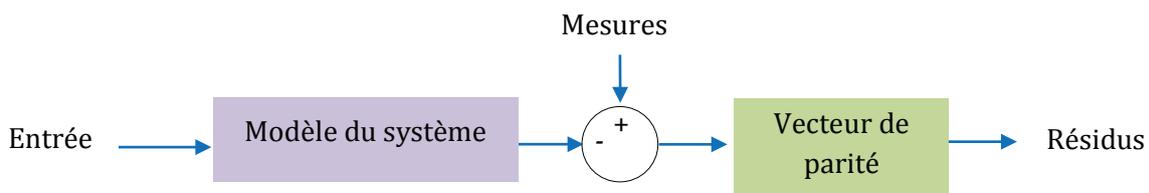


Figure 1. 10. Générateur de résidus dans l'espace de parité

1.5.1.3. Estimation paramétrique

L'estimation paramétrique a été proposée par [Bask_79, Isermann_84] pour la détection et le diagnostic des défauts. Elle est basée sur l'hypothèse que les défauts sont reflétés par les paramètres physiques du système. L'apparition d'un défaut peut être détectée en comparant les paramètres estimés avec ceux du système. Tout écart entre les deux, donnera une indication sur la présence d'un défaut. La figure 1.11 donne le schéma conceptuel de l'identification paramétrique pour le diagnostic. L'estimation paramétrique possède l'avantage de mesurer l'écart, ce qui est utile pour l'analyse des défauts [Frank_96]. Toutefois, un des inconvénients

majeurs de cette méthode réside dans la nécessité d'avoir un système physique excité en permanence.

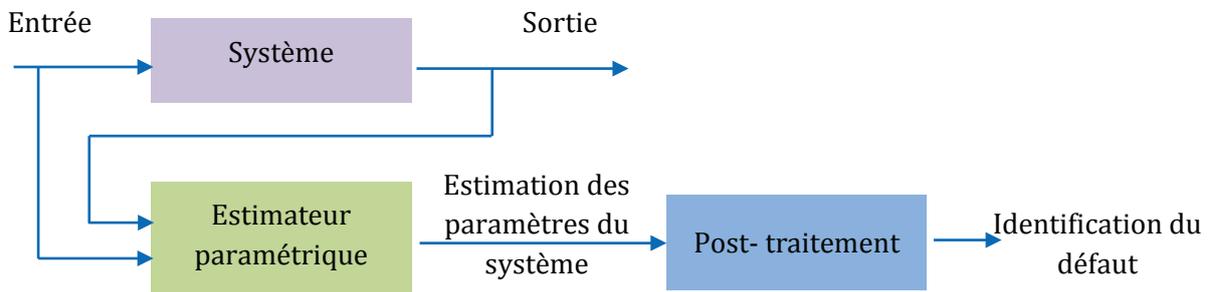


Figure 1. 11. Description schématique d'une estimation paramétrique

1.5.2. Diagnostic orienté données

Contrairement aux méthodes basées sur un modèle physique où la connaissance a priori sur le système est nécessaire, dans les méthodes orientées données, seule la disponibilité des données de mesure est nécessaire. Les méthodes orientées données peuvent être : des systèmes experts, des méthodes statistiques comme l'analyse en composantes principales (ACP), la classification et des méthodes basées sur l'intelligence artificielle comme les réseaux de neurones. La figure 1.12 donne le schéma conceptuel du diagnostic orienté données.

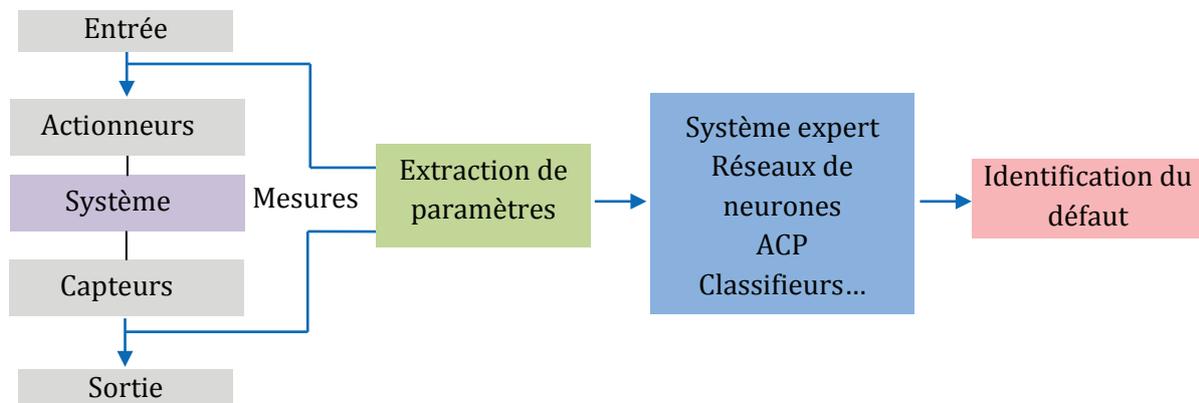


Figure 1. 12. Structure d'un système de diagnostic orienté données

1.5.2.1. Système expert

Un système expert est défini par Angeli [Angeli_04] comme « un programme informatique intelligent qui utilise une base de connaissances et un moteur d'inférence pour résoudre les problèmes qui sont d'une complexité telle qu'ils nécessitent une importante expertise humaine pour leur solution ». Un grand nombre de chercheurs ont travaillé sur l'application des systèmes experts pour des problèmes de diagnostic. Becraft a proposé un cadre intégré comprenant un réseau de neurones et un système expert [Becraft_93]. Un réseau de neurones est utilisé comme un filtre pour diagnostiquer les défauts les plus fréquemment rencontrés dans le système. Une fois que les défauts sont localisés par le réseau neuronal, un système expert analyse le résultat et

confirme le diagnostic ou offre une solution alternative. L'inconvénient d'un système expert est qu'il ne parvient pas à détecter les nouveaux défauts sans signatures connues [Rich_87].

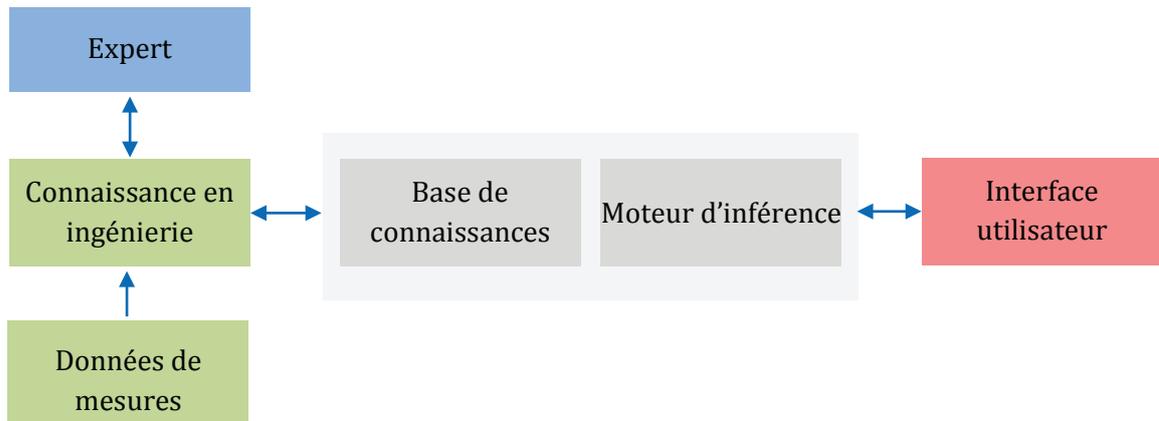


Figure 1. 13. Architecture d'un système

1.5.2.2. Méthodes statistiques multivariées

Les applications des méthodes statistiques multivariées pour le diagnostic des défauts telles que l'analyse en composantes principales (ACP) ont été largement rapportées dans la littérature. Un aperçu de l'utilisation de l'ACP dans le diagnostic des défauts a été donné par [Kresta_91, MacGregor_95 et Wise_96]. L'analyse en composantes principales est une méthode de compression de données. Elle réduit un ensemble de données à quelques éléments d'information, mais qui expliquent l'essentiel des variations du système [Ven(II)_03]. L'ACP combine des outils de l'algèbre linéaire (décomposition en valeurs propres) et l'analyse statistique (variance-covariance) pour construire un modèle représentatif qui peut être utilisé pour le diagnostic des défauts.

1.5.2.3. Réseau de neurones

Les réseaux de neurones, utilisés pour le diagnostic des défauts, peuvent être classés selon : (i) l'architecture du réseau, et (ii) la méthode d'apprentissage [YANG_04]. Il y a eu un certain nombre de documents qui abordent le problème du diagnostic des défauts par des réseaux de neurones. En génie chimique, [Watanabe_89, Ungar_90 et Hoskins_91] ont été parmi les premiers chercheurs à démontrer l'utilité des réseaux de neurones pour le diagnostic. Leur force réside dans l'aptitude à apprendre et à capturer les relations entre l'entrée du réseau de neurones et sa sortie même si ces relations sont inconnues ou difficiles à décrire. Cependant, leur principal inconvénient réside dans l'acquisition et la disponibilité des données d'apprentissage [Ven(III)_03].

1.5.2.4. Classification

Le diagnostic d'un défaut dans un système peut être interprété comme un problème de classification [Ondel(I)_06, Soualhi_11, Casimir(I)_03, Casimir(II)_03]. Dans ce cas, le but est de classer un échantillon de mesure appelé observation (x_i) dans l'une des classe ($\Omega_1, \Omega_2, \dots, \Omega_M$) qui représente un état de santé du système (voir figure 1.14).

Une observation notée $x_i = [x_{i1}, x_{i2}, \dots, x_{iq}]$ est définie dans un espace à q -dimensions où q représente le nombre de paramètres indicateurs de l'état de santé du système.

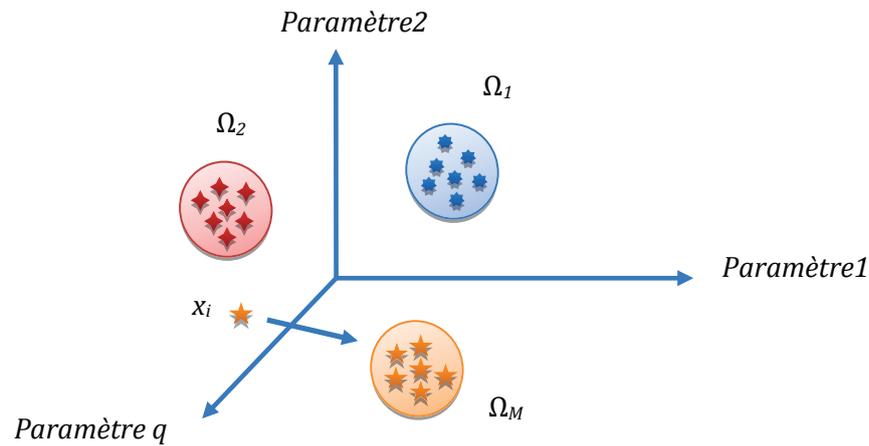


Figure 1. 14. Association d'une observation x_i à l'une des classes ($\Omega_1, \Omega_2, \dots, \Omega_M$)

Les problèmes de la classification ont été abordés par la reconnaissance des formes (Rdf) dans le passé [OCAK_03], [Reng_00], et l'utilisation des réseaux de neurones plus récemment [YANG_04]. Il existe deux types de classifications : la classification supervisée et la classification non supervisée.

La classification supervisée déduit une fonction d'un ensemble d'échantillons d'apprentissage. Chaque échantillon est une paire constituée d'une entrée (par exemple une observation x_i) et une étiquette « o » (c'est à dire la classe d'appartenance de l'observation x_i). Un algorithme d'apprentissage supervisé est alors utilisé afin de produire une fonction d'appartenance entre x_i et o . Cette fonction d'appartenance servira par la suite afin de classer les nouvelles observations.

La classification non supervisée analyse les observations sans tenir compte de leurs étiquettes afin de mettre en évidence les classes regroupant les observations similaires. Le « clustering » est un exemple typique de la classification non supervisée dans lequel les observations sont groupées en classes représentant l'état de santé du système.

La classification supervisée et non supervisée sont deux façons de résoudre les problèmes de diagnostic. Le choix de la classification dépend de la disponibilité de la base d'apprentissage. Si l'on dispose d'observations déjà recueillies sur le système (ensemble d'apprentissage) et pour lesquelles on connaît a priori la classe d'appartenance, c'est-à-dire l'état de santé du système correspondant, on parle de classification en *mode supervisé*. Dans le cas contraire, on parle de classification en *mode non supervisé*.

Une comparaison entre les différentes méthodes utilisées pour le diagnostic des défauts est montrée dans le tableau 1.2. Cette comparaison a pour but d'identifier les points forts et les faiblesses des méthodes de diagnostic mentionnées précédemment. Les recherches effectuées par [Ven(III)_03] et [Soura_00] permettent de définir neuf caractéristiques afin de juger une méthode de diagnostic, comme le montre le tableau 1.2. Il en ressort de cette comparaison que les méthodes à base d'un modèle nécessitent un modèle analytique pour représenter un système. Dans le cas réel, la disponibilité d'un modèle analytique est souvent irréaliste ou inexacte en raison de la complexité du système, ceci peut être causé par des modèles ne reflétant pas de façon précise le comportement du système. Les méthodes orientées données, quant à elles, ne nécessitent pas un modèle analytique pour diagnostiquer un système. Seule la disponibilité des données de mesure est nécessaire. Par conséquent, les méthodes orientées données sont une alternative dans le cas de la non disponibilité d'un modèle analytique. Dans la suite de cette thèse, nous nous intéresserons essentiellement aux méthodes orientées données; plus précisément, au diagnostic par classification supervisée et non supervisée.

Tableau 1. 2. Comparaison des différentes méthodes utilisées pour le diagnostic des défauts

<i>Méthodes critères</i>	<i>Espace de parité ou observateurs</i>	<i>Estimation paramétrique</i>	<i>Systèmes experts</i>	<i>ACP</i>	<i>Réseaux de neurones</i>	<i>Classification</i>
<i>Détection et diagnostic rapide</i>	✓	X	✓	✓	✓	✓
<i>Identification</i>	✓	✓	✓	✓	✓	✓
<i>Robustesse</i>	✓	?	✓	✓	?	?
<i>Identification de nouveaux défauts</i>	✓	X	X	✓	?	✓
<i>Aucun modèle physique requis</i>	X	X	✓	✓	✓	✓
<i>Identification de défauts multiples</i>	✓	✓	✓	X	?	✓
<i>Facilité d'exploitation</i>	X	X	✓	X	X	✓
<i>Temps de calcul</i>	X	X	✓	✓	?	?
<i>Adaptation</i>	X	X	X	X	✓	✓

✓ : Favorable

X : Non favorable

? : dépend de la situation

1.6. Le pronostic

La maintenance préventive conditionnelle nécessite une approche prédictive. Celle-ci fournit au personnel de la maintenance une indication sur l'état futur du système et, idéalement, octroie un délai suffisant pour que le personnel, l'équipement et les pièces de rechange soient organisés, minimisant ainsi les temps d'arrêt et les coûts de maintenance [Zaidi_11, Chen_11, Strangas_11, Mahamad_10, Benke_12]. La figure 1.15 présente les étapes d'un processus de traitement d'un défaut éventuel au niveau d'un système. Ce système est considéré comme étant en état de marche au début, puis, après un certain temps, un défaut naissant se développe dans le système. Au fil du temps, la sévérité du défaut augmente jusqu'à la dégradation totale du système. Si le système est autorisé à poursuivre son fonctionnement, il y a une possibilité de voir apparaître d'autres défauts.

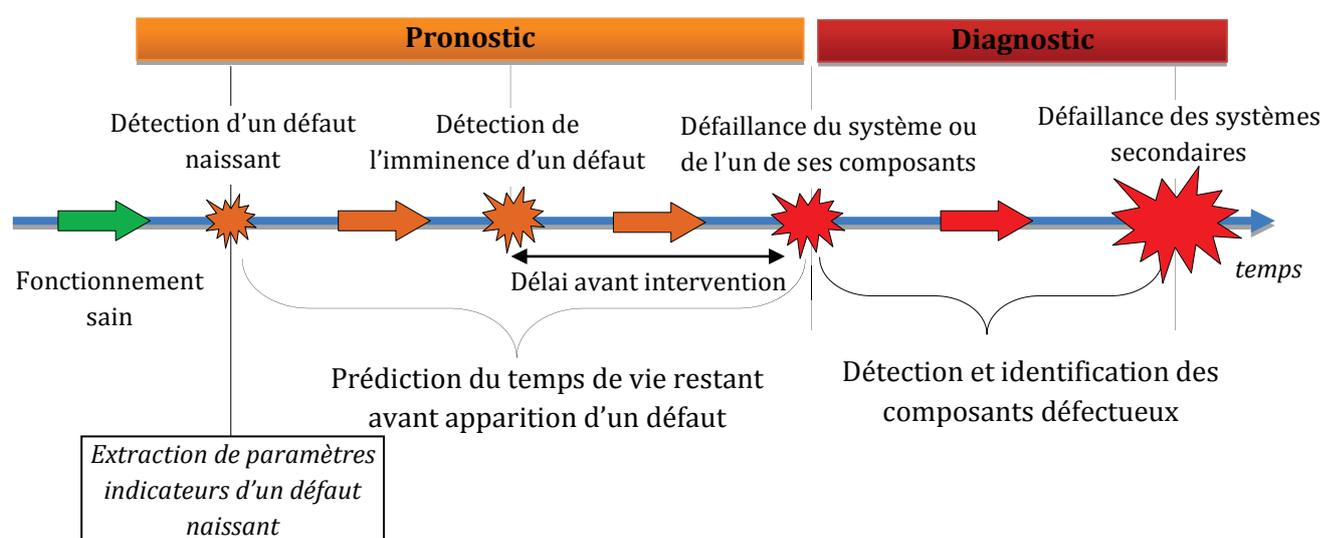


Figure 1. 15. Pronostic Vs diagnostic

Le processus de diagnostic intervient généralement au moment de l'apparition d'un défaut et/ou dans l'intervalle entre la défaillance du système et la défaillance des systèmes secondaires. Toutefois, si un défaut naissant peut être détecté à un stade précoce, alors, les opérations de maintenance pourront être retardées jusqu'à ce que l'état du système évolue vers un état plus dégradé. Cet intervalle, entre la détection d'un défaut naissant et son apparition, définit le domaine (temporel) du pronostic. Fournir un intervalle suffisant, communément appelé temps de vie restant, entre la détection du défaut naissant et la défaillance du système, permet de mieux exploiter le système et réduire d'avantage les opérations de maintenance.

Pour mettre à profit les avantages du pronostic, le personnel de la maintenance doit :

- Disposer de techniques permettant de détecter et d'identifier un défaut naissant ;
- Avoir une estimation fiable du temps qui reste avant l'intervention, à savoir le temps restant avant l'apparition d'un défaut.

Le terme « pronostic » trouve son origine dans le mot grec « progignôskein » qui signifie « savoir à l'avance ». Le terme « savoir à l'avance » ou pronostic peut être défini comme la capacité à prédire avec précision la durée de vie restante d'un composant défaillant du système ou sous-système, et aussi anticiper la prise de décision pour les opérations de maintenance.

L'organisation internationale de normalisation (ISO) 13381-1 [ISO13381] a défini le pronostic comme étant une estimation de la durée de vie restante avant l'apparition d'un défaut et la détection de l'imminence d'un ou plusieurs défauts. Le pronostic détermine si un défaut est imminent et estime le temps restant avant son apparition. Les progrès d'aujourd'hui dans la maintenance préventive conditionnelle ont contribué à améliorer le pronostic sur l'état futur des systèmes. Cette avancée permet non seulement de réduire les coûts liés à la maintenance mais aussi d'augmenter l'efficacité des interventions.

On trouve dans la littérature plusieurs méthodes pour la prédiction de l'état futur des systèmes. Catégoriser ces différentes méthodes dans des classes est assez difficile en raison d'une grande variété d'applications. Vachtsevanos *et al.* [Vacht_06] ont réussi à classer les méthodes liées au pronostic en trois grandes catégories, à savoir, le pronostic basé sur : (i) l'expérience, (ii) sur les données et (iii) sur un modèle. La figure 1.16 résume les diverses méthodes de pronostic qui peuvent être appliquées à différents systèmes et les coûts relatifs à leur exploitation et mise en œuvre.

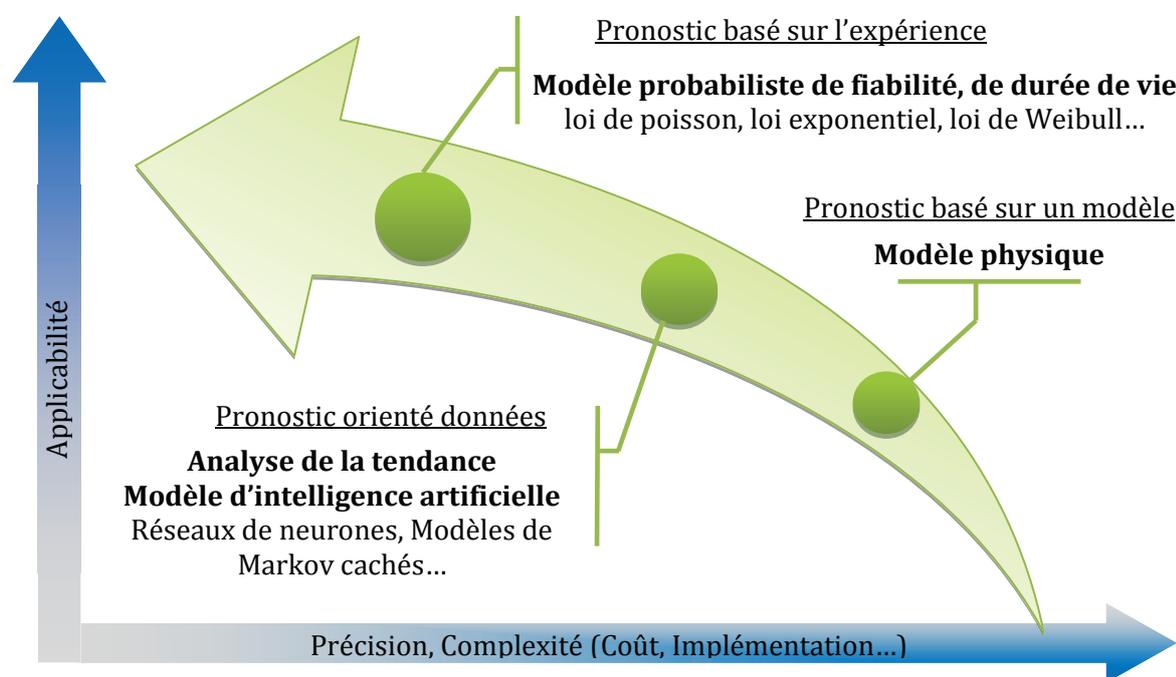


Figure 1. 16. Classification des méthodes de pronostic [Vacht_06]

1.6.1. Pronostic basé sur un modèle physique

Le pronostic basé sur un modèle physique utilise des représentations mathématiques pour intégrer une compréhension physique du processus de dégradation du système [Vacht_06]. Un schéma conceptuel du pronostic basé sur un modèle est représenté sur la figure 1.17. Des fonctions mathématiques, comme les équations différentielles sont développées pour représenter le système [Liu_09]. Ces équations sont ensuite utilisées dans des techniques d'estimations statistiques telles que les filtres de Kalman et les relations de parité dans le but d'estimer l'état du système. Les résidus qui sont la différence entre les prédictions du modèle et les observations du système sont ensuite utilisés pour détecter, isoler et prédire la dégradation du système [Vacht_06, Liu_09]. Les résidus montrent la divergence entre le modèle du système et les observations réelles. L'analyse statistique des résidus est utilisée pour la détection de la

dégradation dans le système suivant lequel la prédiction est réalisée afin d'estimer le temps restant avant l'apparition d'un défaut en fixant un seuil de défaut prédéfini.

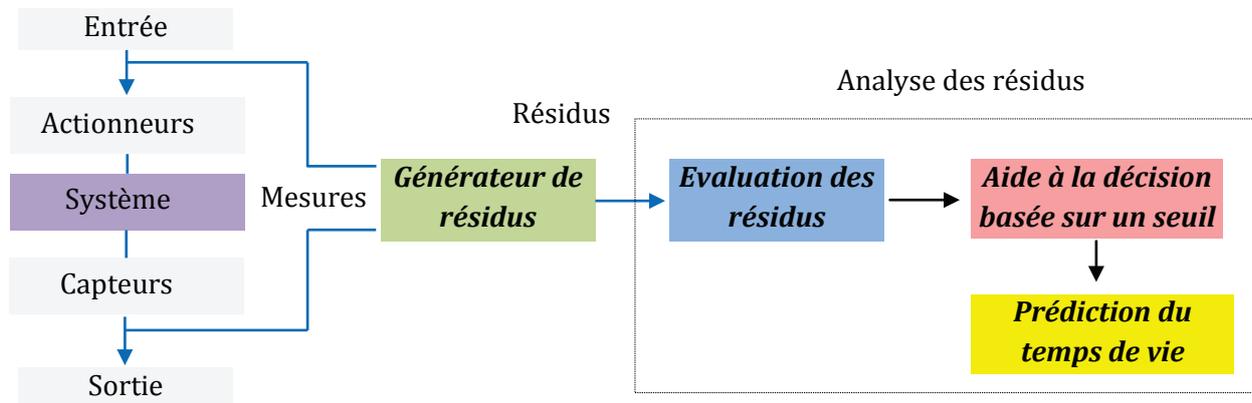


Figure 1. 17. Structure d'un système de pronostic à base d'un modèle

Le premier avantage d'un pronostic basé sur un modèle physique est qu'il ne nécessite pas de données historiques pour estimer l'état futur du système une fois que le modèle est connu. Un autre de ces avantages est qu'il permet de prendre en compte la dégradation causée par des conditions environnementales telles que les charges thermiques, l'humidité, les vibrations et les chocs. Un de ces inconvénients est qu'une représentation mathématique d'un système nécessite une connaissance approfondie des processus physiques sous-jacents qui conduisent à la défaillance du système [Liu_09]. Un autre de ces inconvénients est qu'il est difficile de créer des modèles représentant les multiples processus physiques qui se produisent dans le système [Vacht_06].

1.6.2. Pronostic orienté données

Le pronostic guidé par les données se base sur l'hypothèse que les données de surveillance d'un système restent relativement inchangées jusqu'à l'apparition d'un défaut. Dans cette approche, la surveillance du système est effectuée à l'aide de paramètres de surveillance qui indiquent la dégradation du système. Les données recueillies sont d'abord analysées pour détecter d'éventuelles anomalies. Ensuite, l'évolution de la tendance des paramètres de surveillance est utilisée pour estimer le temps restant avant la défaillance du système en utilisant des algorithmes de prédiction. L'organigramme présenté sur la figure 1.18 résume les étapes de la mise en œuvre d'un pronostic guidé par les données.

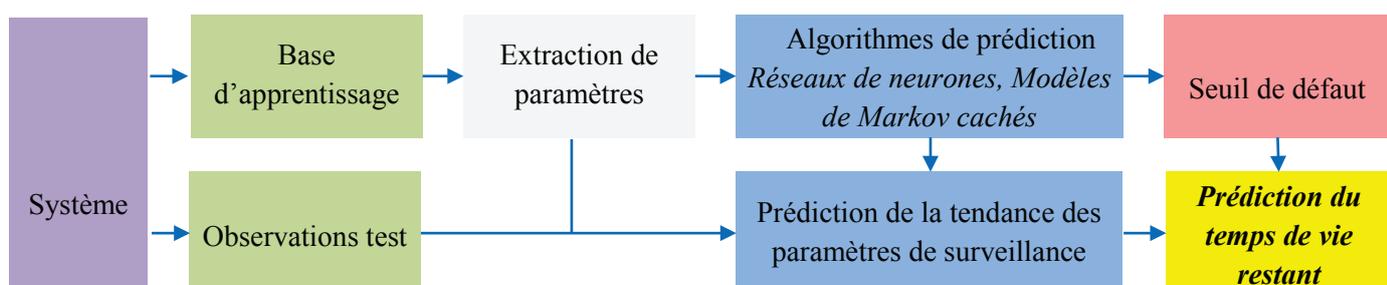


Figure 1. 18. Structure d'un système de pronostic orienté données

Un pronostic orienté données offre de nombreux avantages par rapport à un pronostic basé sur un modèle. Une approche orientée données apprend le comportement du système en se basant sur les données fournies par les capteurs placés sur le système et ne nécessite donc pas de connaissances spécifiques sur lui à des fins de modélisation. De plus, l'approche orientée données peut être appliquée sur des systèmes complexes, où un grand nombre de paramètres de surveillance peuvent être extraits en utilisant des techniques de traitement du signal. Il est également possible de réduire la dimension du problème en limitant les analyses aux paramètres qui contribuent à un comportement anormal dans le système.

1.6.3. Pronostic basé sur l'expérience

Le pronostic basé sur l'expérience se fonde sur la modélisation stochastique des phénomènes de dégradation (loi de poisson, loi exponentiel, loi de Weibull, etc.) à partir d'un ensemble de données recueilli d'un retour d'expérience sur une période de temps significative. Le principal avantage d'un pronostic basé sur l'expérience est qu'il se base sur des fonctions simples plutôt que sur des modèles mathématiques complexes. Il est donc relativement simple à mettre en œuvre et peu coûteux. Aussi, ce type de pronostic ne nécessite pas de connaissance approfondie sur le/les mécanisme(s) physique(s) de dégradation. Le principal inconvénient de cette approche est qu'il est difficile de disposer d'un historique d'expérience passé représentatif de toutes les conditions d'utilisation des systèmes. Les résultats de pronostic sont moins précis que ceux basés sur un modèle et sur les données, en particulier quand les conditions de fonctionnement sont variables.

Le tableau 1.3 permet de juger les trois approches de pronostic par rapport au type d'information nécessaire à leur mise en œuvre [Roemer_05].

Tableau 1. 3. Approches de pronostic et type d'information nécessaire

	Approche		
	Basée sur un modèle	Orientée données	Basée sur l'expérience
Modèle du système	<i>Nécessaire</i>	<i>Pas nécessaire</i>	<i>Pas nécessaire</i>
Historique des défaillances	<i>Utile</i>	<i>Nécessaire</i>	<i>Nécessaire</i>
Conditions passées	<i>Nécessaire</i>	<i>Nécessaire</i>	<i>Utile</i>
Conditions actuelles	<i>Nécessaire</i>	<i>Nécessaire</i>	<i>Utile</i>
Méthodes de reconnaissance des défaillances	<i>Nécessaire</i>	<i>Nécessaire</i>	<i>Pas nécessaire</i>
Historique de maintenance	<i>Utile</i>	<i>Pas nécessaire</i>	<i>Utile</i>
Capteur	<i>Nécessaire</i>	<i>Nécessaire</i>	<i>Pas nécessaire</i>

Il en sort de cette comparaison que l'avantage des méthodes de pronostic orientées données est leur capacité à apprendre le comportement d'un système uniquement à partir des caractéristiques de données surveillées sans nécessiter des connaissances spécifiques sur le système pour des fins de modélisation. Cependant, ces méthodes dépendent fortement des

données historiques du système (par exemple, l'apprentissage) afin de déterminer les corrélations, établir des modèles et évaluer les tendances des données menant à la défaillance.

1.7. Conclusion

Ce chapitre a passé en revue les méthodes actuellement utilisées dans le diagnostic et le pronostic des défauts dans les systèmes industriels. Le diagnostic implique la détection et l'analyse de la cause ou la nature du défaut, alors que le pronostic calcule ou prédit l'état futur à la suite d'une analyse sur des données pertinentes et disponibles. Une méthode de diagnostic et de pronostic efficace peut conduire à la réduction des arrêts imprévus et à l'augmentation de la fiabilité et la sûreté de fonctionnement des systèmes et par conséquent réduit les coûts de la maintenance. L'état de l'art proposé dans le premier chapitre permet de constater que plusieurs méthodes existent pour supporter l'activité de diagnostic et pronostic : méthodes basées sur un modèle, méthodes basées sur l'expérience, méthodes orientées données. Les comparaisons effectuées entre ces méthodes nous amène à privilégier les méthodes orientées données que ce soit pour le diagnostic (par classification supervisée et non supervisée) ou pour le pronostic (réseaux de neurones et modelés de Markov cachés). Même si les méthodes basées sur un modèle permettent d'obtenir de meilleurs résultats, nous ne souhaitons pas dépendre d'un modèle de diagnostic et de pronostic dont la couverture sera nécessairement limitée par leur capacité à intégrer une compréhension physique du processus de défaillance ou de dégradation du système.

L'efficacité de la maintenance préventive conditionnelle est liée à sa faculté à diagnostiquer et à prédire la défaillance d'un système. La phase de diagnostic est ainsi critique et le chapitre suivant lui est consacré.

Chapitre 2

Diagnostic par classification supervisée et non supervisée

Sommaire

2.1. Introduction	25
2.2. Détermination de l'espace de représentation	27
2.2.1. Critère de sélection (J)	29
2.3. Classification	31
2.3.1. La classification non supervisée	31
2.3.2. La classification supervisée	50
2.4. Evaluation des méthodes proposées	60
2.4.1. Algorithme fourmis artificielles	61
2.4.2. Modèles de Markov cachés	65
2.4.3. Comparaison des méthodes de classification	68
2.5. Conclusion	69

2.1. Introduction

Dans le chapitre précédant, nous sommes arrivés à la conclusion que les méthodes orientées données représentent l'une des meilleures alternatives dans le cas de la non disponibilité d'un modèle physique afin de diagnostiquer et pronostiquer l'état de santé du système. Dans ce chapitre, nous nous intéressons à des méthodes de diagnostic basées sur la classification d'une observation dans une classe représentant un mode de fonction du système. Cependant, l'application de la classification au diagnostic se heurte souvent au problème de la connaissance a priori du système étudié. En effet, il est rare de disposer, dans l'historique du système, d'observations (mesures) correspondant à tous les modes de fonctionnement possibles. En général, tout ou une partie des modes de fonctionnement qualifiés d'anormaux n'ont jamais été rencontrés sur le système. C'est dans l'objectif de résoudre cette problématique que ce chapitre présente deux nouvelles méthodes de diagnostic basées sur la classification supervisée et non supervisée. Ces méthodes auront l'avantage, en plus d'effectuer une classification avec un taux d'erreur très faible, de détecter l'apparition de nouveaux modes de fonctionnement.

Les méthodes de classification pour le diagnostic des défauts reposent sur le classement d'observations (objets ou formes) en les comparant à des formes-types. Une observation réalisée sur un système est caractérisée par un ensemble de q -paramètres. Cette observation est représentée par un point dans un espace à q -dimensions appelé *espace de représentation*, défini par les différents paramètres de surveillance. Ces paramètres sont extraits à partir d'analyses effectuées sur les signaux mesurés par les capteurs implantés sur le système (*Chapitre 1, section 1.4*). Les paramètres étant des nombres réels, une observation notée ' x_i ' est définie par un vecteur de paramètres $x_i = [x_{i1}, x_{i2}, \dots, x_{iq}]$ de \mathfrak{R}^q (voir figure 2.1).

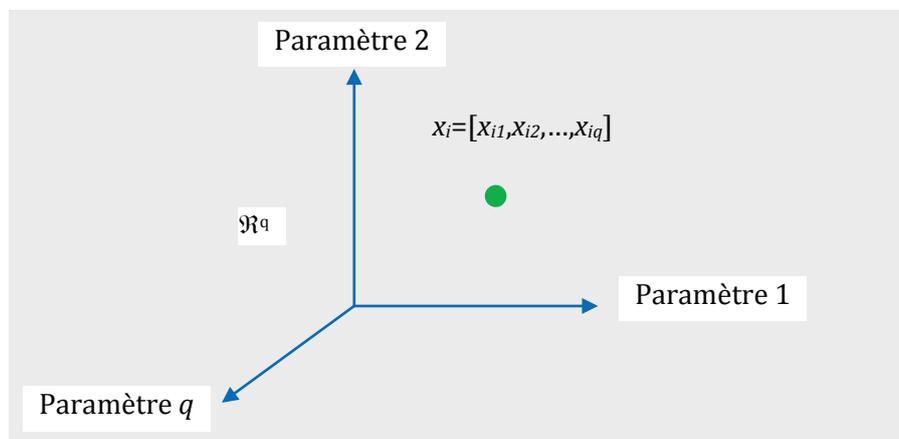


Figure 2. 1. Représentation d'une observation dans un espace à q -dimensions

L'objectif de la classification est d'associer une observation ' x_i ' à une forme type appelée *classe* (Ω_i). Une classe Ω_i $i \in \{1, \dots, M\}$ (voir figure 2.2) regroupe un ensemble d'observations similaires les unes aux autres. Si l'on se place dans le contexte du diagnostic, une classe Ω_i traduira un mode de fonctionnement du système. Dans le cas idéal, une classe qui regroupe plusieurs observations identiques devrait être représentée par un seul point dans l'espace de représentation. Néanmoins, des perturbations (bruit de mesure, précision des capteurs, ...etc.) peuvent apparaître. Ceci implique qu'une observation sera rarement identique à une autre observation d'une même classe. L'influence du bruit sur un ensemble de M -classes ($\Omega_1, \Omega_2, \dots, \Omega_M$)

représentant les M -modes de fonctionnement du système se traduit par des zones dans l'espace de représentation, regroupant des observations semblables (voir figure 2.2). L'objectif sera alors de définir à quel mode de fonctionnement correspond une nouvelle observation x_i recueillie sur le système.

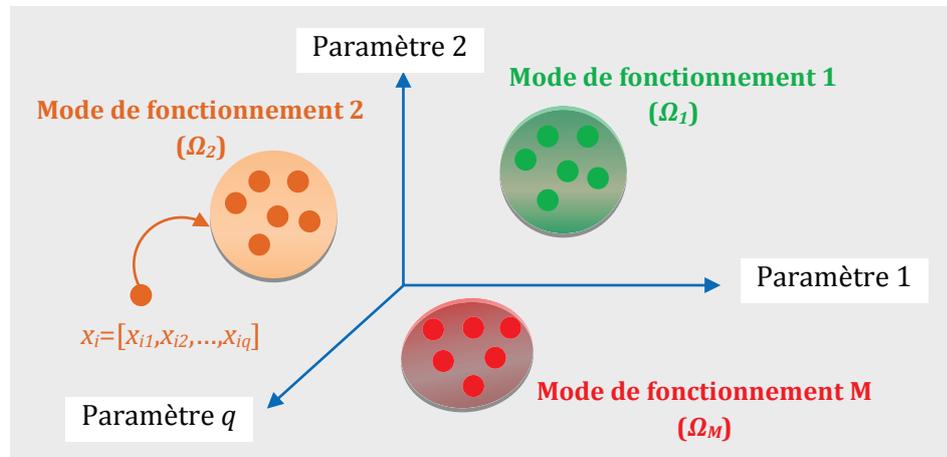


Figure 2. 2. Représentation d'un ensemble de classes dans un espace à q -dimensions

Les méthodes de classification présentées ici, sont habituellement associées à la reconnaissance des formes (RdF). Comme le montre la figure 2.3, un système complet de reconnaissance des formes est associé à plusieurs modules. En faisant abstraction du module d'acquisition et du module de prétraitement, un système de diagnostic par classification est associé à trois modules : construction du vecteur de paramètres, extraction/sélection de paramètres, et classification.

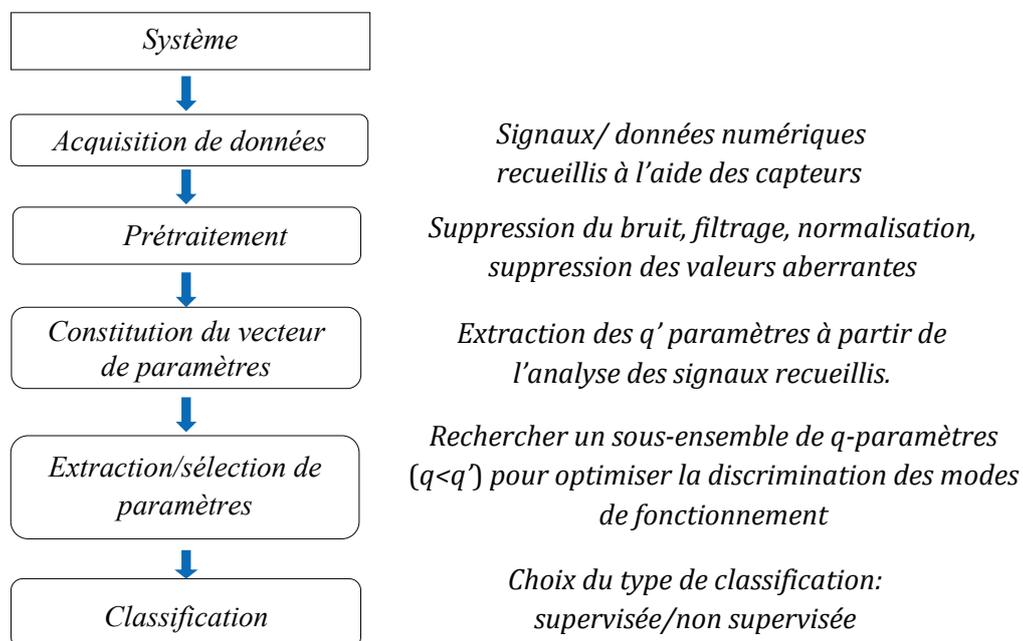


Figure 2. 3. Eléments d'un système de reconnaissance des formes

- **Construction du vecteur de paramètres** : il s'agit, dans ce module de construire un vecteur de paramètres sensible aux défauts du système pour une meilleure discrimination des classes.

Le vecteur de paramètres est constitué d'un ensemble initial de q' -paramètres obtenu en utilisant les différentes techniques d'analyse (analyse temporelle, fréquentielle, temps-fréquence) des signaux de mesure recueillis à partir des capteurs placés sur le système. Ceci permet de caractériser chaque observation à un instant donné par l'ensemble des paramètres.

- **Extraction/sélection de paramètres** : ce module a pour objectif la détermination d'un *espace de représentation* en extrayant à partir d'un ensemble initial de q' -paramètres un sous-ensemble de q -paramètres ($q < q'$) pertinents pour la classification. Les observations contiennent un certain nombre de paramètres redondants ou non pertinents. Les paramètres redondants sont ceux qui ne fournissent pas plus d'informations que les paramètres sélectionnés, et les paramètres non pertinentes ne donnent aucune information utile dans n'importe quel contexte. La détermination de l'espace de représentation peut s'effectuer soit en transformant l'espace initial par une combinaison linéaire des paramètres initiaux, comme l'analyse en composantes principales (ACP) ou alors, par une procédure de sélection qui recherche un sous ensemble de paramètres selon un critère de séparabilité entre les classes et de compacité dans chacune d'elles.

- **Classification** : comme le montre la figure 2.4, la classification peut être considérée comme une tâche d'analyse intelligente d'observations qui peut s'effectuer à travers deux techniques de classification, à savoir, la classification supervisée et non-supervisée. Si l'on dispose d'observations déjà recueillies sur le système (ensemble d'apprentissage) et pour lesquelles on connaît a priori la classe d'origine, c'est-à-dire son mode de fonctionnement correspondant, on parle de classification en *mode supervisé*. Dans le cas contraire, on parle de classification en *mode non supervisé*.

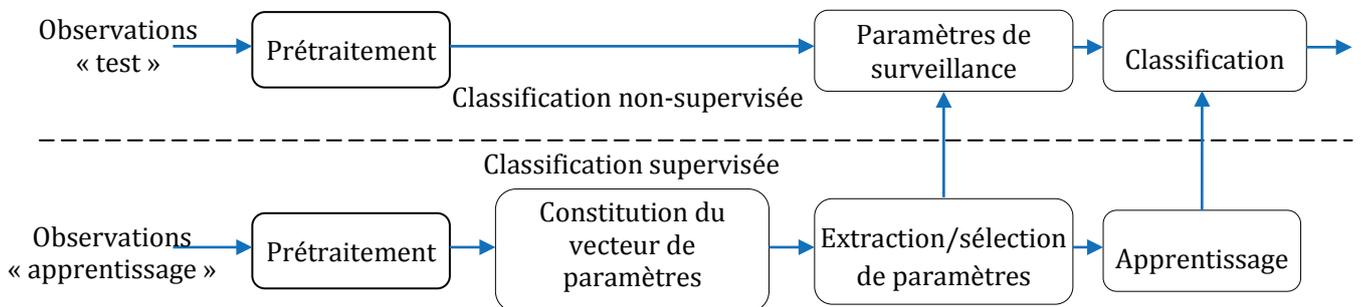


Figure 2. 4. Structure d'une classification supervisée et non-supervisée

Dans la section suivante, une introduction sur la nécessité de sélectionner des paramètres pertinents pour la classification est présentée. Un algorithme sera proposé pour la sélection des paramètres.

2.2. Détermination de l'espace de représentation

La détermination de l'espace de représentation consiste à extraire, à partir d'un ensemble initial de q' -paramètres, un sous-ensemble de q -paramètres ($q < q'$) afin d'obtenir la meilleure représentation possible de l'ensemble des M -classes. Cette réduction peut être réalisée soit par des méthodes de sélection de paramètres soit par des méthodes d'extraction de paramètres.

Les méthodes de sélection de paramètres sont utilisées pour l'identification des paramètres les plus informatifs. Cette procédure supprime certains paramètres non informatifs ou bruyants et réduit la dimension du problème [John_94, Kohavi_97]. Les méthodes de sélection de paramètres nécessitent une connaissance a priori (connaître les classes existantes ainsi que les

observations correspondantes) sur les classes d'observations pour être utilisées de la manière la plus efficace possible. Ces méthodes peuvent être basées sur différentes techniques. Par exemple, certaines procédures de sélection de paramètres sont basées sur des algorithmes d'apprentissage statistiques comme la machine à vecteurs de support [Bradley_98]. Les méthodes d'extraction de paramètres utilisent une combinaison linéaire de l'ensemble des paramètres initiaux plutôt que la sélection. Cette procédure transforme un ensemble de paramètres en un plus petit ensemble, ce qui est souvent plus efficace pour les algorithmes de classification. Parmi les méthodes d'extraction de paramètres, on distingue l'Analyse en Composantes Principales (ACP) ou la transformée de Karhunen Loève discrète. Une explication plus détaillée sur l'ACP est donnée dans la section 2.3.1.4.3.

Dans notre travail, nous nous intéresserons aux méthodes de sélection de paramètres dites sous-optimales, qui présentent l'avantage d'être rapide et de fournir des résultats réalistes. On peut citer entre autres :

- La méthode S.F.S (Sequential Forward Selection) : cette méthode effectue une sélection de paramètres de manière successive. Le vecteur de paramètres est initialisé comme un ensemble vide. À chaque itération, on rajoute le paramètre maximisant le plus un critère noté J (voir section 2.2.1).
- La méthode G.S.F.S (Generalised Sequential Forward Selection) : cette méthode reprend le même principe que la méthode S.F.S, mais cette fois-ci, plusieurs paramètres sont sélectionnés à chaque itération.
- La méthode S.B.S (Sequential Backward Selection) : cette méthode débute par un vecteur de paramètres constitué d'un ensemble de paramètres initiaux. À chaque itération, elle élimine le paramètre pénalisant le plus le critère J .
- La méthode G.S.B.S (Generalised Sequential Backward Selection) : c'est la méthode symétrique de la méthode G.S.F.S. Elle utilise la même procédure que le S.B.S mais en éliminant plus d'un paramètre à chaque itération.

Dans [Kudo_00], les auteurs présentent les avantages et les inconvénients des méthodes de sélection citées précédemment. En accord avec les auteurs, nous utiliserons, dans le cadre de nos travaux, la méthode S.B.S pour sélectionner les q -paramètres portant le plus d'informations.

Le SBS est la plus simple méthode de sélection de paramètres [Doak_92]. Elle part d'un ensemble initial de paramètres et supprime de manière séquentielle un paramètre à la fois en se basant sur un critère J .

Mathématiquement, le SBS peut être formulé comme suit :

Soit Y_0 un vecteur initial de paramètres de dimension q' , $Y_0 = (Y_1, Y_2, \dots, Y_{q'})$ et soit Y le vecteur de paramètres après sélection, de dimension q , $Y = (Y_1, Y_2, \dots, Y_q)$, $Y \subseteq Y_0$. L'objectif est d'optimiser $J(Y)$ qui est un critère de sélection utilisé pour obtenir le vecteur Y . Ceci est décrit par l'algorithme suivant :

Algorithme 2. 1. Sequential Backward Selection (SBS)

- (1) Débuter avec un vecteur initial de paramètres $Y_0 = (Y_1, Y_2, \dots, Y_{q'})$
 - (2) Identifier le paramètre Y_i qui minimise le critère $J(Y_i) : i = \underset{i \in \{1, \dots, q'\}}{\operatorname{argmin}} [J(Y_i)]$
 - (3) $Y_0 \leftarrow Y_0 - \{Y_i\}$
 - (4) Répéter l'étape (2) et (3) jusqu'à ce que le nombre de paramètres q soit atteint
 - (5) $Y \leftarrow Y_0$
-

Le principal inconvénient de l'algorithme SBS est qu'il ne prend pas en considération, lors de sélection, toutes les combinaisons possibles entre les paramètres. Le calcul du critère $J(Y)$, sur toutes les q -combinaisons de paramètres possibles, impose un nombre de solutions égale à :

$$C_{q'}^q = \frac{q!}{(q'-q)! \cdot q!} \quad (2.1)$$

Cette solution est vite dissuasive si la dimension de départ est élevée. Cependant, il est possible de réduire le nombre de solutions si on modifie le SBS. L'idée consiste à sélectionner un nombre q'' ($q < q'' < q'$) de paramètres (les plus pertinents) en se basant sur le critère J , puis sélectionner une combinaison de q -paramètres parmi les q'' -paramètres qui maximise le critère J . Cette stratégie permet de réduire considérablement le nombre de solutions tout en prenant en considération toutes les combinaisons entre les paramètres.

Le nombre de solutions dans ce cas devient égale à :

$$C_{q''}^q = \frac{q!}{(q''-q)! \cdot q!} \quad (2.2)$$

L'algorithme 2.1 devient alors :

Algorithme 2. 2. Optimized Sequential Backward Selection (OSBS)

- (1) Débuter avec un vecteur initial de paramètres $Y_0 = (Y_1, Y_2, \dots, Y_{q'})$
 - (2) Identifier le paramètre Y_i qui minimise le critère $J(Y_i) : i = \text{argmin} [J(Y_i)]_{i \in \{1, \dots, q'\}}$
 - (3) $Y_0 \leftarrow Y_0 - \{Y_i\}$
 - (4) Répéter l'étape (2) et (3) jusqu'à ce que le nombre de paramètres q'' soit atteint
 - (5) $Y' \leftarrow Y_0$
 - (6) Recommencer avec un nouveau vecteur de paramètres $Y' = (Y_1, Y_2, \dots, Y_{q''})$
 - (7) Rechercher un sous-ensemble de q -paramètres $Y = (Y_1, Y_2, \dots, Y_q)$ parmi les q'' -paramètres qui maximise le critère $J(Y)$
-

2.2.1. Critère de sélection (J)

Le critère J utilisé pour la sélection du sous ensemble de q -paramètres doit fournir une dispersion maximale entre les différentes classes (notion de séparabilité) ainsi qu'une faible dispersion des observations dans chaque classe (notion de compacité). Ceci correspond à :

- Une minimisation de la variance intra-classe ;
- Une maximisation de la variance inter-classes.

Les matrices de variance intra-classe et inter-classes traduisent ces deux conditions. La matrice de variance intra-classe fait intervenir les observations de chaque classe ainsi que leurs centres de gravité :

$$\Sigma_w = \frac{1}{N} \sum_{i=1}^M \sum_{j=1}^{n_i} (x_{ij} - g_i) \cdot (x_{ij} - g_i)' \quad (2.3)$$

La matrice de dispersion inter-classes fait intervenir les centres de gravité des classes :

$$\Sigma_B = \frac{1}{N} \sum_{i=1}^M (g_i - g) \cdot (g_i - g)' \quad (2.4)$$

avec :

- g : centre de gravité de l'ensemble des observations ;
- M : nombre de classes ;
- g_i : centre de gravité de la classe Ω_i ;
- x_{ij} : $j^{\text{ème}}$ observation de la classe Ω_i ;
- n_i : nombre d'observations de la classe Ω_i ;
- N : nombre total de l'ensemble d'observations.

Si on considère les matrices de dispersion intra-classe et inter-classes, alors nous pouvons définir deux critères qui permettent à la fois de tenir compte de la notion de compacité et de séparabilité entre les classes :

$$J_1 = \text{trace}(\Sigma_W^{-1} \cdot \Sigma_B) \quad (2.5)$$

$$J_2 = \frac{\det(\Sigma_W + \Sigma_B)}{\det(\Sigma_W)} \quad (2.6)$$

avec : $\text{trace}(\Sigma_W^{-1} \cdot \Sigma_B)$ la somme des vecteurs propres de la matrice $(\Sigma_W^{-1} \cdot \Sigma_B)$

Plus les critères (J_1 et J_2) sont grands, plus les classes sont séparées.

Il existe aussi un autre critère appelé « critère de Fisher » pour la sélection de paramètres. Le critère de Fisher, dans le cas d'un problème à deux classes, s'exprime comme suit :

$$J_3(Y) = \frac{g_1(Y) - g_2(Y)}{n_1 \sigma_1^2(Y) - n_2 \sigma_2^2(Y)} \quad (2.7)$$

Le numérateur de l'expression (2.7) traduit la séparation des classes alors que le dénominateur reflète la compacité des classes. Plus le critère est grand, plus les classes sont séparées, avec :

- $g_i(Y)$ ($i = 1, 2$), centre de gravité de la classe Ω_i en ne considérant que le sous-ensemble de paramètres Y .

$$g_i(Y) = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}(Y) \quad (2.8)$$

- $\sigma_i^2(Y)$: variance de la composante Y des vecteurs de la classe Ω_i .

$$\sigma_i^2(Y) = \frac{1}{n_i} \sum_{j=1}^{n_i} [x_{ij}(Y) - g_i(Y)]^2 \quad (2.9)$$

avec x_{ij} : $j^{\text{ème}}$ observation de la classe Ω_i et n_i : nombre d'observations de la classe Ω_i .

Dans le cas général d'un problème à M classes, $J_3(Y)$ s'exprime :

$$J_3(Y) = \sum_{i=1}^M \sum_{r=1}^{M-1} \frac{g_i(Y) - g_r(Y)}{n_i \sigma_i^2(Y) - n_r \sigma_r^2(Y)} \quad (2.10)$$

Dans la suite de notre travail, l'expression J_1 (2.5) sera utilisée comme critère dans notre méthode de sélection. Il sera noté J .

2.3. Classification

2.3.1. La classification non supervisée

La classification non supervisée, appelée aussi « clustering », est considérée comme un outil qui analyse les observations (données de mesure) inconnues et inexplorées. Il met en évidence des formes cachées en exploitant une certaine corrélation entre les différents paramètres de surveillance en se base sur *une mesure de similarité* entre les observations projetées dans l'espace de représentation.

2.3.1.1. Mesure de similarité

Soit $d()$ une métrique qui calcule la similarité entre trois observations x_1, x_2, x_3 , définies par un vecteur à q -paramètres $x_i = [x_{i1}, x_{i2}, \dots, x_{iq}]$, $1 \leq i \leq 3$. Les observations x_1, x_2 sont dites similaires si et seulement si : $d(x_1, x_2) \ll d(x_1, x_2)$ et $d(x_1, x_2) \ll d(x_1, x_3)$.

Le choix d'une métrique ouvre une large classe de mesures de similarité. Le premier à prendre en compte est la distance de *Minkowsky*, qui est considérée comme l'une des métriques les plus utilisées dans la classification non supervisée.

La distance de *Minkowsky* (d_{Mink}) est définie par :

$$d_{Mink}(x_1, x_2) = \left(\sum_{n=1}^q |x_{1n} - x_{2n}|^m \right)^{\frac{1}{m}} \quad (2.11)$$

La variable m , $m \in \mathbb{N}$, est le paramètre de *Minkowski* utilisé pour ajuster la nature de la métrique. La distance *Minkowsky* peut être considérée comme une distance euclidienne (d_{euc}) si $m=2$ dans la formule (2.11). La distance euclidienne est donnée comme suit :

$$d_{Euc}(x_1, x_2) = \left(\sum_{n=1}^q |x_{1n} - x_{2n}|^2 \right)^{\frac{1}{2}} \quad (2.12)$$

Si le paramètre de *Minkowski* (m) est égal à 1, la formule (2.11) se traduit par la distance de *Manhattan*. La distance de *Manhattan* (d_{Man}) est exprimée par :

$$d_{Man}(x_1, x_2) = \sum_{n=1}^q |x_{1n} - x_{2n}| \quad (2.13)$$

La distance de *Mahalanobis* prend les covariances en compte, ce qui permet de prendre en compte la forme elliptique d'une distribution d'observations (voir figure 2.5). La distance de *Mahalanobis* (d_{Mahal}) est exprimée par :

$$d_{Mahal}(x_1, x_2) = \sqrt{(x_1 - x_2) Cov^{-1} (x_1 - x_2)'} \quad (2.14)$$

où Cov^{-1} désigne la matrice de covariance.

La figure 2.5 montre les contours des métriques : *Euclidienne*, *Manhattan* et *Mahalanobis*. D'ordinaire, la distance *Euclidienne* et la distance de *Manhattan* ne sont pas adaptées pour des classes ayant une forme allongée ou lorsqu'elles sont trop rapprochées les unes des autres. Les contours de la distance de *Mahalanobis* sont des ellipses qui permettent de s'ajuster au mieux à l'enveloppe des observations d'une même classe alors que pour la distance *Euclidienne*, les contours sont des cercles concentriques autour du centre de gravité. La distance de *Mahalanobis* détermine dans quelle mesure une observation, x_i , s'inscrit dans un ensemble de distribution d'observations, compte tenu de la forme et la taille de la distribution.

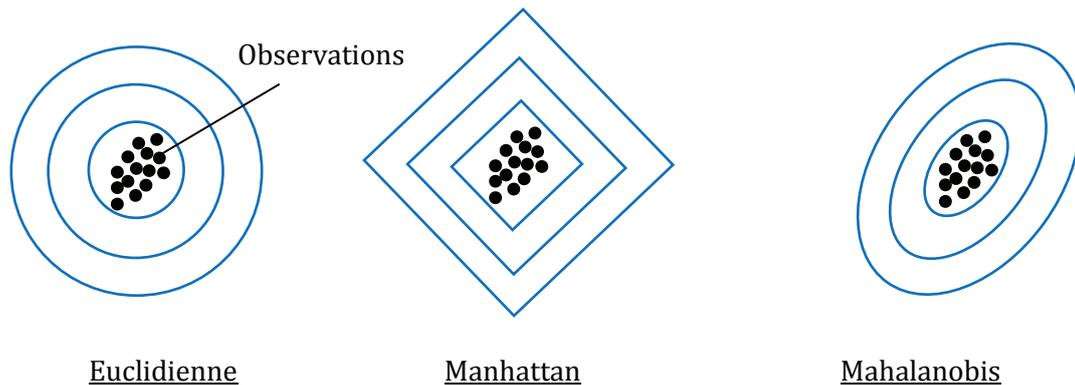


Figure 2. 5. Contours des métriques : Euclidienne, Manhattan et Mahalanobis dans un espace 2D

Les courbes grises dans la figure 2.6 représentent les contours de la distance *Euclidienne* (à gauche) et de la distance de *Mahalanobis* (à droite) dessinés par rapport à deux classes, Ω_1 (*) et Ω_2 (*). Les lignes en noire représentent les limites entre les deux classes. Il faut noter que la limite tracée par la distance de *Mahalanobis* permet une meilleure séparation des classes. Dans la suite de notre travail, nos méthodes de classification seront basées sur la distance de *Mahalanobis*.

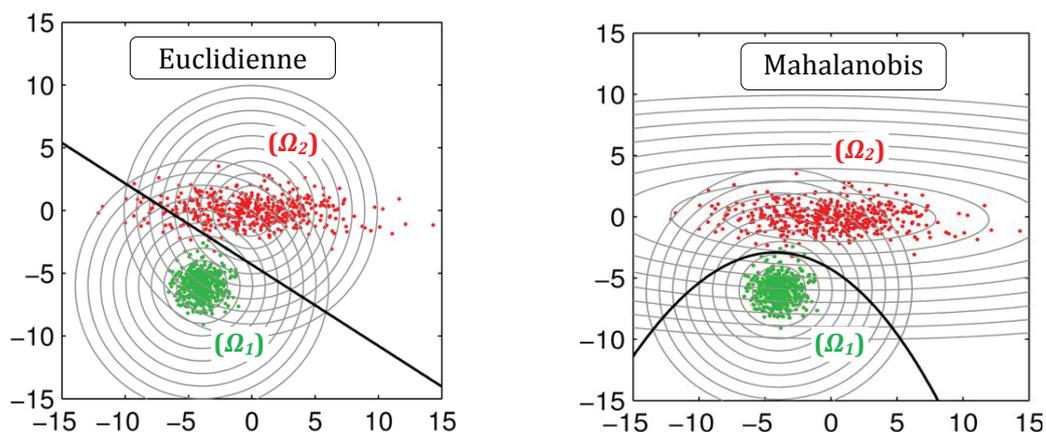


Figure 2. 6. Frontières déclinées par les distances de Mahalanobis et Euclidienne

Il existe plusieurs méthodes de classification non supervisée, également appelées méthodes de coalescence. Si on se limite à l'affectation exclusive, ces méthodes se décomposent en deux parties [Jain_99] :

- Les méthodes de classification hiérarchiques (ascendantes ou descendantes) ;
- Les méthodes de classification par partition.

2.3.1.2. Classification hiérarchique

La classification hiérarchique produit un ensemble hiérarchique de classes selon un critère de liaison. Les classes, qui sont imbriquées, sont liées de manière récursive soit en mode agglomération (Classification Ascendante Hiérarchique CAH) ou en mode division (Classification Descendante Hiérarchique CDH).

Dans la classification ascendante hiérarchique, les classes sont déterminées de manière itérative en quatre étapes :

- Etape 1 : l'algorithme de classification commence par considérer chaque observation comme une classe. La phase d'initialisation est définie par la variable :

$$\Omega_i^\ell = \{x_i\} \quad \forall i \text{ et } \ell = 0 \quad (2.15)$$

où la variable Ω_i^ℓ dénote la classe d'appartenance de l'observation x_i à la $\ell^{\text{ième}}$ itération, avec $\ell = 0$ pour la phase d'initialisation.

- Etape 2 : la similarité entre deux classes est définie par un critère de liaison. Ce critère évalue deux classes à l'aide d'une mesure de similarité (voir section 2.3.1.1) qui évalue la similarité entre deux classes. La similarité « *Sim* » entre deux classes est définie comme suit :

$$Sim(\Omega_l^\ell, \Omega_h^\ell) = \beta(\Omega_l^\ell, \Omega_h^\ell) \quad (2.16)$$

où le critère de liaison est représenté par la fonction β . β peut être un critère de liaison par saut minimal, saut maximal, distance moyenne ou par l'inertie intra-classe représentée par le critère de Ward. La formule (2.16) indique que toutes les observations de la classe (l) doivent être comparées à toutes les observations de la classe (h). Le critère de liaison β retourne une similarité entre les deux classes.

- Etape 3 : sélectionner les paires de classes qui ont une très grande similarité $Sim(\Omega_l^\ell, \Omega_h^\ell)$.
Soit $\Omega_{l^*}^\ell$ et $\Omega_{h^*}^\ell$ deux classes sélectionnées comme suit :

$$[\Omega_{l^*}^\ell, \Omega_{h^*}^\ell] = \underset{l, h \in [1-M]; l \neq h}{arg \max} Sim(\Omega_l^\ell, \Omega_h^\ell) \quad (2.17)$$

- Etape 4 : fusionner les deux classes avec un indice l^* et h^*

$$\Omega_{l^*}^{\ell+1} = (\Omega_{l^*}^\ell \cup \Omega_{h^*}^\ell) \quad (2.18)$$

$$\Omega_{h^*}^{\ell+1} = \{ \} \quad (2.19)$$

Les étapes 2 à 4 sont répétées jusqu'à ce que toutes les classes soient fusionnées en une seule classe. Dans le cas standard, l'agrégation s'arrête lorsqu'un critère d'arrêt est atteint. Dans notre cas, le critère d'arrêt correspond à une dispersion maximale entre les différentes classes ainsi qu'à une faible dispersion des observations dans chaque classe. Les séquences de fusion des classes peuvent être représentées graphiquement par un diagramme en arbre, appelé dendrogramme. Comme le montre la figure 2.7, le dendrogramme est une hiérarchie à plusieurs niveaux avec la fusion de deux classes à chaque niveau.

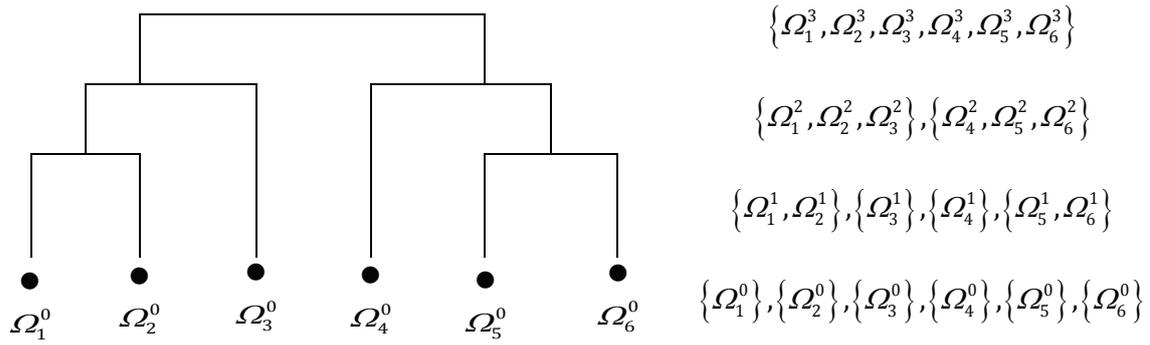


Figure 2. 7. Exemple d'un dendrogramme

2.3.1.2.1. Critère de liaison

2.3.1.2.1.1. Critère de liaison par saut minimal

La fusion des classes est basée sur la similarité entre elles. Cette similarité est calculée à partir d'un critère de liaison. Le critère de liaison par saut minimal $\beta_{saut-min}$ est le premier critère qui peut être pris en compte, il cherche la plus grande similarité entre deux observations x_i et x_j qui se situent dans les classes Ω_l^ℓ et Ω_h^ℓ . Il est exprimé comme suit :

$$\beta_{saut-min}(\Omega_l^\ell, \Omega_h^\ell) = \max \{d^{-1}(x_i, x_j)\} \quad \forall x_i \in \Omega_l^\ell, x_j \in \Omega_h^\ell \text{ et } l \neq h \quad (2.20)$$

La variable x_i désigne l'*i*ème observation et x_j désigne la *j*ème observation. Les distances indiquées dans la section (2.3.1.1) ou toute autre distance dans la littérature peuvent être utilisées pour le calcul du critère de liaison. Cependant, le critère de liaison par saut minimal est mal adapté dans le cas où les classes ont une forme étirée ce qui a pour inconvénient de produire des effets de chaîne comme le montre la figure 2.8.

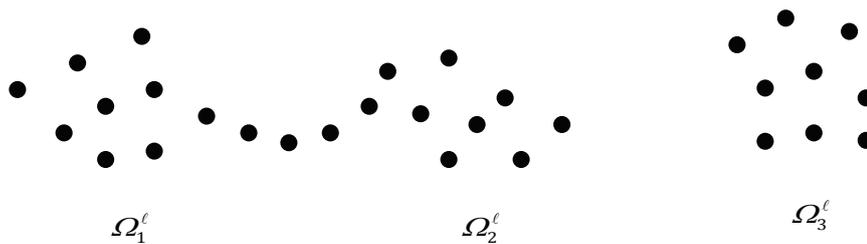


Figure 2. 8. Effet de chaîne

2.3.1.2.1.2. Critère de liaison par saut maximal

Le critère de liaison par saut maximal $\beta_{saut-max}$ calcule une dissimilarité entre les classes en se basant sur la distance maximale entre deux observations qui se situent dans les classes Ω_l^ℓ et Ω_h^ℓ . Il est exprimé comme suit :

$$\beta_{saut-max}(\Omega_l^\ell, \Omega_h^\ell) = \min \{d^{-1}(x_i, x_j)\} \quad \forall x_i \in \Omega_l^\ell, x_j \in \Omega_h^\ell \text{ et } l \neq h \quad (2.21)$$

2.3.1.2.1.3. Critère de liaison par distance moyenne

Le critère de liaison par distance moyenne $\beta_{saut-moy}$ calcule la similarité entre deux classes en trouvant la distance moyenne entre toutes les paires d'observations se situant dans les classes Ω_l^ℓ et Ω_h^ℓ . Il est exprimé comme suit :

$$\beta_{dist-moy}(\Omega_l^\ell, \Omega_h^\ell) = \frac{1}{|\Omega_l^\ell|} \frac{1}{|\Omega_h^\ell|} \sum_{x_i \in \Omega_l^\ell} \sum_{x_j \in \Omega_h^\ell} d(x_i, x_j), \quad l \neq h \quad (2.22)$$

$|\Omega_l^\ell|$ et $|\Omega_h^\ell|$ désignent respectivement le nombre d'observations dans la classe Ω_l^ℓ et Ω_h^ℓ . Ce critère peut être considéré comme un compromis entre $\beta_{saut-min}$ et $\beta_{saut-max}$.

2.3.1.2.1.4. Critère de Ward

Le critère de Ward β_{ward} correspond à l'inertie intra-classe qui résulte de la fusion de Ω_l^ℓ et Ω_h^ℓ . Il est exprimé comme suit :

$$\beta_{ward}(\Omega_l^\ell, \Omega_h^\ell) = \frac{|\Omega_l^\ell| \cdot |\Omega_h^\ell|}{|\Omega_l^\ell| + |\Omega_h^\ell|} d_{Euc}^2(g_l, g_h), \quad l \neq h \quad (2.23)$$

où g_l et g_h sont les centres de gravité des classes Ω_l^ℓ et Ω_h^ℓ , $d_{Euc}^2(g_l, g_h)$ est le carré de la distance Euclidienne entre les centres g_l et g_h .

2.3.1.3. Classification par partition

La classification par partition subdivise un ensemble d'observations en plusieurs sous-ensembles. Chaque sous-ensemble représente une classe contenant au moins une observation. Chaque classe est représentée par une fonction (F) qui affecte les observations à ces classes en fonction leurs similarités.

Soit $\{\Omega_1^\ell, \Omega_2^\ell, \dots, \Omega_M^\ell\}$ un ensemble de M -classes. La fonction (F) qui affecte l'ensemble des N -observations x_i ($i \in \{1, \dots, N\}$) dans les M -classes est donnée comme suit :

$$F(x_i) = \arg \min_{1 \leq k \leq M} \{d(x_i, g_k)\} \quad (2.24)$$

où la métrique $d(x_i, g_k)$ représente une mesure de similarité entre l' i ème observation x_i et le centre de gravité de la classe Ω_k^ℓ ($k \in \{1, \dots, M\}$). La variable ℓ désigne la ℓ ème itération.

L'algorithme K-means de [Forgy_65] est l'algorithme le plus utilisé dans la classification par partition. Il comporte quatre étapes :

- **Etape 1** : choisir un nombre de M -observations parmi l'ensemble des N -observations à partitionner et les considérer comme étant les centres de gravité des M -classes.
- **Etape 2** : assigner chaque observation x_i à son plus proche centre de gravité en utilisant la formule (2.24).

- **Etape 3** : lorsque toutes les observations sont affectées, recalculer les centres de gravité des M -classes par :

$$g_k = \frac{1}{|\Omega_k^\ell|} \sum_{x_i \in \Omega_k^\ell} x_i, \quad 1 \leq k \leq M \quad (2.25)$$

où $|\Omega_k^\ell|$ désigne le nombre d'observations dans la classe Ω_k^ℓ .

- **Etape 4** : répéter les étapes 2 et 3 jusqu'à ce que les centres de gravité des M -classes ne varient plus.

Un exemple montrant l'exécution de l'algorithme K-means est donné dans la figure 2.9. A chaque itération, les observations sont attribuées au centre de gravité le plus proche (montrées en couleur, les observations prennent la même couleur que le centre de gravité de la classe à laquelle elles sont attribuées). Le centre de gravité de chaque classe est ensuite recalculé en fonction de la moyenne des observations attribuées à celle-ci. Le partitionnement final nous donne une bonne séparation entre les deux classes en donnant une couleur distincte pour chacune d'entre elles.

Les observations sont représentées par des points et les centres de gravité des deux classes sont représentés par des croix. (a) représente l'ensemble d'observations. (b) représente des centres de gravité choisis aléatoirement. (c) et (d) illustrent la distribution de l'ensemble des observations après quatre et dix itérations de l'algorithme K-means.

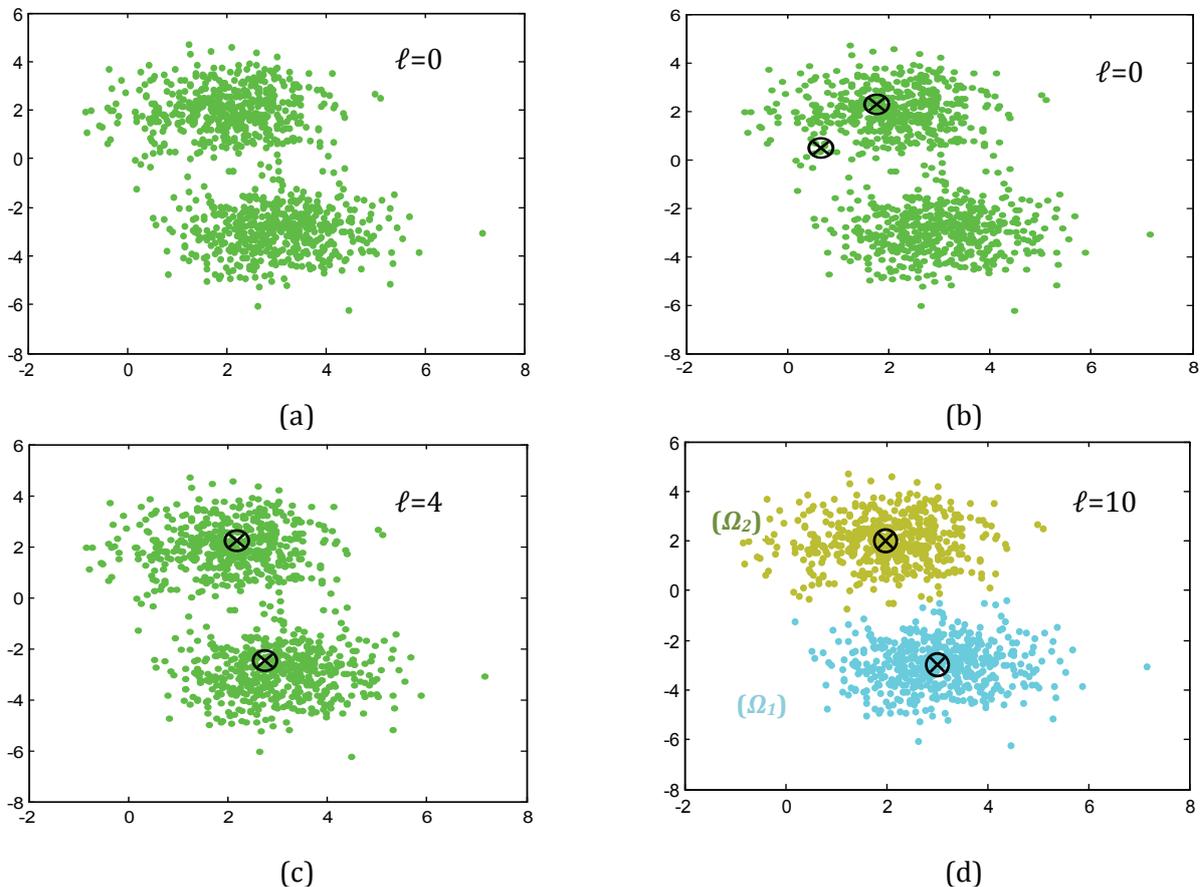


Figure 2. 9. Algorithme K-means pour le partitionnement de deux classes

La prochaine section aborde une approche heuristique qui a trouvé son inspiration dans le comportement des fourmis réelles. Cette technique servira par la suite à l'élaboration de l'algorithme Fourmis Artificielles (FA) permettant de résoudre le problème de la classification mentionné précédemment.

2.3.1.4. Approche heuristique pour l'optimisation de la classification non supervisée

Les sociétés d'insectes sociaux comme les colonies de fourmis sont des systèmes distribués qui, malgré la simplicité de leurs individus, présentent une organisation sociale très structurée. Cette organisation permet aux colonies de fourmis d'accomplir des tâches complexes qui dans certains cas, dépassent la capacité individuelle d'une seule personne. Le comportement le plus reconnu chez les fourmis est leur capacité à travailler en groupe afin de développer une tâche qui ne peut être réalisée par une seule fourmi « agent ». La capacité des fourmis est le résultat d'effets coopératifs. Une coopération ordonnée d'un ou plusieurs agents est plus élevée que le total de leurs effets individuels. Certains chercheurs ont obtenu des résultats prometteurs dans l'exploration de données en s'inspirant du comportement des colonies de fourmis. Le nombre élevé d'agents et la coordination de leurs tâches (exécutés simultanément) procurent aux colonies de fourmis des niveaux élevés de parallélisme et d'auto-organisation. Ces caractéristiques sont souhaitées dans les techniques d'optimisation actuelles [Boryczka_09].

Nous avons construit un nouvel algorithme de classification non supervisée basé sur le comportement d'une colonie de fourmis pour l'exploration et l'analyse d'observations en raison de plusieurs facteurs. Notre approche heuristique améliore les performances des techniques de classification classiques, et qui dans notre cas sont la classification ascendante hiérarchique et la classification par partition (K-means). De plus, cet algorithme permet de déterminer automatiquement le nombre de classes.

Dans un premier temps, nous décrivons le comportement des fourmis pour résoudre des problèmes d'optimisation puis nous dupliquerons ce comportement pour mettre au point l'algorithme Fourmis Artificielles (FA) dédié à la classification.

2.3.1.4.1. Comportement des fourmis

Le rôle d'une fourmi consiste à chercher de la nourriture en empruntant le chemin ayant la plus grande quantité de phéromone. Supposons qu'il existe que deux chemins pour atteindre une source de nourriture (voir figure 2.10).

Au début, une fourmi choisira un chemin au hasard et déposera initialement une certaine quantité de phéromone quel que soit le chemin choisi. La quantité de phéromone déposée sera élevée si le chemin emprunté est le plus court ce qui permettra à d'autres fourmis d'emprunter le chemin le plus court pour récolter plus rapidement la nourriture.

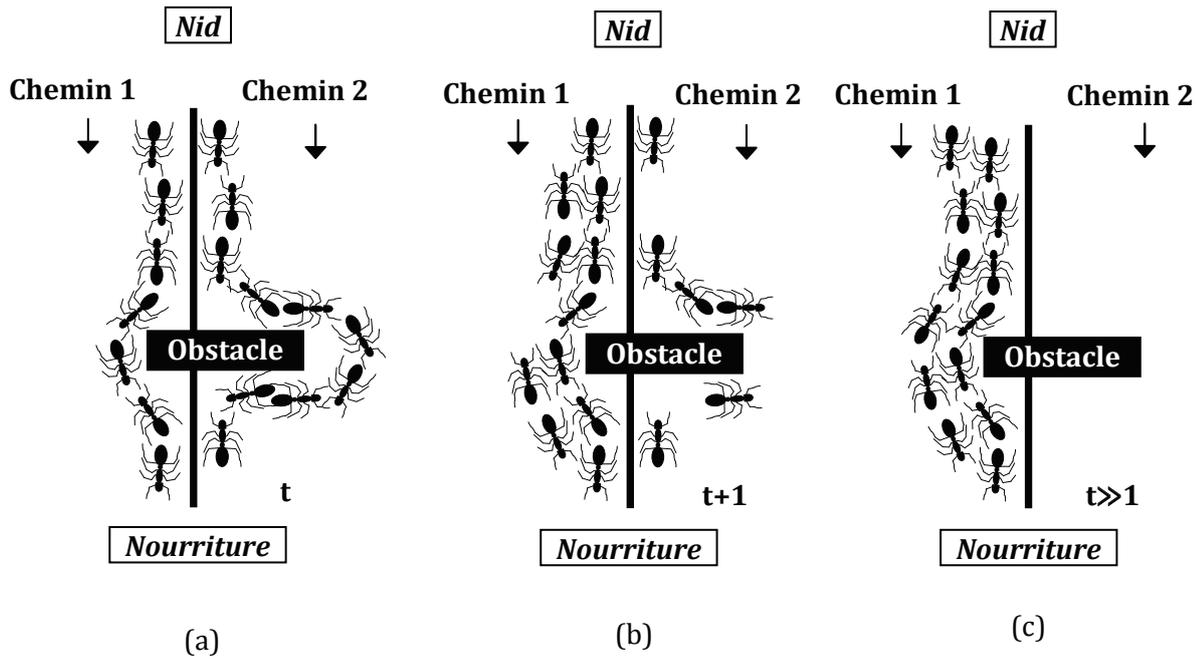


Figure 2. 10. Comportement d'une colonie de fourmis pour trouver le chemin le plus court vers une source de nourriture

(a) Les fourmis choisissent de façon aléatoire le chemin 1 et 2. (b) Les fourmis choisissent le chemin en fonction de la quantité de phéromone déposée par les fourmis précédemment. (c) Toutes les fourmis choisissent la chemin 1 car la quantité de phéromone y est plus élevée.

2.3.1.4.2. Un modèle mathématique sur le comportement des fourmis

Des chercheurs ont réussi à modéliser le comportement des fourmis afin de résoudre des problèmes d'optimisation comme par exemple celui du voyageur de commerce « *Traveling Salesman Problem* » (TSP) proposé par [Dorigo_97]. Le but est de trouver le chemin le plus court entre un nombre de N -villes, et de revenir au point de départ et cela en ne visitant chaque ville qu'une seule fois (voir figure 2.11). Ce problème peut être résolu en s'inspirant du comportement des fourmis cherchant à trouver le chemin le plus court vers une source de nourriture.

Soit Z le nombre de fourmis placées sur des villes à l'instant (t). Chaque fourmi située sur une ville (i) $i \in \{1, \dots, N\}$ choisira la prochaine ville (j) $j \in \{1, \dots, N\}$ qui maximise la probabilité $p_A^t(i, j)$.

$$p_A^t(i, j) = \begin{cases} \frac{[\tau^t(i, j)]^e (\eta_{ij})^f}{\sum_{r \in allowed_A} [\tau^t(i, r)]^e (\eta_{ir})^f} & \text{Si } j \in allowed_A \\ 0 & \text{Sinon} \end{cases} \quad (2.26)$$

avec

- $\eta_{ij} = 1/d(i, j)$ est l'inverse de la distance entre deux villes ;
- e est un paramètre qui contrôle l'importance de la quantité de phéromone déposée ;

- f est un paramètre qui contrôle l'importance de la distance entre les villes ;
- $allowed_A$ $A \in (1, \dots, Z)$: ensemble de villes qui restent à visiter par la fourmi A positionnée sur la ville (i) ;
- $\tau^t(i, j)$ et $\tau^t(i, r)$: quantité de phéromone déposée sur les chemins reliant les villes (i, j) et les villes (i, r) respectivement à l'instant (t) .

Cette probabilité dépend de la distance entre la ville (i) et (j) et la quantité de phéromone déposée entre elles à l'instant (t) [Juang_08]. Après le retour des fourmis à leurs points de départ, elles auront entre temps déposé une quantité de phéromone entre les villes. Cette quantité dépendra de la distance parcourue. Lorsqu'une autre fourmi visitera les villes, elle aura tendance à suivre (privilégier) le chemin ayant la plus grande quantité de phéromone, menant ainsi au chemin le plus court.

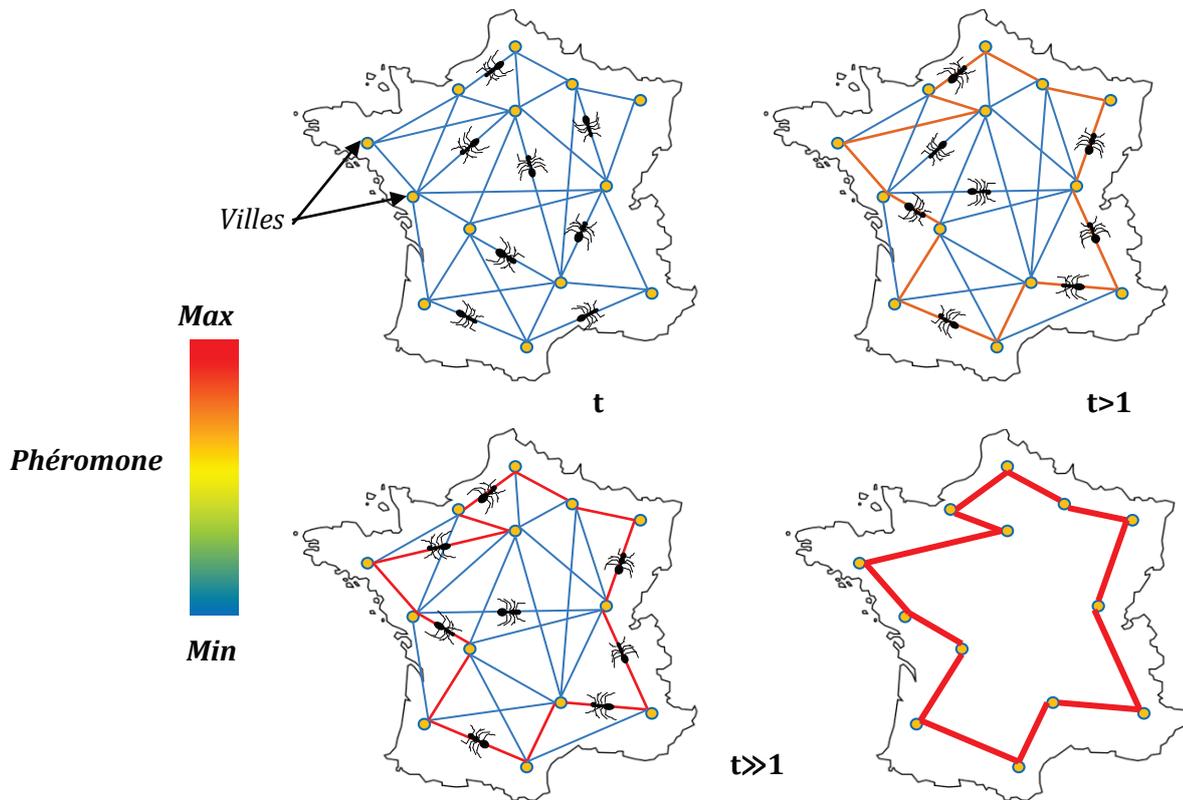


Figure 2. 11. Détermination du chemin le plus court entre N -villes par les fourmis

La quantité de phéromone $\tau^t(i, j)$ déposée sur les chemins reliant les villes (i, j) à l'instant (t) est donnée par :

$$\tau^t(i, j) = (1 - \varepsilon) \cdot \tau^{t-1}(i, j) + \Delta\tau(i, j) \quad (2.27)$$

avec

$$\Delta\tau(i, j) = \sum_{A=1}^Z \Delta\tau_A(i, j) \quad (2.28)$$

où $\varepsilon \in [0-1]$ est un paramètre tel que $(1-\varepsilon)$ représente l'évaporation de la phéromone. $\Delta\tau_A(i, j)$ représente une quantité de phéromone par unité de longueur placée entre les villes (i) et (j) par la fourmi A . La quantité $\Delta\tau_A(i, j)$ est donnée par :

$$\Delta\tau_A(i, j) = \begin{cases} \frac{\xi}{L_A} & \text{Si la fourmi } A \text{ se déplace entre} \\ & \text{les villes } (i) \text{ et } (j) \text{ durant sa tournée} \\ 0 & \text{Sinon} \end{cases} \quad (2.29)$$

avec

- $\xi \in [0-1]$ est une constante liée à la quantité de phéromone déposée par les fourmis ;
- L_A est la distance parcourue par la fourmi A .

À partir de ce modèle, nous allons mettre au point des règles comportementales pour l'algorithme Fourmis Artificielles (FA). Cet algorithme sera par la suite comparé à l'algorithme K-Means et à la classification ascendante hiérarchique ainsi qu'à d'autres méthodes de classification supervisée.

2.3.1.4.3. L'algorithme Fourmis Artificielles (FA)

Principe

La figure 2.12 montre une photo d'une colonie de fourmis à la recherche d'une source de nourriture. Cette colonie débute sa recherche sans avoir de connaissance sur l'emplacement de la source de nourriture dans l'environnement. Le comportement à adopter serait alors d'envoyer un groupe de recherche pour explorer la zone environnante. La fourmi qui trouve la source de nourriture en ramène une partie à la colonie et informe les autres sur l'endroit où se trouve la source de nourriture. Pour explorer la zone environnante, chaque fourmi dépose une quantité de phéromone sur le chemin qu'elle emprunte afin de retrouver plus tard le chemin du nid.

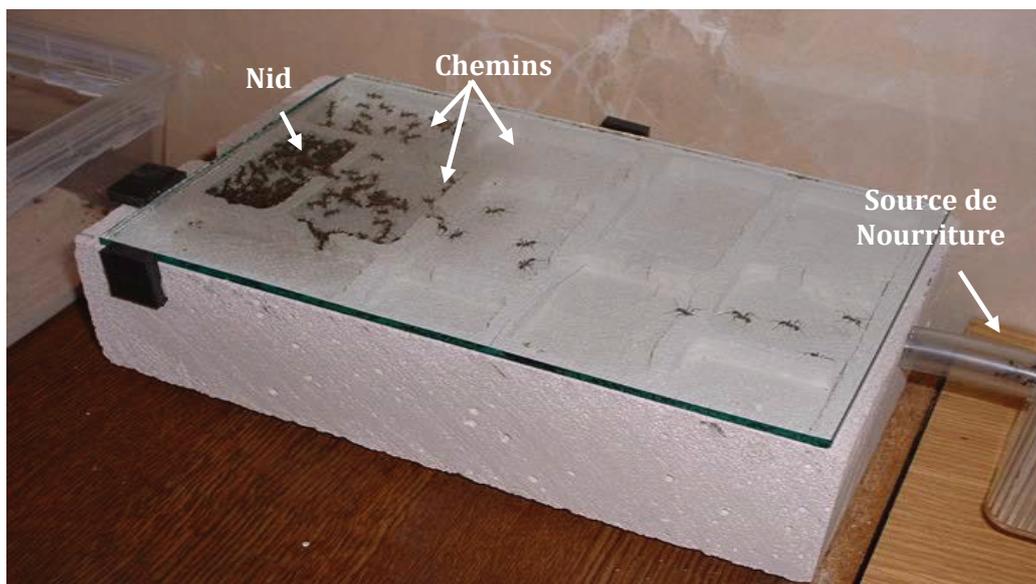


Figure 2. 12. Colonie de fourmis à la recherche de nourriture [Nid_11]

Dans la figure 2.13, la quantité de phéromone déposée par chaque fourmi errante est indiquée en rouge. Cette quantité diffère selon la longueur du chemin emprunté. Une fourmi errante peut alors suivre le même chemin qu'une autre fourmi si la quantité de phéromone déposée sur cette dernière est importante et que la distance qui les sépare est petite (voir figure 2.13(a)). Il arrive parfois que deux fourmis suivant le même chemin se séparent. Cela arrive lorsque la phéromone déposée commence à s'évaporer ce qui entraîne la création de nouveaux chemins (voir figure 2.13(b)).

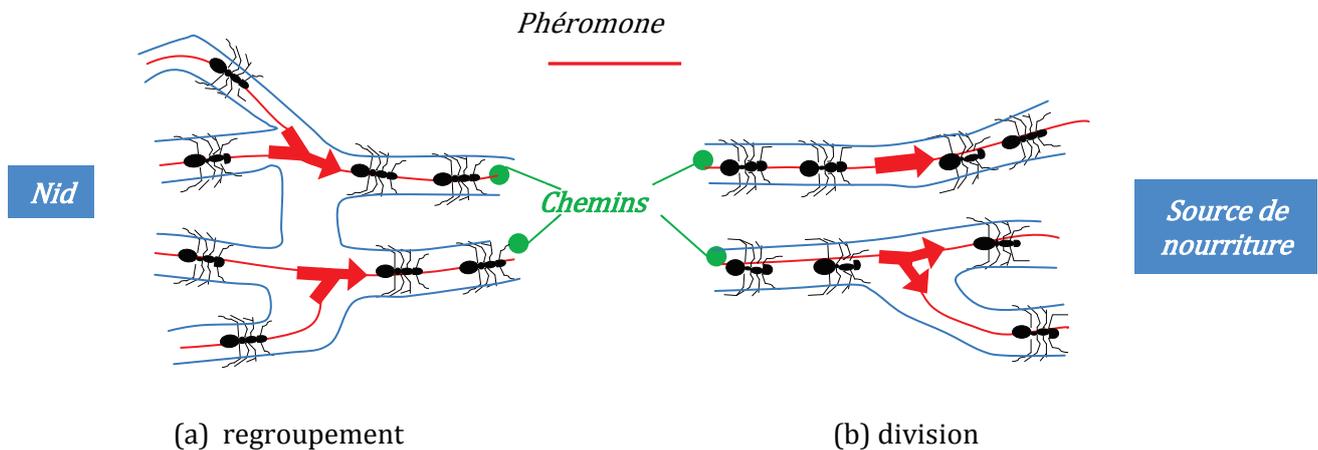


Figure 2. 13. Comportement des fourmis lors de la phase d'exploration

Comme nous pouvons le remarquer, la notion de regroupement entre les fourmis peut être représentée par un regroupement ascendant hiérarchique (classification ascendante hiérarchique) et la notion de division peut représenter le partitionnement des groupes de fourmis (classification par partition).

Nous avons établi un modèle du comportement d'une colonie de fourmis afin de construire un ensemble de M -chemins reliant le nid de fourmis à la source de nourriture dans le but de classer un ensemble de N -observations en s'inspirant de la notion de regroupement entre les fourmis et de la notion de division. Notons que le chemin suivi par une fourmi représente la classe d'appartenance d'une observation.

Le principe de l'algorithme Fourmis Artificielles (FA) est le suivant : considérons une population de N -fourmis transportant un ensemble de N -observations. Chaque observation est décrite par un vecteur à q -paramètres. On cherche alors un regroupement de ces fourmis dans un ensemble de M -classes ($\Omega_1, \Omega_2, \dots, \Omega_M$) de sorte que les observations similaires soient regroupées dans les mêmes classes tandis que les observations dissemblables soient dans des classes séparées, la valeur de M est inconnue. Le regroupement de ces fourmis est obtenu par une succession d'interactions entre elles. Chaque fourmi a des règles de comportement qu'elle suit en fonction des situations qu'elle rencontre :

- Le nid contient les observations à classer, et correspond au point de départ pour les fourmis qui partent vers la source de nourriture ;
- Le chemin emprunté par les fourmis vers la source de nourriture correspond à une classe. L'ensemble des chemins empruntés par les fourmis avant d'arriver à la source de nourriture constitue les M -classes ;

- La recherche d'un chemin vers une source de nourriture correspond à la recherche d'une classe pour les observations ;
- Chaque fourmi qui empreinte un chemin le mémorise en déposant une certaine quantité de phéromone. La quantité de phéromone déposée sur chaque chemin dépend du nombre de fourmis qui l'a emprunté ;
- Le choix d'un chemin pour une fourmi dépend de la similarité entre cette fourmi et toutes les fourmis qui ont emprunté ce chemin.

L'algorithme FA peut se résumer en quatre étapes : les étapes 1 et 2 constituent une optimisation de la classification hiérarchique ascendante. Le critère de liaison par approche fourmis artificielles calcule une similarité entre les chemins suivis par les fourmis en se basant sur la similarité entre les observations portées par ces fourmis.

Dans l'étape 3, une analyse en composantes principales est utilisée. L'ACP consiste à projeter l'ensemble d'observations dans un espace réduit à r -dimensions (avec $r \leq q$). Intuitivement, on cherchera l'espace donnant la meilleure visualisation possible de notre ensemble d'observations. Un bon choix consiste à rechercher la plus grande dispersion (le plus grand étalement) possible des projections dans l'espace choisi. On est amené ainsi à extraire un nouveau système d'axes (axes factoriels) permettant de mieux distinguer les classes. Après réduction de notre système d'axes initial, nous utilisons l'algorithme K-means afin de juger de la qualité des classes obtenues. Cet algorithme permet d'obtenir la meilleure répartition possible des classes et donc du chemin à suivre par les fourmis en tenant compte de la divisibilité et de la compacité des classes.

La quatrième étape consiste à affecter les nouvelles observations x_{new} à l'une des classes obtenues dans l'étape précédente. Un critère est ajouté dans cette étape afin de rejeter les observations non similaires et permettre de détecter de nouvelles classes.

Etape 1

Considérons une population de N -fourmis Ant_i $i \in \{1, \dots, N\}$ transportant N -observations. Ces N -fourmis (situées initialement dans le nid) vont classer les N -observations dans un premier ensemble de m -classes ($m < N$) et cela en essayant de choisir un chemin vers une source de nourriture. Cette étape est effectuée par une fonction de similarité notée $Sim_1 (Ant_i, Ant_j)$. Cette fonction mesure le degré de similarité entre une fourmi (Ant_j) portant une observation x_j qui se trouve dans le nid et une autre fourmi (Ant_i) portant une observation x_i qui se dirige vers la source de nourriture en empruntant un chemin (i). La fonction de similarité $Sim_1 (Ant_i, Ant_j)$ est donnée comme suit :

$$Sim_1 (Ant_i, Ant_j) = [\tau_i] \cdot [d(x_i, x_j)]^{-1} \quad (2.30)$$

où $d(x_i, x_j)$ est la distance entre les observations portées par les fourmis (Ant_i, Ant_j) et τ_i est la quantité de phéromone déposée sur le chemin (i).

Au début, la quantité de phéromone déposée par chaque fourmi est égale à :

$$\tau_i = \xi \quad i \in \{1, \dots, N\} \quad (2.31)$$

Une fois que le groupe de recherche a quitté le nid, la quantité de phéromone déposée par les fourmis sur chaque chemin emprunté est mise à jour par :

$$\tau_i = n_{Ants_i} \cdot \xi \quad (2.32)$$

où $\xi \in [0-1]$ est une constante liée à la quantité de phéromone déposée par les fourmis et n_{Ants_i} est le nombre de fourmis qui a emprunté l' $i^{\text{ème}}$ chemin.

Le premier ensemble de m -classes (classes initiales) est obtenu par l'algorithme suivant :

Algorithme 2. 3. Etape 1 : algorithme simulant le comportement d'une fourmi Ant_j en déplacement lorsqu'elle quitte le nid

- (1) Placer l'ensemble des observations dans le nid
- (2) Evaluer la quantité de phéromone de chaque fourmi par (2.31)
- (3) **Tant qu'il** y a des fourmis dans le nid
- (4) Trouver la fourmi (Ant_j) la plus similaire à la fourmi (Ant_i) en se basant sur la similarité donnée par (2.30)
- (5) **Si** $Sim_1 (Ant_i, Ant_j) \geq S_{Sim}$ % S_{Sim} représente le seuil de similarité %
- (6) Faire emprunter à la fourmi (Ant_j) le même chemin que celui de la fourmi (Ant_i)
Mettre à jour la quantité de phéromone donnée par (2.32)
- (7) **Fin si**
- (8) **Si** aucune fourmi n'emprunte un chemin
- (9) Diminuer le seuil de similarité S_{Sim}
- (10) **Fin si**
- (11) **Fin**

Dans cet algorithme, une fourmi Ant_j suit le chemin parcouru par les fourmis qui lui sont le plus similaires. Mais s'il s'avère que le seuil de similarité (S_{Sim}) est trop élevé, elle ne va pas se déplacer mais rester dans le nid. Plus le seuil de similarité sera bas, plus les fourmis arriveront à créer de nouvelles classes. Cependant, un seuil trop tolérant dès le début entraînera la création rapide d'un trop grand nombre de classes. Il est par conséquent nécessaire de trouver un compromis entre le seuil de similarité et les m -classes créées (voir section 2.4.1).

Dans la première étape, le nombre d'opérations nécessaires pour que les N -fourmis suivent les m -chemins est logarithmique, soit une complexité de $O(N \log N)$. Ceci représente un avantage par rapport à la classification ascendante hiérarchique dont la complexité est de $O(N^2)$.

À l'issue de cette étape, les m -chemins caractérisent les m -classes. Les observations de la classe (Ω_i) ou fourmis empruntant le chemin (i) sont les points de passage ou observations sur le chemin correspondant.

Un exemple sur le comportement de N -fourmis lorsqu'elles quittent le nid vers la source de nourriture en empruntant les m -chemins (m -classes) est illustré dans la figure suivante :

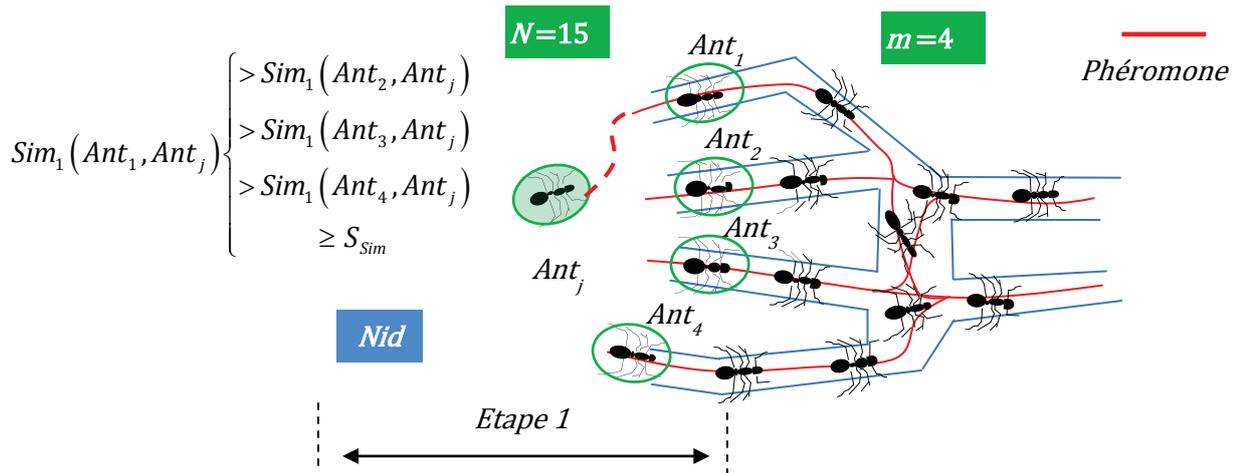


Figure 2. 14. Exemple de Règle de comportement d'une fourmi Ant_j lorsqu'elle se situe dans le nid

- Etape 2

Le nid étant vide, les N -fourmis continuent leurs chemins et tentent d'optimiser la classification des N -observations dans un second ensemble de classes ($m' < m$) et ceci en choisissant un chemin vers la source de nourriture.

L'objectif de cette étape est de fusionner les classes similaires pour en diminuer le nombre. Cela se fait par une seconde fonction de similarité notée $Sim_2 (Ants_i, Ants_j)$. Cette fonction mesure le degré de similarité entre un groupe de fourmis ($Ants_j$) qui a emprunté un chemin (i) et un autre groupe de fourmis ($Ants_i$) qui a emprunté un chemin (j) vers la source de nourriture. Cette fonction retourne une valeur comprise entre 0 (si i et j sont totalement différents) et 1 (si i et j sont identiques). La formule (2.26) est réécrite comme suit :

$$Sim_2 (Ants_i, Ants_j) = 1 - \frac{[\tau_i]^e \cdot [d(g_i, g_j)]^f}{\sum_{A \in \{1, \dots, m\}} [\tau_A]^e \cdot [d(g_A, g_j)]^f} \quad (2.33)$$

avec

- $d(g_i, g_j)$: la distance entre le centre de gravité des observations portées par les groupes de fourmis ($Ants_i, Ants_j$) ;
- $e \in [-1, +1]$: paramètre qui contrôle l'importance de la phéromone déposée par les fourmis sur les chemins (i), (j) et qui dans le cas de la classification contrôle le poids des classes Ω_i et Ω_j ;
- $f \in [0, 1]$: paramètre qui contrôle l'important de la distance qui sépare les groupes de fourmis et qui dans le cas de la classification contrôle la dispersion interclasses.

Le rôle des paramètres e et f est le suivant : un paramètre e fixé à 0 induit une similarité classique entre deux groupes de fourmis sur la base de la distance de Mahalanobis. Un paramètre e supérieur à zéros ($e > 0$) induit une similarité entre deux groupes de fourmis ($Ants_i, Ants_j$) qui discrimine le groupe de fourmis ($Ants_i$) qui a emprunté le chemin ayant accumulé une très grande quantité de phéromone. Ce paramètre incite le groupe de fourmis ($Ants_j$) à explorer d'autres chemins en cherchant (si possible) une meilleure similarité avec d'autres groupes de fourmis et dans le cas de la classification, à explorer d'autres classes pour une meilleure

affectation des observations. Un paramètre e inférieur à zéros ($e < 0$) induit une similarité entre deux groupes de fourmis ($Ants_i, Ants_j$) qui, à l'inverse de la situation précédente, favorise le groupe de fourmis (Ant_i) qui a emprunté le chemin ayant accumulé une très grande quantité de phéromone. Ceci permet d'accélérer le processus de classification. Si le paramètre f est fixé à 0, seule la quantité de phéromone est prise en compte. Par conséquent, un compromis entre la distance et la quantité de phéromone est une bonne approche pour optimiser la classification.

La quantité de phéromone déposée sur chaque chemin emprunté par les groupes de fourmis est mise à jour par :

$$\tau_i^{t+1} = n_{Ants_i} \cdot (1 - \varepsilon) \cdot \tau_i^t \quad (2.34)$$

$i \in \{1, \dots, m'\}$

où $\varepsilon \in [0, 1]$ est le facteur d'évaporation. Il est utilisé pour éviter une accumulation illimitée de la phéromone sur les chemins et dans notre cas, pour limiter le poids de chaque classe. n_{Ants_i} est le nombre de fourmis qui a emprunté l' $i^{\text{ème}}$ chemin. g_i, g_j sont les centres de gravité des observations transportées par le groupe de fourmis $Ants_i, Ants_j$ respectivement.

Les m' -classes obtenues doivent être évaluées en utilisant le critère J (voir section 2.2.1). Ce critère permet de juger du degré de compacité et de séparabilité des m' -classes. Une augmentation du critère J entre l'instant (t), où deux classes (Ω_i, Ω_j) (caractérisées par deux chemins empruntés par deux groupes de fourmis ($Ants_i, Ants_j$)) s'apprêtent à fusionner, et l'instant ($t+1$) où les deux classes sont fusionnées permet d'approuver la fusion $J(t+1) - J(t) > 0$. Dans le cas contraire, les deux groupes de fourmis ($Ants_i, Ants_j$) chercherons d'autres chemins pour satisfaire le critère J.

Le deuxième ensemble de m' -classes est obtenu par l'algorithme suivant :

Algorithme 2. 4. Etape 2 : algorithme simulant le comportement de deux groupes de fourmis
 $Ants_i, Ants_j$ en déplacement vers la source de nourriture

- (12) Evaluer le critère $J(t)$ des chemins empruntés par les fourmis en utilisant la formule (2.5)
 - (13) Evaluer la quantité de phéromone déposée sur chaque chemin emprunté par les fourmis en se référant à la formule (2.34)
 - (14) Sélectionner deux groupes de fourmis $Ants_i, Ants_j$ empruntant deux chemins différents en se référant à la fonction de similarité donnée par (2.33)
 - (15) Faire emprunter aux deux groupes de fourmis $Ants_i, Ants_j$ le même chemin
 - (16) Evaluer le critère $J(t+1)$ du chemin emprunté aux deux groupes de fourmis $Ants_i, Ants_j$
 - (17) **Si** $J(t+1) - J(t) > 0$
 - (18) **Tant que** $J(t+1) - J(t) > 0$
 - (19) Valider le chemin emprunté par les deux groupes de fourmis
 - (20) $J(t) \leftarrow J(t+1)$
 - (21) Mettre à jour la quantité de phéromone de ce chemin en se référant à (2.32)
 - (22) Répéter les étapes 14 à 16
 - (23) **Fin**
 - (24) **Fin**
-

Un exemple sur le comportement de deux groupes de fourmis se dirigeant vers la source de nourriture en empruntant les m' -chemins (m' -classes) est illustré dans la figure suivante :

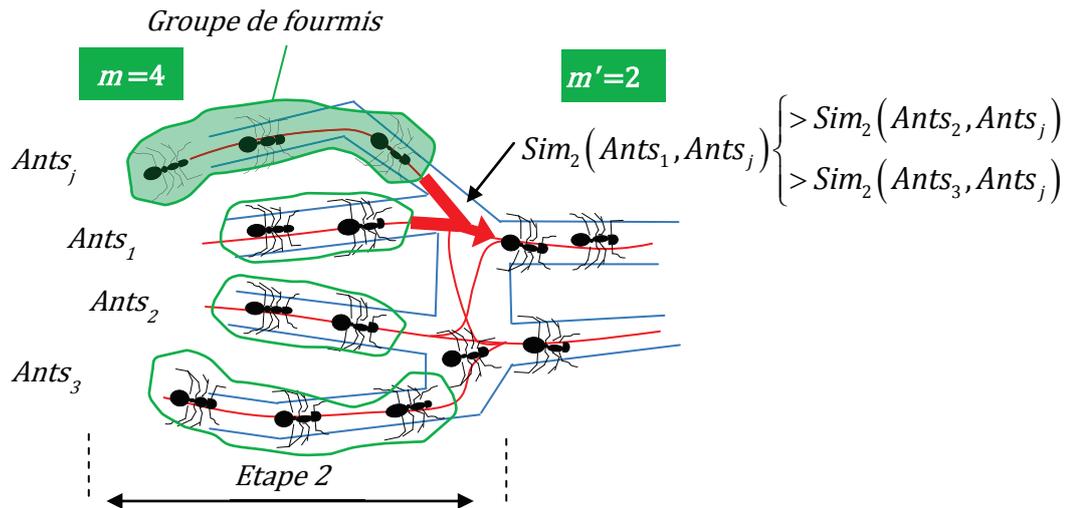


Figure 2. 15. Exemple de comportement de deux groupes de fourmis $Ants_i, Ants_j$ dans le cas où elles empruntent deux chemins similaires

Dans la seconde étape, le nombre d'opérations nécessaires pour que les N -fourmis suivent les m -chemin est polynomiale, soit une complexité de $O(m^2)$ avec $m \ll N$.

- Etape 3

La dernière série de classes ($\Omega_1, \Omega_2, \dots, \Omega_M$) est obtenue par une analyse en composantes principales (ACP) ainsi que par l'algorithme K-means. L'ACP effectue une réduction ainsi qu'une rotation de notre système d'axes (q -paramètres sélectionnés). Cette méthode est couramment utilisée pour réduire le nombre de paramètres. Elle est aussi appelée décomposition en valeurs singulières (SVD), transformée de Karhunen-Loève (KLT), transformée de Hotelling ou méthode de fonction orthogonale empirique, en fonction du domaine d'application.

La réduction de l'espace de représentation implique une perte d'information. Le but est de trouver le sous-espace qui permet de perdre le moins d'information possible. Cela est obtenu en cherchant l'axe qui discrimine le mieux les observations, puis le deuxième axe qui est orthogonal au premier, et ainsi de suite.

Soit x_i une observation de l'espace à q -dimensions, et y_i une observation qui est sa projection orthogonale dans le sous-espace réduit. L'information perdue par la projection peut être mesurée par la distance euclidienne entre l'observation x_i et y_i , $d(x_i - y_i)$, tous les deux représentés dans l'espace initial. On cherche alors la transformation linéaire qui minimise la distance entre une observations x_i et sa projection y_i $i \in \{1, \dots, N\}$. Cela revient à trouver le sous-espace de façon à ce que la projection orthogonale déforme le moins possible le nuage des N -observations. En termes de distance entre observations, le sous-espace recherché est tel que le critère J_{ACP} est minimal.

$$J_{ACP} = \frac{1}{N} \sum_{i=1}^N d(y_i - x_i)^2 \tag{2.35}$$

Cette écriture n'est autre que la forme classique du critère des moindres carrés.

Soit g le centre de gravité d'un ensemble d'observations (voir figure 2.16). D'après le théorème de Pythagore, on a :

$$d(y_i - x_i)^2 = d(x_i - g)^2 - d(y_i - g)^2 \tag{2.36}$$

Par conséquent, minimiser (2.36) revient à maximiser $d(y_i - g)$ puisque $d(x_i - g)$ est constant. De ce fait le sous-espace passera par le centre de gravité g du nuage des N -observations.

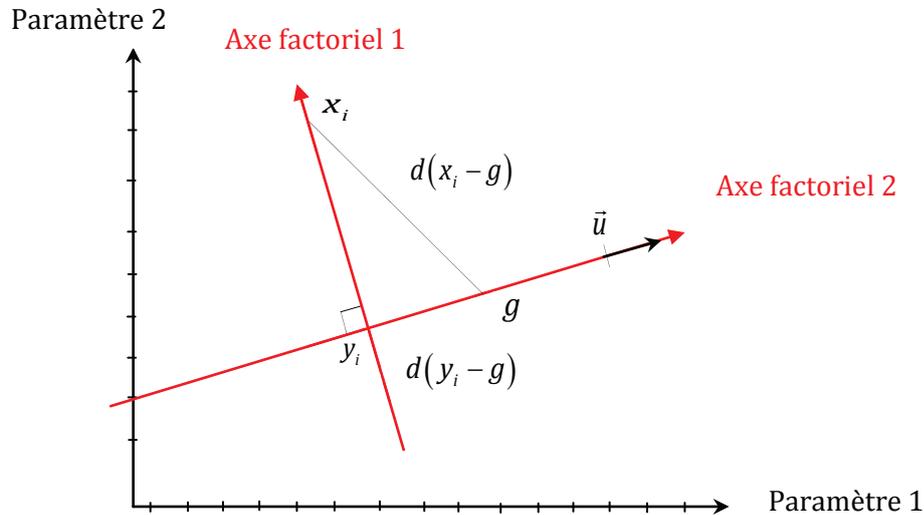


Figure 2. 16. Représentation d'une rotation orthogonale

On peut réécrire J_{ACP} sous la forme :

$$\begin{aligned} J_{ACP} &= -u^T \Lambda u + \frac{1}{N} \sum_{i=1}^N d(x_i - g)^2 \\ &= -J'_{ACP} + \frac{1}{N} \sum_{i=1}^N d(x_i - g)^2 \end{aligned} \tag{2.37}$$

Λ est une matrice carrée ($q \times q$), réelle et symétrique, c'est aussi la matrice d'inertie de l'ensemble d'observations considéré. Ses éléments diagonaux sont les variances de l'ensemble d'observations. Elles valent 1 car les données sont centrées et réduites. Et les éléments non diagonaux sont les covariances. Les vecteurs propres (2.38) de la matrice de variance-covariance Λ sont les vecteurs u_j tels que $\Lambda \cdot u_j = \theta_j \cdot u_j$. Les scalaires θ_j , repris dans la matrice diagonale (2.39), sont les valeurs propres.

$$u = [u_1 \quad u_2 \quad \dots \quad u_r \quad \dots \quad u_q] \tag{2.38}$$

$$\theta = \begin{bmatrix} \theta_1 & & & & 0 \\ & \theta_2 & & & \\ & & \dots & & \\ & & & \theta_r & \\ & & & & \dots \\ 0 & & & & & \theta_q \end{bmatrix} \tag{2.39}$$

Minimiser le critère J_{ACP} revient donc à maximiser le critère J'_{ACP} : $J'_{ACP} = u^T \Lambda u$

On cherche donc u_1 tel que : $u_1 = \max (u^T \Lambda u)$

La solution à ce problème de maximisation est le vecteur propre de Λ associé à la plus grande valeur propre θ_1 . On cherche alors le vecteur u_2 orthogonal à u_1 tel que la dispersion des vecteurs propres projetés sur cet axe soit maximale. Les premiers (r) vecteurs propres, rangés par ordre décroissant des valeurs propres vont constituer la base du sous-espace de dimension (r). Les vecteurs u_j $1 \leq j \leq r$ sont appelés composantes principales. Les composantes principales sont les variables artificielles (axes factoriels) définies par les N valeurs résultantes de la projection de l'ensemble des observations sur le vecteur u_j .

Pour trouver les principales composantes, on peut réaliser les opérations suivantes :

1. Organiser les observations en N colonnes de q -composantes pour obtenir une matrice (X) : $N \times q$;
2. Calculer le centre de gravité de chaque dimension $g = (g_1, g_2, \dots, g_q)'$;
3. Soustraire le centre de gravité à toutes les colonnes de (X) afin d'obtenir des observations centrées; la matrice obtenue est notée \mathcal{P} ;
4. Calculer la matrice de covariance $\Lambda : q \times q$ de \mathcal{P} ;
5. Calculer et trier par ordre décroissant les valeurs propres θ_j et les vecteurs propres u_j de Λ .

La dernière phase de l'étape 3 est la subdivision des classes obtenues par l'algorithme K-means. Le but est de vérifier la qualité des m' -classes obtenues en se basant sur le critère J . Une augmentation du critère J entre l'instant (t), où une classe (Ω_i) (caractérisée par un chemin suivi par un groupe de fourmis ($Ants_i$)) s'apprête à se subdiviser en deux nouvelles classes, et l'instant ($t+1$), où les deux classes sont obtenues, permet d'approuver la division $J(t+1) - J(t) > 0$. Dans ce cas-là, les deux nouvelles classes seront incluses dans les M -classes. Dans le cas contraire, les m' -classes seront interprétées comme étant les M -classes finales.

Les M -classes sont obtenues par l'algorithme suivant :

Algorithme 2. 5. Etape 3 : algorithme simulant le comportement d'un groupe de fourmis $Ants_i$ lorsqu'il décide de se diviser pour arriver à la source de nourriture

- (25) Effectuer la réduction des axes (q -paramètres) en utilisant l'ACP
 - (26) Evaluer le critère $J(t)$ en utilisant la formule (2.5)
 - (27) Sélectionner un chemin parmi les m' -chemins empruntés par les N -fourmis
 - (28) Diviser le chemin sélectionné en deux en utilisant l'algorithme K-means
 - (29) Evaluer le critère $J(t+1)$ des $m'+1$ chemins empruntés par les N -fourmis
 - (30) **Tant que** $J(t+1) - J(t) > 0$
 - (31) Valider les $m'+1$ chemins ($m' \leftarrow m'+1$)
 - (32) $J(t) \leftarrow J(t+1)$
 - (33) Répéter l'étape (27) à (29)
 - (34) **Fin**
 - (35) $M \leftarrow m'$
-

Un exemple sur le comportement de deux groupes de fourmis se dirigeant vers la source de nourriture en empruntant les M -chemins (M -classes) est illustré dans la figure suivante :

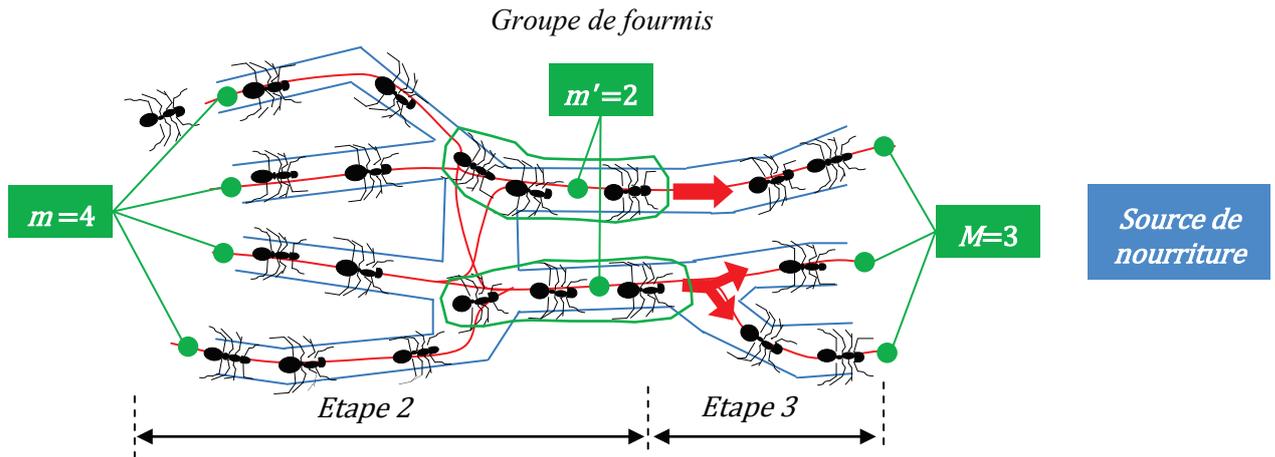


Figure 2. 17. Exemple de comportement de deux groupes de fourmis $Ants_i, Ants_j$ dans le cas où elles empruntent deux chemins différents

- Etape 4

La dernière étape est représentée par une règle de décision. Elle classe chaque nouvelle observation x_{new} dans l'une des M -classes en mesurant la similarité entre la nouvelle observation x_{new} et les M -classes obtenues par les fourmis. La fonction de similarité est donnée comme suit :

$$Sim(\Omega_i, x_{new}) = 1 - \frac{[\tau_i]^e \cdot [d(g_i, x_{new})]^f}{\sum_{A \in \{1, \dots, M\}} [\tau_A]^e \cdot [d(g_A, x_{new})]^f} \quad (2.40)$$

avec :

- g_i : centre de gravité de la classe Ω_i ;
- g_A : centre de gravité de la $A^{ième}$ classe appartenant à l'ensemble des M -classes ;
- τ_i : quantité de phéromone déposée sur le chemin (i) et, dans le cas de la classification, le poids la classe Ω_i .

Les différentes classes sont prédéterminées lors de la phase de conception. Toutefois, dans certains problèmes, de nouvelles classes, qui n'étaient pas présentes lors de la phase de conception, peuvent apparaître lors de la phase de diagnostic. Il arrive parfois d'obtenir des observations mal classées, ce qui conduit à des modèles de classification inexacts. Des exemples d'applications qui sont touchés par cette situation sont :

- Les problèmes de diagnostic dans lequel l'objectif du classifieur est d'identifier un fonctionnement anormal inconnu a priori [Dubuisson_93]. Par exemple, dans le diagnostic des défauts d'une machine [Ypma_99], un essai destructif pour tous les états possibles anormaux peut ne pas être possible ou très coûteux. Il convient donc de les identifier sans leur présence dans la base de connaissance ;
- Les systèmes de reconnaissance comme par exemple la classification des panneaux de signalisation routière. Ici, le classifieur doit non seulement faire la distinction entre des classes de signalisation routière, mais convient également de rejeter des panneaux de signalisation non reconnus [Paclík_06].

Dans notre cas, nous considérons le problème pour lequel il y a un ensemble de classes bien défini, et une ou plusieurs nouvelles classes non définies. L'objectif principal est de maintenir une performance de classification élevée entre les classes connues, et en même temps d'intégrer ces nouvelles classes dans le processus de classification (dues aux changements dans les conditions de fonctionnement reflétés par l'évolution de la répartition de ces classes) en introduisant une règle de décision.

Nous introduisons une règle de décision représentée par un critère noté Cq' . À partir d'un certain nombre N_{\min} d'observations regroupées dans une nouvelle classe, il est possible de détecter l'apparition de modes de fonctionnement inconnus via le critère Cq' . Ce critère permet de déterminer la compacité de ces observations ainsi que leur séparabilité par rapport aux classes déjà identifiées via la matrice de variance intra-classe et inter-classes. Par exemple, dans le cas d'une classe composée de 4 observations, si le critère Cq' est du même ordre de grandeur que les classes déjà identifiées, alors on peut dire qu'il y a apparition d'une nouvelle classe.

Le critère Cq' est donné comme suit :

$$Cq'(\Omega_{new}) = \left(\frac{n_{new}}{N} \right) \text{trace} \left(\frac{(g_{new} - g) \cdot (g_{new} - g)'}{\sum_{j=1}^{n_{new}} (x_j - g_{new}) \cdot (x_j - g_{new})'} \right) \quad (2.41)$$

avec :

- g : centre de gravité des observations constituant l'ensemble des M -classes ;
- g_{new} : centre de gravité de la classe Ω_{new} ;
- x_j : $j^{\text{ième}}$ observation de la classe Ω_{new} ;
- n_{new} : nombre d'observations de la classe Ω_{new} ;
- N : nombre total de l'ensemble d'observations ;

La détection ou l'apparition d'une nouvelle classe peut se résumer ainsi :

Algorithme 2. 6. Etape 4 : détection de nouvelles classes

- (36) Calculer le critère $Cq'(\Omega_i)_{i \in \{1, \dots, M\}}$ des M -classes définies dans l'espace de représentation
 - (37) Calculer la compacité de la nouvelle classe $Cq'(\Omega_{new})$
 - (38) Comparer le critère $Cq'(\Omega_{new})$ avec ceux des M -classes $Cq'(\Omega_i)$
 - (39) **Si** $Cq'(\Omega_{new}) \approx Cq'(\Omega_i)$
 - (40) **Alors** Apparition d'une ou plusieurs nouvelles classes
 - (41) Expertise humaine pour étiqueter la ou les nouvelles classes
 - (42) Intégration de cette ou ces nouvelles classes dans l'ensemble des M -classes
 - (43) **Fin Si**
-

2.3.2. La classification supervisée

À l'inverse de la classification non supervisée, la classification supervisée se base sur les observations ainsi que sur leurs classes d'origine pour identifier les classes d'origine des nouvelles observations. En d'autres termes, les algorithmes de classification supervisée devraient prévoir l'étiquette de chaque nouvelle observation. À cet effet, nous avons mis au

point un nouveau classifieur supervisé basé sur les Modèles de Markov cachés afin de minimiser les erreurs de classification.

La classification supervisée permet de définir une fonction d'appartenance à partir d'un ensemble de classes constitué d'échantillons d'apprentissage. Chaque échantillon est constitué d'une entrée (observation) et d'une valeur de sortie désirée (étiquette de l'observation). La fonction d'appartenance doit alors prévoir une classe d'appartenance correcte pour n'importe quelle entrée. Il faut pour cela effectuer les étapes suivantes :

- Déterminer les signaux et grandeurs représentatifs et discriminants. L'utilisateur doit décider du type de paramètre devant être utilisé pour représenter l'ensemble d'apprentissage. Dans le cas du diagnostic, par exemple, cela pourrait être un paramètre extrait à partir de l'analyse temporelle, fréquentielle ou temps-fréquence (voir section 1.4);
- Construire ou collecter un ensemble d'apprentissage. L'ensemble d'apprentissage doit être représentatif de l'utilisation réelle de la fonction d'appartenance. Ainsi, un ensemble d'entrées est collecté et les sorties correspondantes sont également réunies, soit par une expertise humaine ou à partir des mesures réalisées ;
- Déterminer à partir de l'étape précédente les paramètres représentatifs de l'entrée de la fonction d'appartenance. Le résultat de la sortie dépend fortement de la façon dont l'observation est représentée en entrée. En règle générale, l'entrée est représentée par un vecteur de paramètres qui contient un certain nombre de paramètres qui sont caractéristiques des défauts (voir section 2.2) ;
- Déterminer la structure de la fonction d'appartenance et de l'algorithme d'apprentissage correspondant. Dans notre cas, nous avons choisi d'utiliser les Modèles de Markov cachés et l'algorithme de Baum-Welsh ;
- Évaluer la précision de la fonction d'appartenance. Après ajustement des paramètres de la fonction d'appartenance par un algorithme d'apprentissage, la performance de la fonction résultante doit être mesurée sur un ensemble test qui est séparé de l'ensemble d'apprentissage.

Un algorithme de classification supervisée construit une fonction d'appartenance en se basant sur les classes déjà prédéterminées. La plupart des méthodes référencées dans la littérature utilisent les réseaux de neurones, les réseaux bayésiens et le support vecteur machine (SVM) comme techniques de classification [Wang_04, Babu_08, Xue_09, Zhan_10, Zhen_09, Liu_10]. Dans notre cas, nous nous sommes plutôt orientés vers une approche probabiliste pour faire de la classification supervisée. Cette méthode est basée sur les Modèles de Markov cachés. Un modèle de Markov caché est une classe de modèle probabiliste capable de capturer les propriétés d'une séquence d'observations ordonnée. Il est largement utilisé dans l'analyse bioinformatique des séquences [Krogh_93], du trafic dans le web [Ypma_02], la reconnaissance vocale, [Jelinek_98] et aussi dans la surveillance de l'état de santé des machines électriques [Lebaroud_08]. C'est dans ce contexte là que nous présentons dans la prochaine section une nouvelle méthode basée sur les Modèles de Markov Cachés (MMCs) pour la classification, qui sera plus tard utilisée pour le diagnostic des défauts.

2.3.2.1. Classification par modèles de Markov cachés

Nous définissons dans ce qui suit $(\Omega_1, \Omega_2, \dots, \Omega_M)$ comme étant les M -classes qui représentent les M -modes de fonctionnement (MF) du système. Les M -classes sont obtenues soit par expertise humaine soit par une technique de classification non-supervisée (voir section 2.3.1). Chacune des classes dispose d'un espace géométrique projeté dans un espace à q -dimensions (voir figure 2.18). Ces M -classes sont utilisées pour concevoir un modèle de Markov caché pour chaque mode de fonctionnement (MF) du système. Par exemple, dans le cas d'un moteur asynchrone, les MF étudiés peuvent être un moteur en fonctionnement sain (Ω_1), un moteur défectueux avec des barres cassées (Ω_2), un moteur avec un défaut de roulement (Ω_3)...etc. Une observation x_i est caractérisée par un vecteur de paramètres à q -dimensions, $x_i = [x_{i1}, \dots, x_{iq}]$. L'objectif est alors, compte tenu de l'observation x_i , de décider si elle doit être affectée à $\bar{\lambda}_1$ ou $\bar{\lambda}_2, \dots$, ou $\bar{\lambda}_M$. Le modèle de Markov caché $\bar{\lambda}_k, k \in \{1, \dots, M\}$, pour lequel la probabilité de générer cette observation $P(x_i | \bar{\lambda}_k)$ est maximale, déterminera le mode de fonction du système.

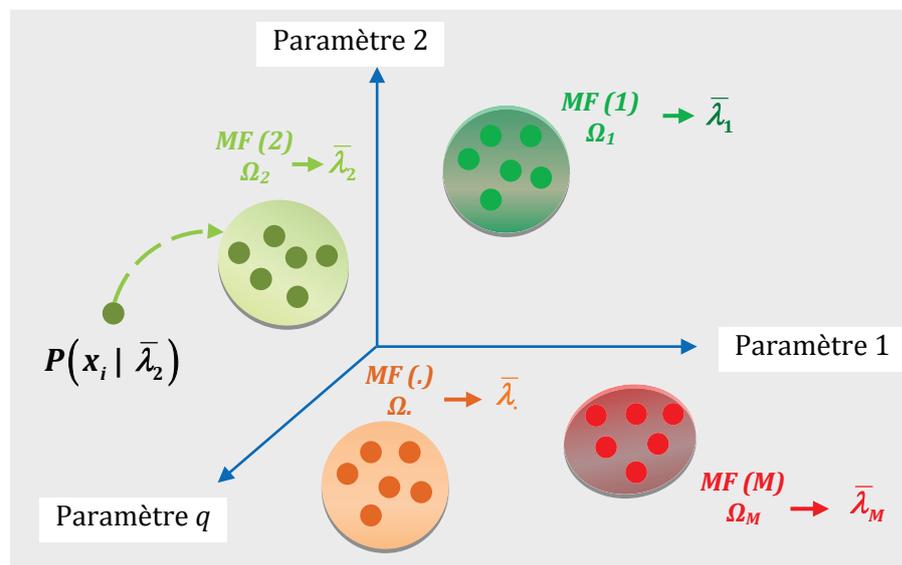


Figure 2. 18. Représentation des modes de fonctionnement d'un système sous forme de classes

2.3.2.2. Proposition

Dans notre approche, le système est décrit par des classes $(\Omega_1, \Omega_2, \dots, \Omega_M)$ qui représentent ses différents modes de fonctionnement (MF). Ces classes sont essentielles pour la modélisation du système et sont appelées *variables d'état*. Lorsque toutes les variables d'état sont connues, le système est dit entièrement défini. Si le processus est dynamique, le système passe d'un MF noté S_i à un autre mode S_j , ce qui entraîne une *transition* entre les MF pendant toute la durée du processus. Ceci permet de déduire qu'un MF peut être décrit en fonction de tous les modes de fonctionnement du système (2.42). Dans ce cas, concevoir un modèle de Markov caché pour un MF du système reviendrait à trouver un modèle qui prend en compte tous les modes de fonctionnement.

Pour mieux comprendre le concept, nous allons prendre un exemple dans lequel nous avons M -classes (M -modes de fonctionnement). Chacune de ces classes est constituée par un certain

nombre d'observations. Supposons que nous disposons d'une nouvelle observation. Cette nouvelle observation ne peut être placée que dans l'une de ces classes et nul par ailleurs. Si q_t désigne la classe où l'observation est placée à l'instant (t), alors le système est dit dans un mode $q_t = S_i$ $i \in \{1, \dots, M\}$ à l'instant (t). Puisque l'observation peut appartenir à l'une des M -classes, alors q_t peut prendre n'importe quelle valeur dans l'intervalle $[S_1, S_2, \dots, S_M]$ à n'importe quel moment $t=1, 2, 3, \dots$. Le mode de fonctionnement du système à l'instant (t) peut être défini par la probabilité conditionnelle d'un mode à l'instant (t) étant donné tous les MF précédents. Ceci se traduit mathématiquement par :

$$P(q_t = S_j | q_{t-1} = S_i, \dots, q_{t-n} = S_k) \quad 1 \leq i, j, k \leq M \quad (2.42)$$

Un processus de Markov est régi par l'hypothèse markovienne, qui stipule qu'un MF du système ne dépend que des dernière n modes, étant donné que tous les modes antérieurs avant les n derniers modes ne sont pas significatifs dans la détermination du MF du système à l'instant (t). Elle est aussi appelée chaîne de Markov d'ordre n . Dans le cas où $n=1$, la chaîne se réduit à une chaîne de Markov de premier ordre, où la description probabiliste est limitée au MF présent et passé. La formule (2.42) devient :

$$P(q_t = S_j | q_{t-1} = S_i) \quad 1 \leq i, j \leq M \quad (2.43)$$

2.3.2.3. Modèle de Markov caché

Dans un processus de Markov, chaque MF correspond à un événement observable physique. Pour être plus précis, dans l'exemple précédent, les MF du système sont définis par des classes, qui sont directement observables. En d'autres termes, les observations sont des fonctions déterministes des MF. Cependant, dans de nombreuses applications, cette approche est trop restrictive car les MF ne sont pas observables. Le processus de Markov est étendu pour inclure le cas où l'observation est une fonction probabiliste du MF. Le modèle qui en résulte est appelé Modèle de Markov Caché (MMC) ou fonction probabiliste des chaînes de Markov [Rabiner_89]. Le MMC est un processus doublement stochastique, avec un processus stochastique sous-jacent (séquence de MF) qui n'est pas directement observable, mais qui peut être observé à travers un autre processus stochastique qui se traduit par une séquence d'observations.

2.3.2.3.1. Paramètres d'un Modèle de Markov caché

Un modèle de Markov caché est un ensemble de modes de fonctionnement finis et reliés par des transitions. Chaque mode est caractérisé par deux ensembles de probabilités : une probabilité de transition sur la base de la chaîne de Markov de premier ordre et une probabilité discrète de sortie qui définit la probabilité d'émettre une observation, compte tenu du mode de fonctionnement du système. Les éléments d'un MMC peuvent être définis comme suit :

- **Nombre de modes de fonctionnement**

Le nombre de MF dans le modèle est noté M . Un ensemble de M -modes cachés est défini par $S = \{S_1, S_2, \dots, S_M\}$. Le MF du modèle à l'instant (t) est donné par $q_t \in S$, $0 \leq t \leq T$, où T est la longueur de la séquence d'observations et q_t représente le MF actuel.

- **Probabilité de transition**

La matrice de probabilité de transition est donnée par $A = \{a_{ij}\}$, $1 \leq i, j \leq M$, où a_{ij} est la probabilité de transition d'un MF (S_i) à l'instant ($t-1$) à un MF (S_j) à l'instant (t).

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i) \quad 1 \leq i, j \leq M \quad (2.44)$$

La probabilité de transition doit satisfaire les contraintes suivantes :

$$\begin{aligned} a_{ij} &\geq 0 & 1 \leq i, j \leq M \\ \sum_{j=1}^M a_{ij} &= 1 \end{aligned} \quad (2.45)$$

- **Probabilité d'observation**

Une matrice de probabilité d'observation notée B ($B = b_j(x_t)$) indique la probabilité d'émettre une observation x_t associée à un mode de fonctionnement (S_j) donné.

$$b_j(x_t) = P(x_t | q_t = S_j) \quad \begin{array}{l} 1 \leq j \leq M \\ 0 \leq t \leq T \end{array} \quad (2.46)$$

avec $0 \leq b_j(x_t) \leq 1$, x_t désigne une observation à l'instant (t).

- **Vecteur de probabilités initiales**

Le vecteur de probabilités initiales $\pi = \{\pi_i\}_{i \in \{1, \dots, M\}}$ indique la probabilité de commencer dans le MF (S_i) à $t=0$.

$$\pi_i = P(q_0 = S_i) \quad 1 \leq i \leq M \quad (2.47)$$

avec :

$$\begin{aligned} 0 &\leq \pi_i \leq 1 \\ \sum_{i=1}^M \pi_i &= 1 \end{aligned} \quad (2.48)$$

La notation suivante est souvent utilisée dans la littérature pour représenter un MMC :

$$\lambda = (A, B, \pi)$$

La spécification des trois ensembles de mesures de probabilité A , B et π sont alors nécessaires afin d'avoir une description complète de chaque mode de fonctionnement du système. Cela revient à décrire chaque MF par un modèle de Markov caché.

Dans la suite de cette thèse, la notation suivante sera utilisée pour décrire un MMC représentant un mode de fonctionnement (k) du système :

$$\lambda_k = (A^{(k)}, B^{(k)}, \pi^{(k)}) \quad 1 \leq k \leq M \quad (2.49)$$

$$A^{(k)} = \{a_{ij}^{(k)}\}, \quad 1 \leq i, j, k \leq M \quad (2.50)$$

$$\pi^{(k)} = \{\pi_i\} \quad (2.51)$$

$$B^{(k)} = b_j(X^k) = 1 - \frac{(X^k - g_j) \text{Cov}^{-1}(E_j) (X^k - g_j)'}{\sum_{\ell=1}^M (X^k - g_\ell) \text{Cov}^{-1}(E_\ell) (X^k - g_\ell)'} \quad (2.52)$$

avec

- k : le numéro de la classe Ω_k ;
- E_j regroupe les observations d'apprentissage de la classe Ω_j ;
- g_j est le centre de gravité de la classe Ω_j ;
- $(.)'$ désigne une matrice transposée. La matrice de covariance Cov est estimée à partir de chaque classe.

2.3.2.3.2. Apprentissage d'une classe par le modèle de Markov caché

Les seules informations disponibles sur le système sont ses différents modes de fonctionnement, autrement dit ses classes. Ainsi, l'approche que nous proposons consiste à modéliser le système en effectuant l'apprentissage de chacune de ses classes par un MMC. La phase d'apprentissage consiste à extraire de chaque classe (k) une séquence d'observations X^k de longueur T , $X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_T^k\}$ (voir figure 2.19). La longueur T correspond à un tirage successif de $T+1$ observations de chacune des classes. Ces séquences d'observations vont nous servir pour ajuster les paramètres de chaque MMC $\bar{\lambda}_k$ de façon à maximiser la probabilité de générer la séquence d'observations X^k étant donné un modèle initial λ_k , $P(X^k | \lambda_k)$.

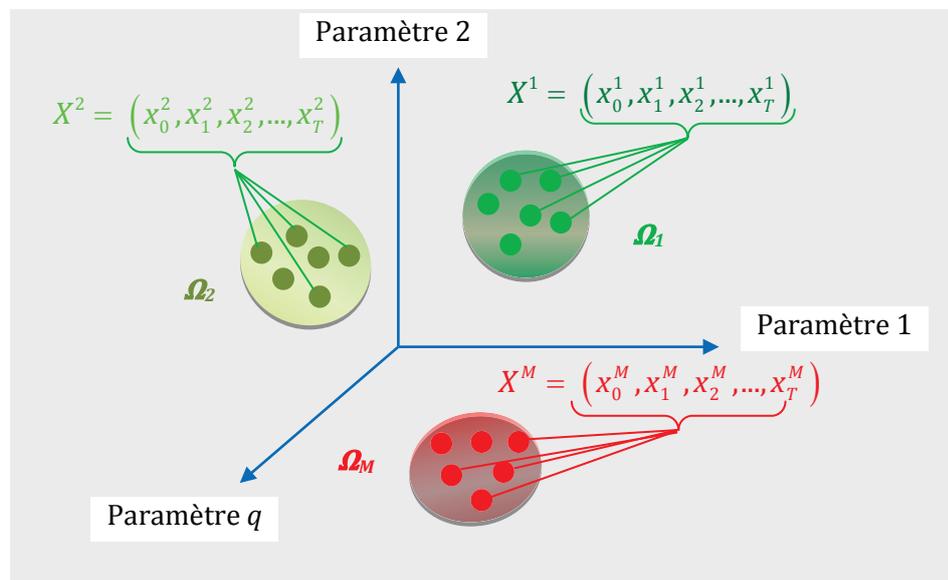


Figure 2. 19. Extraction des séquences d'observations

Comme nous pouvons le remarquer à partir de la figure 2.19, les observations qui constituent les séquences d'observations X^k n'évoluent pas dans le temps puisque $x_0^k \approx x_1^k \approx x_2^k \approx \dots \approx x_T^k$. Si q_t désigne la classe où l'observation est placée à l'instant (t), alors le système est dit dans un mode $q_t = S_k$ et ceci à n'importe quel moment $t=1, 2, 3, \dots$. La formule (2.43) devient :

$$P(q_t = S_k | q_{t-1} = S_k, \bar{\lambda}_k) \approx 1 \quad 1 \leq k \leq M \quad (2.53)$$

Cette formule est bien évidemment valable que dans le cas où les classes sont bien séparées. Cependant, il est tout à fait possible d'être confronté à un problème de chevauchement entre les classes. Par exemple, dans le cas de deux classes chevauchées, les séquences d'observation X^1 et X^2 qui y sont extraites se retrouvent avec des valeurs proches les unes des autres. Dans ce cas-là, la formule (2.53) ne devient plus valable. Ce qui nous incite à déterminer cette probabilité en utilisant un algorithme d'apprentissage.

La réponse au problème d'apprentissage est fournie par les formules de ré-estimation de Baum-Welsh qui, à partir d'un modèle initial λ_k , définissent un algorithme itératif pour maximiser la probabilité $P(X^k | \lambda_k)$ afin d'estimer le modèle $\bar{\lambda}_k$ le plus représentatif de la séquence d'observations X^k . La mise en œuvre de l'algorithme de Baum-Welsh comporte le calcul de deux probabilités différentes. Tout d'abord, la variable « Forward » qui est définie par la probabilité d'une suite partielle d'observations $\{x_0^k, x_1^k, x_2^k, \dots, x_t^k\}$ extraite de la classe (k) se terminant à un certain MF (S_i) à l'instant (t). Ensuite, la variable « Backward » qui désigne la probabilité d'observer la suite partielle $\{x_{t+1}^k, \dots, x_T^k\}$ extraite de la classe (k), qui se termine à l'instant (t) dans un MF (S_i).

La variable Forward

La variable forward notée $\alpha_t^{(k)}(i)$ est définie par :

$$\alpha_t^{(k)}(i) = P(x_0^k, x_1^k, x_2^k, \dots, x_t^k, q_t = S_i | \lambda_k) \quad (2.54)$$

L'algorithme Forward se détermine ainsi :

1. Initialisation :

$$\alpha_0^{(k)}(i) = \pi_i^{(k)} b_i(x_0^k) \quad (2.55)$$

2. Récurrence :

$$\alpha_t^{(k)}(j) = b_j(x_t^k) \sum_{i=1}^M \alpha_{t-1}^{(k)}(i) a_{ij}^{(k)} \quad \begin{array}{l} 1 \leq i, j, k \leq M \\ 1 \leq t \leq T \end{array} \quad (2.56)$$

avec :

- $b_j(x_t^k) = P(x_t^k | q_t = S_j)$: probabilité d'émettre une observation x_t extraite de la classe (k) associée à un mode de fonctionnement (S_j) à l'instant (t) ;
- $a_{ij}^{(k)}$: probabilité de transition d'un MF (S_i) à l'instant (t-1) à un autre MF (S_j) à l'instant (t) étant donné le modèle λ_k .

La variable Backward

La variable Backward notée $\beta_t^{(k)}(i)$ est définie par :

$$\beta_t^{(k)}(i) = P(x_{t+1}^k, \dots, x_T^k | q_t = S_i, \lambda_k) \quad (2.57)$$

L'algorithme Backward se détermine ainsi :

1. Initialisation :

$$\beta_t^{(k)}(i) = 1 \quad (2.58)$$

2. Récurrence :

$$\beta_t^{(k)}(i) = \sum_{j=1}^M \beta_{t+1}^{(k)}(j) b_j(x_{t+1}^k) a_{ij}^{(k)} \quad \begin{array}{l} 1 \leq i, j, k \leq M \\ 0 \leq t \leq T-1 \end{array} \quad (2.59)$$

Le but est de maximiser la probabilité $P(X^k | \lambda_k)$. Ceci demande d'utiliser les variables forward et backward ainsi que la probabilité de transiter d'un MF (S_i) à l'instant (t) vers un MF (S_j) à l'instant ($t+1$) étant donné la séquence d'observations X^k et un modèle λ_k initial, $\xi_t^{(k)}(i, j) = P(q_{t+1} = S_j | q_t = S_i, X^k, \lambda_k)$:

$$\xi_t^{(k)}(i, j) = \frac{P(q_t = S_i, q_{t+1} = S_j, X^k | \lambda_k)}{P(X^k | \lambda_k)} \quad (2.60)$$

En utilisant les variables forward et backward, on obtient :

$$\xi_t^{(k)}(i, j) = \frac{\alpha_t^{(k)}(i) a_{ij}^{(k)} b_j(x_{t+1}^k) \beta_{t+1}^{(k)}(j)}{\sum_{i=1}^M \alpha_t^{(k)}(i) \beta_t^{(k)}(i)} \quad (2.61)$$

Considérons également la variable $\gamma_t^{(k)}(i)$:

$$\gamma_t^{(k)}(i) = P(q_t = S_i | X^k, \lambda_k) \quad (2.62)$$

La variable $\gamma_t^{(k)}(i)$ exprime la probabilité de se trouver dans un MF (S_i) à l'instant (t) étant donné la séquence d'observations X^k , et le modèle λ_k .

Exprimée en fonction de $\xi_t^{(k)}(i, j)$, $\gamma_t^{(k)}(i)$ donne :

$$\gamma_t^{(k)}(i) = \sum_{j=1}^M \xi_t^{(k)}(i, j) \quad (2.63)$$

Si on somme $\gamma_t^{(k)}(i)$ sur l'ensemble des instants (t), on obtient une quantité que l'on peut interpréter comme l'espérance du nombre de fois où le MF (S_i) est utilisé pour générer la séquence X^k . De même, si on somme $\xi_t^{(k)}(i, j)$ sur l'ensemble des instants (t), on obtient une quantité que l'on peut interpréter comme l'espérance du nombre de fois où la transition $S_i \rightarrow S_j$ est utilisée pour générer la séquence X^k . On a donc :

$$\sum_{t=0}^{T-1} \gamma_t^{(k)}(i) = \text{espérance du nombre de passage dans le MF } S_i \quad (2.64)$$

$$\sum_{t=0}^{T-1} \xi_t^{(k)}(i, j) = \text{espérance du nombre de transition du MF } S_i \text{ vers le MF } S_j \quad (2.65)$$

En combinant les formules (2.64) et (2.65) on peut alors proposer un ensemble d'estimateurs des paramètres $\bar{\lambda}_k$:

$$\bar{\pi}_i^{(k)} = \text{probabilité de se trouver dans un MF } S_i \text{ à l'instant } t = 0, \gamma_0^{(k)}(i) \quad (2.66)$$

$$\begin{aligned} \bar{a}_{ij}^{(k)} &= \frac{\text{espérance du nombre de transition du MF } (S_i) \text{ vers le MF } (S_j)}{\text{espérance du nombre de passage dans le MF } (S_i)} \\ &= \frac{\sum_{t=0}^{T-1} \xi_t^{(k)}(i, j)}{\sum_{t=0}^{T-1} \gamma_t^{(k)}(i)} \quad 1 \leq i, j, k \leq M \end{aligned} \quad (2.67)$$

Une fois la mise à jour terminée, nous obtenons un ensemble de nouveaux paramètres pour chaque modèle $\bar{\lambda}_k = (\bar{A}^{(k)}, \bar{\pi}^{(k)})$. L'apprentissage par l'algorithme de Baum-Welch peut se résumer ainsi :

Algorithme 2. 7. Algorithme d'apprentissage de Baum-Welch

- (1) Données : $X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_T^k\}$
 - (2) Choisir un paramétrage initial $(A^{(k)}, \pi^{(k)})$
 - (3) **Répéter**
 - (4) **Pour** chaque $x_t^k \in X^k$
 $0 \leq t \leq T$
 - (5) Calculer les valeurs des variables $\alpha_t^{(k)}(i)$ à l'aide de l'algorithme de forward
 - (6) Calculer les valeurs des variables $\beta_t^{(k)}(i)$ à l'aide de l'algorithme de backward
 - (7) Calculer les valeurs des variables $\xi_t^{(k)}(i, j)$ à l'aide de la formule (2.61)
 - (8) Calculer les valeurs des variables $\gamma_t^{(k)}(i)$ à l'aide de la formule (2.63)
 - (9) Calculer et stocker les valeurs des expressions (2.64) et (2.65)
 - (10) **Fin pour**
 - (11) Ré-estimer les paramètres $(A^{(k)}, \pi^{(k)})$ en utilisant les formules (2.67) et (2.66)
 - (12) **Jusqu'à** un critère d'arrêt.
-

2.3.2.4. Classification d'une observation par un modèle de Markov caché

Etant donné une nouvelle observation x_i et un modèle de Markov $\bar{\lambda}_k = (\bar{A}^{(k)}, \bar{\pi}^{(k)})$, nous avons

besoin de calculer la probabilité $P(x_i | \bar{\lambda}_k)$ afin de déterminer le modèle le plus représentatif (classe d'origine) de cette nouvelle observation. Afin d'utiliser les formules précédentes pour le calcul de cette probabilité, nous gênerons une séquence d'observations d'une longueur T, $x_i = \{x_0, x_1, x_2, \dots, x_T\}$ qui a la même longueur que celles utilisées pour l'apprentissage, avec $x_0 = x_1 = x_2 = \dots = x_T = x_i$.

La probabilité $P(x_i | \bar{\lambda}_k)$ est obtenue en additionnant le produit de $P(x_i, S | \bar{\lambda}_k)$ par $P(x_i | S, \bar{\lambda}_k)$ sur toutes les séquences de MF possibles $S = \{S_1, S_2, \dots, S_M\}$:

$$P(x_i | \bar{\lambda}_k) = \sum_S P(x_i | S, \bar{\lambda}_k) \cdot P(S | \bar{\lambda}_k) \tag{2.68}$$

La probabilité $P(x_i | \bar{\lambda}_k)$ peut être calculée comme suit :

$$P(x_i | \bar{\lambda}_k) = \sum_{j=1}^M \alpha_t^{(k)}(j) \beta_t^{(k)}(j) = \sum_{j=1}^M \alpha_T^{(k)}(j), \quad \begin{matrix} 1 \leq i \leq N \\ 1 \leq k \leq M \end{matrix} \tag{2.69}$$

La classification par les modèles de Markov cachés est représentée dans la figure 2.20.

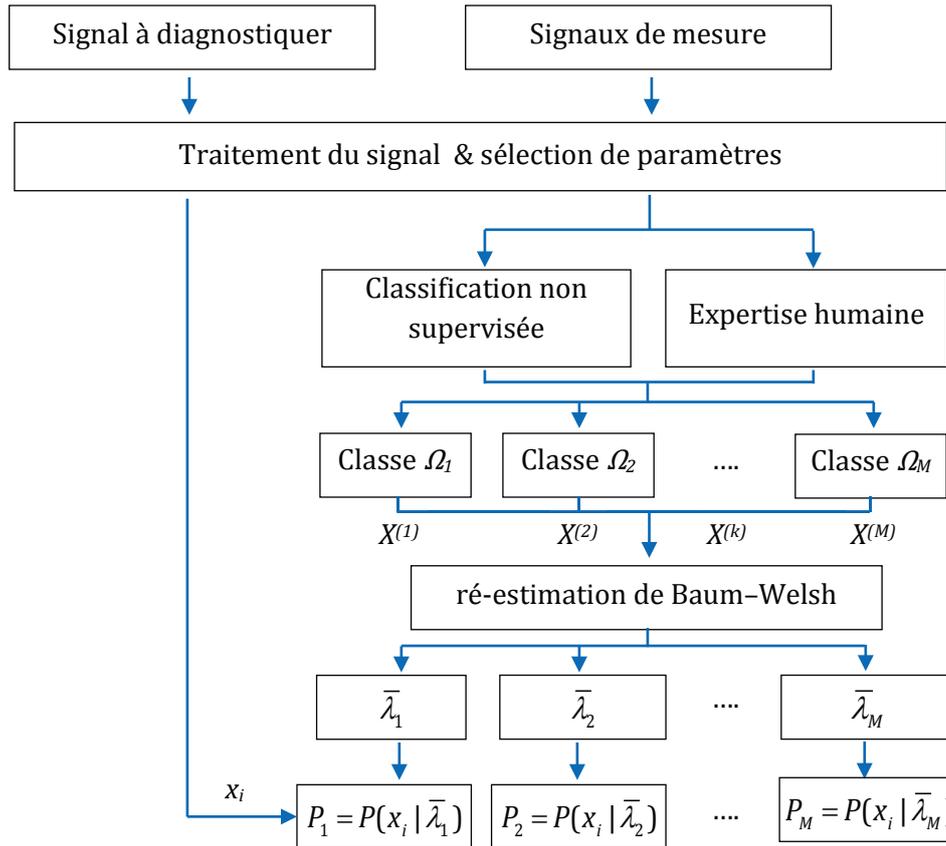


Figure 2. 20. Organigramme d'une classification supervisée par MMCs

L'architecture proposée est divisée en quatre parties. La première partie extrait des signaux de mesure à partir des capteurs placés sur les composants susceptibles de générer un défaut. La deuxième partie effectue un traitement du signal sur les signaux mesurés afin d'en extraire des paramètres représentatifs des défauts étudiés. À ce niveau, une sélection de paramètres est nécessaire afin de diminuer le nombre de paramètres calculés et éviter une information surdéterminée qui diminuerait les performances du classifieur. La troisième partie projette les observations mesurées sur un espace à q -dimensions. Le regroupement de ces classes est effectué soit par expertise humaine ou par des méthodes de classification non supervisée. Ces classes représentent les modes de fonctionnement du système sous surveillance. La quatrième et dernière partie effectue l'apprentissage de chacune de ces classes par un modèle de Markov caché. Les modèles obtenus sont ensuite utilisés pour classer les nouvelles observations. Le MMC pour lequel la probabilité de générer cette nouvelle observation est maximale déterminera le mode de fonctionnement (la classe) du système.

2.4. Evaluation des méthodes proposées

Nous présentons dans cette section les résultats obtenus par l'algorithme FA ainsi que celui des modèles de Markov cachés sur une base de données test appelée « Iris ». L'algorithme FA et les MMCs seront ensuite comparés à d'autres méthodes de classification supervisée et non-supervisée.

Lorsque nous utilisons un classificateur pour le diagnostic, les données d'apprentissage, en général, ne représentent pas parfaitement les données réelles que nous avons à classer. La base de données peut être incomplète ou ne pas couvrir tous les modes de fonctionnement du système. Nous avons testé nos méthodes sur le problème de classification des données d'Iris [Fisher_36]. Cet ensemble de données, comme montré dans la figure 2.21, est composé d'un ensemble de 150 observations correspondant à trois classes de fleurs : l'iris setosa, l'iris versicolor et l'iris virginica. Quatre paramètres sont utilisés pour caractériser chaque type de fleur : la largeur et la longueur des pétales et sépales.

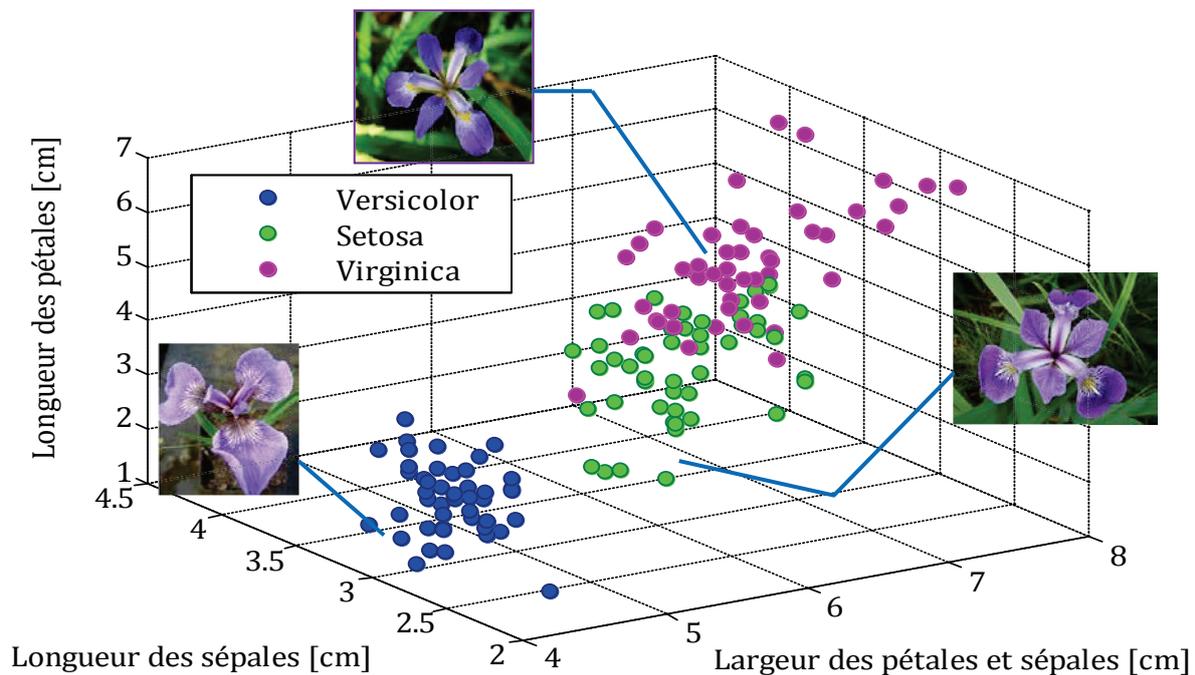


Figure 2. 21. Répartition des observations d'iris

L'objectif est d'évaluer les performances de l'algorithme FA et des MMCs ainsi que l'influence du manque d'informations sur les résultats de classification. À cette fin, nous simulons un ensemble d'observations incomplet en utilisant que 50% des observations de l'iris comme base d'apprentissage. Les observations restantes seront utilisées pour mesurer les performances (taux d'erreur de classification) de l'algorithme FA et des MMCs par rapport au K-means, le système neuro-flou (adaptive neuro fuzzy inference system) et le réseau de neurones non bouclé «Feed-Forward».

Le réseau de neurones Feed-Forward est un réseau de neurones artificiels formé par une couche d'entrée qui contient les neurones imposés de l'extérieur, un nombre de couches cachées et une couche de sortie qui donne le résultat désiré. Dans le cas d'un problème de classification supervisée, la couche de sortie donne la classe d'appartenance des observations. Ce type de réseau requiert une étape d'apprentissage où les poids de chaque neurone sont optimisés, et une étape de validation [Svozil_97]. L'interface graphique *nprtool* (Neural Network Pattern Recognition tool Graphical User Interface) de Matlab © 2010b est utilisée afin d'effectuer

l'apprentissage du réseau de neurones. L'apprentissage est effectué par la méthode de descente du gradient. Le réseau en question est composé de deux couches cachées avec des fonctions de type sigmoïdes dans les deux couches cachées et la couche sortie. Il est appelé *patternnet* (Pattern recognition network) dans Matlab ®.

La méthode « Mountain clustering/ANFIS » proposée par Chiu [Chiu_94] est une méthode de classification supervisée basée sur un algorithme de classification non supervisée. Les centres de gravité des classes sont utilisés pour initialiser un système neuro-flou qui est formé par ANFIS, un système d'inférence flou à base de réseaux de neurones adaptatifs, proposé par Jang [Jang_93]. Une explication plus détaillée sur le système ANIFS est donnée dans le chapitre 3, section 3.3.1.

2.4.1. Algorithme fourmis artificielles

Reprenons les différentes étapes de l'algorithme fourmis artificielles :

- *Etape 1 (voir section 2.3.1.4.3)*

La figure 2.22 montre l'évolution du critère J en fonction de différentes valeurs de seuils de similarité (S_{sim}).

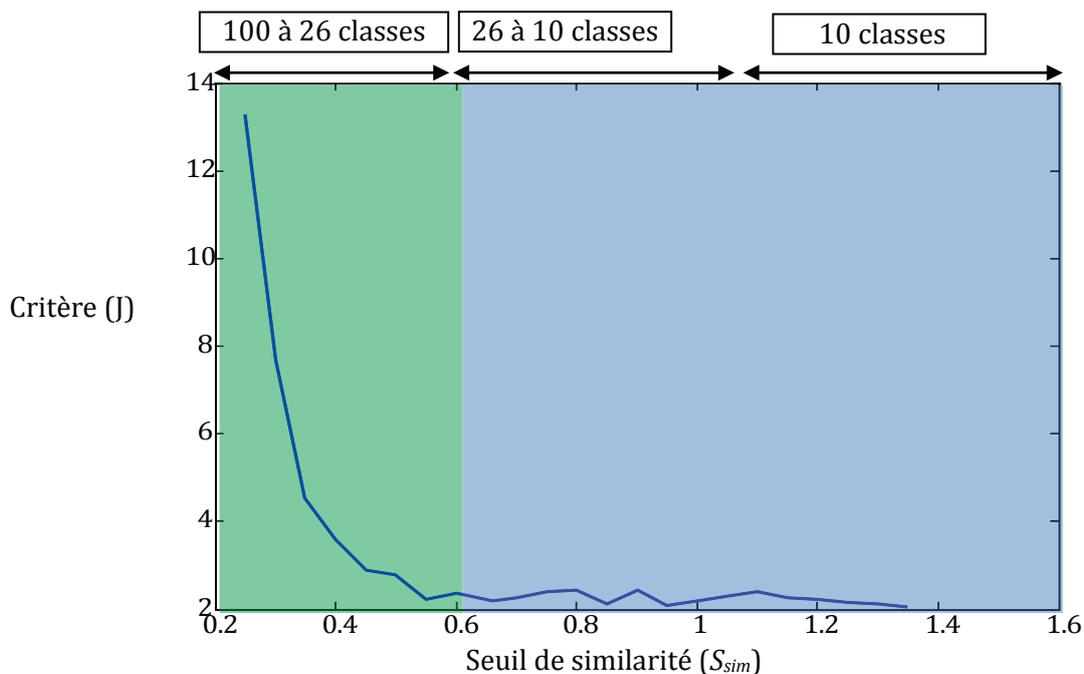


Figure 2. 22. Evolution du critère J en fonction du seuil de similarité

Il faut noter qu'une bonne classification correspond à un critère J élevé et ceci pour un nombre de classes assez faible. La formule (2.5) permet d'interpréter ceci par une grande variance inter-classes et une faible variance intra-classe. À partir de la figure 2.22, nous avons distingué deux zones. Dans la première zone, le critère J décroît de 13,3 à 2,19. Cette décroissance correspond à un regroupement d'observations qui varie entre 100 classes pour $J = 13,3$ et 26 classes pour $J = 2,19$. Il est normal que le critère J pour 100 classes soit élevé (ici 13,3) car la plupart des classes ne contiennent qu'une seule observation. La variance intra-classe dans ce cas est voisine de 0 ce qui donne une valeur élevée pour J. En revanche, plus le nombre de classes diminue plus J diminue car la variance intra-classe augmente. Ceci nous place dans la deuxième zone où $J \in [2,19 - 2,43]$. Un seuil de similarité fixé à 0,25 risque d'être

pénalisant en temps de calcul surtout pour déterminer les paramètres (e et f) de l'algorithme (voir étape 2). Un seuil de similarité fixé à 1,1 ne nous permet pas d'optimiser la classification car le nombre de classes ne change pas au-delà de cette valeur. Le seuil de similarité retenu pour augmenter le critère J est fixé à 0,6, ce qui nous permettra par la suite d'améliorer la classification.

Un seuil de similarité fixé à 0,6 permet à la première partie de l'algorithme fourmis artificielles d'obtenir un premier ensemble de m -classes, avec $m=26$.

- Etape 2 (voir section 2.3.1.4.3)

La deuxième partie de l'algorithme fourmis artificielles nécessite différents paramètres à régler. Elle comprend la quantité de phéromone ξ , l'évaporation de la phéromone ε et les paramètres e et f qui contrôlent les poids des classes ainsi que la dispersion interclasses.

Le paramètre ε est fixé à 0,01 afin limiter le poids de chaque classe. Le paramètre ξ est arbitrairement fixé à 1. La détermination des paramètres e et f peut être obtenue à partir du calcul de J sur un maillage des valeurs de e et f (voir formule (2.5)). À partir de valeurs de e et f choisies dans l'intervalle $[-1, 1]$, le critère J le plus élevé définit la meilleure valeur du paramètre e et f . Pour J avoisinant les 100%, on obtient : $e = 0,5$ et $f = 0,8$ (voir figure 2.23).

Cette deuxième étape permet à l'algorithme des fourmis artificielles de rassembler les classes précédentes en un nombre de classes plus restreint (ici 2).

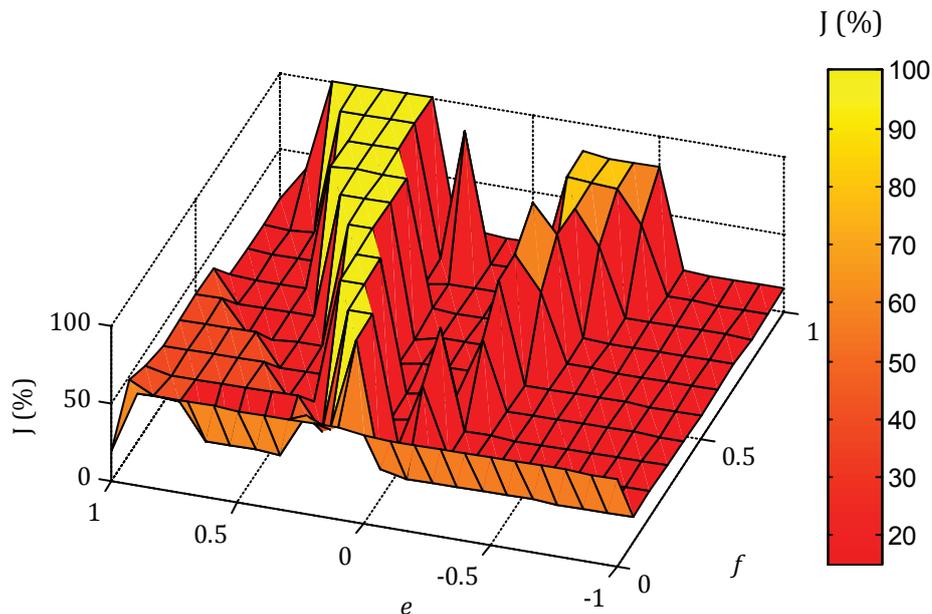


Figure 2. 23. Détermination des paramètres e et f à partir du critère J

- Etape 3 (voir section 2.3.1.4.3)

La détermination des paramètres e et f , dans la deuxième étape de l'algorithme des fourmis artificielles, a permis d'obtenir un deuxième ensemble de m' -classes égal à 2. Ce deuxième ensemble de classes est ensuite subdivisé en un ensemble de M -classes en faisant appel à l'analyse en composantes principales puis à l'algorithme K-means. L'analyse en composantes principales est utilisée afin de réduire la dimension du vecteur de paramètres ce qui a pour

avantage d'obtenir une meilleure distinction entre les classes en augmentant la dispersion interclasses. Les figures 2.24 et 2.25 montrent la projection des observations avant et après l'ACP.

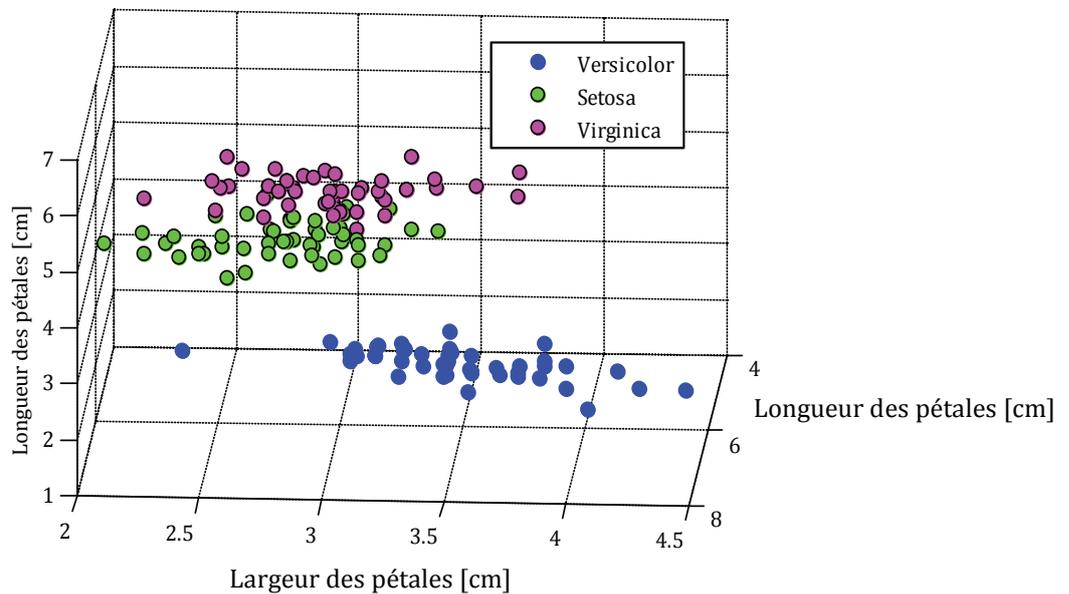


Figure 2. 24. Répartition des observations d'iris avant l'ACP

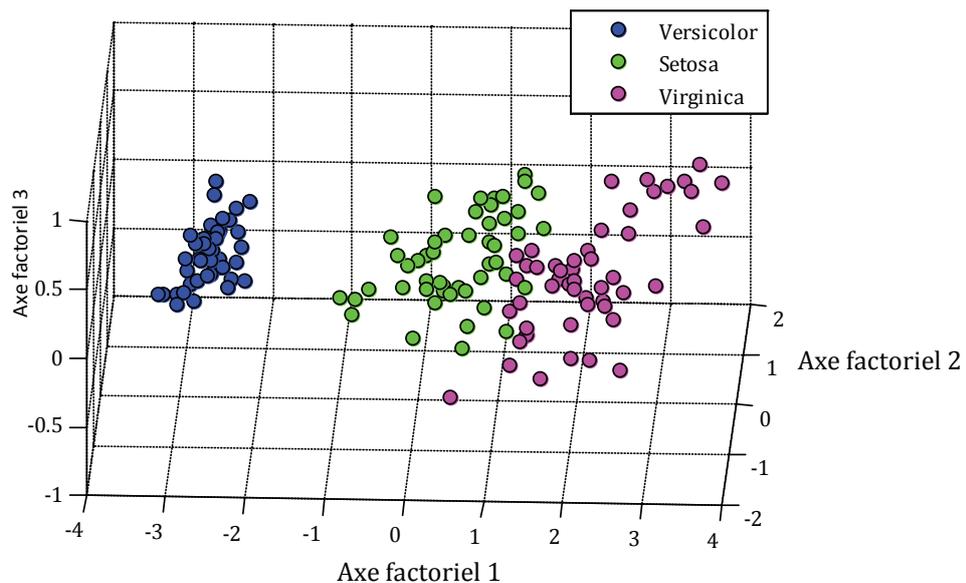


Figure 2. 25. Répartition des observations d'iris après une ACP

La première classe obtenue par l'algorithme des fourmis artificielles regroupe les observations de la fleur « Versicolor ». Cette classe est compacte et donc non séparable par l'algorithme K-means. Ceci est confirmé par le calcul du critère J indiquant que la subdivision de cette classe induit une diminution de 12% au niveau du critère J. La seconde classe regroupe les observations de la fleur « Setosa » et « Virginica ». Si on se réfère à la figure 2.24, nous pouvons observer un fort chevauchement entre les deux types de fleurs. Cependant, la réduction du vecteur de paramètres par l'ACP permet d'avoir une meilleure distinction entre les deux types de fleurs comme le montre la figure 2.25. La subdivision de cette classe par l'algorithme K-means induit une augmentation de 19% au niveau du critère J. Ceci permet d'obtenir le dernier ensemble de M -classes, avec $M=3$, qui correspond aux trois iris.

Résultats de classification pour les approches non-supervisées

Erreur de classification

Un critère de qualité (E_c %) est utilisé afin d'évaluer la répartition des observations obtenues par nos méthodes. Ce critère mesure le taux d'erreur de classification en calculant le rapport entre le nombre d'observations mal classées par rapport aux classes réelles et le nombre total d'observations utilisées. Le taux d'erreur de classification est formulé comme suit :

$$E_c \% = 100 \times \frac{\text{Observations mal classées}}{\text{Nombre total d'observations utilisées}} \quad (2.70)$$

La répartition des observations dans les classes obtenues par l'algorithme des fourmis artificielles est donnée comme suit, avec $E_c=1,33\%$.

Classement obtenu	Ω_1	50 0%	0 0%	0 0%
	Ω_2	0 0%	48 0%	0 0%
	Ω_3	0 0%	2 1,33%	50 0%
		Ω_1	Ω_2	Ω_3
	Classement réel			

La répartition des observations dans les classes obtenues par l'algorithme K-means est donnée comme suit, avec $E_c=10,66\%$.

Classement obtenu	Ω_1	50 0%	0 0%	0 0%
	Ω_2	0 0%	48 0%	14 9,33 %
	Ω_3	0 0%	2 1,33%	36 0%
		Ω_1	Ω_2	Ω_3
	Classement réel			

La répartition des observations dans les classes obtenues par la classification ascendante hiérarchique avec le critère de Ward est donnée comme suit avec $E_c=10,66\%$.

Classement obtenu	Ω_1	50 0%	0 0%	0 0%
	Ω_2	0 0%	48 0%	14 9,33 %
	Ω_3	0 0%	2 1,33%	36 0%
		Ω_1	Ω_2	Ω_3
	Classement réel			

2.4.2. Modèles de Markov cachés

Pour la classification par les modèles de Markov cachés, nous devons déterminer 03 MMCs correspondant aux 03 classes d'iris. Dans ce cas, les modes de fonctionnement du système sont respectivement : l'iris Versicolor (S_1), l'iris Setosa (S_2) et l'iris Virginica (S_3). À partir de chaque classe, nous extrayons une séquence d'observations de longueur ($T=3$). La longueur T correspond à un tirage successif de quatre observations de chacune des classes. Nous verrons par la suite (section 2.4.3) que la longueur minimale d'une séquence d'observations pour l'apprentissage des MMCs est égale à 3.

Les séquences d'observations utilisées pour estimer les paramètres des MMCs peuvent être formulées par : $X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_T^k\}$, avec $1 \leq k \leq 3$ et $T = 3$. Les paramètres initiaux de chaque MMC (matrice de transition $A^{(k)}$ et vecteur d'état initial $\pi^{(k)}$), à partir des quels nous commencerons l'apprentissage, sont définis comme suit :

$$A^{(1,2,3)} = \begin{matrix} & \Omega_1 & \Omega_2 & \Omega_3 \\ \Omega_1 & \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \\ \Omega_2 & \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \\ \Omega_3 & \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \end{matrix} \quad (2.71)$$

$$\pi^{(1,2,3)} = \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \\ \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \end{matrix}$$

La matrice de probabilité d'observations pour les séquences X^1, X^2, X^3 est donnée comme suit :

$$b_j(X^k) = 1 - \frac{(X^k - g_j) Cov^{-1}(E_j) (X^k - g_j)'}{\sum_{\ell=1}^M (X^k - g_\ell) Cov^{-1}(E_\ell) (X^k - g_\ell)'} \quad (2.72)$$

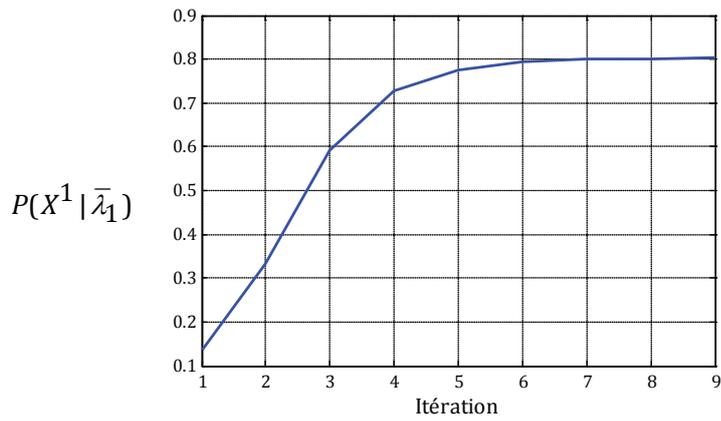
avec

- E_j : ensemble d'observations d'apprentissage appartenant à la classe Ω_j ;
- g_j : centre de gravité de la classe Ω_j ;
- $(.)'$: matrice transposée. La matrice de covariance Cov est estimée à partir de chaque classe.

Les figures 2.26 (a, b, c) montrent l'évolution de la probabilité $P(X^k | \bar{\lambda}_k)$ durant la phase d'apprentissage et cela pour chaque MMC.

Par exemple, pour la figure 2.26 (a), la phase d'apprentissage commence avec une probabilité $P(X^1 | \bar{\lambda}_1)$ égale à 0,13. Ceci indique que la matrice de transition A^1 et le vecteur d'état initial π^1 que nous avons initialement choisi ne représentent pas la séquence d'observations extraite de la classe Ω_1 . Pour augmenter cette probabilité, nous utilisons les formules de re-estimation données dans la section 2.3.2.3.2. Nous obtenons ainsi un accroissement de la probabilité $P(X^1 | \bar{\lambda}_1)$ jusqu'à 0,8 en 9 itérations. Au-delà de 9 itérations la probabilité $P(X^1 | \bar{\lambda}_1)$ ne varie plus. Ce résultat indique que les paramètres du modèle $\bar{\lambda}_1$ représentent parfaitement le mode de fonctionnement étudié (l'iris Versicolor).

Les figures 2.27 (a, b, c) montrent les probabilités générées par l'affectation des 150 observations d'iris lorsque les 03 MMCs ($\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3$) sont utilisés pour la classification.

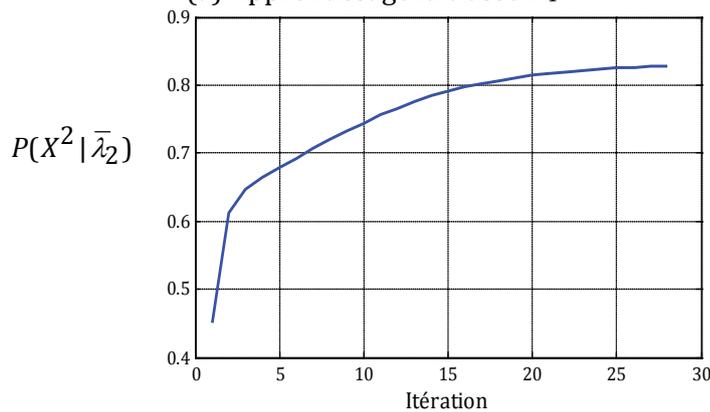


$$\bar{A}^1 = \begin{bmatrix} 0,9748 & 0,0198 & 0,0054 \\ 0,9762 & 0,0186 & 0,0051 \\ 0,9761 & 0,0187 & 0,0052 \end{bmatrix}$$

$$\bar{\pi}^1 = [0,9868 \quad 0,0103 \quad 0,0030]$$

(a) Apprentissage la classe Ω_1

MMC de l'iris Versicolor

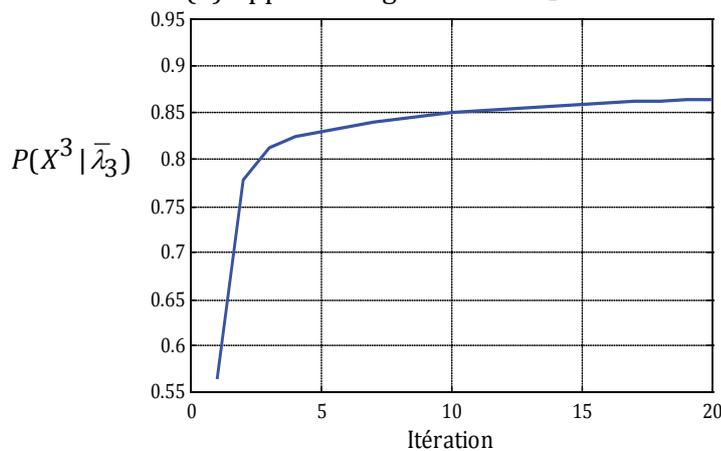


$$\bar{A}^2 = \begin{bmatrix} 0,0001 & 0,6147 & 0,3852 \\ 0,0001 & 0,6098 & 0,3901 \\ 0,0001 & 0,6093 & 0,3906 \end{bmatrix}$$

$$\bar{\pi}^2 = [0,0003 \quad 0,6360 \quad 0,3637]$$

(b) Apprentissage la classe Ω_2

MMC de l'iris Setosa



$$\bar{A}^3 = \begin{bmatrix} 0,0002 & 0,4126 & 0,5871 \\ 0,0002 & 0,4126 & 0,5857 \\ 0,0002 & 0,4140 & 0,5858 \end{bmatrix}$$

$$\bar{\pi}^3 = [0,0009 \quad 0,4038 \quad 0,5953]$$

(c) Apprentissage la classe Ω_3

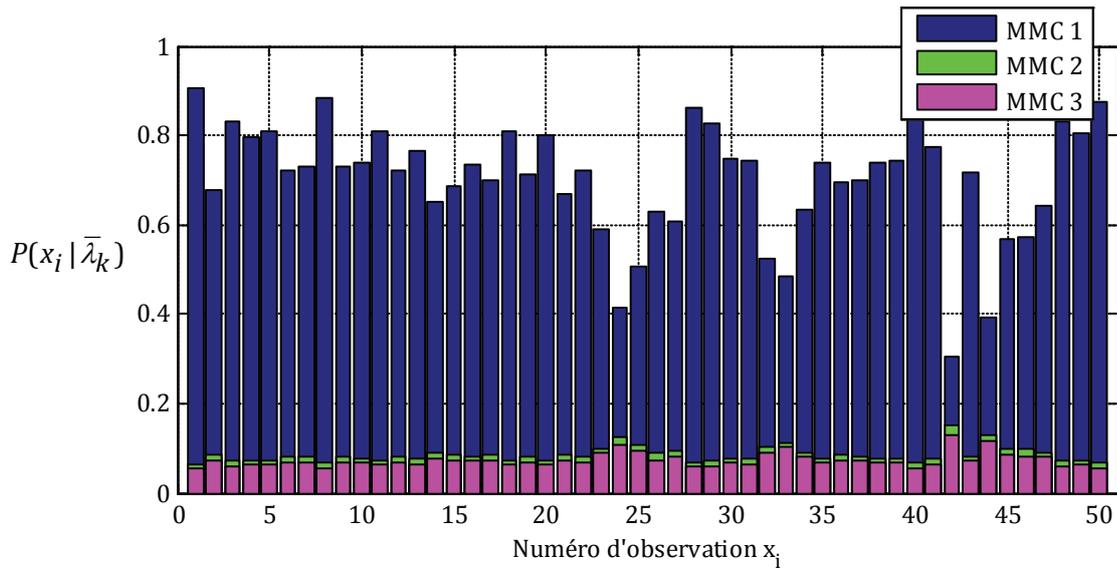
MMC de l'iris Virginica

Figure 2. 26. Processus d'apprentissage des classes par ré-estimation de Baum-Welsh

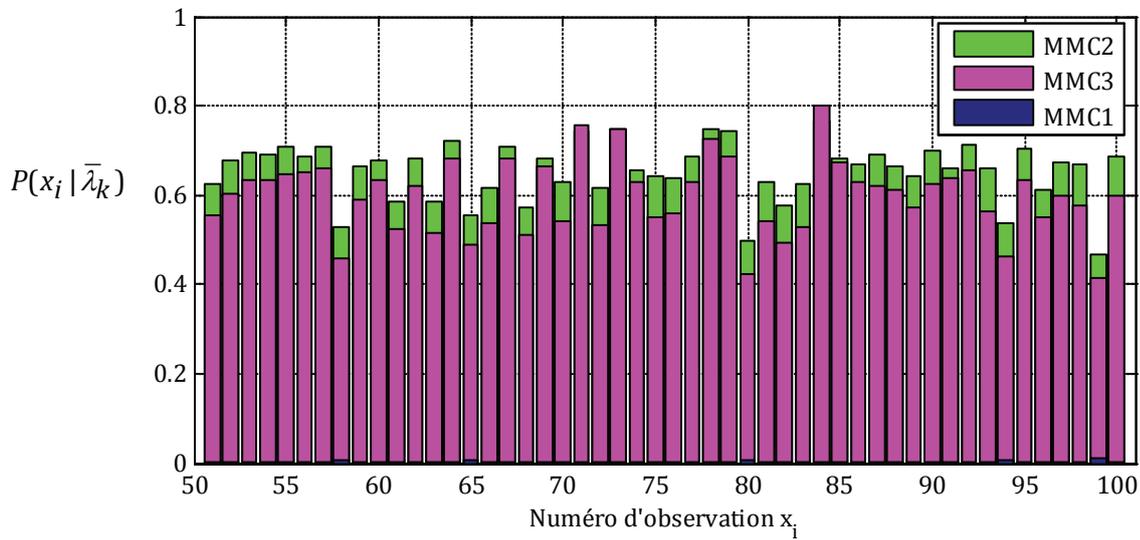
Les 50 premières observations représentent la fleur d'iris Versicolor. Les probabilités générées par l'affectation de ces 50 observations varient entre 0,35 et 0,9 pour le modèle $\bar{\lambda}_1$ et entre 0,08 et 0,15 pour $\bar{\lambda}_2$ et $\bar{\lambda}_3$. Ceci nous permet d'affecter les 50 premières observations à la classe Ω_1 .

Les 50 observations suivantes représentent la fleur d'iris Setosa. Les probabilités générées par l'affectation de ces 50 observations varient entre 0,5 et 0,78 pour le modèle $\bar{\lambda}_2$ et entre 0,4 et 0,8 pour le modèle $\bar{\lambda}_3$ indiquant un fort chevauchement entre la classe Ω_2 et Ω_3 . Pour le modèle $\bar{\lambda}_1$ les probabilités sont quasi nulles. Ceci nous permet d'affecter 48 observations à la classe Ω_2 (Voir figure 2.27(b)).

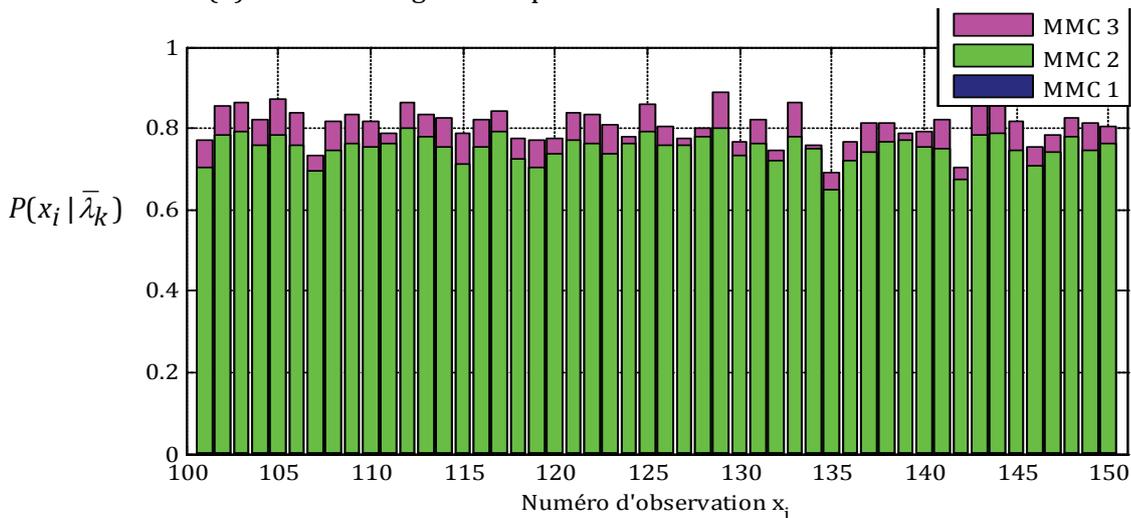
Les 50 dernières observations de la fleur d'iris Virginica sont toutes affectées à la classe Ω_3 .



(a) Probabilités générées par l'affectation des 50 observations du Versicolor



(b) Probabilités générées par l'affectation des observations du Setosa



(c) Probabilités générées par l'affectation des observations du Virginica

Figure 2. 27. Classification des observations d'iris par les modèles de Markov cachés

La répartition des observations dans les classes obtenues par les MMCs est donnée comme suit, avec $E_c=1,33\%$.

Classement obtenu	Ω_1	50 0%	0 0%	0 0%
	Ω_2	0 0%	48 0%	0 0%
	Ω_3	0 0%	2 1,33%	50 0%
		Ω_1	Ω_2	Ω_3

Classement réel

La répartition des observations dans les classes obtenues par la méthode «Mountain clustering/ANFIS» est donnée comme suit, avec $E_c=2,66\%$.

Classement obtenu	Ω_1	50 0%	0 0%	0 0%
	Ω_2	0 0%	48 0%	2 1,33%
	Ω_3	0 0%	2 1,33%	48 0%
		Ω_1	Ω_2	Ω_3

Classement réel

La répartition des observations dans les classes obtenues par le réseau de neurones « Feed-Forward » est donnée comme suit avec $E_c=2\%$.

Classement obtenu	Ω_1	50 0%	0 0%	0 0%
	Ω_2	0 0%	48 0%	1 0,67%
	Ω_3	0 0%	2 1,33%	49 0%
		Ω_1	Ω_2	Ω_3

Classement réel

2.4.3. Comparaison des méthodes de classification

Le tableau 2.1 résume les résultats obtenus par les méthodes de classification mentionnées précédemment. Les algorithmes ayant le plus faible taux d'erreur de classification sont indiqués en gras. En ce qui concerne la classification non-supervisée, il apparait que l'algorithme fourmis artificielles surpasse le K-means ainsi que la classification ascendante hiérarchique en termes d'erreur de classification. Ceci est normal car l'algorithme « fourmis artificielles » est une

optimisation de la classification ascendante hiérarchique couplée avec une analyse en composantes principales et le K-means. Il utilise une fonction de similarité basée sur les paramètres e et f pour une meilleure affectation des observations et utilise le critère J pour déterminer le nombre de classes et juger de la qualité du partitionnement.

Pour la classification supervisée, les MMCs indiquent le plus faible taux d'erreur de classification et ceci pour une longueur d'observation $T=2$. Il apparait clairement que plus la longueur d'observations est grande, meilleur est la classification. Ceci étant, dans la suite de notre thèse, la longueur de la séquence d'observations sera fixée à $T=3$ afin de garantir une meilleure classification des observations. En plus d'un faible E_c , la complexité des MMCs est assez faible $O(M^2T)$; ceci est démontré dans les figures 2.26 (a, b, c) avec un nombre d'itérations assez faible (<30). L'un de ses autres avantages est la lecture de ses résultats. En effet, pour le MLFF et ANFIS, le résultat de classification est donné sous forme d'une étiquette (un nombre entier dans notre cas) qui indique la classe d'appartenance d'une observation. Pour les MMCs, le résultat de classification est donné sous forme d'une probabilité qui varie entre 0 et 1. Ceci nous permet d'évaluer le degré d'appartenance d'une observation aux modèles construits et permet ainsi de rejeter les observations dont la probabilité est inférieure à un seuil défini.

Tableau 2. 1. Résultats de classification

Algorithme	Paramètres de configuration	Erreur de classification E_c %
Classification non-supervisée		
Fourmis artificielles	$e = 0,5$ $f = 0,8$ $\xi = 1$ $\varepsilon = 0,01$	1,33 %
K-means	03 classes	10,66 %
Classification ascendante hiérarchique (CAH)	Critère de ward Saut-min Saut- max	10,66% 32% 16,66%
Classification supervisée		
Modèles de Markov caches (MMCs)	$T=2$ $T=3$	1,33 %
Multi-Layer Feed-Forward network (MLFF)	Hidden neurones $n = 10$	2 %
Mountain clustering/ANFIS	Cluster radius : $R = 0,8, 0,9, 1,0$	2,67 %

2.5. Conclusion

Les méthodes de classification (supervisée et non supervisée) présentées dans ce chapitre s'inscrivent dans le cadre des méthodes basées sur les données de mesure. L'objectif de ces méthodes est de classer des observations mesurées en les comparant à des observations types. Une observation est représentée par un point dans un espace multidimensionnel à q -paramètres. Les observations types sont en réalité des ensembles de points occupant des zones géométriques de l'espace de représentation appelées « classes » où sont regroupées des

observations similaires. En termes de diagnostic, les classes correspondent aux modes de fonctionnement du système et le fait de classer une nouvelle observation revient à identifier un mode de fonctionnement du système. L'une des limitations des méthodes de surveillance consiste à travailler avec une base de données limitée sur les modes de fonctionnement. Cela pose la question de la qualité du classement ?. Les méthodes de classification proposées pour répondre à cette problématique s'effectuent en deux étapes :

- Définition de l'espace de représentation : c'est une étape très importante, puisqu'elle influence directement la qualité du système de diagnostic. Elle consiste à choisir les paramètres les plus représentatifs vis-à-vis des modes de fonctionnement du système. Ceci s'effectue par une extraction ou une sélection de paramètres pertinents. Pour la sélection de paramètres, nous avons retenu le SBS que nous avons amélioré afin de prendre en compte la corrélation entre les paramètres et ainsi assurer une meilleure sélection.
- Dans le cas d'une classification supervisée, nous avons mis au point une nouvelle fonction d'appartenance pour optimiser l'affectation des nouvelles observations. Cette fonction est basée sur les modèles de Markov cachés. Dans le cas d'une classification non supervisée, nous avons proposé un nouvel algorithme basé sur le comportement des fourmis réelles pour optimiser la classification. Les résultats obtenus montrent que les deux méthodes permettent d'améliorer la classification.

Dans le chapitre suivant nous décrirons la mise au point de nouvelles méthodes de pronostic en se basant les méthodes de diagnostic montrées précédemment.

Chapitre 3

Du diagnostic au pronostic

Sommaire

3.1. Introduction	72
3.2. Prédiction de l'imminence d'un défaut par les modèles de Markov cachés	72
3.2.1. Méthode basée sur l'identification du processus de dégradation du système	73
3.2.2. Méthodes basées sur l'analyse des séquences d'observations	80
3.3. Estimation du temps de vie restant	86
3.3.1. Prédiction de séries temporelles par ANFIS	89
3.3.2. Prédiction de séries temporelles par le neurone neo-flou (NNF)	92
3.3.3. Mesure de la performance des modèles de prédiction	94
3.3.4. ANFIS et NNF pour la prédiction de séries temporelles chaotiques	95
3.4. Conclusion	98

3.1. Introduction

Le chapitre précédant a démontré comment il était possible de diagnostiquer les différents modes de fonctionnement d'un système en utilisant des méthodes de classification. Ces méthodes ne se limitent pas qu'à diagnostiquer les défauts, elles peuvent aussi contribuer au pronostic des défauts. En effet, le pronostic peut être défini comme une extension du problème de diagnostic. Le pronostic consiste à prédire l'évolution d'une situation de défaut jusqu'à la défaillance totale du système. Ceci peut se traduire par la prédiction de l'imminence d'un défaut et l'estimation de sa durée de vie restante. Baruah et Chinnam ont souligné que les modèles de Markov cachés, utilisés dans le domaine du diagnostic, peuvent être appliqués dans le domaine du pronostic comme dans les machines d'usinage par exemple [Baruah_05]. [Sam_08] *et all* ont présenté le modèle ANFIS (système d'inférence flou à base de réseaux de neurones adaptatifs) dédié au pronostic des machines. Ces chercheurs ont combiné les réseaux de neurones et la logique floue en prenant avantage des deux méthodes (la capacité d'apprentissage de la première et la lisibilité et la flexibilité de la seconde) afin d'améliorer les performances du pronostic. L'objectif de ce chapitre est de dégager des méthodes de prédiction capables de détecter l'imminence d'un défaut et d'estimer le temps restant avant son apparition. Sur cette base, nous retenons les modèles de Markov cachés et les systèmes neuro-flous comme techniques fondatrices pour nos méthodes de pronostic.

3.2. Prédiction de l'imminence d'un défaut par les modèles de Markov cachés

Nous définissons dans ce qui suit $(\Omega_1, \Omega_2, \dots, \Omega_M)$ comme étant les M -classes qui représentent les M -stades de dégradation d'un système. La première classe (Ω_1) correspond à un système avec un fonctionnement sain et la dernière classe (Ω_M) correspond à un système dégradé (défaillant). Les classes intermédiaires correspondent à un système avec plusieurs stades de dégradation. Ces M -classes, dont le nombre M est inconnu, seront déterminées par notre méthode de classification non supervisée (voir section, 2.3.1.4.3). Chacune de ces classes dispose d'un espace géométrique (espace de représentation) composé d'un ensemble d'observations projeté dans un espace à q -dimensions (voir figure 3.1).

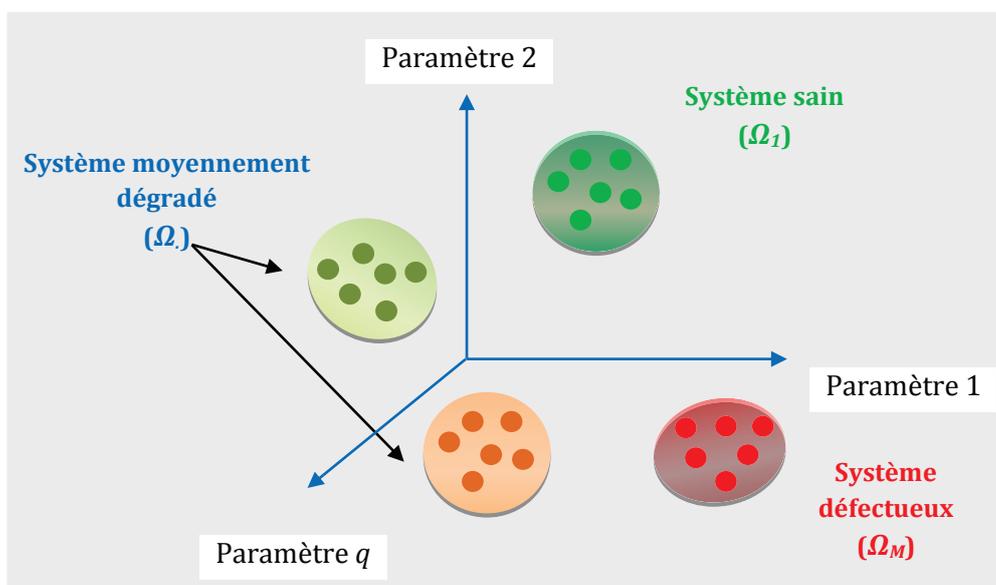


Figure 3. 1. Visualisation des différents stades de dégradation d'un système par des classes

Ces observations sont récoltées à partir de l'historique de dégradation (depuis un fonctionnement sain jusqu'à l'apparition du défaut) d'un ou plusieurs systèmes identiques. Les M -classes sont utilisées pour concevoir des méthodes de pronostic basées sur les modèles de Markov cachés afin de détecter l'imminence du prochain stade de dégradation, et donc de l'imminence d'un défaut du système. Nous avons pour cela proposé trois méthodes :

- Une méthode basée sur l'identification du processus de dégradation du système à l'étude ;
- Deux méthodes basées sur l'analyse des séquences d'observations.

3.2.1. Méthode basée sur l'identification du processus de dégradation du système

Le premier « pronostiqueur » que nous proposons repose sur l'historique de dégradation de N -systèmes identiques. Il a pour but de prédire l'imminence du prochain stade de dégradation d'un nouveau système en se basant sur des modèles de Markov cachés construits à partir des historiques de dégradation de ces N -systèmes. Dans ce cas, le processus de dégradation de chaque système (k) est modélisé par un MMC $(\bar{\lambda}_k)_{k \in \{1,2,\dots,N\}}$. Les paramètres d'un modèle $(\bar{\lambda}_k)$ sont déterminés en effectuant un apprentissage d'une séquence d'observations $X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_t^k\}$ qui représente le processus de dégradation d'un système (k) jusqu'à l'instant (t). L'apprentissage d'un MMC à partir d'une séquence d'observations X^k est obtenu par les formules de ré-estimation de Baum-Welch (voir section, 2.3.2.3.2).

Pour une nouvelle séquence d'observations $X = \{x_0, x_1, x_2, \dots, x_t\}$, mesurée à des intervalles de temps réguliers à partir d'un nouveau système, si $q_{t+1} = S_j$ $j \in \{1, \dots, M\}$ désigne le stade de dégradation d'un système à l'instant ($t+1$), et si le stade de dégradation du système $q_{t+1} = S_j$ est différent de son état à l'instant (t), $q_t = S_i$, alors nous pouvons prédire l'imminence du prochain stade de dégradation en utilisant la formule (3.3).

La première étape consiste à identifier le MMC $\bar{\lambda}_k$ $k \in \{1, 2, \dots, N\}$ le plus représentatif de la nouvelle séquence d'observations X en utilisant la formule (3.1).

$$\bar{\lambda}_k = \arg \max_{1 \leq k \leq N} [P(x_0, x_1, \dots, x_t | \bar{\lambda}_k)] = \arg \max_{1 \leq k \leq N} \left[\sum_{i=1}^M \alpha_t^{(k)}(i) \beta_t^{(k)}(i) \right] \quad (3.1)$$

Le MMC $\bar{\lambda}_k$ pour lequel la probabilité de générer la nouvelle séquence d'observations X est maximale déterminera le processus de dégradation le plus similaire à celui du nouveau système.

L'identification du modèle $\bar{\lambda}_k$ permet de calculer la probabilité $P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1} = S_j | \bar{\lambda}_k)$ qui traduit la probabilité que la nouvelle séquence d'observations $\{x_0, x_1, x_2, \dots, x_t, x_{t+1}\}$ arrive à un stade S_j à l'instant ($t+1$), avec x_{t+1} inconnue. Cela revient à déterminer parmi les M -stades de dégradation S_j $j \in \{1, \dots, M\}$ celui qui maximise cette probabilité. Si l'état futur du système $q_{t+1} = S_j$ est différent de son état à l'instant (t) $q_t = S_i$, alors nous pouvons prédire l'imminence du prochain stade de dégradation en utilisant la formule suivante :

$$\text{Stade de dégradation du système } S_j \text{ à } (t+1) = \arg \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} (P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1} = S_j | \bar{\lambda}_k)) \quad (3.2)$$

Un algorithme est proposé afin de prédire le stade de dégradation du système à $(t+1)$. Nous définissons pour cela la variable $\delta_t(i)$ comme étant la probabilité Forward normalisée du stade S_i à l'instant (t) .

Le stade de dégradation du système S_j à $(t+1)$ est donné par :

$$\text{Stade de dégradation du système } S_j \text{ à } (t+1) = \arg \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \left(\sum_{i=1}^M \delta_t^{(k)}(i) a_{ij}^{(k)} \right) \quad (3.3)$$

avec $\delta_{t+1}^{(k)}(i) = b_j(x_{t+1}) \alpha_t^{(k)}(i)$.

Démonstration

D'après la formulation de forward et backward, pour une séquence d'observations $X = \{x_0, \dots, x_T\}$ qui évolue jusqu'à l'instant (T) on a :

$$\begin{aligned} P(x_0, x_1, x_2, \dots, x_T | \bar{\lambda}_k) &= \sum_{j=1}^M P(x_0, x_1, x_2, \dots, x_t, q_t = S_j | \bar{\lambda}_k) P(x_{t+1}, \dots, x_T | q_t = S_j, \bar{\lambda}_k), & 1 \leq i, j \leq M \\ & & 1 \leq k \leq N \\ &= \sum_{j=1}^M \alpha_t^{(k)}(j) \beta_t^{(k)}(j) \end{aligned} \quad (3.4)$$

Ceci permet d'en déduire que pour une séquence d'observations $X = \{x_0, \dots, x_t, x_{t+1}\}$ (avec $T=t+1$), on aura :

$$P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1} = S_j | \bar{\lambda}_k) = \alpha_{t+1}^{(k)}(j) \quad (3.5)$$

avec $\beta_{t+1}^{(k)}(j) = 1$

La formule (2.56) nous permet de réécrire (3.5) sous la forme :

$$P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1} = S_j | \bar{\lambda}_k) = \alpha_{t+1}^{(k)}(j) = b_j(x_{t+1}) \sum_{i=1}^M \alpha_t^{(k)}(i) a_{ij}^{(k)} \quad (3.6)$$

À partir de (3.6), la probabilité pour que la séquence d'observations $\{x_0, x_1, x_2, \dots, x_t, x_{t+1}\}$ arrive au stade de dégradation S_j à l'instant $(t+1)$ est donnée en fonction des probabilités de transition $a_{ij}^{(k)}$, de la variable forward $\alpha_t^{(k)}(i)$ à l'instant (t) ainsi que de la probabilité d'émettre une observation à $(t+1)$ associée à un stade S_j , $b_j(x_{t+1})$. Pour déterminer le stade de dégradation du système $q_{t+1} = S_{j \in \{1, \dots, M\}}$ à $(t+1)$, il suffit de déterminer le stade S_j qui maximise l'expression (3.6) :

$$\text{Stade de dégradation du système } S_j \text{ à } (t+1) = \arg \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \left(b_j(x_{t+1}) \sum_{i=1}^M \alpha_t^{(k)}(i) a_{ij}^{(k)} \right) \quad (3.7)$$

Il faut noter que nous ne disposons pas de l'observation x_{t+1} . Cela n'empêche pas de déterminer l'expression (3.7) en normalisant la probabilité $b_j(x_{t+1}) = 1$ ($b_j(x_{t+1}) = 1$). Cette normalisation est interprétée par le fait que la probabilité d'émettre une observation à $(t+1)$ est maximale pour tous les stades de dégradation S_j possibles du système. Nous obtenons ainsi :

$$P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1} = S_j | \bar{\lambda}_k) = \sum_{i=1}^M \delta_t^{(k)}(i) a_{ij}^{(k)} \quad (3.8)$$

avec $\delta_0^{(k)}(i) = \pi_i^{(k)} b_i(x_0)$ et $\delta_{t+1}^{(k)}(j) = b_j(x_{t+1}) \sum_{i=1}^M \delta_t^{(k)}(i) a_{ij}^{(k)}$

d'où l'obtention de la formule(3.3).

L'algorithme pour la prédiction de l'imminence du prochain stade de dégradation à $(t+1)$ est donné comme suit :

1. *Initialisation* : $\delta_0^{(k)}(i) = \pi_i b_i(x_0)$

2. *Récurrance* :

$$\text{stade de dégradation du système } S_j \text{ à } (t+1) = \arg \max_{\substack{1 \leq j \leq M \\ 1 \leq k \leq N}} \left(\sum_{i=1}^M \delta_t^{(k)}(i) a_{ij}^{(k)} \right) \quad \begin{matrix} 1 \leq i, j \leq M \\ 1 \leq k \leq N \end{matrix}$$

$$\delta_{t+1}^{(k)}(j) = b_j(x_{t+1}) \sum_{i=1}^M \delta_t^{(k)}(i) a_{ij}^{(k)} \quad t \geq 1$$

Nous montrons dans la figure 3.2 l'organigramme de la première méthode de pronostic basée sur l'identification du processus de dégradation.

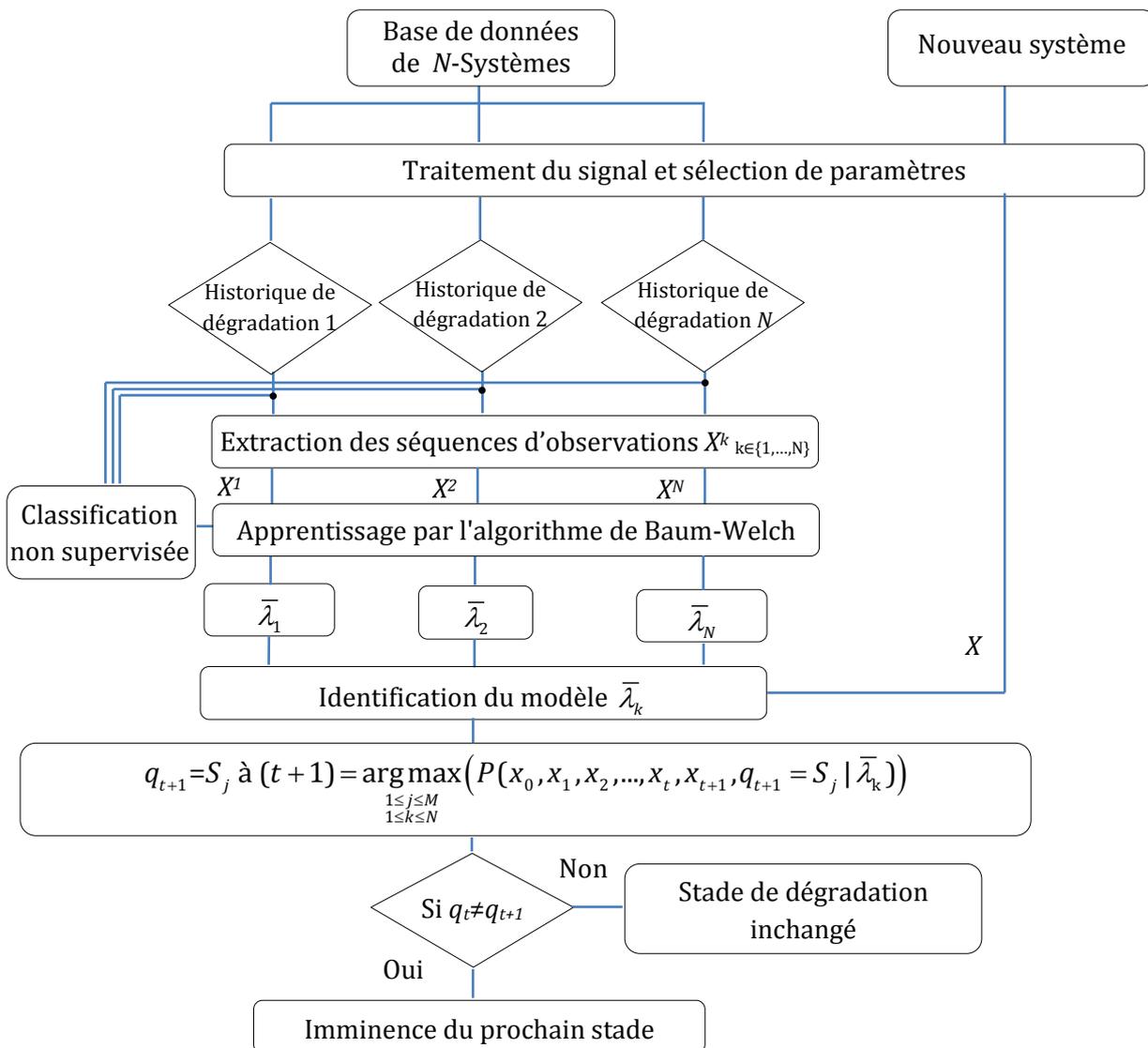


Figure 3. 2. Organigramme de la première méthode de pronostic basée sur l'indentification du processus de dégradation

Illustration

Nous présentons dans cette section un exemple d'application de la première méthode de pronostic en se basant sur l'historique de dégradation de 10 moteurs de même type installés sur une chaîne de production pour une durée (t), estimée en mois. L'espace de représentation est composé d'un ensemble de q -paramètres indicateurs de l'état de dégradation des 10 moteurs. Les mesures effectuées sur ces moteurs permettent de distinguer trois stades de dégradation : un stade de dégradation mineur, un stade de dégradation moyen et un stade de dégradation avancé (voir figure 3.3). Nous pouvons en déduire que le processus de dégradation de chaque moteur passe par ces trois stades de dégradation comme illustré dans la figure 3.4.

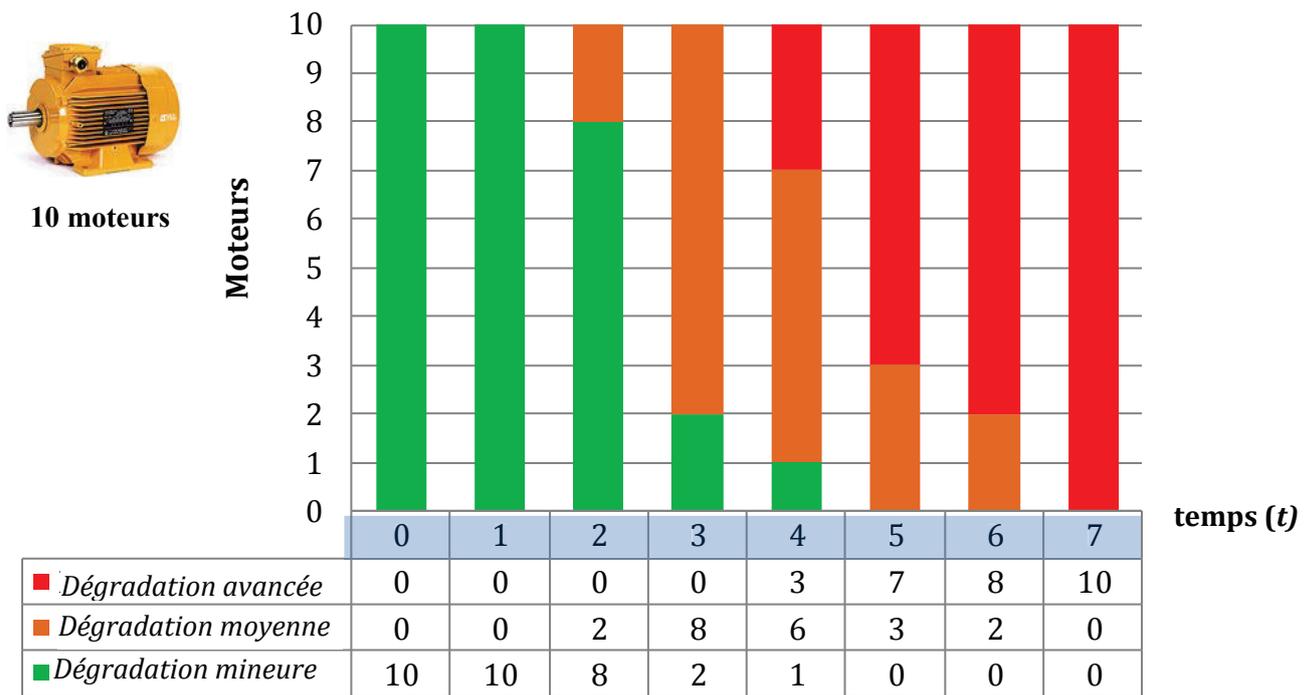


Figure 3. 3. Évolution de la dégradation en fonction du temps (mois)

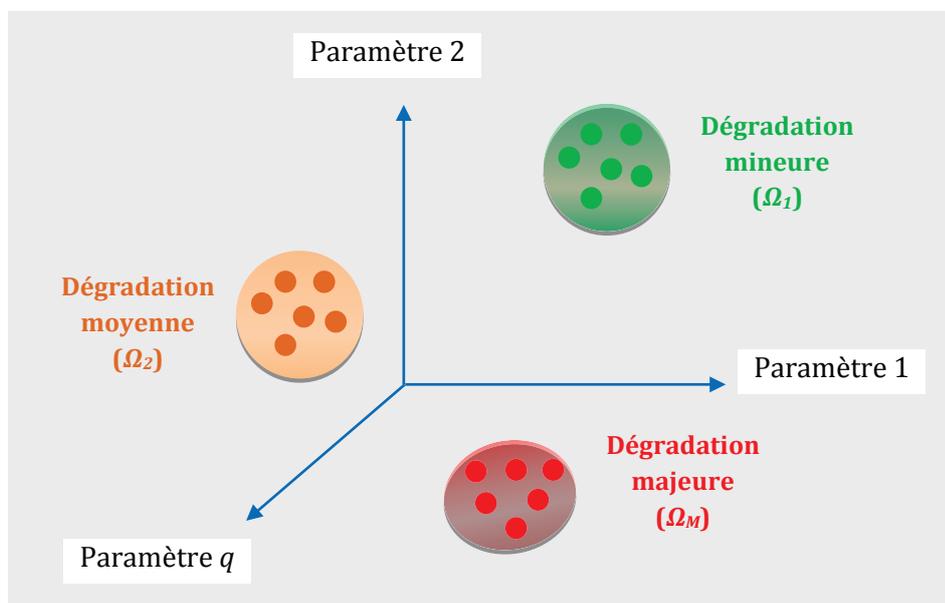


Figure 3. 4. Représentation des stades de dégradation des moteurs

Chaque stade de dégradation est représenté par une classe composée d'un ensemble d'observations projeté dans l'espace à q -dimensions. La dégradation débutera bien sûr par un stade de dégradation minime en passant par un stade de dégradation moyen pour au final terminer par un stade de dégradation avancé. La séquence d'observations extraite de chaque moteur est formulée comme suit :

$$X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_t^k\}, \quad \begin{matrix} 0 \leq t \leq 7 \\ 1 \leq k \leq 10 \end{matrix} \quad (3.9)$$

avec $x_t^k = [x_{t,1}^k, x_{t,2}^k, x_{t,3}^k, \dots, x_{t,q}^k]$ désignant une observation mesurée sur le moteur (k) à l'instant (t) décrite par le vecteur de paramètres.

Considérons une nouvelle séquence d'observations $X^{11} = \{x_0^{11}, x_1^{11}, x_2^{11}, x_3^{11}, x_4^{11}, x_5^{11}, x_6^{11}\}$ extraite de mesures effectuées sur un nouveau moteur jusqu'à l'instant $t=6$, numéroté 11, identique aux précédents moteurs. L'objectif du pronostiqueur sera d'évaluer l'imminence du prochain stade de dégradation à l'instant $t+1=7$.

Pour pronostiquer l'imminence d'un stade de dégradation sur un nouveau moteur, il nous faut tout d'abord construire un modèle de Markov caché pour chacun des dix moteurs en s'appuyant sur leurs historiques de dégradation. Pour cela, il nous faut : définir les trois stades de dégradation, notons les $S = \{S_1, S_2, S_3\}$, la matrice de transition A (3×3) entre ces stades et aussi une matrice d'observation B donnée par la formule(3.10).

Prenons comme exemple le moteur 1, dont l'indice $k=1$ et dont le processus de dégradation, représentée sur la figure 3.5, est la même que celui du moteur 11.

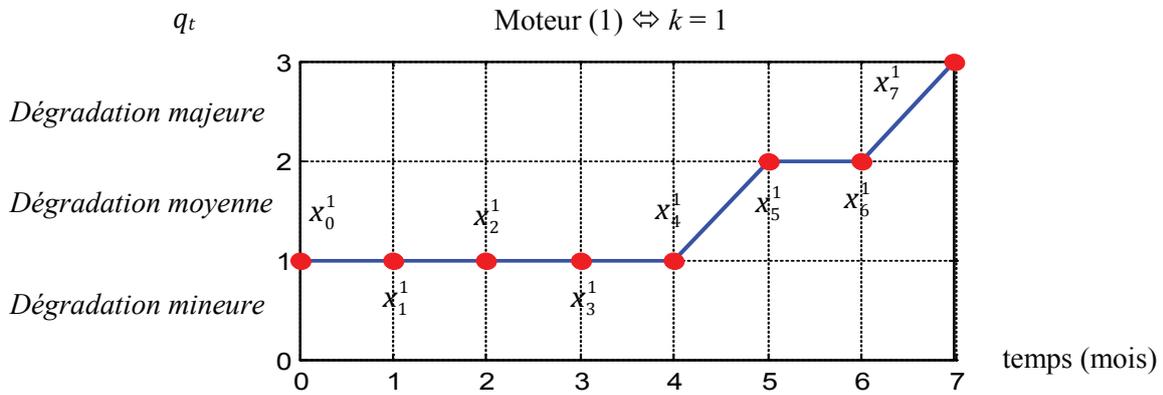


Figure 3. 5. Évolution de la dégradation (q_t) du moteur n°1

La séquence d'observations qui en est extraite jusqu'à l'instant $t=6$ est formulée comme suit :

$$X^1 = \{x_0^1, x_1^1, x_2^1, x_3^1, x_4^1, x_5^1, x_6^1\}$$

La matrice d'observation B est donnée par (3.10) et est résumée dans le tableau 3.1.

$$b_j(x_t^k) = P(x_t^k | q_t = S_j) = 1 - \frac{(x_t^k - g_j) \text{Cov}^{-1}(E_j) (x_t^k - g_j)'}{\sum_{\substack{1 \leq j \leq 3 \\ k=1 \\ 0 \leq t \leq 6}} (x_t^k - g_\ell) \text{Cov}^{-1}(E_\ell) (x_t^k - g_\ell)'} \quad (3.10)$$

avec

- E_j : ensemble d'observations extraites de la classe Ω_j ;
- g_j : centre de gravité de la classe Ω_j ;
- $(.)'$ désigne une matrice transposée. La matrice de covariance Cov est estimée à partir de chaque classe.

Tableau 3. 1. Matrice d'observation B de la séquence d'observations X^I

Moteur 1 ($k=1$)	x_0^1	x_1^1	x_2^1	x_3^1	x_4^1	x_5^1	x_6^1
Dégradation mineur $b_1(x_t^1) = P(x_t^1 q_t = S_1)$	0,8	0,7	0,6	0,5	0,5	0,3	0,1
Dégradation moyenne $b_1(x_t^1) = P(x_t^1 q_t = S_2)$	0,2	0,2	0,3	0,4	0,4	0,5	0,5
Dégradation majeure $b_1(x_t^1) = P(x_t^1 q_t = S_3)$	0	0,1	0,1	0,1	0,1	0,2	0,4

Les paramètres initiaux $(A^{(1)}, \pi^{(1)})$ du modelé λ_1 qui représentera le processus de dégradation du moteur 1 sont donnés comme suit :

$$A^{(1)} = \begin{matrix} & \Omega_1 & \Omega_2 & \Omega_3 \\ \Omega_1 & \begin{pmatrix} 0,5 & 0,25 & 0,25 \end{pmatrix} \\ \Omega_2 & \begin{pmatrix} 0 & 0,5 & 0,5 \end{pmatrix} \\ \Omega_3 & \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \end{matrix}, \quad \pi^{(1)} = \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \\ \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \end{matrix}$$

Notons que les probabilités de transition a_{32}, a_{31}, a_{21} sont fixées à zéros. Dans notre cas, la classe Ω_1 correspond à un système sain et Ω_2, Ω_3 correspondent à des dégradations progressives du système. Dans ce cas, il est normal de fixer à zéros a_{32}, a_{31}, a_{21} car nous ne prenons pas en compte le fait qu'un système dégradé puisse revenir à un fonctionnement normal.

A partir de la formulation de forward-backward (2.67) et (2.66), on obtient une ré-estimation de la matrice de transition $\bar{A}^{(1)}$ et du vecteur d'état initial $\bar{\pi}^{(1)}$:

$$\bar{A}^{(1)} = \begin{matrix} & \Omega_1 & \Omega_2 & \Omega_3 \\ \Omega_1 & \begin{pmatrix} 0,7 & 0,2 & 0,1 \end{pmatrix} \\ \Omega_2 & \begin{pmatrix} 0 & 0,65 & 0,35 \end{pmatrix} \\ \Omega_3 & \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \end{matrix}, \quad \bar{\pi}^{(1)} = \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \\ (0,7 & 0,2 & 0,1) \end{matrix}$$

Pour appliquer la méthode de pronostic, il nous faut d'abord identifier le MMC $\bar{\lambda}_k$ pour lequel la probabilité de générer la séquence d'observations X^{II} est maximale.

L'identification du MMC est obtenue en utilisant la formule (3.1) qui est donnée comme suit :

$$\bar{\lambda}_k = \arg \max_{1 \leq k \leq 10} \left[P(x_0^{11}, x_1^{11}, x_2^{11}, x_3^{11}, x_4^{11}, x_5^{11}, x_6^{11} | \bar{\lambda}_k) \right] = \arg \max_{1 \leq k \leq 10} \left[\sum_{i=1}^M \alpha_t^{(k)}(i) \beta_t^{(k)}(i) \right] \quad (3.11)$$

La matrice d'observation B qui correspond à cette nouvelle séquence d'observations X^{II} est donnée dans le tableau 3.2.

Tableau 3. 2. Matrice d'observation B de la séquence d'observations X^{11}

Moteur 11 ($k=11$)	x_0^{11}	x_1^{11}	x_2^{11}	x_3^{11}	x_4^{11}	x_5^{11}	x_6^{11}
Dégradation mineur $b_1(x_t^{11}) = P(x_t^{11} q_t = S_1)$	0,75	0,6	0,5	0,45	0,35	0,3	0,1
Dégradation moyenne $b_1(x_t^{11}) = P(x_t^{11} q_t = S_2)$	0,2	0,3	0,35	0,4	0,45	0,5	0,47
Dégradation majeure $b_1(x_t^{11}) = P(x_t^{11} q_t = S_3)$	0,05	0,1	0,15	0,15	0,2	0,2	0,43

Le modèle $\bar{\lambda}_1$ du moteur n°1, qui est le plus représentatif de la séquence d'observations X^{11} avec une probabilité $P(X^{11} | \bar{\lambda}_1) = 1,63 \times 10^{-3}$, permet de prédire l'imminence du prochain stade de dégradation à $(t+1=7)$ en calculant la probabilité que la séquence d'observations X^{11} arrive à un stade S_j à l'instant $(t+1)$. Cela revient à déterminer parmi les M -stades de dégradation S_j $j \in \{1, \dots, M\}$, celui qui maximise la formule (3.12).

$$P(x_0^{11}, x_1^{11}, x_2^{11}, x_3^{11}, x_4^{11}, x_5^{11}, x_6^{11}, x_{t+1}^{11}, q_{t+1} = S_j | \bar{\lambda}_1) = \sum_{i=1}^M \delta_t^{(1)}(i) a_{ij}^{(1)} \quad (3.12)$$

Le tableau 3.3 résume les variables Forward normalisées obtenues pour chaque stade de dégradation et cela jusqu'à l'instant $t=6$.

Tableau 3. 3. Variables Forward normalisées obtenues de la séquence d'observations X^{11} et cela jusqu'à l'instant $t=6$

	Dégradation mineur	Dégradation moyenne	Dégradation majeure
$t=0$	$\delta_0^1(1) = 0,525$	$\delta_0^1(2) = 0,04$	$\delta_0^1(3) = 5 \times 10^{-3}$
$t=1$	$\delta_1^1(1) = 0,2205$	$\delta_1^1(2) = 0,0393$	$\delta_1^1(3) = 7,15 \times 10^{-3}$
$t=2$	$\delta_2^1(1) = 0,077$	$\delta_2^1(2) = 0,0243$	$\delta_2^1(3) = 6,44 \times 10^{-3}$
$t=3$	$\delta_3^1(1) = 0,0242$	$\delta_3^1(2) = 0,0125$	$\delta_3^1(3) = 2,26 \times 10^{-3}$
$t=4$	$\delta_4^1(1) = 5,9 \times 10^{-3}$	$\delta_4^1(2) = 5,83 \times 10^{-3}$	$\delta_4^1(3) = 1,18 \times 10^{-3}$
$t=5$	$\delta_5^1(1) = 1,23 \times 10^{-3}$	$\delta_5^1(2) = 2,48 \times 10^{-3}$	$\delta_5^1(3) = 7,012 \times 10^{-4}$
$t=6$	$\delta_6^1(1) = 8,61 \times 10^{-5}$	$\delta_6^1(2) = 8,73 \times 10^{-4}$	$\delta_6^1(3) = 8,23 \times 10^{-4}$

Pour $t+1=7$, on a :

$$S_j(t+1) = \arg \max \left[\sum_{i=1}^M \delta_t^{(1)}(i) a_{ij}^{(1)} \right] \Leftrightarrow S_j(t+1=7) = \arg \max_{1 \leq j \leq M} \left[\sum_{i=1}^M \delta_6^{(1)}(i) a_{ij}^{(1)} \right]$$

$$\begin{aligned} \text{pour } q_{t+1} = S_1 \Rightarrow \sum_{i=1}^M \delta_6^{(1)}(i) a_{i1}^{(1)} &= \delta_6^{(1)}(1) a_{11}^{(1)} + \delta_6^{(1)}(2) a_{21}^{(1)} + \delta_6^{(1)}(3) a_{31}^{(1)} \\ &= (8,61 \times 10^{-5}) \times 0,7 + (8,73 \times 10^{-4}) \times 0 + (8,32 \times 10^{-4}) \times 0 \\ &= 6,02 \times 10^{-5} \end{aligned}$$

$$\begin{aligned} \text{pour } q_{t+1} = S_2 \Rightarrow \sum_{i=1}^M \delta_6^{(1)}(i) a_{i2}^{(1)} &= \delta_6^{(1)}(1) a_{12}^{(1)} + \delta_6^{(1)}(2) a_{22}^{(1)} + \delta_6^{(1)}(3) a_{32}^{(1)} \\ &= (8,61 \times 10^{-5}) \times 0.2 + (8,73 \times 10^{-4}) \times 0.65 + (8,32 \times 10^{-4}) \times 0 \\ &= 5,84 \times 10^{-4} \end{aligned}$$

$$\begin{aligned} \text{pour } q_{t+1} = S_3 \Rightarrow \sum_{i=1}^M \delta_6^{(1)}(i) a_{i3}^{(1)} &= \delta_6^{(1)}(1) a_{13}^{(1)} + \delta_6^{(1)}(2) a_{23}^{(1)} + \delta_6^{(1)}(3) a_{33}^{(1)} \\ &= (8,61 \times 10^{-5}) \times 0.1 + (8,73 \times 10^{-4}) \times 0.35 + (8,32 \times 10^{-4}) \times 1 \\ &= 1,14 \times 10^{-3} \end{aligned}$$

$$S_j(t+1=7) = \arg \max_{1 \leq j \leq M} \left[\sum_{i=1}^M \delta_6^{(1)}(i) a_{ij}^{(1)} \right] = \arg \max \{ 6,02 \times 10^{-5} \quad 5,84 \times 10^{-4} \quad 1,14 \times 10^{-3} \} \rightarrow q_{t+1} = S_3$$

La probabilité que la séquence d'observations X^{11} arrive au stade S_3 à l'instant $(t+1=7)$, étant donné un modèle $\bar{\lambda}_1$, est supérieure à celle des stades S_1 et S_2 . Ceci nous permet de prédire l'imminence du troisième stade de dégradation avant son apparition (voir figure 3.6).

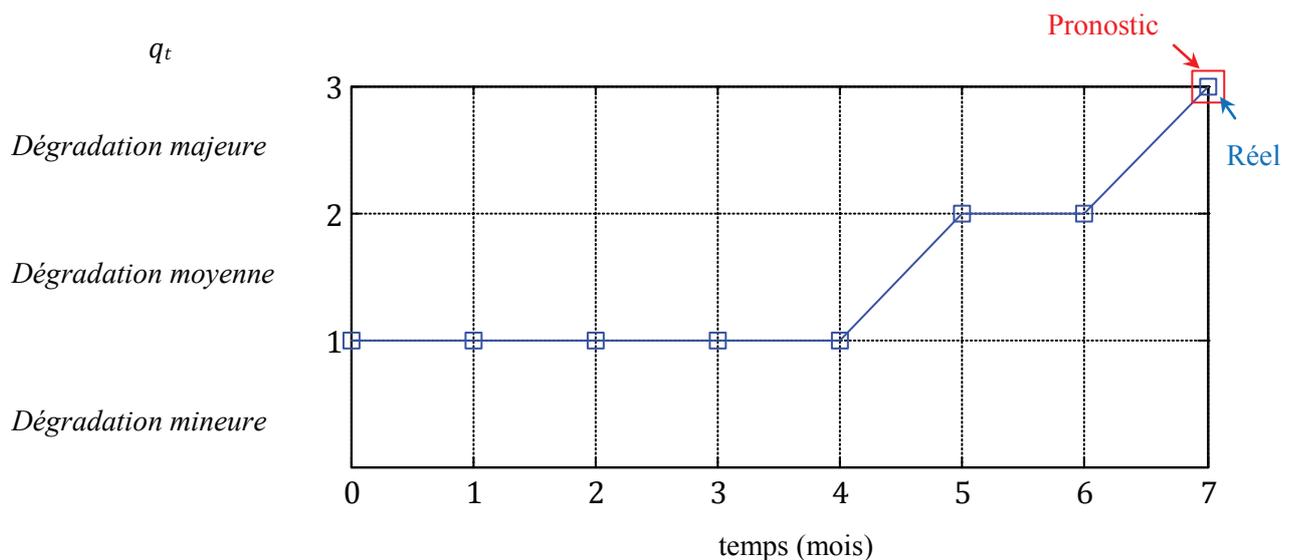


Figure 3. 6. .Évolution de la dégradation (q_t) du moteur n°11

3.2.2. Méthodes basées sur l'analyse des séquences d'observations

L'objectif de la première méthode est de détecter l'imminence du prochain stade de dégradation d'un système en se basant sur l'historique de dégradation d'autres systèmes du même type. L'applicabilité de cette méthode n'est valable que dans le cas où nous disposons d'un système qui se dégrade de façon similaire à ceux étudiés. Cependant, la disponibilité d'une telle base de données dans la plupart des applications est difficile à obtenir voir impossible. Pour résoudre ce problème, nous proposons deux nouvelles méthodes basées sur l'analyse des séquences d'observations pour la détection de l'imminence de l'état futur du système. Ces deux nouvelles méthodes ne requièrent qu'une base de données d'un système faisant apparaître tous les stades de dégradation. Elles ne requièrent que l'historique de dégradation d'un seul système

qui devra être du même type que celui étudié et sa base de données devra inclure tous les stades de dégradation.

Modèle de Markov évolutif

La deuxième méthode que nous proposons consiste à calculer la probabilité que la séquence d'observations $X=\{x_0, x_1, x_2, \dots, x_t\}$ arrive à un stade S_j à l'instant $(t+1)$, étant donné un modèle noté $\bar{\lambda}_t$, $P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1}=S_j | \bar{\lambda}_t)$, avec x_{t+1} inconnue. La différence par rapport à la méthode n°1 réside dans le fait que la construction du MMC $\bar{\lambda}_t$ à l'instant (t) est obtenue à partir de la séquence d'observations X elle-même. Cela revient donc à ré-estimer à chaque pas de temps le modèle de Markov caché correspondant à chaque nouvelle séquence d'observations. Ceci est illustré dans la figure 3.7.

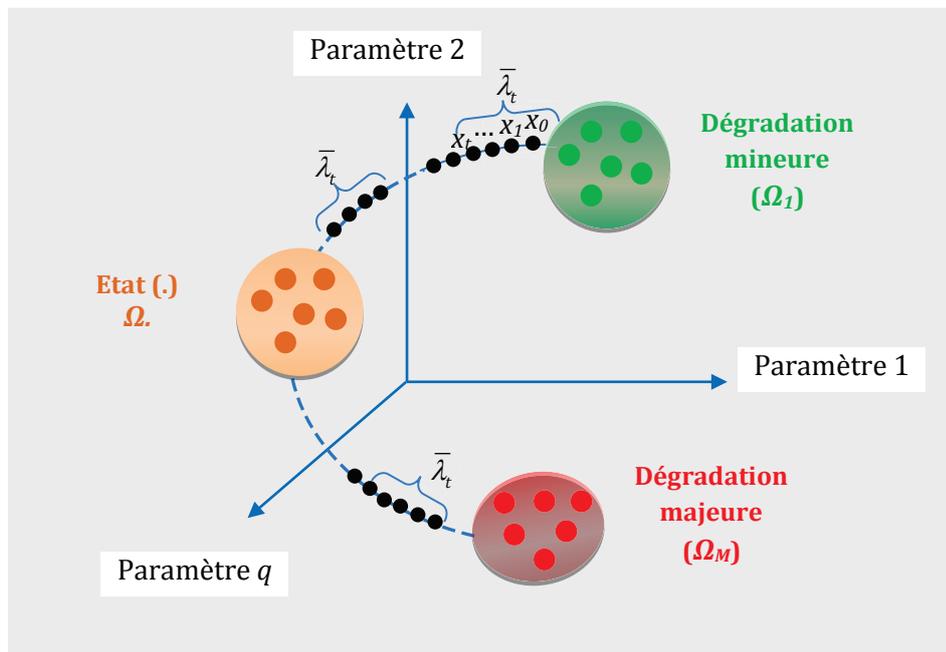


Figure 3. 7. Construction d'un modèle de Markov évolutif

L'imminence du prochain stade de dégradation $q_{t+1}=S_j$ $j \in \{1, \dots, M\}$ est obtenue par le calcul de la probabilité $P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1}=S_j | \bar{\lambda}_t)$ et la détermination du stade S_j qui maximise cette probabilité (3.13).

$$\begin{aligned} \text{Stade de dégradation du système } S_j \text{ à } (t+1) &= \arg \max_{1 \leq j \leq M} \left(P(x_0, x_1, x_2, \dots, x_t, x_{t+1}, q_{t+1} = S_j | \bar{\lambda}_t) \right) \\ &= \arg \max_{1 \leq j \leq M} \left(\sum_{i=1}^M \delta_t(i) a_{ij} \right) \end{aligned} \quad (3.13)$$

Nous montrons dans la figure 3.8 l'organigramme de la deuxième méthode de pronostic basée sur le modèle de Markov évolutif.

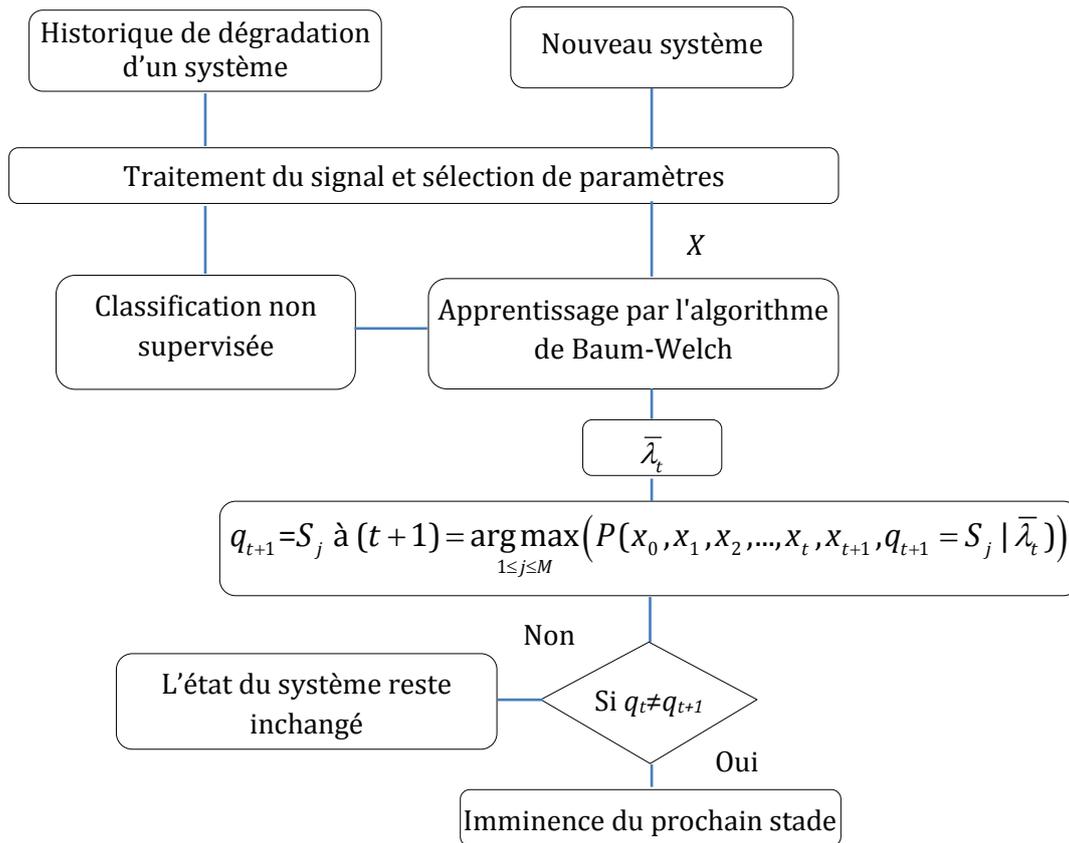


Figure 3. 8. Organigramme de la deuxième méthode de pronostic basée sur un modèle de Markov évolutif

Retour à l'exemple :

L'exemple précédent est utilisé afin d'illustrer l'applicabilité du modèle de Markov évolutif pour prédire l'imminence du prochain stade de dégradation du moteur n°11. Nous allons pour cela considérer la disponibilité de l'historique de dégradation du moteur n°1 qui fait apparaître tous les stades de dégradation sous forme de classes, comme illustré sur la figure 3.4. La matrice d'observation B donnée dans tableau 3.2 va nous servir à déterminer le modèle $\bar{\lambda}_t$ qui représente le mieux la séquence d'observations X^{11} .

A partir de la formulation de forward-backward (2.67) et (2.66), on obtient une ré-estimation de la matrice de transition $\bar{A}^{(t)}$ et du vecteur d'état initial $\bar{\pi}^{(t)}$ du modèle $\bar{\lambda}_t$:

$$\bar{A}^{(t)} = \begin{matrix} & \Omega_1 & \Omega_2 & \Omega_3 \\ \Omega_1 & \begin{pmatrix} 0,69 & 0,21 & 0,1 \\ 0 & 0,66 & 0,34 \\ 0 & 0 & 1 \end{pmatrix} \\ \Omega_2 & \\ \Omega_3 & \end{matrix}, \quad \bar{\pi}^{(t)} = \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \\ (0,69 & 0,21 & 0,1) \end{matrix}$$

Le tableau 3.4 résume les variables Forward normalisées obtenues pour chaque stade de dégradation à partir de la séquence d'observation X^{11} jusqu'à l'instant $t=6$.

Tableau 3. 4. Variables Forward normalisées obtenues de la séquence d'observations X^{11}

	Dégradation mineur	Dégradation moyenne	Dégradation majeure
$t=0$	$\delta_0(1)=0,5325$	$\delta_0(2)=0,04$	$\delta_0(3)=4,5 \times 10^{-3}$
$t=1$	$\delta_1(1)=0,226$	$\delta_1(2)=0,0398$	$\delta_1(3)=6,6 \times 10^{-3}$
$t=2$	$\delta_2(1)=0,08$	$\delta_2(2)=0,025$	$\delta_2(3)=6,07 \times 10^{-3}$
$t=3$	$\delta_3(1)=0,0255$	$\delta_3(2)=0,013$	$\delta_3(3)=3,26 \times 10^{-3}$
$t=4$	$\delta_4(1)=6,33 \times 10^{-3}$	$\delta_4(2)=6,156 \times 10^{-3}$	$\delta_4(3)=1,99 \times 10^{-3}$
$t=5$	$\delta_5(1)=1,34 \times 10^{-3}$	$\delta_5(2)=2,66 \times 10^{-3}$	$\delta_5(3)=9,3 \times 10^{-4}$
$t=6$	$\delta_6(1)=9,51 \times 10^{-5}$	$\delta_6(2)=9,49 \times 10^{-4}$	$\delta_6(3)=8,4 \times 10^{-4}$

Pour $t+1=7$, on a :

$$S_j(t+1) = \arg \max \left[\sum_{i=1}^M \delta_t(i) a_{ij} \right] \Leftrightarrow S_j(t+1=7) = \arg \max \left[\sum_{i=1}^M \delta_6(i) a_{ij} \right]$$

$$\text{pour } q_{t+1} = S_1 \Rightarrow \sum_{i=1}^M \delta_6(i) a_{i1} = \delta_6(1) a_{11} + \delta_6(2) a_{21} + \delta_6(3) a_{31}$$

$$= (9,51 \times 10^{-5}) \times 0,71 + (9,49 \times 10^{-4}) \times 0 + (8,4 \times 10^{-4}) \times 0$$

$$= 6,75 \times 10^{-5}$$

$$\text{pour } q_{t+1} = S_2 \Rightarrow \sum_{i=1}^M \delta_6(i) a_{i2} = \delta_6(1) a_{12} + \delta_6(2) a_{22} + \delta_6(3) a_{32}$$

$$= (9,51 \times 10^{-5}) \times 0,2 + (9,49 \times 10^{-4}) \times 0,66 + (8,4 \times 10^{-4}) \times 0$$

$$= 6,45 \times 10^{-4}$$

$$\text{pour } q_{t+1} = S_3 \Rightarrow \sum_{i=1}^M \delta_6(i) a_{i3} = \delta_6(1) a_{13} + \delta_6(2) a_{23} + \delta_6(3) a_{33}$$

$$= (9,51 \times 10^{-5}) \times 0,1 + (9,49 \times 10^{-4}) \times 0,34 + (8,4 \times 10^{-4}) \times 1$$

$$= 1,17 \times 10^{-3}$$

$$S_j(t+1=7) = \arg \max_{1 \leq j \leq M} \left[\sum_{i=1}^M \delta_6(i) a_{ij} \right] = \arg \max \{ 6,75 \times 10^{-5} \quad 6,45 \times 10^{-4} \quad 1,17 \times 10^{-3} \} \rightarrow q_{t+1} = S_3$$

La probabilité que la séquence d'observations X^{11} arrive au stade S_3 à l'instant $(t+1=7)$, étant donné un modèle $\bar{\lambda}_1$, est supérieure à celle des stades S_1 et S_2 . Ceci nous permet de prédire l'imminence du troisième stade de dégradation avant son apparition.

Modèles de Markov cachés

Les deux méthodes de pronostic que nous avons proposées précédemment reposent sur le même principe, modéliser le processus de dégradation d'un système par un seul modèle de Markov caché. La troisième méthode, quant à elle, s'inspire de la méthode de diagnostic présentée dans la section 2.3.2.4. Le principe reste le même que celui du diagnostic, modéliser un MMC $\bar{\lambda}_k$, $1 \leq k \leq M$ à partir d'une séquence d'observations X^k extraite d'une classe (k) . La séquence d'observations X^k , dans ce cas, est formulée comme suit :

$$X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_T^k\}, \quad 1 \leq k \leq M \quad (3.14)$$

avec $x_0^k, x_1^k, x_2^k, \dots, x_T^k$ une séquence d'observations qui correspond à un tirage successif de T+1 observations de la classe (k). Le pronostic, dans ce cas, consiste à détecter l'imminence du prochain stade de dégradation par la probabilité de générer la séquence d'observations, $\{x_0, x_1, x_2, \dots, x_t, x_{t+1}\}$ étant donné un modèle $\bar{\lambda}_k$, avec x_{t+1} inconnue.

Le MMC $\bar{\lambda}_k$ pour lequel la probabilité donnée par la formule (3.15) est maximale déterminera l'imminence du prochain stade de dégradation du système.

$$\text{Stade de dégradation du système } \bar{\lambda}_k \text{ à } (t+1) = \arg \max_{1 \leq k \leq M} (P(x_0, x_1, x_2, \dots, x_t, x_{t+1} | \bar{\lambda}_k)) \quad (3.15)$$

A partir des formules (3.4) et (3.5), la probabilité de générer la séquence d'observations $\{x_0, \dots, x_t, x_{t+1}\}$ étant donné un modèle $\bar{\lambda}_k$ est donnée comme suit :

$$\begin{aligned} P(x_0, x_1, \dots, x_t, x_{t+1} | \bar{\lambda}_k) &= \sum_{j=1}^M P(x_0, x_1, x_2, \dots, x_{t+1}, q_{t+1} = S_j | \bar{\lambda}_k) = \sum_{j=1}^M \alpha_{t+1}^{(k)}(j) \\ &= \sum_{j=1}^M b_j(x_{t+1}) \sum_{i=1}^M \alpha_t^{(k)}(i) a_{ij}^{(k)} \quad 1 \leq i, j \leq M \end{aligned} \quad (3.16)$$

A partir de la formule (3.16), l'imminence du prochain stade de dégradation est donnée en fonction de la probabilité de transition $a_{ij}^{(k)}$, la variable Forward $\alpha_t^{(k)}(i)$ à l'instant (t), du modèle $\bar{\lambda}_k$ et la probabilité d'émettre une nouvelle observation associée à un stade de dégradation S_j à (t+1). Comme pour les précédentes méthodes, nous ne disposons pas de l'observation x_{t+1} . Cependant, il est possible d'estimer la valeur de la formule (3.16) en remplaçant la probabilité $b_j(x_{t+1})$ par un coefficient noté C_{sj} . Comme pour la probabilité $b_j(x_{t+1})$, ce coefficient est compris entre "0" et "1". Un coefficient de "0" signifie que la probabilité d'émettre une nouvelle observation associée à un stade de dégradation S_j est faible. Un coefficient de "1" signifie qu'il y a une forte probabilité d'associer la nouvelle observation à un stade de dégradation S_j . Ainsi, l'imminence du prochain stade de dégradation peut être obtenue en donnant un coefficient C_{sj} égal à la probabilité d'émettre une observation à l'instant (t) ($C_{sj} = b_j(x_t)$) si le prochain stade de dégradation est le même que celui à l'instant (t+1) ($q_t = S_i = q_{t+1} = S_j$) et un coefficient de "1" ($C_{sj} = 1$) pour les futurs stades de dégradation ($q_t = S_i \neq q_{t+1} = S_j$). Cette stratégie favorise la transition d'un état à l'instant (t) à un nouvel état à l'instant (t+1). Nous obtenons :

$$P(x_0, x_1, \dots, x_t, x_{t+1} | \bar{\lambda}_k) = \sum_{j=1}^M C_{sj} \sum_{i=1}^M \alpha_t^{(k)}(i) a_{ij}^{(k)} \quad 1 \leq i, j, k \leq M \quad (3.17)$$

avec

$$C_{s_j} = \begin{cases} 1 & \text{si } q_{t+1} \neq q_t \\ b_j(x_t) & \text{si } q_{t+1} = q_t \end{cases}$$

Retour à l'exemple :

L'exemple précédent est utilisé afin d'illustrer l'applicabilité des modèles de Markov cachés pour prédire l'imminence du prochain stade de dégradation du moteur n°11. Nous allons pour cela considérer que l'historique de dégradation du moteur n°1 qui fait apparaître tous les stades de dégradation sous forme de classes comme illustré sur la figure 3.4. À partir de chaque classe, nous extrayons une séquence d'observations qui correspond à un tirage successif de T+1=7 observations de la classe (k).

Les paramètres initiaux de chaque MMC (matrice de transition $A^{(k)}$ et vecteur d'état initial $\pi^{(k)}$) à partir des quels nous commencerons l'apprentissage, sont définis comme suit :

$$A^{(1,2,3)} = \begin{matrix} & \Omega_1 & \Omega_2 & \Omega_3 \\ \Omega_1 & \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix} \quad \pi^{(1,2,3)} = \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \\ \begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix} \end{matrix} \quad (3.18)$$

Notons que les probabilités de transition a_{32} , a_{31} , a_{21} sont fixées à zéros. Dans notre cas, la classe Ω_1 correspond à un système sain et Ω_2, Ω_3 correspondent à des dégradations progressives du système. Dans ce cas, il est normal de fixer à zéros a_{32} , a_{31} , a_{21} car nous ne prenons pas en compte le fait qu'un système dégradé puisse revenir à un fonctionnement normal.

A partir de la formulation de forward-backward (2.67) et (2.66), on obtient une ré-estimation de la matrice de transition $\bar{A}^{(k)}$ et du vecteur d'état initial $\bar{\pi}^{(k)}$ du modèle $\bar{\lambda}_k$:

$$\begin{matrix} \bar{A}^1 = \begin{bmatrix} 0,99 & 0,01 & 0 \\ 0 & 0,99 & 0,01 \\ 0 & 0 & 1 \end{bmatrix} & \bar{A}^2 = \begin{bmatrix} 0 & 0,99 & 0,01 \\ 0 & 0,99 & 0,01 \\ 0 & 0 & 1 \end{bmatrix} & \bar{A}^3 = \begin{bmatrix} 0 & 0,01 & 0,99 \\ 0 & 0,01 & 0,99 \\ 0 & 0 & 1 \end{bmatrix} \\ \bar{\pi}^1 = [0,99 \quad 0,01 \quad 0] & \bar{\pi}^2 = [0 \quad 0,99 \quad 0,01] & \bar{\pi}^3 = [0 \quad 0,01 \quad 0,99] \end{matrix}$$

Les variables Forward obtenues pour chaque stade de dégradation, à partir de la séquence d'observations X^{11} jusqu'à l'instant $t=6$, sont données comme suit :

	Dégradation mineur	Dégradation moyenne	Dégradation majeure
$k=1$	$\alpha_6^{(1)}(1)=0$	$\alpha_6^{(1)}(2)=6,05 \times 10^{-5}$	$\alpha_6^{(1)}(3)=2,67 \times 10^{-6}$
$k=2$	$\alpha_6^{(2)}(1)=0$	$\alpha_6^{(2)}(2)=4,7 \times 10^{-3}$	$\alpha_6^{(2)}(3)=2 \times 10^{-4}$
$k=3$	$\alpha_6^{(3)}(1)=0$	$\alpha_6^{(3)}(2)=0$	$\alpha_6^{(3)}(3)=2,7 \times 10^{-4}$

Pour $t+1=7$, on a $C_{s1}=0,1$, $C_{s2}=0,47$, $C_{s3}=1$ et :

$$\begin{aligned} \lambda_k \hat{a}(t+1) &= \arg \max_{1 \leq k \leq 3} \left[\sum_{j=1}^M C_{sj} \sum_{i=1}^M \alpha_t^{(k)}(i) a_{ij}^{(k)} \right] = \arg \max_{1 \leq k \leq 3} \left[\sum_{j=1}^M C_{sj} \sum_{i=1}^M \alpha_6^{(k)}(i) a_{ij}^{(k)} \right] \\ &= \arg \max \{ 2,67e^{-6} \quad 2,42e^{-3} \quad 2,7e^{-3} \} \rightarrow \lambda_k \hat{a}(t+1=7) = \lambda_3 \end{aligned}$$

La probabilité de générer la séquence d'observations X^{11} jusqu'à l'instant $(t+1=7)$, étant donné le modèle $\bar{\lambda}_3$, est supérieure à celle de $\bar{\lambda}_1$ et $\bar{\lambda}_2$. Ceci nous permet de prédire l'imminence du troisième stade de dégradation avant son apparition.

Nous montrons dans la figure 3.9 l'organigramme de la troisième méthode de pronostic basée sur les modèles de Markov cachés.

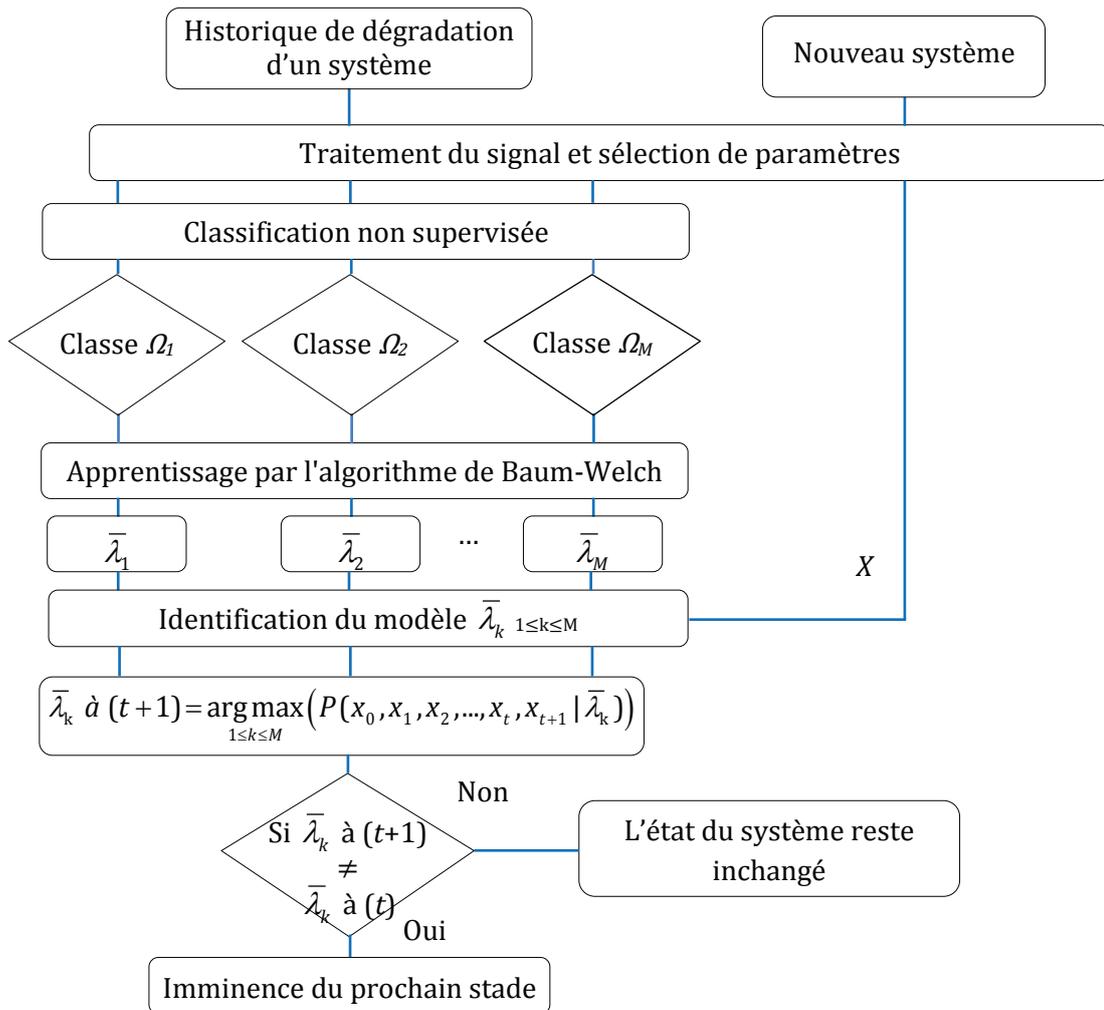


Figure 3. 9. Organigramme de la troisième méthode de pronostic

Le tableau 3.5 permet de juger les trois méthodes de pronostic par rapport à leur complexité et au type d'information nécessaire à leur mise en œuvre.

Tableau 3. 5. Avantage et inconvénient des méthodes de pronostic proposées

	Avantage	Inconvénient
Méthode (1) Identification par MMC	Complexité de $O(M^2TN)$	Nécessite l'historique de dégradation de N -systèmes
Méthode (2) MMC évolutif	Nécessite que l'identification des stades de dégradation	Complexité de $O(M^2t^2N)$
Méthode (3) MMCs	Complexité de $O(M^2TM)$ Nécessite que l'identification des stades de dégradation	-

3.3. Estimation du temps de vie restant

Les méthodes de pronostic présentées précédemment sont utilisées pour prédire l'imminence des différents stades de dégradations, et donc de l'imminence d'un défaut du système.

L'information fournie par ces pronostiqueurs permet d'aider les techniciens présents sur le terrain à prendre des décisions afin de réparer ou remplacer le système avant sa défaillance. Ils acquièrent les variations des paramètres de surveillance dans le temps d'un système, analysent les données en temps réel et donnent une prédiction sur l'imminence d'un défaut. L'organisation internationale de normalisation (ISO) 13381-1 a défini le pronostic comme étant une estimation de la durée de vie restante avant l'apparition d'un défaut et la détection de l'imminence d'un ou plusieurs défauts. Les recherches effectuées dans [Tob_11] montrent qu'il est relativement difficile de prédire le temps de vie restant d'un système avec les modèles de Markov cachés. En effet, une limitation inhérente d'un modèle de Markov caché est qu'il ne fournit pas une représentation adéquate sur la durée d'un état. Cependant, il est possible d'estimer le temps de vie restant d'un système en prédisant la tendance (l'évolution) d'un paramètre de surveillance (voir figure 3.10) par le biais de la prédiction de séries temporelles. La différence entre l'instant (t), où on débute le pronostic, et l'instant ($t+RUL$), où on diagnostique l'apparition d'un éventuel défaut en se basant sur la prédiction de la tendance des paramètres de surveillance, permet d'avoir une estimation du temps de vie restant (RUL) avant l'apparition du défaut.

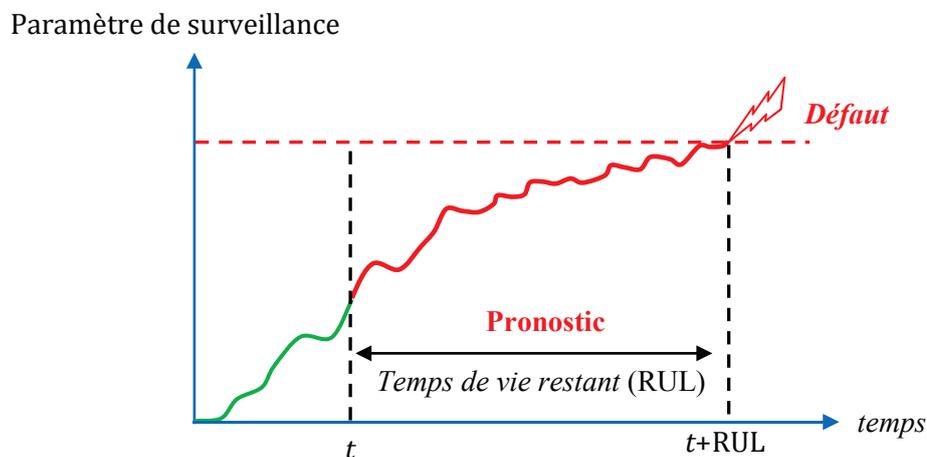


Figure 3. 10. Détermination du temps de vie restant par le suivi de la tendance d'un paramètre de surveillance

La prédiction de séries temporelles est une technique qui établit un modèle de prédiction sur un paramètre de surveillance à partir d'une série temporelle d'observations. Elle utilise un modèle de prévision pour prédire à court et à long termes l'évolution de la dégradation du système sous surveillance.

La prédiction de séries temporelles peut être formulée comme suit : soit une série temporelle d'observations $\{x_t, x_{t-r}, x_{t-2r}, x_{t-3r} \dots x_{t-(n-1)r}\}$, nous avons besoin de connaître la valeur de l'observation (x_{t+p}) à l'instant ($t+p$), où p représente l'horizon de prédiction, (r) désigne le pas de prédiction et ($n-1$) définit le nombre de pas précédents. La prédiction de x_{t+p} notée \hat{x}_{t+p} peut être obtenue grâce à deux approches :

- La première approche consiste à prédire la valeur de l'observation \hat{x}_{t+p} à partir d'une série temporelle d'observations $\{x_t, x_{t-r}, x_{t-2r}, x_{t-3r} \dots x_{t-(n-1)r}\}$ mesurée jusqu'à un instant (t). Cette approche nécessite un apprentissage pour chaque horizon de prédiction (p).

$$\hat{x}_{t+p} = f(x_t, x_{t-r}, x_{t-2r}, x_{t-3r}, \dots, x_{t-(n-1)r}) \quad (3.19)$$

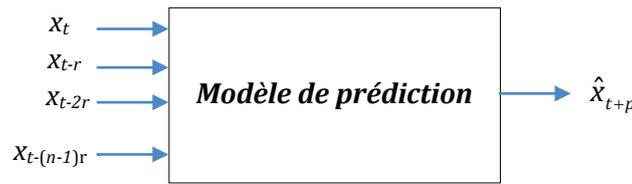


Figure 3. 11. Prédiction à un pas dans le futur

- La deuxième approche consiste à prédire la valeur de l'observation \hat{x}_{t+p} par la mise en cascade de plusieurs modèles de prédiction afin de réaliser des prévisions à court terme ($t+1$) à partir des prédictions précédentes. La première prédiction devient l'entrée du second modèle et ainsi de suite. Pour cette approche, un seul apprentissage est nécessaire.

$$\hat{x}_{t+1} = f(x_t, x_{t-r}, x_{t-2r}, x_{t-3r}, \dots, x_{t-(n-1)r}) \tag{3.20}$$

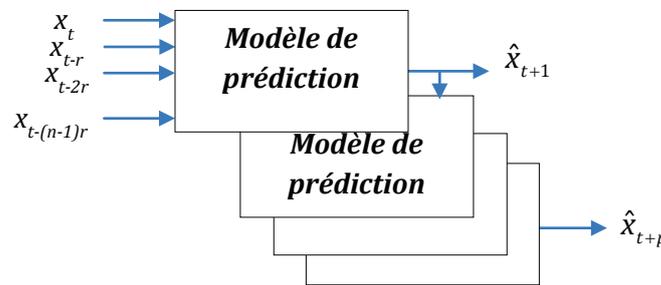


Figure 3. 12. Prédiction à plusieurs pas dans le futur (multi-pas)

Un des modèles de prédiction le plus utilisé est le Réseau de Neurones Artificiels (RNA). L'idée d'utiliser les réseaux de neurones artificiels comme modèle de prédiction a commencé en 1964, lorsque Hu a utilisé le réseau adaptatif de Widrow [Widrow_85] pour faire des prévisions climatiques [Hu_64]. Cependant, le manque d'algorithmes d'apprentissage à cette époque limita la poursuite de cette étude. Depuis les années 80, la recherche dans le domaine du pronostic par les RNAs s'est accélérée. La figure 3.13 montre l'évolution des RNAs vers des systèmes hybrides (systèmes neuro-flous).

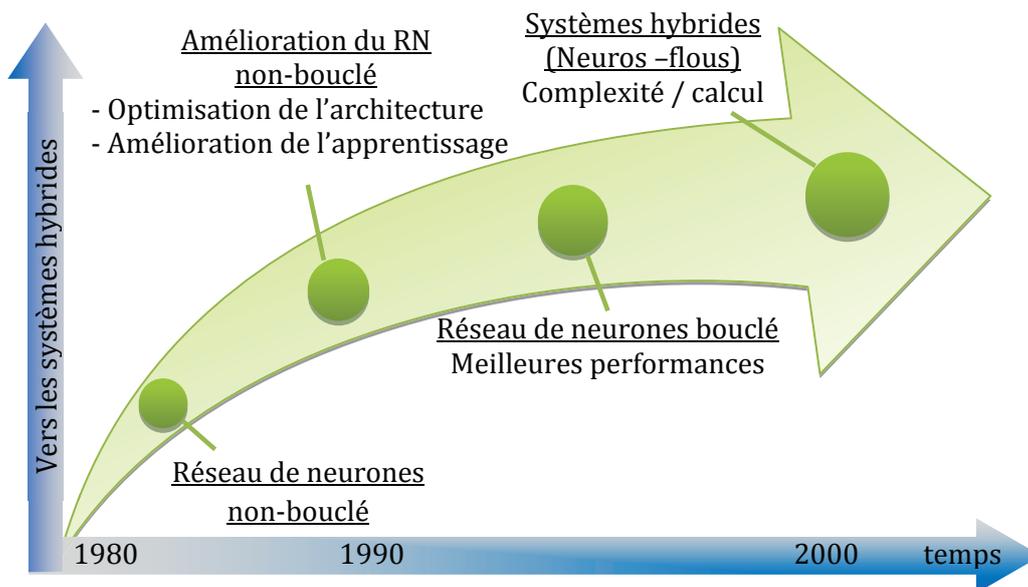


Figure 3. 13. Vers les systèmes neuro-flous [Gouriveau_07]

Il existe principalement deux types de RNAs : les réseaux de neurones non bouclés et les réseaux de neurones bouclés [Gouriveau_07]. Ces deux types de RNAs ont été utilisés avec succès dans le domaine du pronostic. Ils ont néanmoins quelques limitations :

- Les RNAs sont des boîtes noires et il n'est pas possible d'expliquer et d'analyser la relation entre les entrées et les sorties;
- L'architecture des RNAs reste à améliorer.

Ainsi, pour remédier à ces faiblesses, des recherches ont souligné l'importance des systèmes combinant l'aspect connectique des RNAs avec des techniques de raisonnement. Ceci a mené à l'élaboration d'un système hybride nommé système neuro-flou (SNF). L'utilisation combinée des réseaux de neurones et la logique floue permet de prendre les avantages des deux méthodes : la capacité d'apprentissage de la première et la facilité de compréhension linguistique du système à base de règles floues de la seconde.

L'utilisation des systèmes neuro-flous pour le pronostic de séries temporelles doit prendre en compte l'aspect temporel. C'est dans ce contexte que nous proposons dans la section suivante un modèle de réseau adaptatif équivalent aux systèmes neuro-flous. Le modèle proposé est le prédicteur ANFIS pour (Adaptive neuro fuzzy inference system), qui signifie système d'inférence flou basé sur le réseau adaptatif.

3.3.1. Prédiction de séries temporelles par ANFIS

Par souci de simplicité, nous supposons que le système neuro-flou à base de réseaux de neurones adaptatifs (ANFIS) possède un ensemble d'entrées $\{x_t, x_{t-1}, x_{t-2}, x_{t-3}\}$ sous forme d'une série temporelle (avec $n=4$ et $r=1$ où r désigne le pas de prédiction, n désigne le nombre d'entrées et $n-1$ définit le nombre de pas précédents.) et une sortie notée \hat{x}_{t+p} avec $p=1, 2, \dots$. Si deux ensembles flous (définis par des fonctions d'appartenance du type triangle, trapèze ou gaussienne) sont associés à chaque variable d'entrée, alors le système présente 16 règles floues ($4^2=16$). Ces règles sont du type Sugeno du premier ordre [Takagi_85]. Elles sont données comme suit :

$$\begin{aligned} \text{Règle}(j): \quad & \text{Si } (x_t \text{ est } A_0^j) \text{ et } (x_{t-1} \text{ est } A_1^j) \text{ et } (x_{t-2} \text{ est } A_2^j) \text{ et } (x_{t-3} \text{ est } A_3^j) \\ \text{Alors } \hat{x}_{t+p}^j = & f_j(x_t, x_{t-1}, x_{t-2}, x_{t-3}) = c_0^j x_t + c_1^j x_{t-1} + c_2^j x_{t-2} + c_3^j x_{t-3} + c_4^j, \quad j=1, 2, \dots, 16 \end{aligned} \quad (3.21)$$

où ($j = 16$) est le nombre de règles, x_{t-i} pour ($i = 0, 1, 2, 3$) sont les entrées du modèle et \hat{x}_{t+p}^j est la sortie de la règle (j). A_i^j est un ensemble flou caractérisé par des fonctions d'appartenance. $\mu_{A_i^j}(x_{t-i})$ et $\{c_0^j, c_1^j, c_2^j, c_3^j, c_4^j\}$ sont les coefficients de la règle (j).

L'architecture d'ANFIS correspondante est représentée sur la figure 3.14, où les nœuds de la même couche ont des fonctions similaires.

La structure du prédicteur ANFIS contient cinq couches à l'exclusion de la couche d'entrée. Elles sont définies comme suit :

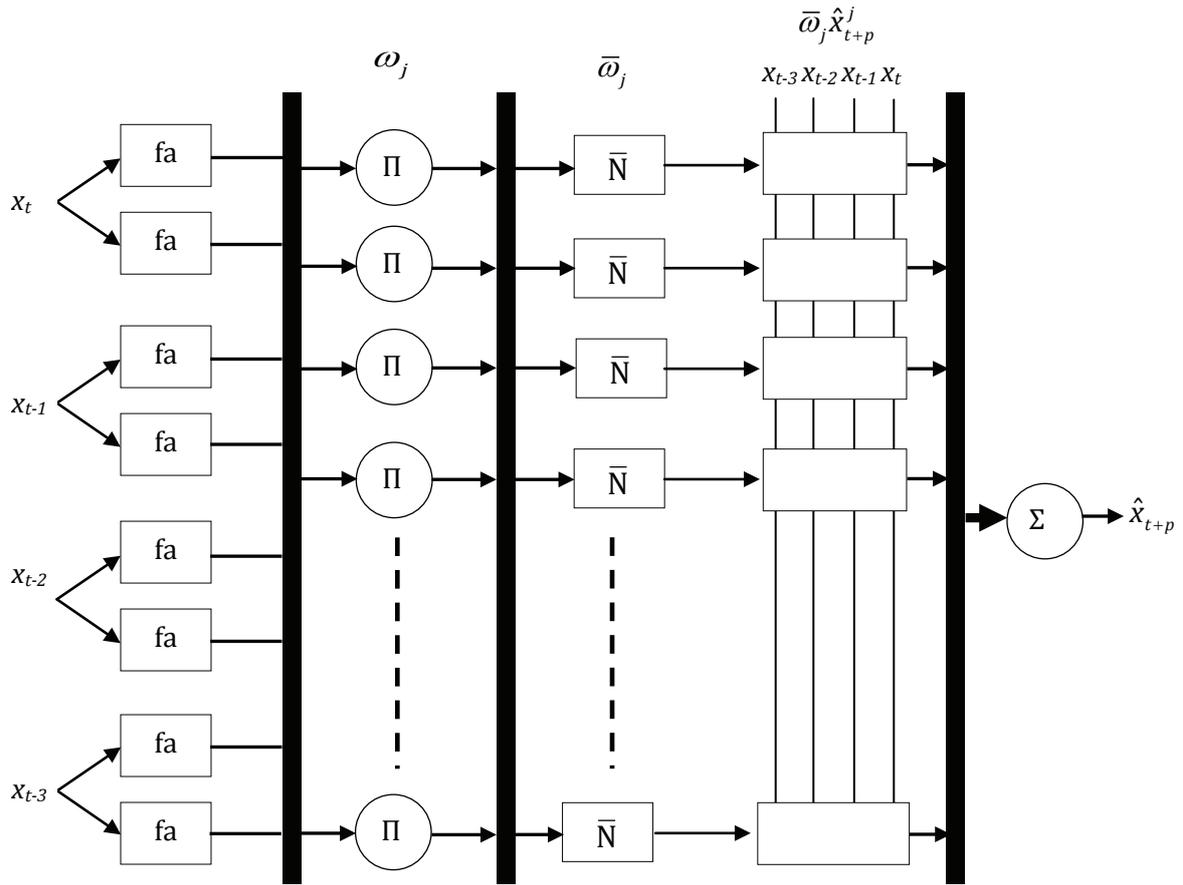


Figure 3. 14. Architecture du prédicteur ANFIS avec quatre entrées et deux fonctions d'appartenance associées à chaque entrée

Couche 1 (Fuzzyfication) : chaque nœud dans cette couche est considéré comme un nœud adaptatif. Il est associé à une fonction d'appartenance (fa) qui évalue le degré d'appartenance du $j^{\text{ième}}$ terme de la partition floue et de l' $i^{\text{ième}}$ variable d'entrée A_i^j . La fonction d'appartenance pour la variable d'entrée A_i^j peut être n'importe quelle fonction (triangle, trapèze ou gaussienne). Par exemple, pour une fonction gaussienne, on a :

$$\mu_{A_i^j}(x_{t-i}) = \exp\left(-\left[\frac{x_{t-i} - m_i^j}{b_i^j}\right]^2\right) \tag{3.22}$$

avec $\{m_i^j, b_i^j\}$, l'ensemble des paramètres des gaussiennes (paramètres prémisses) et x_{t-i} est l'entrée vers le nœud (i) ($i = 0, 1, 2, 3$).

Couche 2 (Pondération des règles floues) : chaque nœud (j) dans cette couche est considéré comme un nœud fixe étiqueté Π . Il est associé à une règle floue et permet de définir le poids de cette règle à l'aide d'une T-norme. La T-norme la plus utilisée dans ce cas est le produit :

$$\omega_j = \mu_{A_0^j}(x_t) \cdot \mu_{A_1^j}(x_{t-1}) \cdot \mu_{A_2^j}(x_{t-2}) \cdot \mu_{A_3^j}(x_{t-3}) \tag{3.23}$$

Couche 3 (Normalisation) : chaque nœud (j) dans cette couche est considéré comme un nœud fixe étiqueté \bar{N} . Cette couche a pour but de normaliser les différents poids.

$$\bar{\omega}_j = \frac{\omega_j}{\sum_{j=1}^{16} \omega_j} \quad (3.24)$$

Couche 4 (Defuzzyfication) : chaque nœud (j) dans cette couche est un nœud adaptatif qui calcule la contribution de chaque règle (j) par une fonction du premier ordre des entrées (approche de Takagi-Sugeno).

$$\bar{\omega}_j \hat{x}_{t+p}^j = \bar{\omega}_j \cdot (c_0^j x_t + c_1^j x_{t-1} + c_2^j x_{t-2} + c_3^j x_{t-3} + c_4^j) \quad (3.25)$$

avec $\bar{\omega}_j$ la sortie du $j^{\text{ème}}$ nœud de la couche précédente. Quant à $\{c_0^j, c_1^j, c_2^j, c_3^j, c_4^j\}$, ce sont les coefficients de cette combinaison linéaire et sont appelés paramètres conséquents.

Couche 5 (Somme). Le nœud de cette couche est un nœud unique et fixe étiqueté Σ , qui définit la sortie comme étant la somme de tous les signaux entrants.

$$\hat{x}_{t+p} = \sum_{j=1}^{16} \bar{\omega}_j \hat{x}_{t+p}^j = \sum_{j=1}^{16} \bar{\omega}_j \cdot (c_0^j x_t + c_1^j x_{t-1} + c_2^j x_{t-2} + c_3^j x_{t-3} + c_4^j) \quad (3.26)$$

L'objectif du prédicteur ANFIS est de former des réseaux adaptatifs capables d'approximer des fonctions inconnues par le biais d'une base d'apprentissage. Cette démarche implique l'utilisation d'un algorithme hybride qui combine la méthode de descente du gradient et la méthode d'estimation des moindres carrés. La méthode de descente du gradient est utilisée pour régler les paramètres non-linéaires prémisse $\{m_i^j, b_i^j\}$ et la méthode des moindres carrés pour identifier les paramètres linéaires conséquents $\{c_0^j, c_1^j, c_2^j, c_3^j, c_4^j\}$. Comme on peut le voir à partir de la figure 3.14, les nœuds circulaires sont des nœuds fixes sans paramètres variables, et les nœuds carrés sont des nœuds adaptatifs avec des paramètres variables (les paramètres sont modifiés lors de la phase d'apprentissage). La procédure d'apprentissage comporte deux étapes : dans la première étape, la méthode des moindres carrés est utilisée pour identifier les paramètres linéaires conséquents, tandis que les paramètres antécédents (fonctions d'appartenance) sont supposés être fixes pour le cycle en cours à travers l'ensemble d'apprentissage. Ensuite, les signaux d'erreur se propagent vers l'arrière. La méthode de descente du gradient est utilisée pour mettre à jour les paramètres prémisses en minimisant l'erreur moyenne entre la valeur de sortie désirée (prédiction) et celle générée par ANFIS, tandis que les paramètres linéaires conséquents restent fixes. Une explication plus détaillée sur l'algorithme hybride d'apprentissage peut être trouvée dans [Jang_93].

L'inconvénient du système neuro-flou ANFIS est lié à une convergence lente de la procédure d'apprentissage lorsque le nombre de fonctions d'appartenance est grand et/ou lorsque le nombre d'entrées est grand. Pour remédier à ce problème, nous avons utilisé un nouveau système neuro-flou appelé « neurone neo-flou » NNF pour la prédiction de séries temporelles. Introduite par Takeshi Yamakawa [Yamak_92, Uchino_97, Yamak_99], l'architecture d'un NNF est assez proche d'un neurone formel à n entrées. Cependant, à la place des poids synaptiques habituels, elle contient des synapses non linéaires $SYN_i, i = 0, 1, \dots, n-1$. Les utilisateurs du NNF admettent que parmi ses avantages : le taux élevé d'apprentissage, la simplicité de calcul, la possibilité de trouver le minimum global du critère d'apprentissage en temps réel et aussi qu'il est caractérisé par des règles floues « Si-Alors ».

3.3.2. Prédiction de séries temporelles par le neurone neo-flou (NNF)

La structure d'un neurone neo-flou est montrée dans la figure 3.15 où les poids synaptiques ne sont pas des fonctions constantes mais des fonctions non-linéaires des entrées. Ces dernières sont représentées par des modèles de logique floue basés sur une collection de règles « Si – Alors ». Cette structure dispose d'un point de sommation qui somme les sorties du modèle associé à chaque entrée.

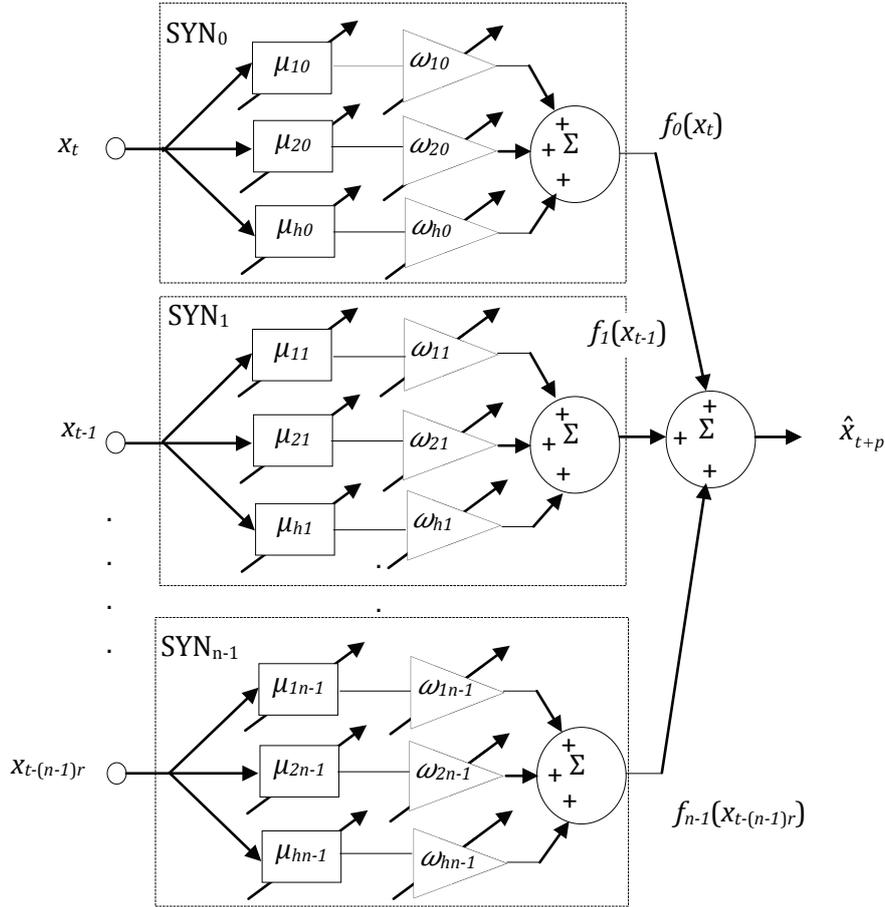


Figure 3. 15. Architecture du prédicteur NNF

La sortie du NNF, pour une série temporelle $\{x_t, x_{t-r}, x_{t-2r}, x_{t-3r} \dots x_{t-(n-1)r}\}$, est donnée par la formule suivante :

$$\hat{x}_{t+p} = \sum_{i=0}^{n-1} f_i(x_{t-ir}) \tag{3.27}$$

où x_{t-ir} est la $i^{\text{ème}}$ entrée ($i=0, 1, 2, \dots, n-1$), \hat{x}_{t+p} est la sortie du système, $f_i(x_{t-ir})$ est une fonction non linéaire qui transforme la $i^{\text{ème}}$ variable d'entrée sous la forme :

$$f_i(x_{t-ir}) = \sum_{j=1}^h \omega_{ji} \cdot \mu_{ji}(x_{t-ir}) \tag{3.28}$$

où μ_{ji} est une fonction d'appartenance floue et ω_{ji} est un poids synaptique.

La disposition des fonctions d'appartenance dans le cas d'une forme triangulaire est montrée dans la figure 3.16.

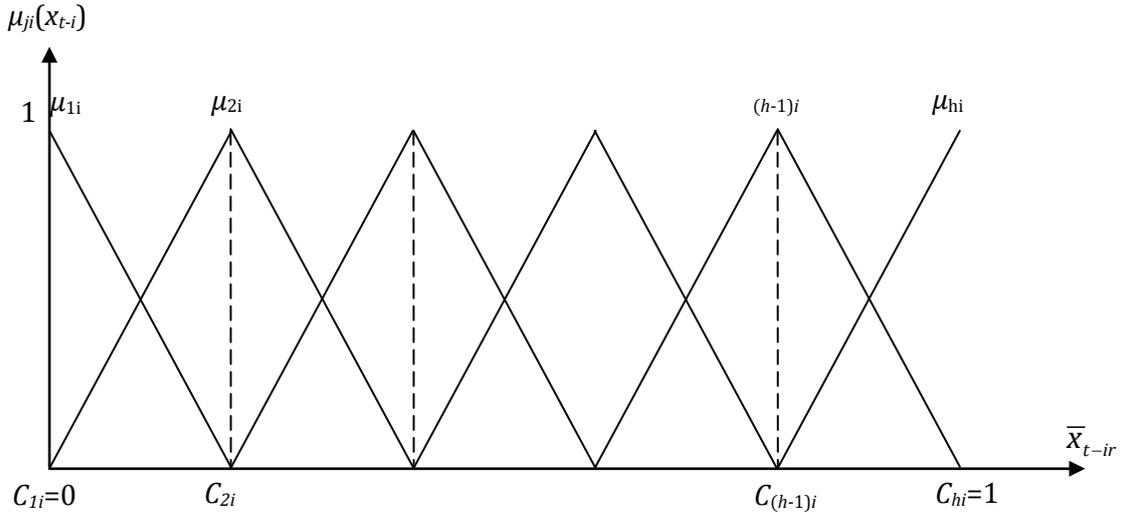


Figure 3. 16. Fonctions d'appartenance triangulaires

Pour des variables d'entrée normalisées \bar{x}_{t-ir} ($0 \leq \bar{x}_{t-ir} \leq 1$), les fonctions d'appartenance peuvent être exprimées par la formule suivante.

$$\mu_{ji}(x_{t-ir}) = \begin{cases} \frac{x_{t-ir} - c_{j-1,i}}{c_{ji} - c_{j-1,i}} & , \text{ Si } x_{t-ir} \in [c_{j-1,i}, c_{ji}] \\ \frac{c_{j+1,i} - x_{t-ir}}{c_{j+1,i} - c_{ji}} & , \text{ Si } x_{t-ir} \in [c_{ji}, c_{j+1,i}] \\ 0 & \text{ Sinon} \end{cases} \quad (3.29)$$

où c_{ji} sont les centres des fonctions d'appartenance. Habituellement, ils sont équidistants sur un intervalle de [0-1].

Une variable d'entrée x_{t-ir} active seulement deux fonctions d'appartenance voisines et la somme de ces deux fonctions est égale à 1 :

$$\mu_{ji}(x_{t-ir}) + \mu_{j+1,i}(x_{t-ir}) = 1 \quad (3.30)$$

Ainsi, la fonction non linéaire $f_i(x_{t-ir})$ peut être donnée sous cette forme simplifiée :

$$f_i(x_{t-ir}) = \omega_{ji} \cdot \mu_{ji}(x_{t-ir}) + \omega_{j+1,i} \cdot \mu_{j+1,i}(x_{t-ir}) \quad (3.31)$$

3.3.2.1. Apprentissage d'un NNF (neurone neo-fou)

Lorsqu'une série temporelle $\{x_{t-(n-1)r}(\ell) \dots x_{t-3r}(\ell), x_{t-2r}(\ell), x_{t-r}(\ell), x_t(\ell)\}$ (où $\ell=1,2,\dots$ est le numéro d'itération) est insérée à l'entrée du neurone neo-flou, la sortie de ce neurone peut être exprimée par la formule suivante :

$$\hat{x}_{t+p}(\ell) = \sum_{i=0}^{n-1} f_i(x_{t-ir}(\ell)) = \sum_{i=0}^{n-1} \sum_{j=1}^h \omega_{ji}(\ell-1) \mu_{ji}(x_{t-ir}(\ell)), \quad r \geq 1 \quad (3.32)$$

Cette sortie est déterminée à la fois par les fonctions d'appartenance $\mu_{ji}(x_{t-ir})$ et les poids synaptiques ajustables $\omega_{ji}(\ell-1)$, qui ont été obtenus à partir de la précédente itération ($\ell-1$). Par conséquent, il y a $h \times n-1$ poids synaptiques à déterminer.

Le critère d'apprentissage est représenté par la fonction d'erreur quadratique :

$$E(\ell) = \frac{1}{2} (x_{t+p}(\ell) - \hat{x}_{t+p}(\ell))^2 = \frac{1}{2} e(\ell+1)^2 = \frac{1}{2} \left(x_{t+p}(\ell) - \sum_{i=0}^{n-1} \sum_{j=1}^h \omega_{ji}(\ell-1) \mu_{ji}(x_{t-ir}(\ell)) \right)^2 \quad (3.33)$$

L'erreur quadratique est minimisée par l'algorithme de descente du gradient (retro propagation) qui permet la mise à jour des poids par la formule suivante :

$$\begin{aligned} \omega_{ji}(\ell+1) &= \omega_{ji}(\ell) + \eta e(\ell+1) \mu_{ji}(x_{t-ir}(\ell+1)) \\ &= \omega_{ji}(\ell) + \eta \left(x_{t+p}(\ell+1) - \sum_{i=0}^{n-1} \sum_{j=1}^h \omega_{ji}(\ell) \mu_{ji}(x_{t-ir}(\ell+1)) \right) \mu_{ji}(x_{t-ir}(\ell+1)) \end{aligned} \quad (3.34)$$

avec η : le paramètre d'apprentissage qui détermine la rapidité de convergence.

3.3.3. Mesure de la performance des modèles de prédiction

La performance d'un modèle de prédiction peut être jugée par sa capacité prédictive par les critères suivants :

- Erreur quadratique moyenne RMSE (Root Mean Square Error)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{t+i} - \hat{x}_{t+i})^2} \quad (3.35)$$

- Erreur moyenne absolue en pourcentage MAPE (Mean Absolute Pourcentage Error)

$$MAPE(\%) = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_{t+i} - \hat{x}_{t+i}}{x_{t+i}} \right| \times 100 \quad (3.36)$$

où N est le nombre total d'échantillons test. x_{t+i} et \hat{x}_{t+i} sont les valeurs réelles et prédites de l' $i^{\text{ème}}$ échantillon test.

La plupart des chercheurs se reportent essentiellement à l'erreur moyenne absolue MAPE (Mean Absolute Pourcentage Error) exprimée en pourcentage pour juger de la qualité d'un système prédictif. Le MAPE est préféré parce qu'il tient compte de l'ordre de grandeur du paramètre de mesure. Par contre, des études indiquent qu'un problème de prédiction est clairement non linéaire, et que des valeurs proches de zéro ($x_{t+i} \approx 0$) peuvent donner des résultats erronés (MAPE $\nearrow \infty$). Pour cette raison, les mesures basées sur l'erreur quadratique moyenne RMSE sont parfois suggérées car elles pénalisent les grandes erreurs. La RMSE permet d'éviter le cas « non déterminée » du point de vue mathématique, qui peut survenir lors de l'apparition d'une valeur ($x_{t+i} \approx 0$) trop proche de 0.

3.3.4. ANFIS et NNF pour la prédiction de séries temporelles chaotiques

Cette section a pour but de tester le NNF et ANFIS pour prédire la série temporelle de Mackey-Glass (MG) [Mackey_77]. Cette série temporelle est générée par l'équation différentielle suivante :

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1+x^{10}(x-\tau)} - 0.1x(t) \quad (3.37)$$

La série temporelle de MG est chaotique, il n'y a donc pas de période clairement définie. C'est un problème très connu dans la communauté des chercheurs pour tester les modèles de prédiction. Pour obtenir les données de la série temporelle, nous avons appliqué la méthode de Kunge kutta d'ordre 4 sur la formule (3.37). Nous supposons ici que : $x(0)=1.2$, $\tau=17$ et $x(t) = 0$ pour $t > 0$. La forme de la série temporelle de MG est donnée dans la figure 3.17.

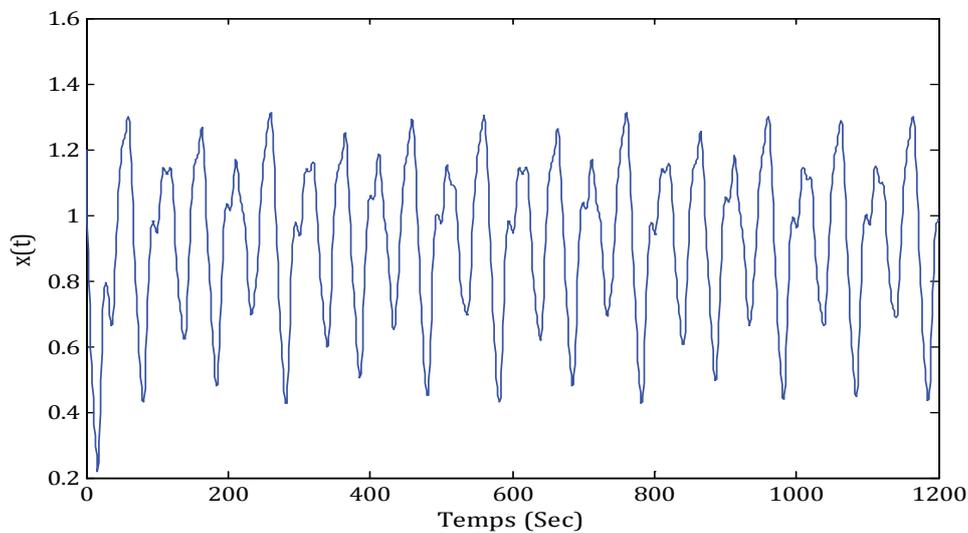


Figure 3. 17. Série temporelle de Mackey-Glass

Nous avons besoin d'utiliser des valeurs connues de la série temporelle de MG jusqu'à l'instant (t) pour donner une prédiction à un moment donné dans le futur, disons, à ($t+p$). La méthode standard pour ce type de prédiction est de créer une série temporelle $\{x_{t-(n-1)r}, \dots, x_{t-3r}, x_{t-2r}, x_{t-r}, x_t\}$ pour prédire la valeur future x_{t+p} . Suivant les paramètres classiques de prévision de la série temporelle de MG, nous posons $p=6, r=6$ et $n=4$. À chaque pas de temps t , les données d'entrée pour le NNF et ANFIS sont une série temporelle de la forme : $\{x_t, x_{t-6}, x_{t-12}, x_{t-18}\}$. La sortie du NNF et ANFIS correspond à la prédiction \hat{x}_{t+6} .

Nous choisissons un intervalle de données compris entre 118 et 1117 afin d'effectuer l'apprentissage et le test du NNF et d'ANFIS. Nous obtenons une base de données de 1000 échantillons. Les 500 premiers échantillons sont utilisés pour l'apprentissage tandis que les 500 restants sont utilisés comme base de test. Pour le système ANFIS, deux fonctions d'appartenance en forme de cloche sont affectées à chaque entrée. Ceci nous ramène à 16 règles floues et 104 paramètres à ajuster (24 paramètres prémisses et 80 paramètres conséquents). Les fonctions d'appartenance initiales qui sont espacées de manière égale et couvrant la totalité de l'espace d'entrée sont données dans la figure 3.18. Les fonctions d'appartenance après la phase d'apprentissage sont données dans la figure 3.19. Pour le NNF, 55 fonctions d'appartenance

équidistantes de type « triangle » sont affectées à chaque entrée. Ceci nous donne un total de 220 poids synaptiques à ajuster.

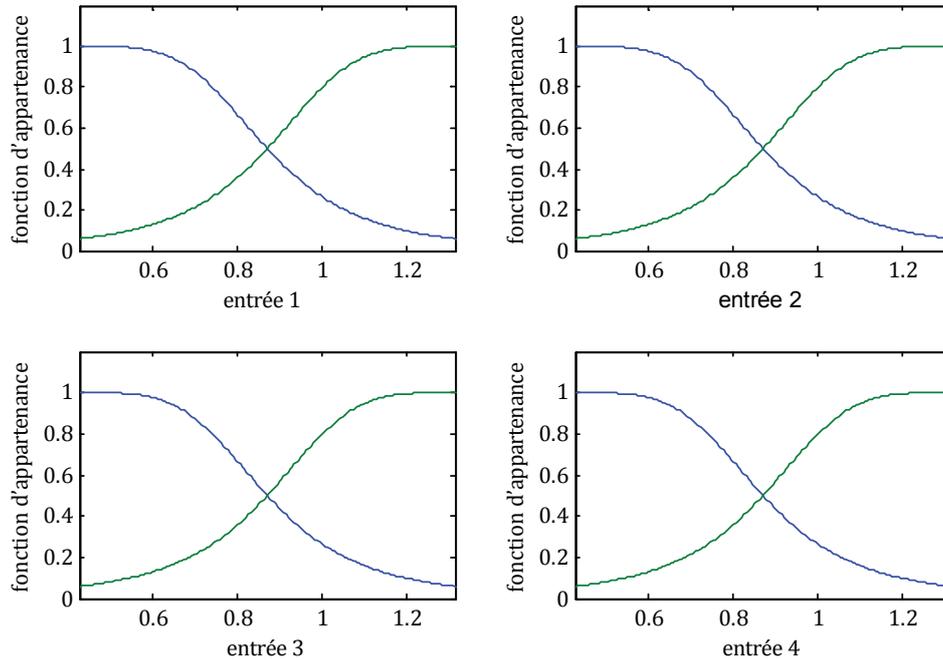


Figure 3.18. Fonctions d'appartenance initiales

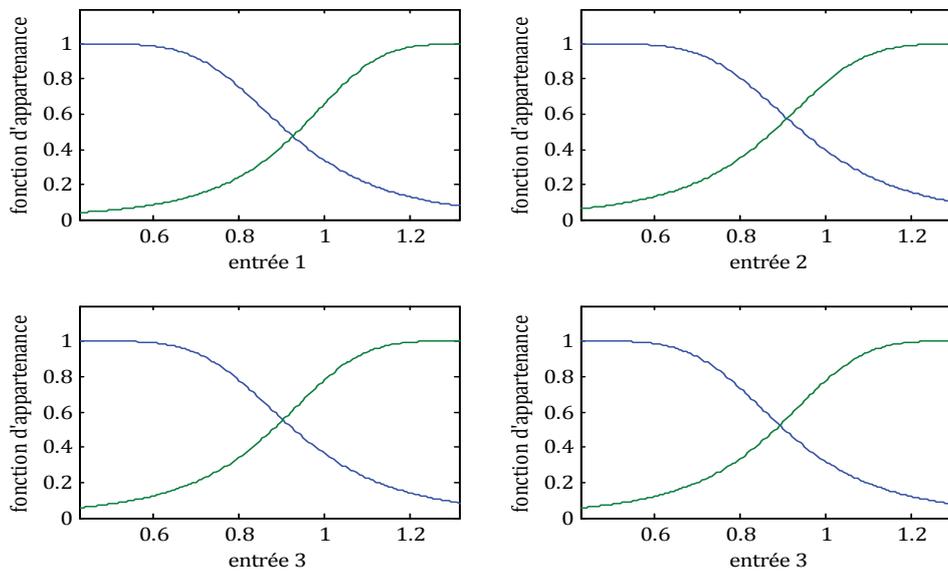


Figure 3.19. Fonctions d'appartenance après la phase d'apprentissage

Pour le système ANFIS, nous pouvons noter que les fonctions d'appartenance ne changent pas significativement après la phase d'apprentissage. Bien évidemment, le fitting se fait par des paramètres linéaires tandis que les paramètres non linéaires permettent les derniers ajustements.

L'erreur quadratique moyenne normalisée (NRMSE) des échantillons prédits par le modèle NNF et ANFIS ainsi que d'autres modèles de prédiction est donnée dans le tableau 3.6. L'erreur

quadratique moyenne normalisée est définie comme l'erreur quadratique moyenne divisée par l'écart type de la série temporelle cible.

Nous pouvons constater à partir du tableau 3.6 que les deux modèles proposés pour prédire la série temporelle de MG à court terme donnent de très bons résultats.

Tableau 3. 6. Erreur de prédiction pour les 500 échantillons test avec un pas de prédiction $p=6$

Méthode	NRMSE _{test}
Linear model prediction	0,55
Cascade correlation NN	0,06
6th-order polynomial	0,04
Genetic algorithm & Fuzzy System [Kim_97]	05 fonctions d'appartenance → 0,049 07 fonctions d'appartenance → 0,042 09 fonctions d'appartenance → 0,038
ANFIS	0,0077
NNF	0,031

Les figures 3.20 et 3.21 montrent le résultat de la prédiction des 500 échantillons test ainsi que l'erreur de prédiction pour le modèle NNF et ANFIS.

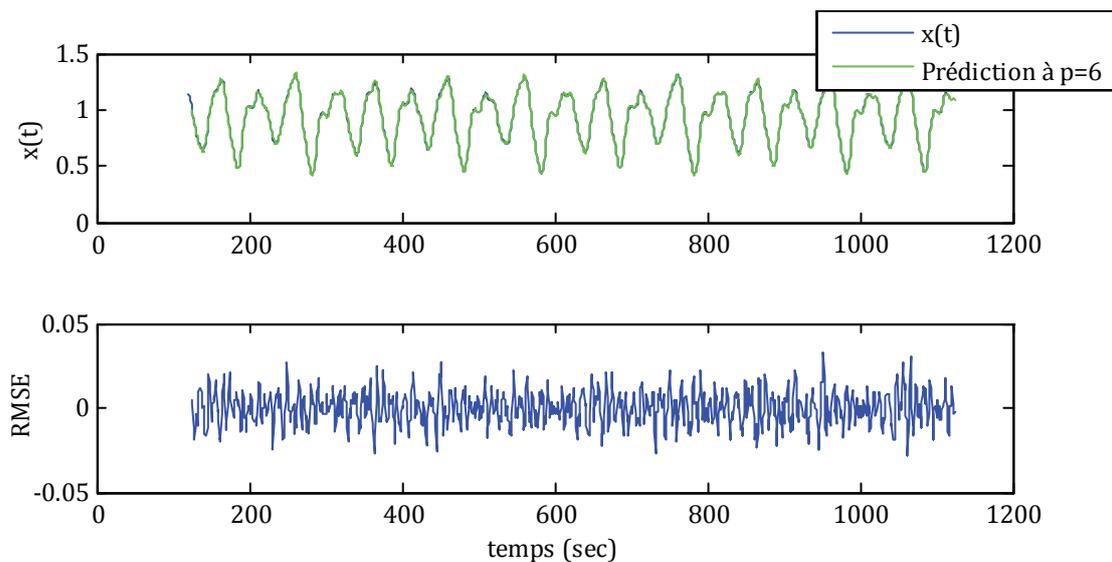


Figure 3. 20. Résultats de prédiction pour le NNF

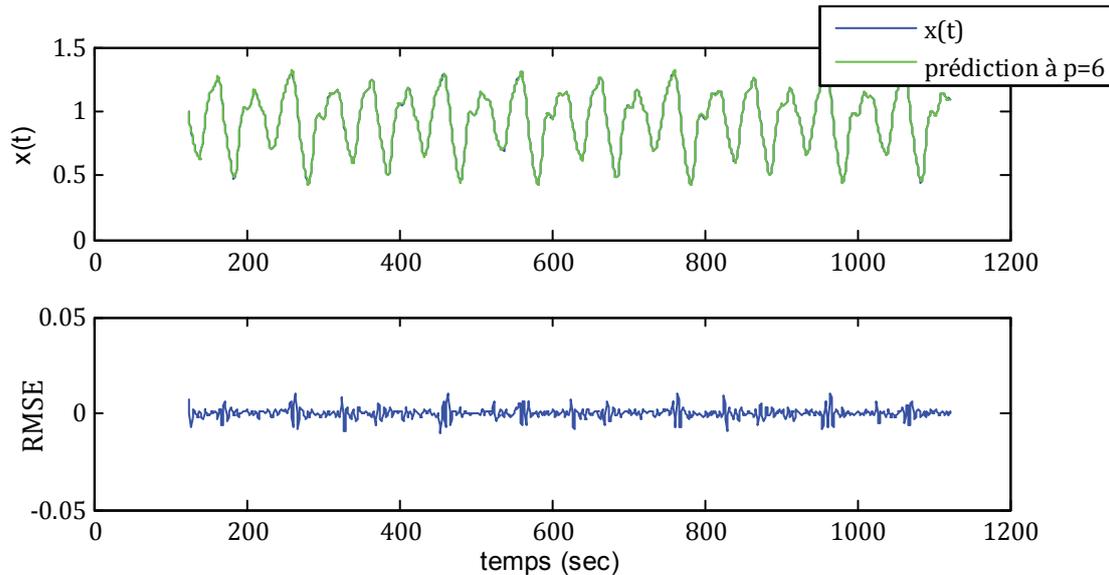


Figure 3. 21. Résultats de prédiction pour ANFIS

3.4. Conclusion

Comme mentionné précédemment, le pronostic d'un système signifie la prédiction de l'imminence d'un défaut et l'estimation de sa durée de vie restante. C'est dans ce contexte là que nous avons proposé cinq méthodes de pronostic. Pour la *première méthode*, le processus de dégradation de chaque système est modélisé par un modèle de Markov caché. Pour une séquence d'observations mesurée, la solution consiste à identifier parmi les MMCs construits celui qui représente le mieux cette séquence. L'identification du MMC qui représente le mieux la séquence d'observations permet de prédire l'imminence du prochain stade de dégradation, et donc le défaut du système étudié. Dans la *deuxième méthode* (modèle de Markov évolutif), le calcul de la probabilité qu'une séquence d'observations arrive à un stade de dégradation à l'instant $(t+1)$, étant donné le MMC modélisé à partir de cette même séquence d'observations, nous permet aussi de prédire l'imminence d'un défaut. La *troisième méthode* prédit l'imminence d'un défaut en modélisant non pas le processus de dégradation d'un système, mais en modélisant chaque stade de dégradation.

La *quatrième* et la *cinquième méthode* utilisent le modèle du neurone neo-flou et d'ANFIS pour la prédiction de la tendance des paramètres de surveillance. La classification des observations prédites par ces modèles dans une classe représentant un fonctionnement défectueux permet d'estimer le temps de vie restant avant l'apparition d'un défaut. Le neurone neo-flou constitue un outil qui offre de grands avantages pour la modélisation de systèmes complexes grâce à la simplicité de sa structure qui est constituée d'un seul neurone. Alors que dans les réseaux de neurones artificiels, il est nécessaire de changer le nombre de couches, le nombre de neurones dans chaque couche, et la fonction d'appartenance pour trouver la structure qui obtient un bon ajustement pour le problème à modéliser; dans le système neuronal neo-flou, il n'est nécessaire de modifier que le nombre de partitions floues des variables d'entrée, ce qui permet ainsi de trouver la structure la plus appropriée avec une plus grande facilité.

Le modèle du neurone neo-flou et d'ANFIS peuvent être utilisés pour la prédiction à court terme et comme nous le verrons dans le chapitre suivant, pour la prédiction à long terme.

Chapitre 4

Application au diagnostic de la machine asynchrone et au pronostic des roulements

Sommaire

4.1. Introduction	100
4.2. Présentation de la machine asynchrones à cage d'écurueil	100
4.3. Diagnostic des défauts d'une machine asynchrone	103
4.3.1. Présentation du banc de test diagnostic	103
4.3.2. Modes de fonctionnement étudiés	104
4.3.3. Construction du vecteur de paramètre par analyse des courants statoriques	106
4.3.4. Diagnostic d'une machine alimentée par onduleur	116
4.4. Pronostic des roulements	137
4.4.1. Présentation du banc de test pronostic	137
4.4.2. Validation des méthodes de pronostic	139

4.1. Introduction

Les deux chapitres précédents ont permis de présenter des méthodes de diagnostic et de pronostic en vue d'optimiser la maintenance préventive conditionnelle. Ce chapitre porte sur l'application de ces méthodes aux entraînements électriques : sur les défauts d'une machine asynchrone et sur ceux de roulements à rouleaux intégrés dans un actionneur asynchrone.

Dans la deuxième section, les défauts les plus courants pouvant apparaître dans la machine asynchrone sont rappelés. Ceux-ci peuvent provoquer des pannes et par conséquent des arrêts intempestifs préjudiciables au bon fonctionnement du système ainsi qu'à la sécurité des personnes.

Dans la troisième section, sont présentés le banc moteur ainsi que les différents modes de fonctionnement étudiés. Une étude complète sur la recherche de paramètres indicateurs de défauts en vue de construire le vecteur initial de paramètres est présentée. Ces paramètres sont calculés en utilisant l'analyse temporelle et fréquentielle des signaux électriques (tension et courant). Deux cas de figure sont étudiés dans cette section :

- Le premier concerne le diagnostic d'une machine asynchrone avec apparition d'un nouveau mode de fonctionnement ;
- Le second concerne le diagnostic d'une machine asynchrone avec une connaissance limitée sur chacun de ses modes de fonctionnement.

Une étude sur le choix du vecteur de paramètres, c'est-à-dire le choix du sous-ensemble de paramètres (au nombre de q) caractéristiques des modes de fonctionnements de la machine est présentée. Ce vecteur de paramètres permet aux méthodes de classification non-supervisée (fourmis artificielles) et supervisée (modèles de Markov cachés) de classer les nouvelles observations en vue de diagnostiquer la machine. Les résultats de diagnostic obtenus y sont présentés ainsi qu'une analyse sur les observations rejetées afin de mettre en évidence l'apparition d'un ou plusieurs nouveaux modes de fonctionnement encore inconnus de la base d'apprentissage.

La quatrième section est consacrée au pronostic des roulements à rouleaux. Un descriptif du banc de test y est présenté ainsi qu'une étude complète sur la recherche de paramètres indicateurs de défauts. Ces paramètres sont calculés en utilisant l'analyse temporelle et fréquentielle des signaux vibratoires. Un vecteur de paramètres est construit en vue d'identifier les différents stades de dégradation des roulements. À l'issue de cette étape, deux méthodes de pronostic basées sur les modèles de Markov cachés seront appliquées en vue de détecter l'imminence des prochains stades de dégradation et deux autres méthodes basées sur le système neuro-flou seront utilisées afin de prédire, à court et à long termes, l'évolution des paramètres de surveillance en vue d'estimer le temps restant avant l'apparition des prochains stades de dégradation. Cette section met en évidence l'efficacité des méthodes de pronostic utilisées pour suivre l'évolution de la dégradation des roulements.

4.2. Présentation de la machine asynchrones à cage d'écureuil

La machine asynchrone est une machine électrique utilisée dans presque toutes les applications industrielles en raison de sa fiabilité, sa simplicité de construction et sa robustesse. Elle est utilisée dans des équipements industriels comme les pompes, les convoyeurs, les

presses, ascenseurs... etc. Elle est également utilisée dans des zones à risque comme la pétrochimie, les centrales au gaz naturel et les industries minières, et dans des conditions d'environnement sévères trouvées dans les silos, les broyeurs et dans les équipements pour les centrales au charbon. Il en résulte une détérioration précoce de cette machine entraînant des défauts mécaniques et/ou électriques. Par conséquent, la maintenance préventive conditionnelle de ces machines est essentielle pour assurer la continuité de fonctionnement.

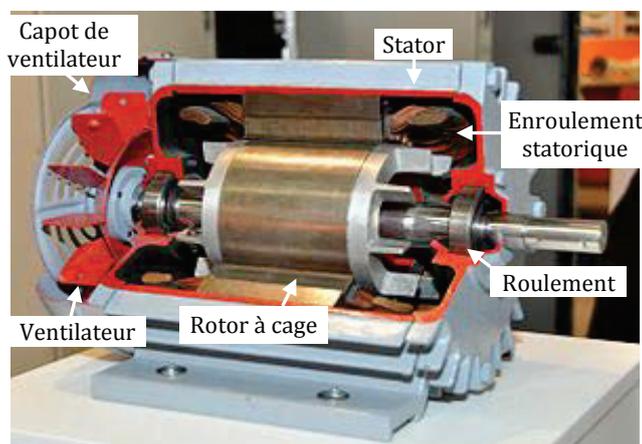


Figure 4. 1. Constitution générale d'une machine asynchrone à cage d'écureuil

Des études¹ menées sur la machine asynchrone ont classé les mécanismes de défauts les plus communs selon les défauts liés au stator, les défauts liés au rotor, les défauts de roulements et d'autres défauts (voir figure 4.2).

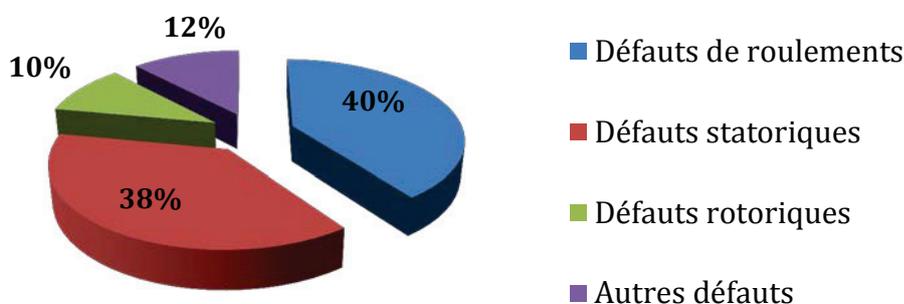


Figure 4. 2. Les pourcentages de défauts communs dans les machines asynchrones [Albrecht_86,Repo_85].

La figure 4.3 présente un diagramme qui répertorie une liste non-exhaustive des défaillances pouvant apparaître sur la machine asynchrone en définissant leurs lieux d'apparition ainsi que leurs causes possibles.

¹ Ces études ont été menées par le groupe "fiabilité du moteur" de la société "IEEE-Industry Applications" (IAS), qui a étudié 1141 moteurs, et l'institut de recherche sur l'énergie électrique (EPRI), qui a étudié 6312 moteurs au-dessus de 148 KW.

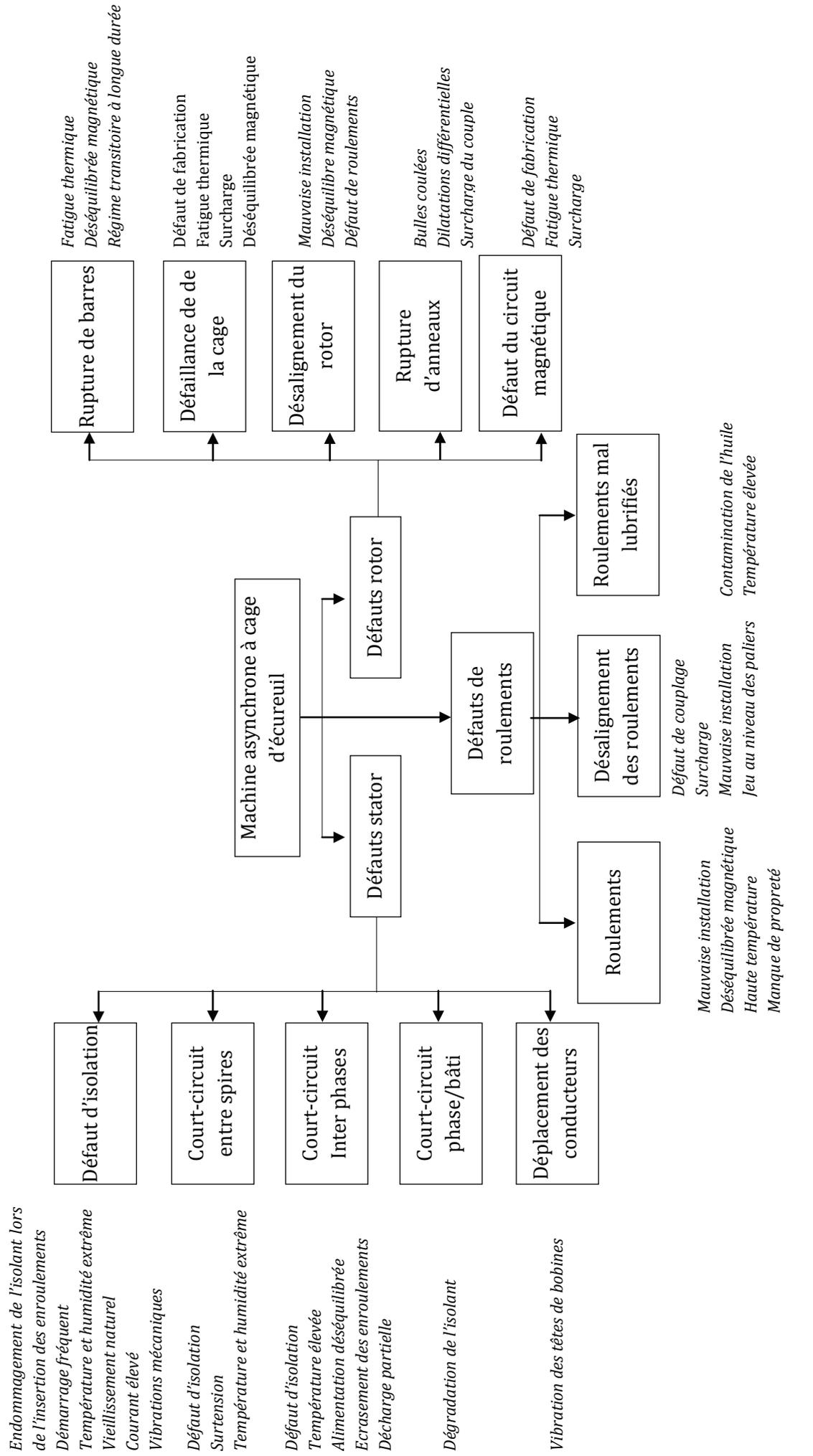


Figure 4. 3. Diagramme des défaillances sur la machine asynchrone à cage d'écoreuil

4.3. Diagnostic des défauts d'une machine asynchrone

4.3.1. Présentation du banc de test diagnostic

Le banc de test diagnostic, qui est présenté dans la figure 4.4, a été utilisé dans le cadre des travaux de thèse de [Ondel(II)_06] et [Casimir(III)_03]. Ce banc est composé d'un moteur asynchrone triphasé à cage d'écurueil de type Leroy Somer LS 132S, IP 55, classe F, T°C normalisée à 40°C, régime nominal S1. Les caractéristiques de ce moteur sont données dans le tableau 4.1.

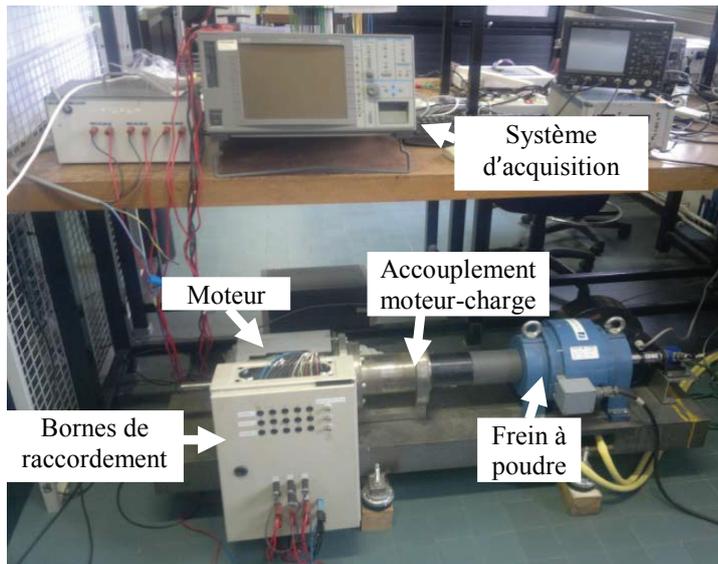


Figure 4. 4. Banc de test diagnostic sur une machine asynchrone de 5.5 Kw

Tableau 4. 1. Caractéristiques de la machine asynchrone

Puissance utile nominale	P_n	5,5 Kw
Fréquence d'alimentation	f_s	50 Hz
Vitesse nominale	N_n	1440 tr/min
Tension nominale entre phases	U_n	400 V
Courant nominal	I_n	11,4 A
Facteur de puissance	$\cos \varphi$	0,84
Nombre de paires de pôles	n_p	2
Résistance stator par phase	R_s	1,315 Ohms
Nombre d'encoches au rotor	N_r	28
Nombre d'encoches au stator	N_e	48

Les enroulements au stator sont couplés en étoile. Le moteur est chargé par un frein à poudre. Son couple maximal (100 Nm) est atteint à la vitesse nominale. Ce frein est dimensionné pour dissiper une puissance maximale de 5 kW.

Le système d'acquisition dispose de :

- Huit entrées différentielles dont la plage de la fréquence d'échantillonnage va jusqu'à 10 MHz avec une quantification sur 14 bits. Ces entrées serviront à mesurer les trois tensions ainsi que les trois courants d'alimentation de la machine ;

- Huit entrées en mode commun dites voies lentes sur lesquelles la fréquence d'échantillonnage peut atteindre 1 MHz avec une quantification sur 16 bits. Une de ces entrées sera utilisée pour mesurer la vitesse de rotation.

Les acquisitions sont réalisées en régime permanent sur une durée de 5 secondes, avec une fréquence d'échantillonnage de 20 kHz (soit 100000 points pour chacun des signaux mesurés).

La figure 4.5 représente le dispositif mis en place pour l'acquisition des signaux : ensemble machine-charge et acquisition des signaux.

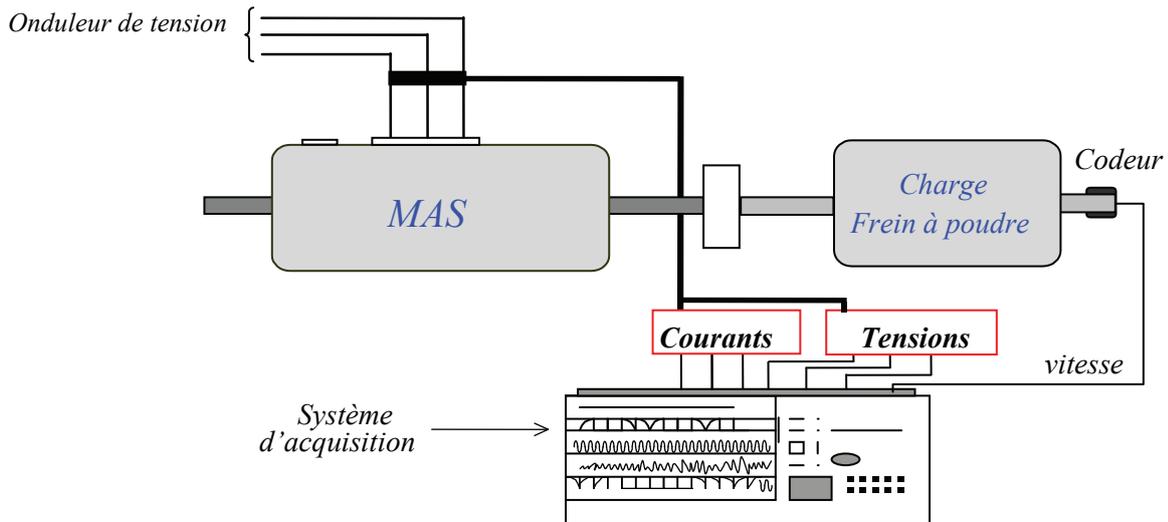


Figure 4. 5. Acquisition des signaux

4.3.2. Modes de fonctionnement étudiés

Les modes de fonctionnement ayant servi à valider la procédure de diagnostic sont repris dans le tableau 4.2. Les défauts ont été créés au rotor et aux roulements de la machine asynchrone.

Tableau 4. 2. Modes de fonctionnement servant à constituer l'ensemble d'apprentissage et l'ensemble test

Mode de fonctionnement	Types de défauts	Taux de charge	
		(Pour une puissance nominale : $P_u = 5.5 \text{ kW}$)	Alimentation
Machine saine	Aucun défaut	0%, 25%, 50%, 75%, 100%	onduleur
1 barre cassée au rotor	Défaut rotor (barres cassées)	0%, 25%, 50%, 75%, 100%	onduleur
3 barres cassées au rotor			
Usure progressive (vieillessement naturel)	Défaut roulement	25%, 50%, 75%, 100%	onduleur

Toutes ces mesures ont été réalisées dans le but de valider la mise en place de nos méthodes de diagnostic. Une partie de ces mesures va servir à construire l'ensemble d'apprentissage. Cet

ensemble sera utilisé pour modéliser les modèles de Markov cachés et identifier les q -paramètres optimaux.

Une autre partie des mesures servira pour la *phase d'exploitation*. De ce fait, ces mesures seront parfaitement disjointes de l'ensemble d'apprentissage et serviront ainsi à tester l'efficacité de nos méthodes.

Au niveau du rotor, les barres ont été détériorées ou endommagées par perçage. Les roulements (naturellement) usés ont été fournis par la société SECCO.

Les données obtenues sur la machine alimentée par onduleur permettront de montrer l'efficacité de nos méthodes. Elles mettront aussi en évidence l'importance de l'utilisation des signatures dites « temporelles » pour obtenir une bonne séparation des modes de fonctionnement.

La figure 4.6 montre une partie du schéma de principe de l'approche utilisée pour la maintenance préventive conditionnelle. Une description détaillée de chacun des blocs est donnée dans le chapitre 1 (section 1.3). Le procédé de diagnostic commence tout d'abord par identifier les défauts susceptibles de provoquer l'arrêt de la machine. Dans le cas de la machine asynchrone, nous avons identifié des défauts rotoriques, des défauts statoriques, des défauts de roulement et des défauts liés au problème d'excentricité. Différents types de capteurs peuvent être utilisés afin de mesurer les signaux caractéristiques (marqueurs) de ces défauts. Nous utiliserons dans notre cas des capteurs de courant, tension et de vitesse. Diverses techniques de traitement des signaux peuvent être appliquées à ces signaux pour extraire des paramètres sensibles à la présence des défauts. Nous utiliserons dans notre cas des paramètres extraits du domaine temporel et fréquentiel. Une technique de sélection de paramètres (Optimized Sequential Backward Selection) sera ensuite utilisée afin de ne garder que les paramètres permettant de mieux distinguer les modes de fonctionnement étudiés. Enfin, dans la phase diagnostic de défaut, la classification d'une observation dans l'un des modes de fonctionnement permet d'identifier le mode fonctionnement du système. Nos méthodes de classification interviennent dans cette étape.

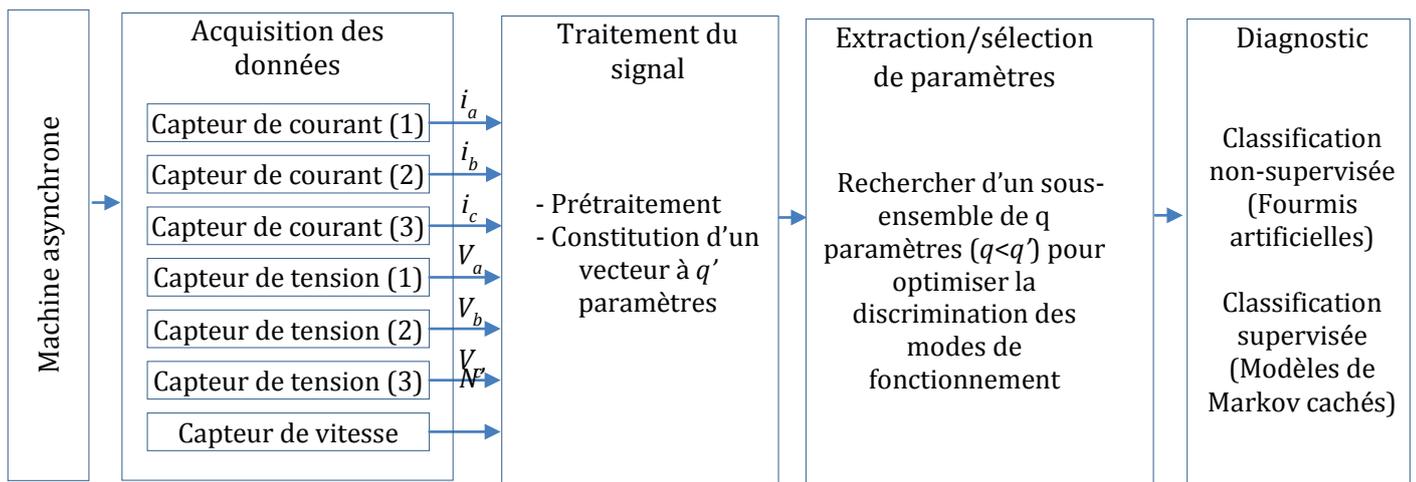


Figure 4. 6. Schéma de principe du processus de diagnostic

4.3.3. Construction du vecteur de paramètre par analyse des courants statoriques

4.3.3.1. Analyse fréquentielle

4.3.3.1.1. Prétraitement des courants statoriques

La signature fréquentielle spécifique d'un défaut, comme par exemple les « barres cassées », est liée à l'apparition de composantes fréquentielles du courant en relation avec le glissement. Cette technique, connue sous le nom de MCSA (Motor Current Signature Analysis), est largement employée depuis ces dernières années pour le diagnostic des défauts de la machine asynchrone. Elle consiste à réaliser une analyse fréquentielle (FFT) directement sur les courants statoriques i_a , i_b ou i_c afin d'identifier les fréquences caractéristiques d'un défaut spécifique. Cette technique repose sur l'analyse fréquentielle de chaque courant et implique donc un temps de calcul important. Pour résoudre ce problème, nous avons opté pour l'analyse spectrale de la composante symétrique directe.

Le système triphasé est la superposition de trois systèmes équilibrés de même fréquence, à savoir, un système triphasé direct, un système triphasé inverse et un système homopolaire : \underline{I}_d , \underline{I}_i et \underline{I}_0 avec :

\underline{I}_d représentant la composante symétrique directe ;

\underline{I}_i représentant la composante symétrique inverse ;

\underline{I}_0 représentant la composante homopolaire.

Ainsi, l'expression de la composante symétrique directe, notée \underline{I}_d , est obtenue à partir de la combinaison des trois courants statoriques i_a , i_b , i_c et de la matrice de Fortescue :

$$\underline{I}_d = i_a + \underline{a} \cdot i_b + \underline{a}^2 \cdot i_c \quad (4.1)$$

où l'opérateur \underline{a} représente un déphasage de $\frac{2\pi}{3}$: $\underline{a} = e^{j\frac{2\pi}{3}}$.

Pour la suite, l'analyse spectrale sera réalisée sur la partie réelle de \underline{I}_d , notée $\text{Re}[\underline{I}_d]$.

4.3.3.1.2. Ruptures de barres

Lors de l'apparition d'une rupture de barre, des harmoniques de flux sont produits et induisent des harmoniques de courant dans l'enroulement statorique aux fréquences :

$$f = f_s(1 \pm 2 \cdot g) \quad (4.2)$$

avec f_s : fréquence d'alimentation (fréquence fondamentale = supply frequency) et g : le glissement.

La figure 4.7 représente une analyse fréquentielle réalisée sur la partie réelle de \underline{I}_d .

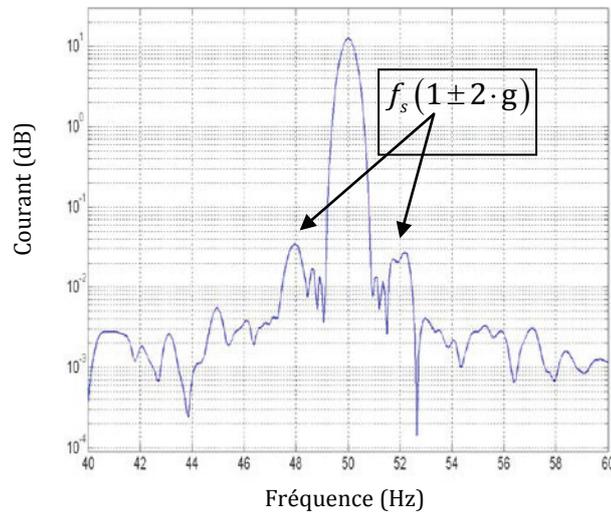


Figure 4. 7. Analyse fréquentielle pour une machine avec 1 barre cassée à pleine charge. FFT sur $\text{Re}[L_d]$

Le glissement (g) est exprimé en fonction de la vitesse de rotation de la machine N' en tr/min et aussi en fonction de la vitesse de synchronisme N_s . Il est donné comme suit :

$$g = \frac{N_s - N'}{N_s} \quad (4.3)$$

4.3.3.1.3. Excentricité

La variation de l'épaisseur de l'entrefer entraîne des variations dans la densité du flux d'entrefer. Les enroulements statoriques sont électriquement modifiés, ce qui entraîne des effets dans le spectre des courants aux fréquences f_{exc} .

Pour une excentricité statique, on a :

$$f_{exc} = f_s \cdot \left[1 \pm \frac{n \cdot N_r (1-g)}{n_p} \right] \quad (4.4)$$

Pour une excentricité dynamique, on a :

$$f_{exc} = f_s \cdot \left[1 \pm 2 \cdot n (1-g) \right] \quad (4.5)$$

Pour une excentricité mixte, on a :

$$f_{exc} = f_s \cdot \left[1 \pm \frac{n \cdot (1-g)}{n_p} \right] \quad (4.6)$$

avec :

- N_r : nombre d'encoches au rotor ;
- f_s : fréquence d'alimentation ;
- $n = 1, 2, 3, 4 \dots$;
- n_p : nombre de pair de pôles ;
- g : glissement.

4.3.3.1.4. Défauts de roulements

Une fréquence caractéristique f_c peut être associée à chaque type de défaut de roulement. Cette fréquence correspond à la périodicité à laquelle une anomalie apparaît en raison de l'existence du défaut. Elle est donnée comme suit :

$$f_c = f_s \pm n \cdot f_{i,e,b} \quad (4.7)$$

avec :

- $n = 1, 2, 3, 4, \dots$;
- f_i : fréquence caractéristique de la bague intérieure ;
- f_e : fréquence caractéristique de la bague extérieure ;
- f_b : fréquence caractéristique de la bille.

Les fréquences caractéristiques f_i , f_e , f_b dépendent de la géométrie du roulement et de la fréquence rotorique f_r . Un calcul détaillé de ces fréquences peut être trouvé dans [Schoen_95]. Elles sont données comme suit :

$$f_e = \frac{n_b}{2} \cdot f_r \cdot \left[1 - \frac{DB}{DP} \cdot \cos \psi \right] \quad (4.8)$$

$$f_i = \frac{n_b}{2} \cdot f_r \cdot \left[1 + \frac{DB}{DP} \cdot \cos \psi \right] \quad (4.9)$$

$$f_b = \frac{DP}{DB} \cdot f_r \cdot \left[1 - \frac{DB^2}{DP^2} \cdot \cos^2 \psi \right] \quad (4.10)$$

avec :

- f_r : fréquence de rotation du rotor ;
- ψ : angle de contact ;
- n_b : nombre de billes ;
- DB : diamètre de bille ;
- DP : diamètre primitif du roulement.

4.3.3.1.5. Défauts au niveau du stator

Dans les travaux de recherche de [Casimir(III)_03] et [Ondel(II)_06], l'impédance directe (Z_d) et la puissance moyenne (P_d) sont considérés comme étant de bons paramètres pour détecter les déséquilibres d'alimentation au niveau du stator. Si on considère les transformées en fréquence, $I_d(f)$ et $V_d(f)$ de $\text{Re}[I_d]$ et $\text{Re}[V_d]$, l'impédance directe Z_d , peut être calculée à partir du rapport entre le fondamental de $V_d(f)$ et celui de $I_d(f)$:

$$Z_d = \frac{V_d(f=f_s)}{I_d(f=f_s)} \quad (4.11)$$

Si nous désignons par $I_d(f)$ la transformée en fréquence de $\text{Re}[I_d]$. La densité de puissance spectrale du courant est :

$$S_d(f) = I_d(f) \cdot \overline{I_d(f)} = |I_d(f)|^2 \quad (4.12)$$

Alors, la puissance moyenne du signal vaut :

$$P_d = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{\infty} S_d(f) df \quad (4.13)$$

4.3.3.1.6. Autres paramètres

Les fréquences engendrées par l'alimentation de l'onduleur sont rajoutées dans le vecteur de paramètres et sont résumées dans le tableau 4.3.

Tableau 4. 3. Fréquences contenues dans le vecteur de paramètres et induites par l'alimentation onduleur

Fréquences	Dénomination
$5f_s$	Harmonique 5 du fondamental
$7f_s$	Harmonique 7 du fondamental
$9f_s$	Harmonique 9 du fondamental
$11f_s$	Harmonique 11 du fondamental
$13f_s$	Harmonique 13 du fondamental
$17f_s$	Harmonique 17 du fondamental
$19f_s$	Harmonique 19 du fondamental
$23f_s$	Harmonique 23 du fondamental
$25f_s$	Harmonique 25 du fondamental
$29f_s$	Harmonique 29 du fondamental

Afin d'analyser la sensibilité de ces fréquences, nous avons utilisé deux modes de fonctionnement différents pour une machine alimentée par onduleur :

- Machine saine, à vide et en charge ;
- Machine avec 4 barres cassées au rotor, à vide et en charge.

Dans les figure 4.8 et 4.9, les amplitudes des harmoniques correspondent à une moyenne des 15 acquisitions réalisées par mode de fonctionnement. L'analyse fréquentielle (FFT) est réalisée sur la partie réelle de la composante I_d et elle est calculée sur les 100000 points du signal.

A vide, on remarque que les harmoniques de rang (5, 7, 11, 13, 17, 19, 23, 25, 29) respectivement $5f_s$, $7f_s$, $11f_s$, $13f_s$, $17f_s$, $19f_s$, $23f_s$, $25f_s$, $29f_s$ permettent de différencier le défaut rotor de la machine saine. Ces harmoniques permettent aussi la même différenciation en charge. Il est donc intéressant de rajouter ces fréquences dans le vecteur de paramètres initial car l'analyse des résultats a mis en évidence la possibilité de différencier les modes de fonctionnement aussi bien à vide qu'en charge.

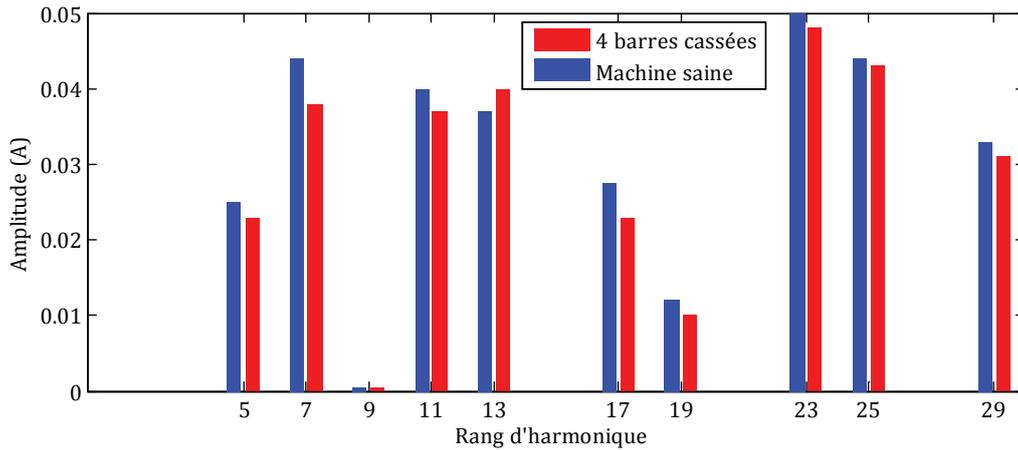


Figure 4. 8. Variation de l’amplitude correspondant aux fréquences recherchées. Fonctionnement à vide

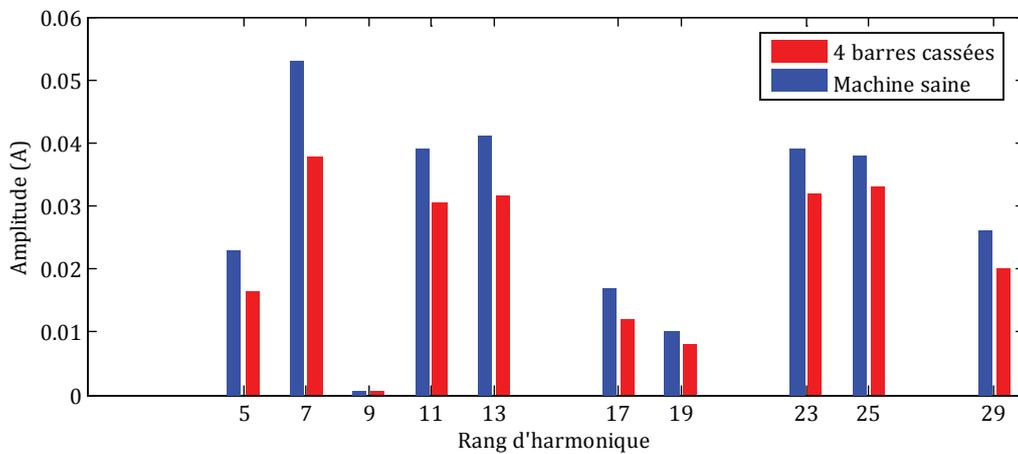


Figure 4. 9. Variation de l’amplitude correspondant aux fréquences recherchées. Fonctionnement en charge

4.3.3.2. Analyse temporelle

4.3.3.2.1. Prétraitement des courants statoriques

Une représentation bidimensionnelle des composantes triphasées des courants statoriques peut être utilisée afin de détecter les défauts d’une machine asynchrone. Les composantes qui découle de cette représentation, notées I_α, I_β, I_0 , sont exprimée comme suit :

$$\begin{bmatrix} I_\alpha \\ I_\beta \\ I_0 \end{bmatrix} = [C] \cdot \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} \tag{4.14}$$

avec : [C], la matrice de Concordia.

4.3.3.2.2. Ruptures de barres

Considérons le module des composantes biphasées du courant stator, noté $I_{\alpha\beta}$, en fonction des composantes en quadratures :

$$I_{\alpha,\beta} = \left\| \overline{I_{\alpha,\beta}} \right\| = \sqrt{I_{\alpha}^2 + I_{\beta}^2} \quad (4.15)$$

Considérons aussi la valeur efficace ($I_{\alpha,\beta}$), notée $I_{\alpha,\beta \text{ eff}}$:

$$I_{\alpha,\beta \text{ eff}} = \left[\frac{1}{N_p} \sum_{\ell=1}^{N_p} I_{\alpha,\beta}^2(\ell) \right]^{1/2} \quad (4.16)$$

La caractéristique $I_{\beta}(I_{\alpha})$ normalisée par rapport à $I_{\alpha,\beta \text{ eff}}$ montre une forme circulaire centrée sur l'origine comme illustrée dans la figure 4.10. Cette figure permet la détection de défauts de barres en calculant la déformation ($\delta_{\omega} \delta_{\beta}$) de cette caractéristique.

Les deux paramètres, notés ($\delta_{\omega} \delta_{\beta}$) sont les mesures de la déformation de cette caractéristique à $\varphi=0$ et à $\varphi=\frac{\pi}{2}$, autrement dit dans les axes α et β :

$$\begin{cases} \delta_{\alpha} = \left| \max(I_{\alpha,\beta}) - \min(I_{\alpha,\beta}) \right|_{\varphi=0} \\ \delta_{\beta} = \left| \max(I_{\alpha,\beta}) - \min(I_{\alpha,\beta}) \right|_{\varphi=\frac{\pi}{2}} \end{cases} \quad (4.17)$$

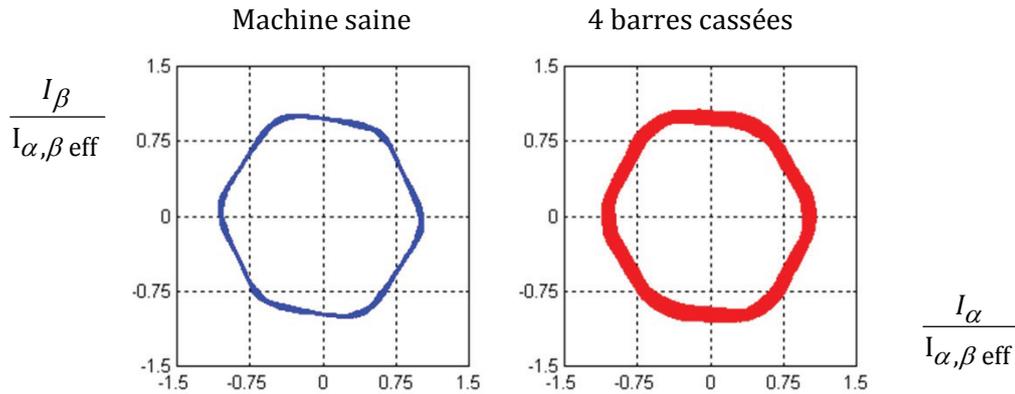


Figure 4. 10. Caractéristique $I_{\beta}(I_{\alpha})$ normalisée par rapport à $I_{\alpha,\beta \text{ eff}}$. Fonctionnement en charge

La figure 4.11 illustre l'évolution temporelle des courants en alimentation réseau, dans le cas d'un défaut rotorique (4 barres cassées dans cet exemple) par rapport à un fonctionnement sain. Il est possible d'extraire à partir de l'évolution temporelle des composantes en quadrature I_{α} et I_{β} les écarts-types ($\sigma_{\omega} \sigma_{\beta} \sigma_s$) correspondant aux trois signaux ($I_{\alpha}, I_{\beta}, I_{\alpha,\beta}$) à l'aide de leurs valeurs moyennes notées respectivement m_{α}, m_{β} et $m_{\alpha,\beta}$:

$$\sigma_{\alpha} = \left[\frac{1}{N_p} \sum_{\ell=1}^{N_p} (I_{\alpha}(\ell) - m_{\alpha})^2 \right]^{1/2} \quad (4.18)$$

$$\sigma_{\beta} = \left[\frac{1}{N_p} \sum_{\ell=1}^{N_p} (I_{\beta}(\ell) - m_{\beta}) \right]^{1/2} \quad (4.19)$$

$$\sigma_s = \left[\frac{1}{N_p} \sum_{\ell=1}^{N_p} (I_{\alpha,\beta}(\ell) - m_{\alpha,\beta}) \right]^{1/2} \quad (4.20)$$

avec N_p : le nombre de points dans les signaux.

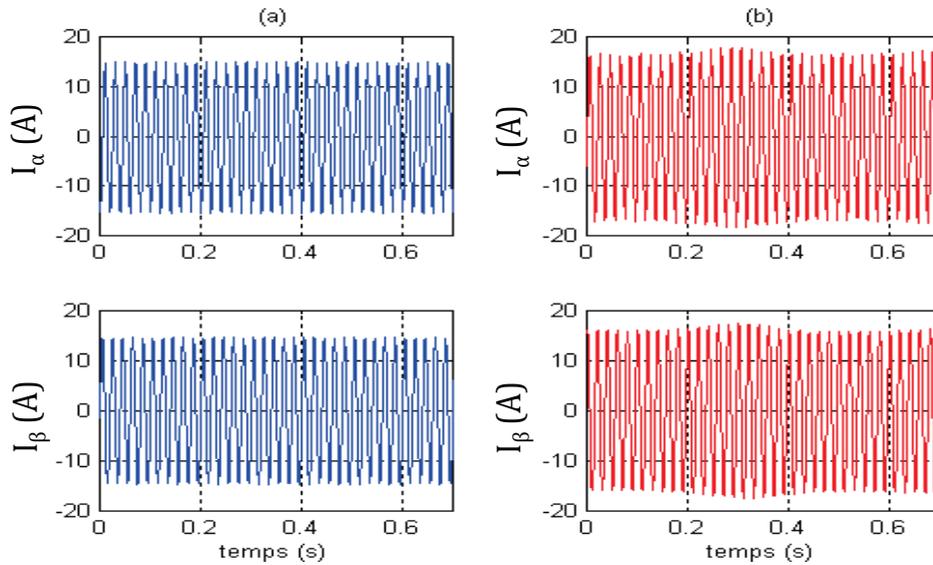


Figure 4. 11. Evolution des composantes I_{α} et I_{β} des courants stators pour un rotor sain (a) et pour un rotor ayant 4 barres cassées (b)

4.3.3.2.3. Défauts au niveau du stator

Les travaux de recherche de [Casimir(III)_03] montrent qu'il est possible d'extraire trois paramètres indicateurs du déséquilibre d'alimentation stator à partir des puissances réactives et actives. Les deux premiers paramètres, notés m_q et m_p , représentent respectivement la moyenne de la puissance réactive et active normalisés par la valeur RMS du module de la puissance apparente $S = \sqrt{P^2 + Q^2}$ et notées Q' et P' .

$$m_p = \frac{1}{N_p} \sum_{\ell=1}^{N_p} P'(\ell) \quad (4.21)$$

$$m_q = \frac{1}{N_p} \sum_{\ell=1}^{N_p} Q'(\ell) \quad (4.22)$$

avec :

$$\begin{aligned} P &= V_{\alpha} \cdot I_{\alpha} + V_{\beta} \cdot I_{\beta} \\ Q &= V_{\beta} \cdot I_{\alpha} - V_{\alpha} \cdot I_{\beta} \end{aligned} \quad P' = \frac{P}{S} \text{ et } Q' = \frac{Q}{S}$$

[Casimir(III)_03] montre que la puissance active moyenne m_p et la puissance réactive moyenne m_q peuvent aussi être utilisées comme indicateurs de défauts de roulements.

Le troisième paramètre, noté ζ , est obtenu en calculant la dispersion du nuage de points formé par la caractéristique normalisée $Q'(P')$ (voir figure 4.12) qui représente l'évolution de la puissance apparente consommée par la machine à chaque instant. Il est donné comme suit :

$$\zeta = \sum_{\ell=1}^{N_p} (x_{\ell} - m_{p,q})(x_{\ell} - m_{p,q})' \tag{4.23}$$

avec $m_{p,q} = [m_p, m_q]$, le centre de gravité de l'ensemble et x_k un point de coordonnées $(P'(\ell), Q'(\ell))$, N_p est le nombre de points qui compose le nuage.

Là encore, [Casimir(III)_03] montre que la dispersion des points de la caractéristique $Q'(P')$ peut être utilisée comme indicateur de rupture de barres ou de portion d'anneaux.

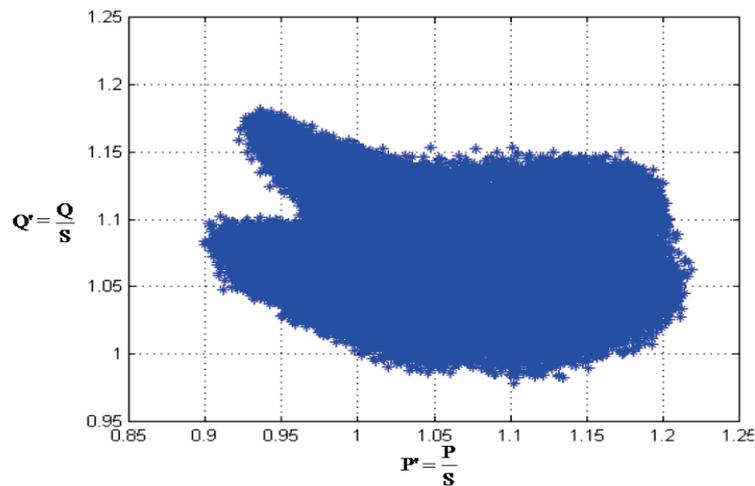


Figure 4. 12. Caractéristique normalisée $Q'(P')$ obtenue pour la machine saine, pour une charge de 3 kW en alimentation réseau

4.3.3.2.4. Excentricité

Le facteur de puissance $\cos \varphi$ peut être un bon indicateur pour la détection d'excentricité au rotor [Chang_94].

Le tableau 4.4 résume l'ensemble des q' -paramètres qui vont être utilisés pour le diagnostic de la machine asynchrone.

Υ_i	Dénomination		Υ_i	Dénomination	
1	m_p	Puissance active moyenne	17	$N_r f_r - 2f_s - 2gf_s$	Amplitudes autour de $N_r f_r \pm 2f_s$
2	m_q	Puissance réactive moyenne	18	$N_r f_r - 2f_s$	
3	ζ	Dispersion des points de la caractéristique $Q'(P')$	19	$N_r f_r - 2f_s + 2gf_s$	
4	$f_s(1-4g)$	Amplitudes autour de f_s	20	$N_r f_r + 2f_s - 2gf_s$	
5	$f_s(1-2g)$		21	$N_r f_r + 2f_s$	
6	$f_s(1+2g)$		22	$N_r f_r + 2f_s + 2gf_s$	
7	$f_s(1+4g)$		23	$N_r f_r - 3f_s - 2gf_s$	
8	$f_s - f_r$	Amplitudes autour de $f_s \pm f_r$	24	$N_r f_r - 3f_s$	Amplitudes autour de $N_r f_r \pm 3f_s$
9	$f_s + f_r$		25	$N_r f_r - 3f_s + 2gf_s$	
10	$3f_s$	Harmonique 3 du fondamental	26	$N_r f_r + 3f_s - 2gf_s$	
11	$N_r f_r - f_s(1+2g)$	Amplitudes autour de $N_r f_r \pm f_s$	27	$N_r f_r + 3f_s$	
12	$N_r f_r - f_s$		28	$N_r f_r + 3f_s + 2gf_s$	
13	$N_r f_r - f_s(1-2g)$		29	$5f_s$	
14	$N_r f_r + f_s(1-2g)$		30	$7f_s$	Harmonique 7 du fondamental
15	$N_r f_r + f_s$		31	$9f_s$	Harmonique 9 du fondamental
16	$N_r f_r + f_s(1+2g)$	32	$11f_s$	Harmonique 11 du fondamental	

Υ_i	Dénomination		Υ_i	Dénomination	
33	$13f_s$	Harmonique 13 du fondamental	46	cos φ	Déphasage entre la tension et le courant d'une phase
34	$17f_s$	Harmonique 17 du fondamental			
35	$19f_s$	Harmonique 19 du fondamental			
36	$23f_s$	Harmonique 23 du fondamental			
37	$25f_s$	Harmonique 25 du fondamental			
38	$29f_s$	Harmonique 29 du fondamental			
39	P_d	Puissance moyenne de la composante directe des courants			
40	δ_α	Déformation de la caractéristique $I_\beta(I_\alpha)$ dans l'axe α			
41	δ_β	Déformation de la caractéristique $I_\beta(I_\alpha)$ dans l'axe β			
42	σ_α	Ecart - type de la composante I_α des courants			
43	σ_β	Ecart - type de la composante I_β des courants			
44	σ_s	Ecart - type du module de $I_\alpha + jI_\beta$			
45	Z_d	Impédance directe			

Tableau 4. 4. Paramètres calculés constituant le vecteur initial à q'-paramètres

4.3.4. Diagnostic d'une machine alimentée par onduleur

4.3.4.1. Diagnostic avec apparition d'un nouveau mode de fonctionnement

Les modes de fonctionnement qui vont servir à valider la procédure de diagnostic sont présentés dans le tableau 4.5. Chacun de ces modes est représenté par une classe dans l'espace de représentation (voir chapitre 2, figure 2.2). Les défauts ont été créés au rotor de la machine asynchrone et ceci pour un fonctionnement à vide et en charge.

Dans cet exemple, compte tenu des données disponibles, les classes prennent en compte le niveau de charge pour montrer l'applicabilité de nos méthodes. Par ailleurs, ces méthodes seraient applicables avec des données normalisées avec le niveau de charge.

Tableau 4. 5. Modes de fonctionnement étudiés

Mode de fonctionnement	Types de défauts	Niveau de charge	
		(Pour une puissance nominale : $P_u = 5.5 \text{ kW}$)	Alimentation
Machine saine	Aucun défaut	0%, 100%	onduleur
1 barre cassée au rotor	Défaut rotor (barres cassées)	0%, 100%	onduleur

Comme indiqué dans la figure 4.13, la barre cassée a été créée artificiellement en perçant une barre de la cage d'écureuil. L'objectif est de vérifier les performances de nos méthodes de classification sur la base d'un ensemble d'apprentissage incomplet. À cette fin, nous supposons la non disponibilité d'informations sur le mode défectueux à 1 barre cassée à 0% comme indiqué dans le tableau 4.6.



Figure 4. 13. Barre cassée au niveau du rotor de la machine asynchrone à cage d'écureuil

Tableau 4. 6. Composition de l'ensemble d'apprentissage

Mode de fonctionnement	Classe	Niveau de charge	Nombre d'observations
Machine saine	Ω_1	0 %	8
Machine saine	Ω_2	100%	8
1 barre cassée	Ω_3	100%	8
1 barre cassée	Ω_4	0%	inconnu

L'ensemble d'apprentissage est constitué de trois modes de fonctionnement représentés par 03 classes. Chaque classe est composée de 8 observations par niveau de charge, ce qui fait au

final, que les classes sont composées de 24 observations. Ces classes vont nous servir par la suite afin de déterminer l'espace de représentation.

4.3.4.1.1. Détermination de l'espace de représentation (sélection des q -paramètres)

L'algorithme OSBS (Optimized Sequential Backward Selection) utilise le calcul d'un critère sur les classes regroupant l'ensemble d'apprentissage afin de sélectionner les q -paramètres. Le critère utilisé est la trace des valeurs propres de $\Sigma_W^{-1} \cdot \Sigma_B$. Il représente le rapport entre la dispersion inter-classes et la dispersion intra-classe.

$$J = \text{trace}(\Sigma_W^{-1} \cdot \Sigma_B) \quad (4.24)$$

avec Σ_W et Σ_B : Les matrices de variance intra-classe et inter-classes.

Le vecteur de paramètres initial regroupe l'ensemble des q' -paramètres calculés précédemment (voir tableau 4.4). La première étape consiste à sélectionner un ensemble de q'' -paramètres parmi les q' -paramètres initiaux en éliminant les paramètres qui minimisent le critère J . L'étape suivante consistera à sélectionner les sous-espaces à q -paramètres qui maximisent le critère J . Il ne s'agit pas d'éliminer le paramètre le moins pertinent, mais de conserver le sous-ensemble qui caractérise le mieux l'espace de représentation en accord avec la dimension (q) désirée (Chapitre 2, section 2.2). Le mode de sélection des q'' -paramètres suit la procédure suivante : à l'aide de l'ensemble d'apprentissage, et en partant de l'ensemble initial à q' -paramètres, le critère (4.24) est calculé pour chaque paramètre. Ce calcul est une évaluation de la capacité de chaque paramètre à discriminer les classes. Un paramètre (Y_i) aura un pouvoir de séparation d'autant meilleur que le critère J qui lui est associé est grand.

Le critère J indique d'après la figure 4.14 qu'il y a 13 paramètres qui, pris séparément, ont un pouvoir discriminant important ($q''=13$). La figure 4.14 montre clairement que les paramètres Y_1 , Y_2 et Y_{29} ont les pouvoirs discriminants les plus élevés. Mais ce résultat ne signifie pas pour autant qu'il faille limiter la dimension (q) du vecteur de paramètres à trois. En effet, celui-ci doit être défini sur la base d'un apport collectif des q'' -paramètres discriminant.

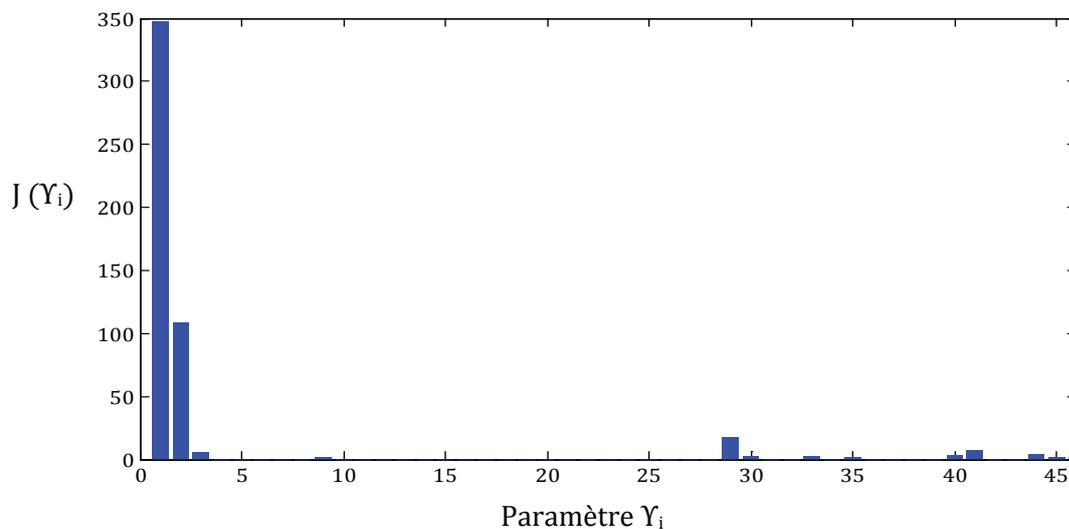


Figure 4. 14. Variation du critère J en fonction de chaque paramètre

Donc d'après le résultat de la figure 4.14, la dimension du vecteur q -paramètres peut aller de 1 à 13. La figure 4.15 illustre la réalisation de la sélection des sous-ensembles des q -paramètres les plus pertinents à partir de l'ensemble des q'' -paramètres.

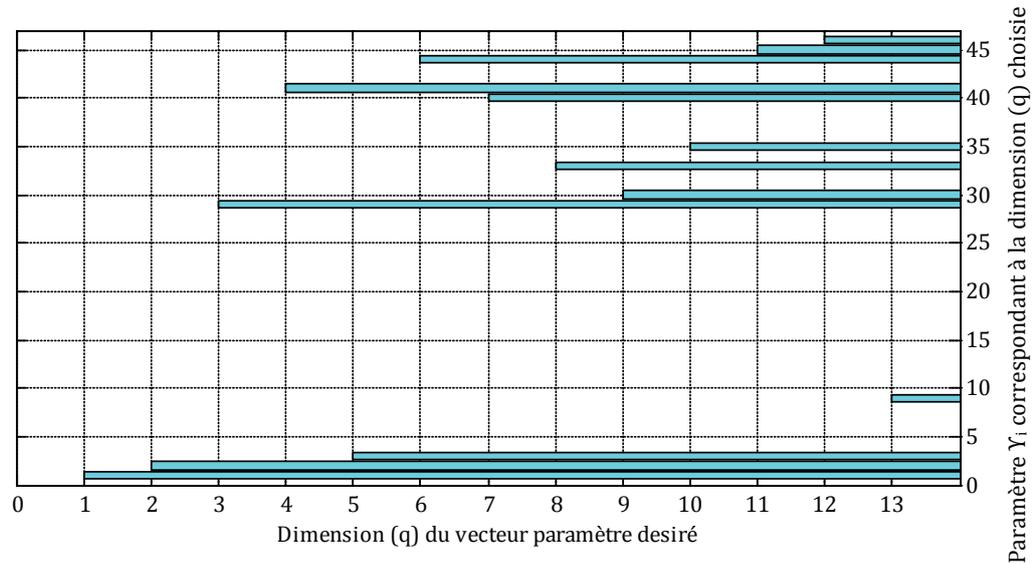


Figure 4. 15. Sélection de l'ensemble q -paramètres par l'algorithme OSBS

A partir de la figure 4.15, le meilleur vecteur de paramètres, parmi les 13 solutions possibles, est celui qui donne les meilleures performances au sens du nombre d'observations mal classées de l'ensemble d'apprentissage. La figure 4.16 donne le nombre d'observations mal classées dans toutes les configurations du vecteur de paramètres possibles. Cette analyse a été réalisée à l'aide de l'algorithme fournis artificielles.

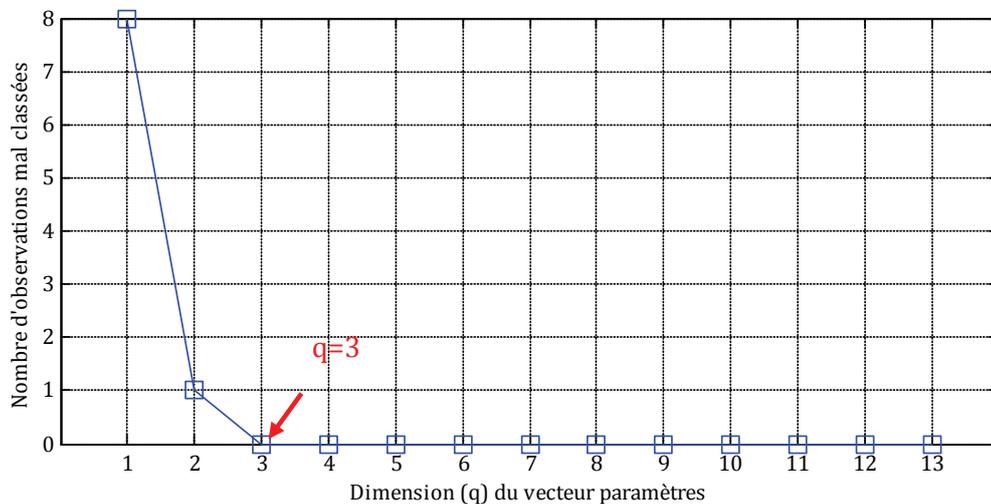


Figure 4. 16. Choix de la dimension (q) du vecteur de paramètres

Nous pouvons remarquer que la procédure de sélection est optimale lorsque $q=3$ (avec 0% d'erreur de classification), le sous-ensemble de paramètres constituant le vecteur de paramètres est composé des paramètres Y_1 , Y_2 et Y_{29} , c'est-à-dire :

- La puissance active moyenne (m_p) ;
- La puissance réactive moyenne (m_q) ;
- L'harmonique 5 du fondamental ($5f_s$).

Ces trois paramètres, $[\gamma_1, \gamma_2, \gamma_{29}]$, suffisent à obtenir une très bonne séparation des classes comme le montre la figure 4.17.

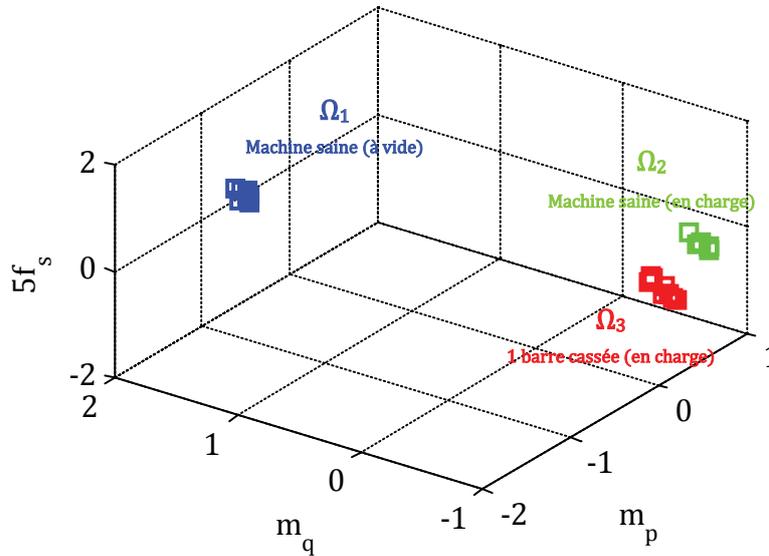


Figure 4. 17. Position des classes de l’ensemble d’apprentissage dans le sous- espace de dimension 3. Les paramètres sont sélectionnés avec l’algorithme OSBS

4.3.4.1.2. Validation des méthodes de classification

L’ensemble test utilisé pour nos méthodes de classification est donné dans le tableau suivant :

Tableau 4. 7. Composition de l’ensemble test

Mode de fonctionnement	Classe	Niveau de charge	Nombre d’observations
Machine saine	Ω_1	0 %	8
Machine saine	Ω_2	100%	8
1 barre cassée	Ω_3	100%	8
1 barre cassée	Ω_4	0%	8

Comme nous pouvons le voir, l’ensemble test est composé de 8 observations par niveau de charge et par mode de fonctionnement, ce qui fait au final que l’ensemble test est constitué de 32 observations.

Il s’agit maintenant de valider nos méthodes de classification mises en place sur l’ensemble test en se basant sur le vecteur de paramètres $[\gamma_1, \gamma_2, \gamma_{29}]$.

4.3.4.1.2.1. Algorithme fourmis artificielles (FA)

- Etape 1 (Chapitre 2, section 2.3.1.4.3)

La première étape de l’algorithme FA permet de constituer un premier ensemble de m -classes en se basant sur la similarité entre les observations donnée par la formule (2.30) ainsi que sur un seuil de similarité qui reste à déterminer.

La figure 4.18 montre l'évolution du critère J en fonction des différentes valeurs de seuil de similarité (S_{sim}). Nous pouvons distinguer à partir de cette figure trois zones bien distinctes qui indiquent l'évolution du critère J (formule (4.24)). Dans la première zone, S_{sim} décroît de 239 à 91. Cette décroissance correspond à un regroupement d'observations qui varie entre 12 classes pour $J = 239$ et 7 classes pour $J = 91$. Il est normal que le critère J pour 12 classes soit égal à 239 car la plus part des classes ne contiennent qu'une seule observation. La variance intra-classe dans ce cas est voisine de 0 ce qui donne une valeur élevée pour J. En revanche, plus le nombre de classes diminue plus J diminue car la variance intra-classe augmente. Ceci nous place dans la deuxième zone avec $J \in [91-48]$ et la troisième zone avec $J \in [48-13]$. Un seuil de similarité fixé à 0,1 risque d'être pénalisant en temps de calcul surtout pour déterminer les paramètres (e et f) de l'algorithme (voir étape 2). Un seuil de similarité fixé à 0,4 ne nous permet pas d'optimiser la classification car le nombre de classes ne change pas au-delà de cette valeur. Le seuil de similarité sera donc fixé à 0,2 afin d'augmenter le critère J, ce qui nous permettra par la suite d'améliorer la classification.

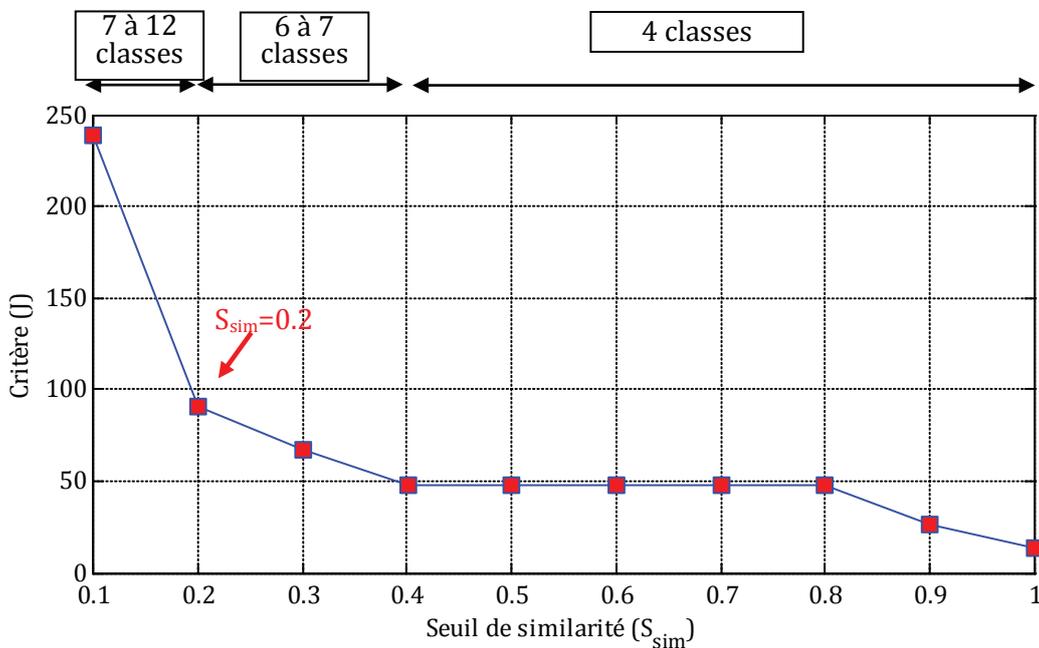


Figure 4. 18. Evolution du critère J en fonction du seuil de similarité

Un seuil de similarité fixé à 0,2 permet à la première étape de l'algorithme fournis artificielles d'obtenir un premier ensemble de m -classes, avec $m=7$.

- Etape 2 (Chapitre 2, section 2.3.1.4.3)

La deuxième étape permet de rassembler les classes précédentes en un nombre de classes plus restreint. Il faut pour cela définir les valeurs des différents paramètres de l'algorithme FA. Comme indiqué dans la section (2.3.1.4.3) du chapitre 2, il faut définir la valeur de la quantité de phéromone ξ , l'évaporation de la phéromone ε et aussi les paramètres e et f qui définissent l'influence relative du poids de chaque classe et la dispersion inter-classes. Le paramètre ε est fixé à 0,01 afin limiter le poids de chaque classe. Le paramètre ξ est arbitrairement fixé à 1. La détermination des paramètres e et f est obtenue grâce au suivi du critère J donné par la formule (4.24). Les valeurs de e et f sont initialement choisies dans l'intervalle $[-1,1]$. Le critère J le plus

élevé définit les meilleurs paramètres e et f , tel que : pour J avoisinant les 28 %, $e = -0,9$ et $f = 0,6$ (voir figure 4.19).

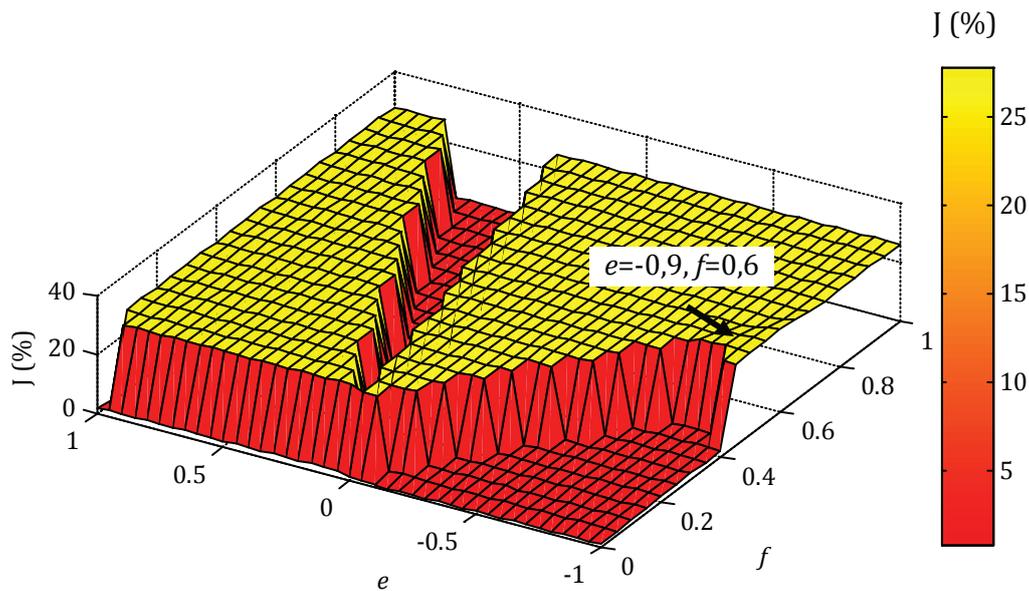


Figure 4.19. Détermination des paramètres e et f à partir du critère J

- Etape 3 (Chapitre 2, section 2.3.1.4.3)

La détermination des paramètres e et f permet à la deuxième étape de l'algorithme FA d'obtenir un deuxième ensemble de m' -classes, avec $m'=4$. Une analyse en composantes principales est effectuée afin de changer l'axe d'orientation des paramètres et permettre ainsi une meilleure visualisation des classes. La figure 4.20 montre la position des classes de l'ensemble test dans le sous - espace de dimension 3 obtenu par l'ACP.

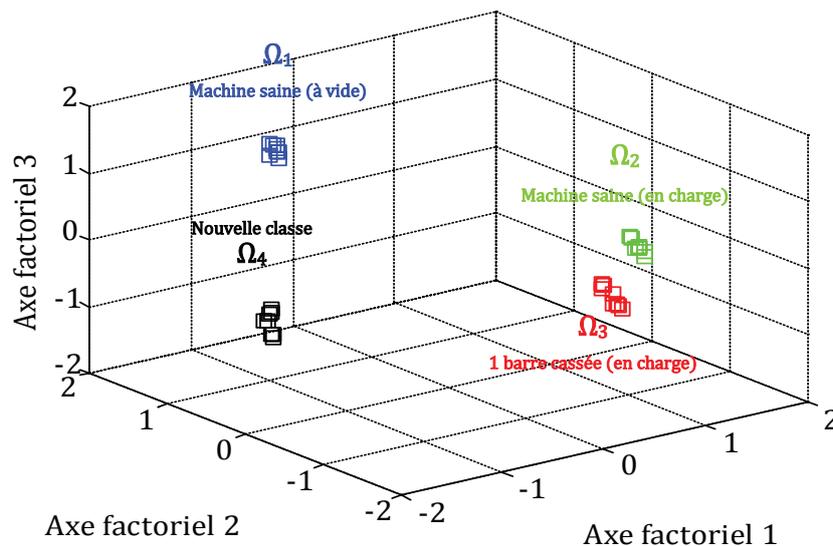


Figure 4.20. Position des classes dans le sous - espace de dimension 3. Machine saine (à vide) «□», machine saine (en charge) «□», 1 barre cassée (à vide) «□», 1 barre cassée (en charge)

«□»

Nous pouvons constater à partir de cette figure que les classes obtenues sont compactes et donc non séparables en sous-classes par l’algorithme K-means. Ceci est confirmé par le calcul du critère J qui indique que la séparation de ces classes induit une diminution de 50% au niveau du critère J. L’utilisation de l’ensemble d’apprentissage permet l’identification des classes ($\Omega_1, \Omega_2, \Omega_3$) et la détection d’une nouvelle classe notée Ω_4 . Les observations qui constituent cette classe sont à ce stade considérées comme rejetées. Une analyse sera effectuée par la suite pour confirmer l’apparition d’un nouveau mode de fonctionnement.

- Analyse des nouvelles observations

Les observations regroupées dans la nouvelle classe peuvent représenter un mode inconnu de fonctionnement. Pour déterminer si ces observations caractérisent l’apparition d’une nouvelle classe, c’est-à-dire d’un nouveau mode de fonctionnement, la formule (2.41) est utilisée. Elle est donnée comme suit :

$$Cq'(\Omega_{new}) = \left(\frac{n_{new}}{N} \right) trace \left(\frac{(g_{new} - g) \cdot (g_{new} - g)'}{\sum_{j=1}^{n_{new}} (x_j - g_{new}) \cdot (x_j - g_{new})'} \right) \tag{4.25}$$

avec :

- g : centre de gravité des observations constituant l’ensemble des M -classes ;
- g_{new} : centre de gravité de la classe Ω_{new} ;
- x_j : $j^{ième}$ observation de la classe Ω_{new} ;
- n_{new} : nombre d’observations de la classe Ω_{new} ;
- N : nombre total de l’ensemble d’observations.

Cette formule indique que le critère Cq' pour la classe Ω_4 est du même ordre de grandeur que les classes déjà identifiées (voir figure 4.21). Nous pouvons alors confirmer qu’il y a apparition d’un nouveau mode de fonctionnement.

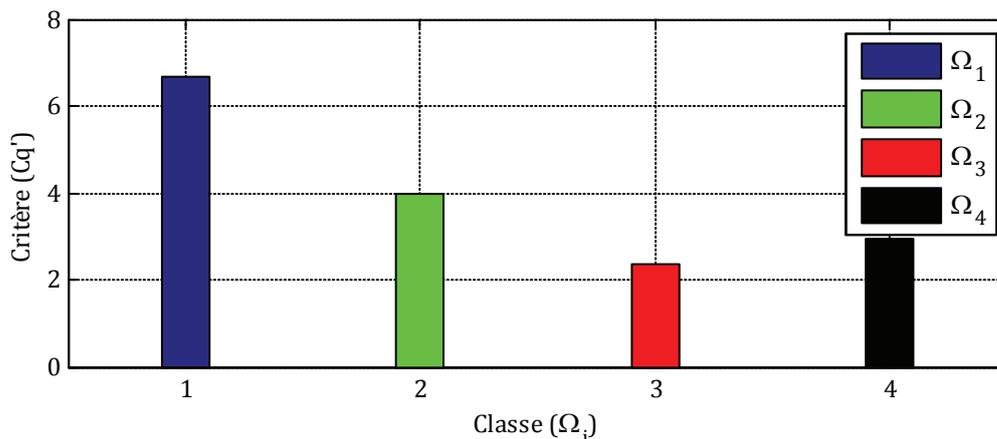


Figure 4. 21. Détection de l’apparition d’une nouvelle classe

La comparaison des classes obtenues après classification par rapport aux classes réelles est illustrée ci-dessous. On obtient une erreur de classification $E_c=0\%$.

Classement obtenu	Ω_1	8 0%	0 0%	0 0%	0 0%
	Ω_2	0 0%	8 0%	0 0%	0 0%
	Ω_3	0 0%	0 0%	8 0%	0 0%
	Ω_4	0 0%	0 0%	0 0%	8 0%
		Ω_1	Ω_2	Ω_3	Ω_4

Classement réel

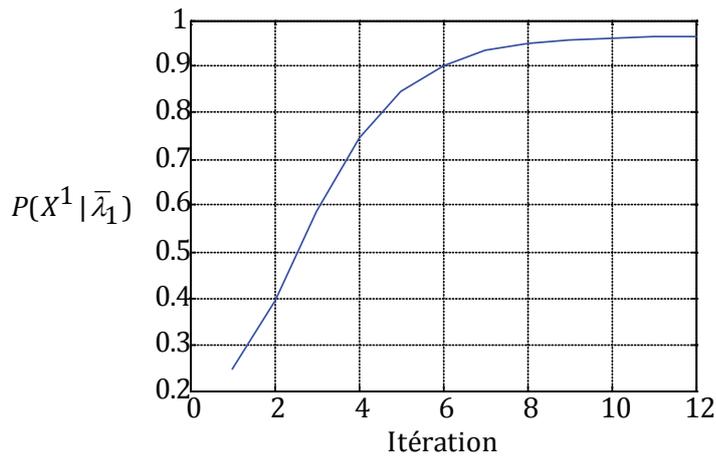
4.3.4.1.2.2. Classification par les modèles de Markov cachés

Pour la classification par modèles de Markov cachés, nous devons déterminer 03 MMCs correspondant aux 03 modes de fonctionnement connus : machine saine (à vide) (S_1), machine saine (en charge) (S_2) et machine avec une barre cassée (en charge) (S_3). À partir de chaque classe, nous extrayons une séquence d'observations de longueur ($T = 3$) (Chapitre 2, section 2.3.2.3.2). La longueur T correspond à un tirage successif de $T+1$ observations de chacune des classes. Les séquences d'observations utilisées pour estimer les paramètres des MMCs peuvent être formulées par : $X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_T^k\}$, avec $1 \leq k \leq 3$ et $T = 3$. x_j^k désigne la $j^{\text{ème}}$ observation extraite de la classe Ω_k . Les paramètres initiaux de chaque MMC (matrice de transition $A^{(k)}$ et vecteur d'état initial $\pi^{(k)}$), à partir des quels nous commencerons l'apprentissage, sont définis comme suit :

$$A^{(1,2,3)} = \begin{matrix} & \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \end{matrix} \\ \begin{matrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{matrix} & \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix} \quad \pi^{(1,2,3)} = \begin{matrix} & \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \end{matrix} \\ \begin{pmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \end{pmatrix} \end{matrix} \quad (4.26)$$

Notons que les probabilités de transition a_{31}, a_{32} sont fixées à zéros. Ceci se traduit par le fait que nous imposons une direction de transition qui débute par la classe Ω_1 et qui se termine par la classe Ω_3 . Dans notre cas, les classes Ω_1 et Ω_2 correspondent à une machine saine à deux niveaux de charge et Ω_3 correspond à une machine avec 1 barre cassée en charge. Dans ce cas, il est normal de fixer à zéros a_{31}, a_{32} car nous ne prenons pas en compte le fait qu'un système défectueux puisse revenir à un fonctionnement normal.

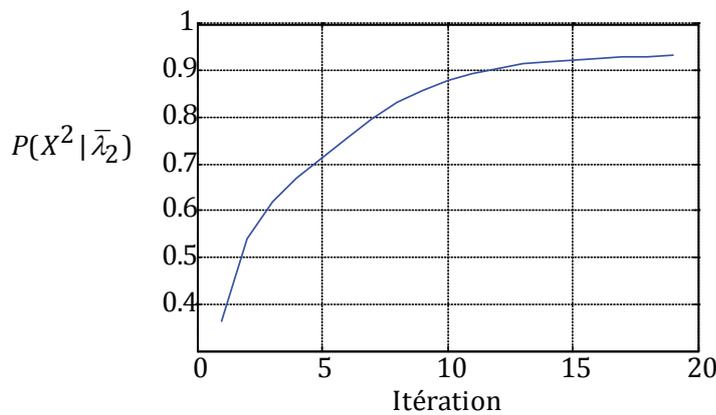
Les figures 4.22 (a, b, c) montrent l'évolution de la probabilité $P(X^k | \bar{\lambda}_k)$ durant la phase d'apprentissage et cela pour chaque MMC. Nous pouvons remarquer, par exemple, que la probabilité de générer la séquence d'observations X^1 , étant donné les paramètres ré-estimés du modèle $\bar{\lambda}_1$, avoisine les 0,95. Ce résultat indique que les paramètres du modèle $\bar{\lambda}_1$ représentent correctement le mode de fonctionnement étudié.


 (a) Apprentissage la classe Ω_1

$$\bar{A}^1 = \begin{bmatrix} 0,9997 & 0,0003 & 0 \\ 0,9997 & 0,0003 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

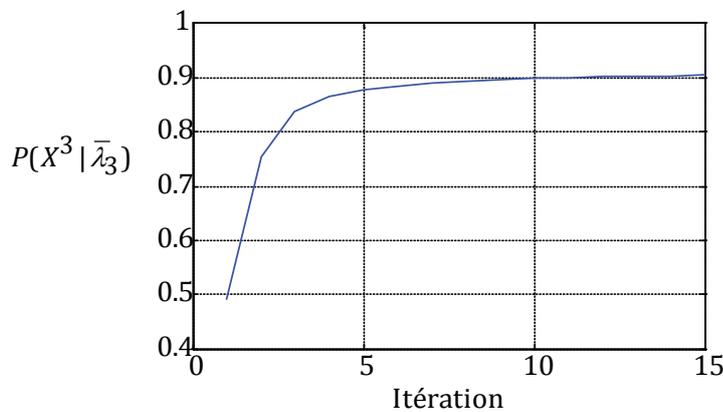
$$\bar{\pi}^1 = [0,9997 \quad 0,0003 \quad 0]$$

MMC de la machine saine (à vide)


 (b) Apprentissage la classe Ω_2

$$\bar{A}^2 = \begin{bmatrix} 0 & 0,9975 & 0,0025 \\ 0 & 0,9972 & 0,0028 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\bar{\pi}^2 = [0 \quad 0,9999 \quad 0,0001]$$

 MMC de la machine saine
(en charge)

 (c) Apprentissage la classe Ω_3

$$\bar{A}^3 = \begin{bmatrix} 0 & 0,0234 & 0,9766 \\ 0 & 0,0236 & 0,9764 \\ 0 & 0 & 1 \end{bmatrix}$$

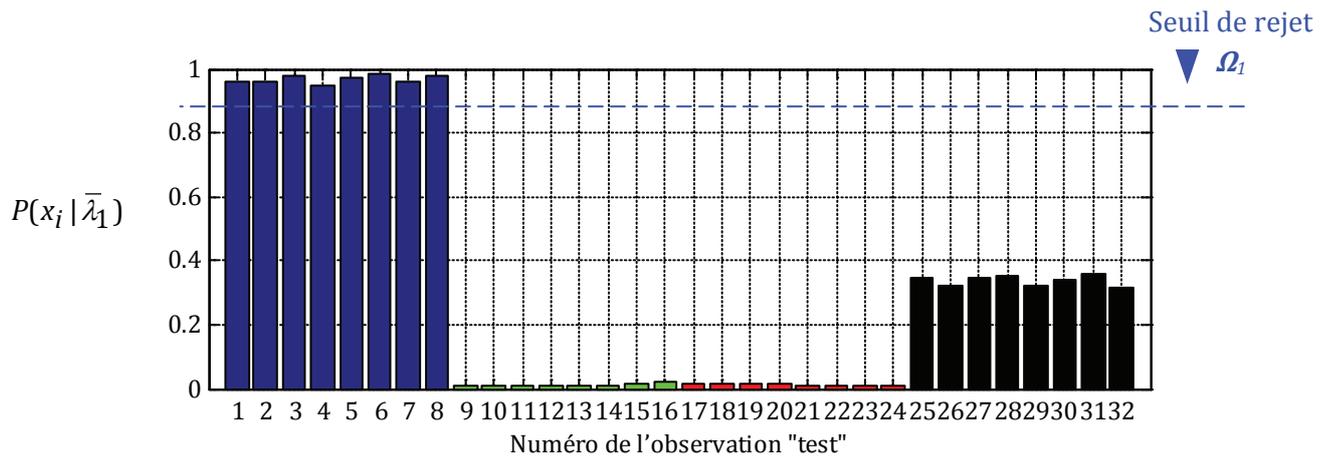
$$\bar{\pi}^3 = [0 \quad 0,0167 \quad 0,9833]$$

 MMC de la machine avec une barre
cassée (en charge)

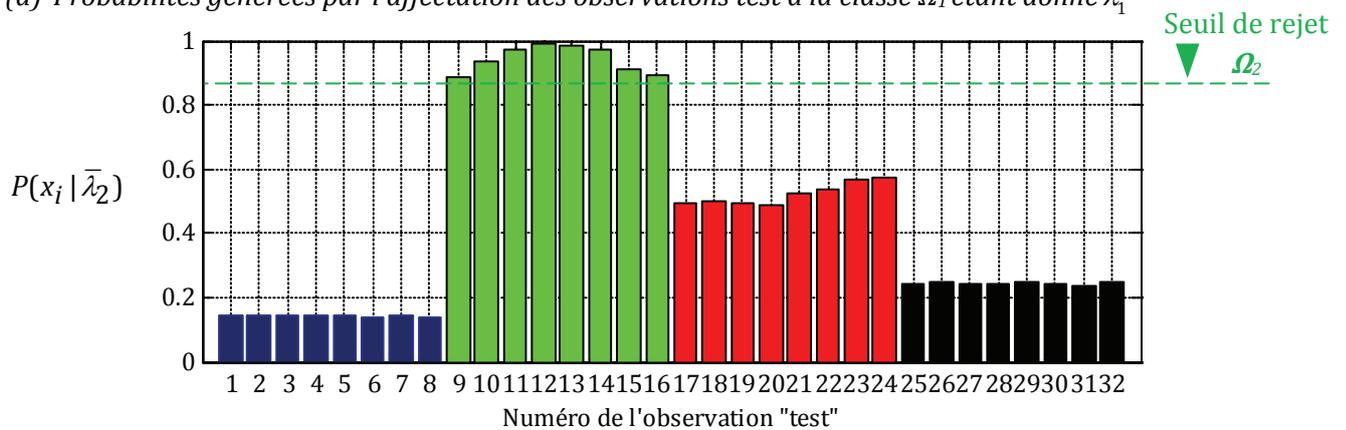
Figure 4.22. Processus d'apprentissage des classes par ré-estimation de Baum-Welsh

Les matrices de transition qui ont été identifiées pour chaque mode de fonctionnement indiquent que la probabilité $P(q_t = S_k | q_{t-1} = S_k, \bar{\lambda}_k) \approx 1$ pour $1 \leq k \leq 3$. En se référant à la formule (2.53), définie dans le chapitre 2, cela signifie une bonne séparation entre les classes $(\Omega_1, \Omega_2, \Omega_3)$ (voir figure 4.17).

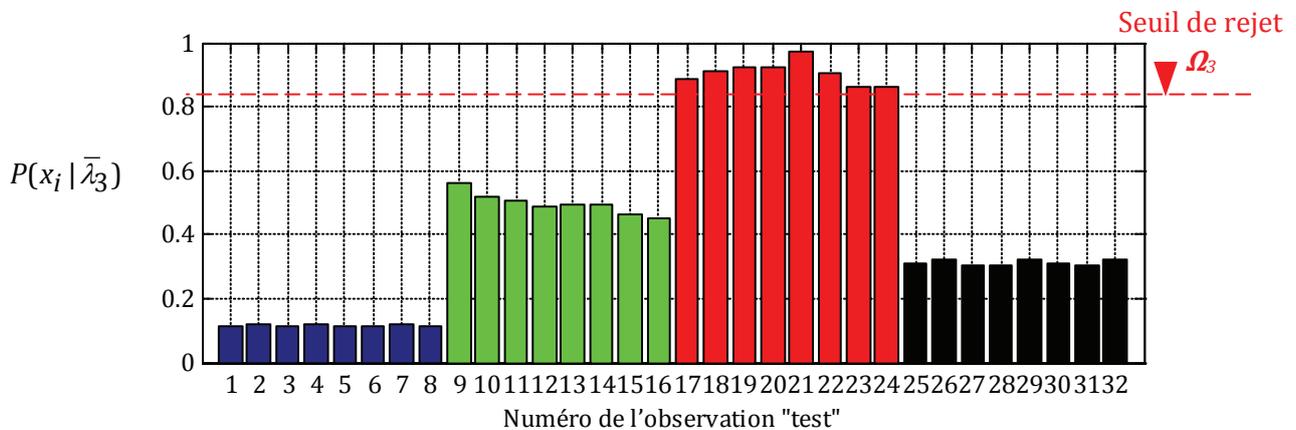
Les figure 4.23 (a, b, c) montrent les probabilités générées par l'affectation de l'ensemble d'observations « test » aux classes $(\Omega_1, \Omega_2, \Omega_3)$ étant donné les MMCs $\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_3$:



(a) Probabilités générées par l'affectation des observations test à la classe Ω_1 étant donné $\bar{\lambda}_1$



(b) Probabilités générées par l'affectation des observations test à la classe Ω_2 étant donné $\bar{\lambda}_2$



(c) Probabilités générées par l'affectation des observations test à la classe Ω_3 étant donné $\bar{\lambda}_3$

Figure 4. 23. Classification de l'ensemble des observations test par modèles de Markov cachés

Les probabilités générées par l'affectation des 32 observations « test » à la classe Ω_1 varient entre 0,0106 et 0,979 avec 25 % de ces observations affectées avec une probabilité supérieure à 0,94. Afin d'éviter de mal classer ces observations, nous fixons un seuil de rejet égal à 0,94 afin de n'affecter que les observations qui correspondent le plus à la classe Ω_1 . Ceci nous permet d'affecter avec certitude 8/32 observations à la classe Ω_1 .

Les probabilités générées par l'affectation des 32 observations à la classe Ω_2 varient entre 0,14 et 0,99 avec 25% de ces observations affectées avec une probabilité supérieure à 0,86. Là aussi, afin d'éviter de mal classer ces observations, nous fixons un seuil de rejet égal à 0,85 afin de

n'affecter que les observations qui correspondent le plus à la classe Ω_2 . Ceci permet d'affecter avec certitude 8/32 observations à la classe Ω_2 .

Les probabilités générées par l'affectation des 32 observations à la classe Ω_3 varient entre 0,15 et 0,97, avec 25% de ces observations affectées avec une probabilité supérieure à 0,85. Afin d'éviter de mal classer ces observations, nous fixons un seuil de rejet égal à 0,84 afin de n'affecter que les observations qui correspondent le plus à la classe Ω_3 . Ceci permet d'affecter avec certitude 8/32 observations à la classe Ω_3 .

Les observations numérotées de 25 à 32 indiquent une machine avec une barre cassée (à vide). Les probabilités générées par l'affectation de ces 8 observations varient entre 0,24 et 0,34 pour les modèles $\bar{\lambda}_1, \bar{\lambda}_2$ et $\bar{\lambda}_3$. Ces faibles valeurs ne permettent pas leurs affectations aux classes ($\Omega_1, \Omega_2, \Omega_3$) à cause de leurs faibles probabilités d'appartenance. Ces observations sont considérées à ce stade comme étant rejetées. Cependant, il est possible d'effectuer la même analyse que celle de l'algorithme FA qui montrera bien l'apparition d'un nouveau mode fonctionnement (voir figure 4.21).

La répartition des observations sur les classes obtenues par les MMCs est donnée comme suit, avec $E_c=0\%$.

Classement obtenu	Ω_1	8 0%	0 0%	0 0%	0 0%
	Ω_2	0 0%	8 0%	0 0%	0 0%
	Ω_3	0 0%	0 0%	8 0%	0 0%
	Ω_4	0 0%	0 0%	0 0%	8 0%
		Ω_1	Ω_2	Ω_3	Ω_4
		Classement réel			

4.3.4.2. Diagnostic avec une connaissance limitée sur les modes de fonctionnement

Les modes de fonctionnement qui vont servir à valider la deuxième procédure de diagnostic sont présentés dans le tableau 4.8. Les défauts ont été créés au rotor et aux roulements de la machine asynchrone et ceci pour un fonctionnement à plusieurs niveaux de charge.

Tableau 4. 8. Modes de fonctionnement étudiés

Mode de fonctionnement	Types de défauts	Niveau de charge		Alimentation
		(Pour une puissance nominale : $P_u = 5,5$ kW)		
Machine saine	Aucun défaut	25%, 50%, 75%, 100%		onduleur
3 barres cassées au rotor	Défaut rotor (barres cassées)	25%, 50%, 75%, 100%		onduleur
Usure progressive (vieillessement naturel)	Défaut roulement	25%, 50%, 75%, 100%		onduleur

Les 3 barres cassées ont été créées artificiellement en perçant 3 barres de la cage d'écureuil, comme indiqué dans la figure 4.24 (b) [Ondel(II)_06], [Casimir(III)_03]. L'usure obtenue sur les roulements est une usure naturelle d'utilisation (roulements fournis par la société SECCO) (voir figure 4.24 (a)) [Ondel(II)_06].

L'objectif est de vérifier les performances de nos méthodes de classification sur la base d'un ensemble d'apprentissage incomplet. À cette fin, nous supposons avoir une base d'apprentissage limitée pour tous les modes de fonctionnement et cela pour un niveau de charge de 25 et 50%, comme indiqué dans le tableau 4.9.



(a) vieillissement des roulements



(b) 3 barres cassées

Figure 4. 24. Défauts générés au niveau du rotor et des roulements de la machine asynchrone à cage d'écureuil

Tableau 4. 9. Composition de l'ensemble d'apprentissage

Mode de fonctionnement	Classe	Niveau de charge	Nombre d'observations
Machine saine	Ω_1	25 %	inconnu
		50 %	inconnu
		75 %	8
		100%	8
Default rotor (3 barres cassées)	Ω_2	25 %	inconnu
		50 %	inconnu
		75 %	8
		100%	8
Défaut roulement (vieillessement naturel)	Ω_3	25 %	inconnu
		50 %	inconnu
		75 %	8
		100%	8

L'ensemble d'apprentissage est constitué de trois modes de fonctionnement représentés par 3 classes. Chaque classe est composée de 16 observations, 8 observations à 75% et 8 observations à 100%. Les classes sont donc composées de 48 observations. Ces classes vont nous servir pour la suite afin de déterminer l'espace de représentation.

Afin de représenter un mode de fonctionnement par une seule classe regroupant tous les niveaux de charge, nous allons effectuer une normalisation des signaux de mesure. [Ondel(II)_06] a montré qu'il était possible de représenter un mode de fonctionnement à plusieurs niveaux de charges par une seule classe et ceci en effectuant une normalisation des signaux de courant i_a, i_b, i_c et de tension V_a, V_b, V_c . Ces derniers sont normalisés à partir de la valeur de l'amplitude fondamentale de chaque signal. C'est nouveaux signaux, notés i_{an}, i_{bn}, i_{cn} et V_{an}, V_{bn}, V_{cn} , vont permettre d'extraire le nouveau vecteur initial de paramètres porteur des signatures des défauts. La normalisation des courants et des tensions est formulée comme suit:

- Courant

$$i_{a_n} = \frac{i_a}{i_{a_1}} \quad ; \quad i_{b_n} = \frac{i_b}{i_{b_1}} \quad ; \quad i_{c_n} = \frac{i_c}{i_{c_1}} \tag{4.27}$$

- Tension

$$V_{a_n} = \frac{V_a}{V_{a_1}} \quad ; \quad V_{b_n} = \frac{V_b}{V_{b_1}} \quad ; \quad V_{c_n} = \frac{V_c}{V_{c_1}} \tag{4.28}$$

où l'indice 1 représente la valeur de 1^{er} harmonique.

4.3.4.2.1. Détermination de l'espace de représentation (sélection des q -paramètres)

Comme expliqué précédemment, le vecteur initial de paramètres regroupe l'ensemble des q' -paramètres calculés à partir du tableau 4.4. La première étape consiste à sélectionner un ensemble de q'' -paramètres en éliminant les paramètres qui minimisent le critère J. L'étape suivante consistera à sélectionner les sous-ensembles à q -paramètres qui maximisent le critère J. À l'aide de l'ensemble d'apprentissage, et en partant de l'ensemble initial à q' -paramètres, le critère donné par le formule (4.24) est calculé pour chaque paramètre (voir figure 4.25).

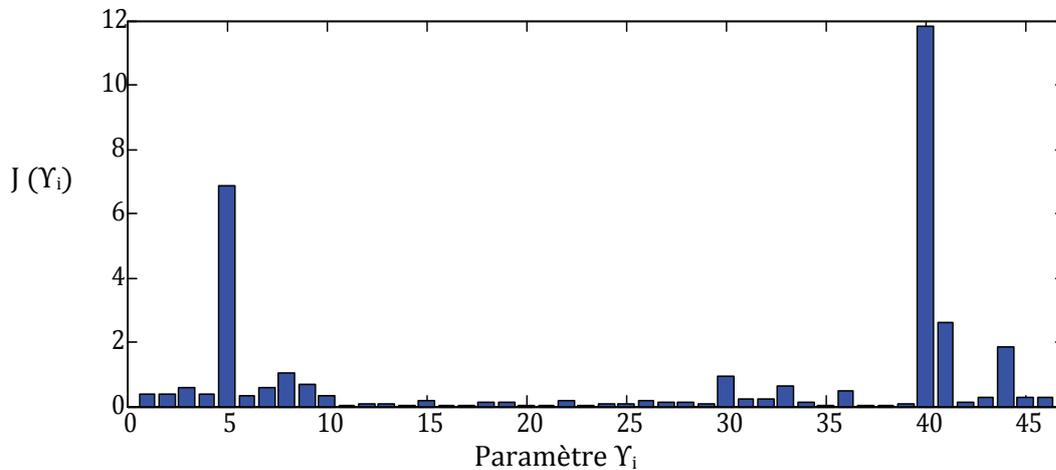


Figure 4. 25. Variation du critère J en fonction de chaque paramètre

Le critère J indique d'après la figure 4.25 que 15 paramètres, pris séparément, ont un pouvoir discriminant élevé. Ce résultat montre que la dimension (q) du vecteur de paramètres peut aller de 1 à 15. En fonction de ces 15 paramètres, l'algorithme OSBS nous donnera une sélection de q -paramètres qui maximisent le critère J .

La figure 4.26 illustre la réalisation de la sélection des sous-ensembles de q -paramètres les plus pertinents à partir de l'ensemble des q -paramètres. Le meilleur vecteur de paramètres, parmi les 15 solutions possibles, est celui qui donne les meilleures performances au sens du nombre d'observations mal classées de l'ensemble d'apprentissage.

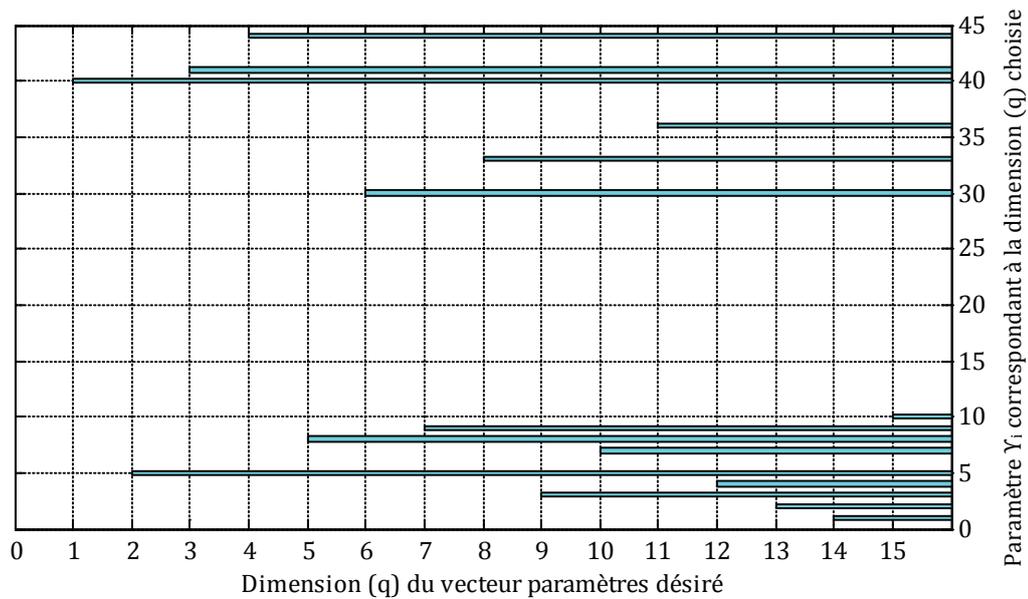


Figure 4.26. Sélection des ensembles de q -paramètres par la méthode OSBS

La figure 4.27 indique le nombre d'observations mal classées dans toutes les configurations du vecteur de paramètres possibles. Cette analyse a été réalisée à l'aide de l'algorithme FA.

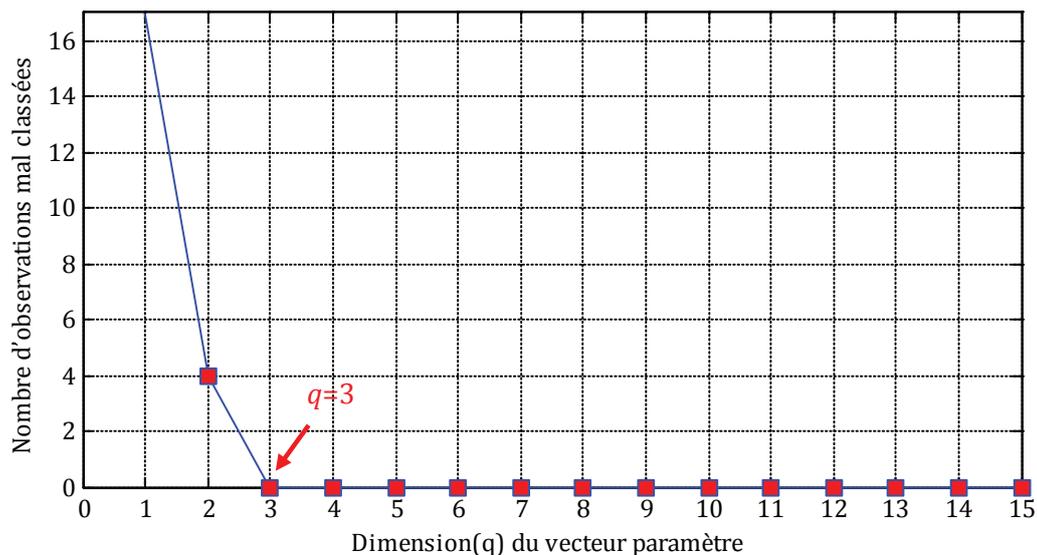


Figure 4.27. Choix de la dimension (q) du vecteur de paramètres

Nous pouvons remarquer que la procédure de sélection est optimale lorsque $q=3$ (avec 0% d'erreur de classification), le sous-ensemble de paramètres constituant le vecteur de paramètre est composé des paramètres Y_5 , Y_{40} et Y_{41} , c'est à dire :

- L'amplitude autour de la fréquence $f_s (f_s(1-2g))$;
- La déformation δ_α de la caractéristique $I_\alpha(I_\beta)$ dans l'axe α ;
- La déformation δ_β de la caractéristique $I_\beta(I_\alpha)$ dans l'axe β .

Les indications fournies dans la section 4.3.3.2 permettent de confirmer l'efficacité des deux paramètres δ_α et δ_β pour la détection des ruptures de barres. Il en est de même pour l'amplitude de l'harmonique de rang 5 (voir figures 4.8 et 4.9). En revanche, les paramètres indiquant les défauts de roulements (la puissance moyenne active $Y_1 (m_p)$ et la puissance réactive $Y_2 (m_q)$), même s'ils sont sélectionnés parmi les 15 paramètres les plus discriminants de l'ensemble q' -paramètres, ne permettent pas de distinguer les trois modes de fonctionnement.

Les trois paramètres, $[Y_5, Y_{40}, Y_{41}]$, suffisent à obtenir une très bonne séparation des classes comme le montre la figure 4.28.

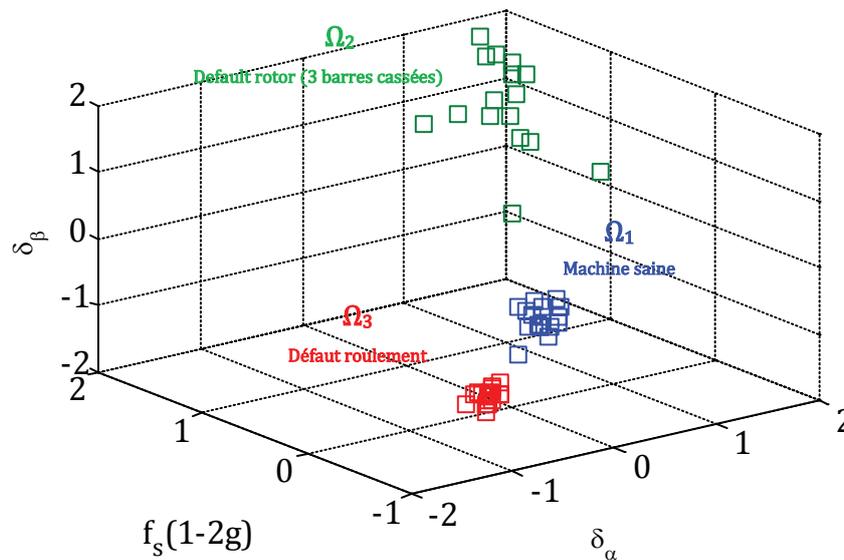


Figure 4. 28. Position des classes de l'ensemble d'apprentissage dans le sous espace de dimension 3. Les paramètres sont sélectionnés avec l'algorithme OSBS

4.3.4.2.2. Validation des méthodes de classification

L'ensemble « test » utilisé pour nos méthodes de classification est donné dans le tableau 4.10.

Tableau 4. 10. Composition de l'ensemble test

Mode de fonctionnement	Classe	Niveau de charge	Nombre d'observations
Machine saine	Ω_1	25 %	8
		50 %	8
		75 %	8
		100%	8
Defaut rotor (3 barres cassées)	Ω_2	25 %	8
		50 %	8
		75 %	8
		100%	8
Défaut roulement (vieillessement naturel)	Ω_3	25 %	8
		50 %	8
		75 %	8
		100%	8

Comme nous pouvons le voir, l'ensemble test est composé de 8 observations par niveau de charge et par mode de fonctionnement. L'ensemble test est donc constitué de 96 observations. Il s'agit maintenant de valider nos méthodes de classification mises en place.

4.3.4.2.2.1. Algorithme fournis artificielles

- Etape 1

La figure 4.29 montre l'évolution du critère J en fonction des différentes valeurs de seuil de similarité (S_{sim}). Nous pouvons distinguer à partir de cette figure trois zones qui indiquent l'évolution du critère J. Dans la première zone, S_{sim} décroît de 226 à 37. Cette décroissance correspond à un regroupement d'observations qui varie entre 35 classes pour $J = 226$ et 19 classes pour $J = 37$. La variance intra-classe dans ce cas est voisine de 0 ce qui donne une valeur élevée pour J. En revanche, plus le nombre de classes diminue plus J diminue car la variance intra-classe augmente. Ceci nous place dans la deuxième zone avec $J \in [37-9,7]$ et la troisième zone avec $J \in [9,7-2,7]$. Un seuil de similarité fixé à 0,1 risque d'être pénalisant en temps de calcul surtout pour déterminer les paramètres (e et f) de l'algorithme (voir étape 2). Un seuil de similarité fixé à 0,4 ne nous permet pas d'optimiser la classification car le nombre de classes ne change pas au-delà de cette valeur. Le seuil de similarité sera donc fixé à 0,2.

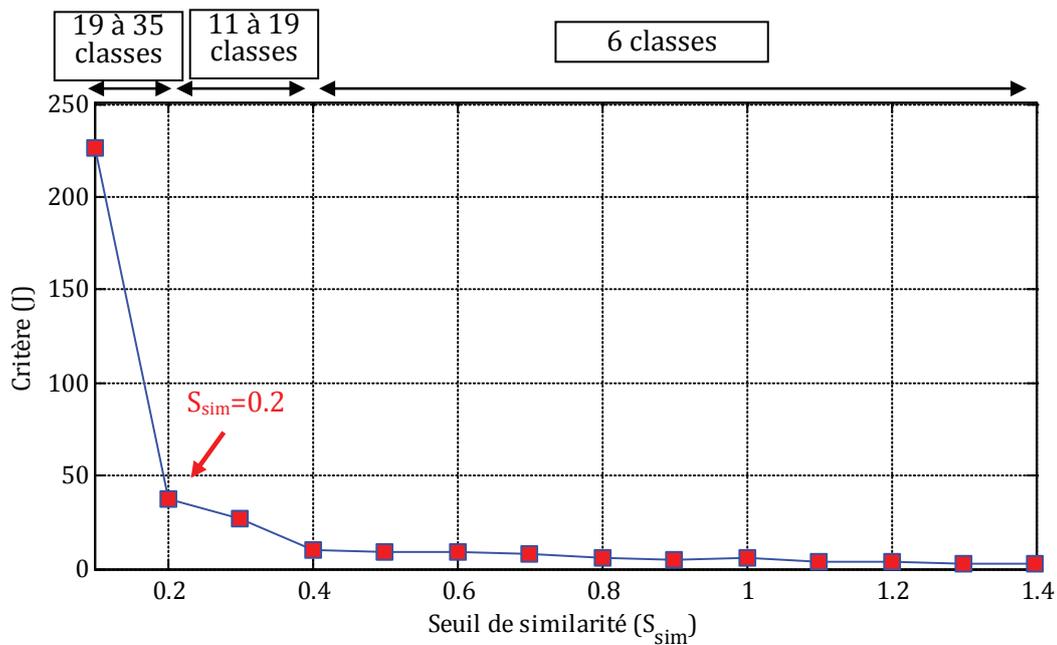


Figure 4. 29. Evolution du critère J en fonction du seuil de similarité

Un seuil de similarité fixé à 0,2 permet à la première étape de l'algorithme FA d'obtenir un premier ensemble de m -classes, avec $m= 19$.

- Etape 2

La deuxième étape permet de rassembler les classes précédentes en un nombre de classes plus restreint. Le paramètre ϵ est fixé à 0,01 afin de limiter le poids de chaque classe. Le paramètre ξ est arbitrairement fixé à 1. Les valeurs de e et f sont initialement choisies dans l'intervalle $[-1,-1]$. Le critère J le plus élevé définit les meilleurs paramètres e et f , tel que : pour J avoisinant les 99%, $e = -0,4$ et $f = 0,4$ (voir figure 4.30).

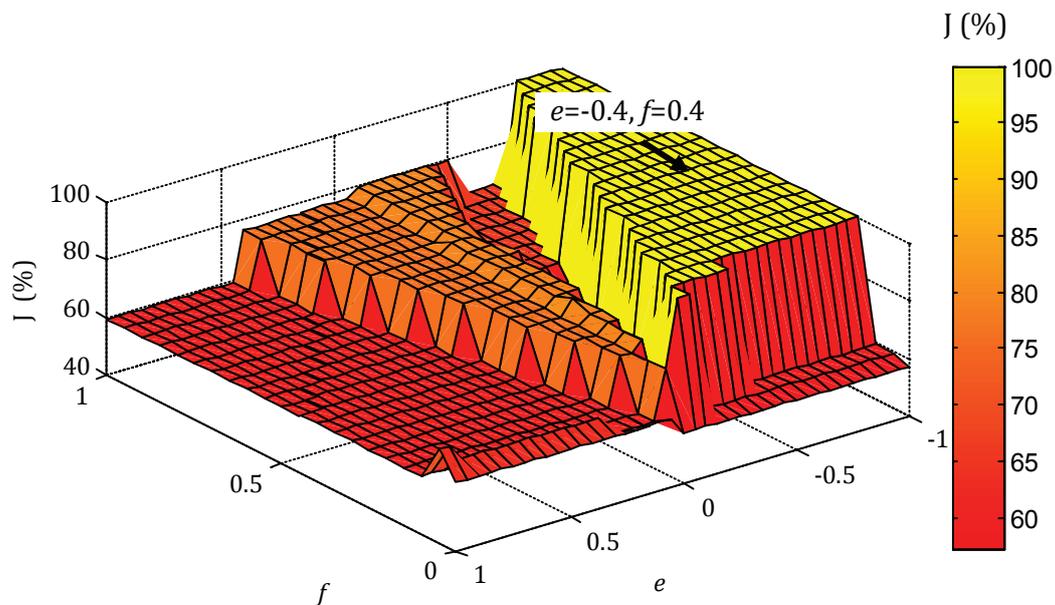


Figure 4. 30. Détermination des paramètres e et f à partir du critère J

- Etape 3

La détermination des paramètres e et f permet à la deuxième étape de l'algorithme FA d'obtenir un deuxième ensemble de m' -classes, avec $m'=2$. Ceci est dû au fait qu'il y a un fort chevauchement entre les classes comme le montre la figure 4.31. Une analyse en composantes principales est effectuée afin de changer l'axe d'orientation des paramètres et permettre ainsi une meilleure visualisation et discrimination des classes. La figure 4.32 montre la position des M -classes de l'ensemble test dans le sous-espace de dimension 3 après l'ACP.

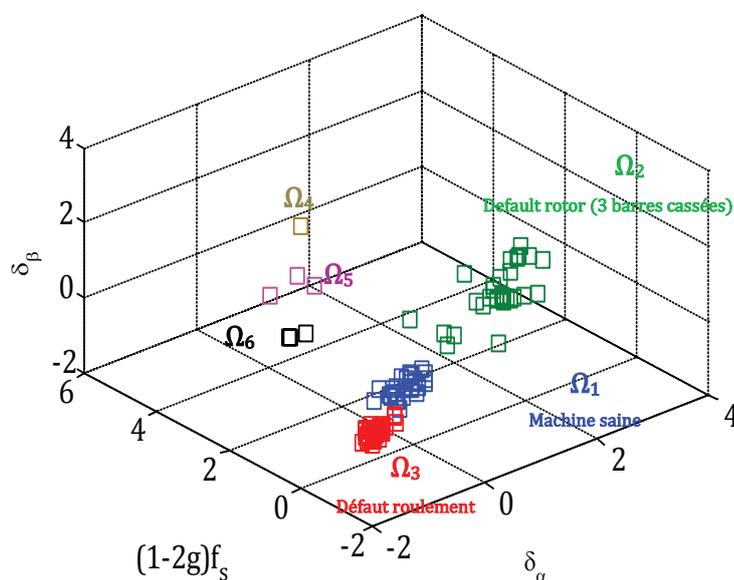


Figure 4. 31. Position des classes obtenues par l'algorithme FA dans le sous - espace de dimension 3 avant l'ACP

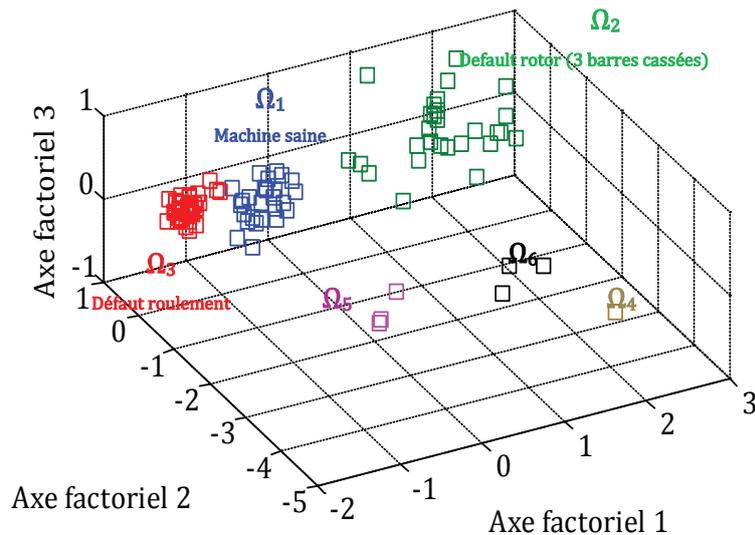


Figure 4. 32. Position des classes obtenues par l'algorithme FA dans le sous - espace de dimension 3 après l'ACP

L'utilisation de l'algorithme K-means indique que la séparation de ces classes induit une augmentation de 90 % du critère J. On obtient finalement un ensemble de six classes ($\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5, \Omega_6$). L'utilisation de l'ensemble d'apprentissage permet l'identification des classes ($\Omega_1, \Omega_2, \Omega_3$) et la détection de trois nouvelles classes notées ($\Omega_4, \Omega_5, \Omega_6$). Les observations qui constituent ces classes sont à ce stade considérées comme rejetées. Une analyse sera effectuée par la suite pour confirmer l'apparition d'un nouveau mode de fonctionnement.

- Analyse des nouvelles observations

Les observations regroupées dans les nouvelles classes peuvent représenter des nouveaux modes inconnus de fonctionnement. Pour déterminer si de nouvelles classes sont apparues, la formule (2.41) est utilisée. Cette formule indique que le critère Cq' pour les classes ($\Omega_4, \Omega_5, \Omega_6$) n'est pas du même ordre de grandeur que les classes déjà identifiées (voir figure 4.33). Ceci permet de rejeter ces observations.

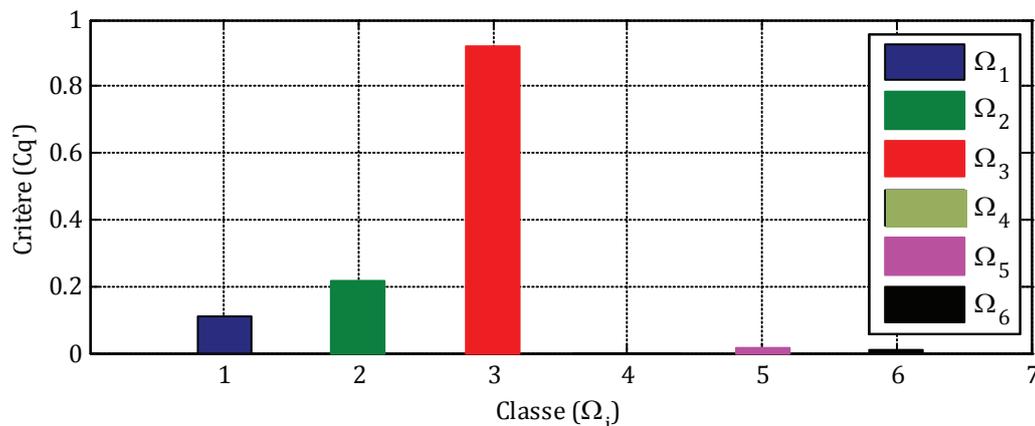


Figure 4. 33. Détection de l'apparition d'une nouvelle classe

La répartition des observations sur les classes obtenues par l'algorithme FA est donnée comme suit, avec un $E_c=3,1\%$ et un taux de rejet de $7,2\%$.

	Ω_1	Ω_2	Ω_3	
Classement obtenu	Ω_1	26 0%	0 0%	0 0%
	Ω_2	0 0%	28 0%	0 0%
	Ω_3	3 3,1%	0 0%	32 0%
	Rejet	3 3,1%	4 4,16%	0 0%
	Ω_1	Ω_2	Ω_3	
	Classement réel			

4.3.4.2.2. Classification par les modèles de Markov cachés

Pour la classification par modèles de Markov cachés, nous devons déterminer 03 MMCs correspondant aux 03 modes de fonctionnement connus : machine saine (S_1), défaut rotor avec 3 barres cassées (S_2) et défaut de roulements (S_3). À partir de chaque classe, nous extrayons une séquence d'observations de longueur ($T=3$) (Chapitre 2, section 2.3.2.3.2). La longueur T correspond à un tirage successif de $T+1$ observations de chacune des classes.

Les séquences d'observations utilisées pour estimer les paramètres des MMCs peuvent être formulées par : $X^k = \{x_0^k, x_1^k, x_2^k, \dots, x_T^k\}$, avec $1 \leq k \leq 3$ et $T = 3$. Les paramètres initiaux de chaque MMC (matrice de transition $A^{(k)}$ et le vecteur d'état initial $\pi^{(k)}$), à partir des quels nous commencerons l'apprentissage, sont définis comme suit :

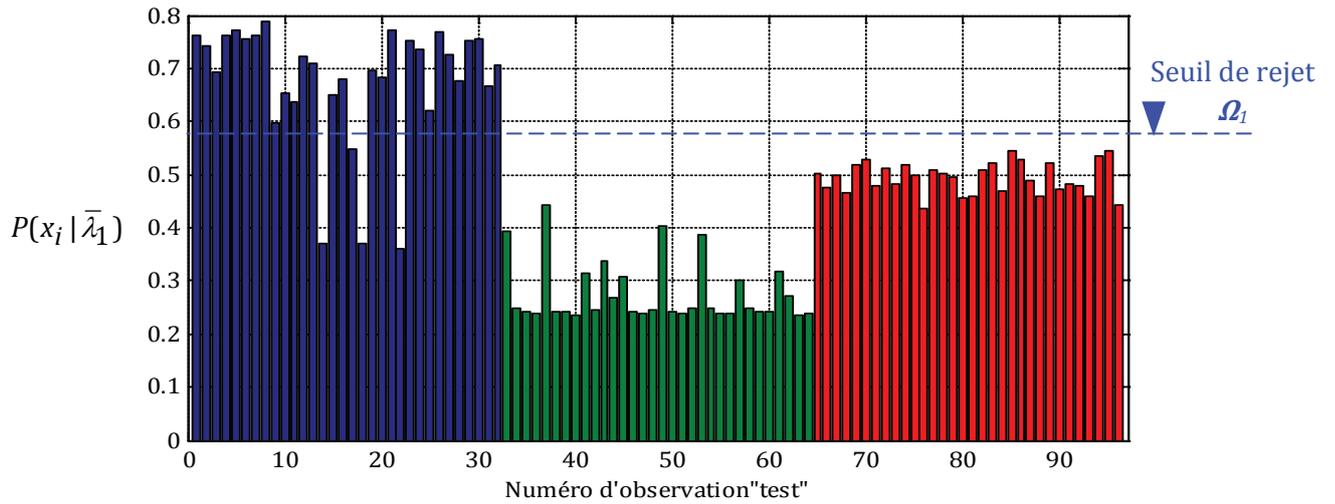
$$A^{(1,2,3)} = \begin{matrix} & \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 \end{matrix} \\ \begin{matrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{matrix} & \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix} \end{matrix} \quad \pi^{(1,2,3)} = \begin{matrix} & \Omega_1 & \Omega_2 & \Omega_3 \\ (1/3 & 1/3 & 1/3) \end{matrix} \quad (4.29)$$

Notons que les probabilités de transition a_{31}, a_{21} sont fixées à zéros. Dans notre cas, la classe Ω_1 correspond à une machine saine et Ω_2, Ω_3 correspondent à une machine défectueuse avec des défauts de barres et de roulements. Dans ce cas, il est normal de fixer à zéros a_{31}, a_{21} car nous ne prenons pas en compte le fait qu'un système défectueux puisse revenir à un fonctionnement normal.

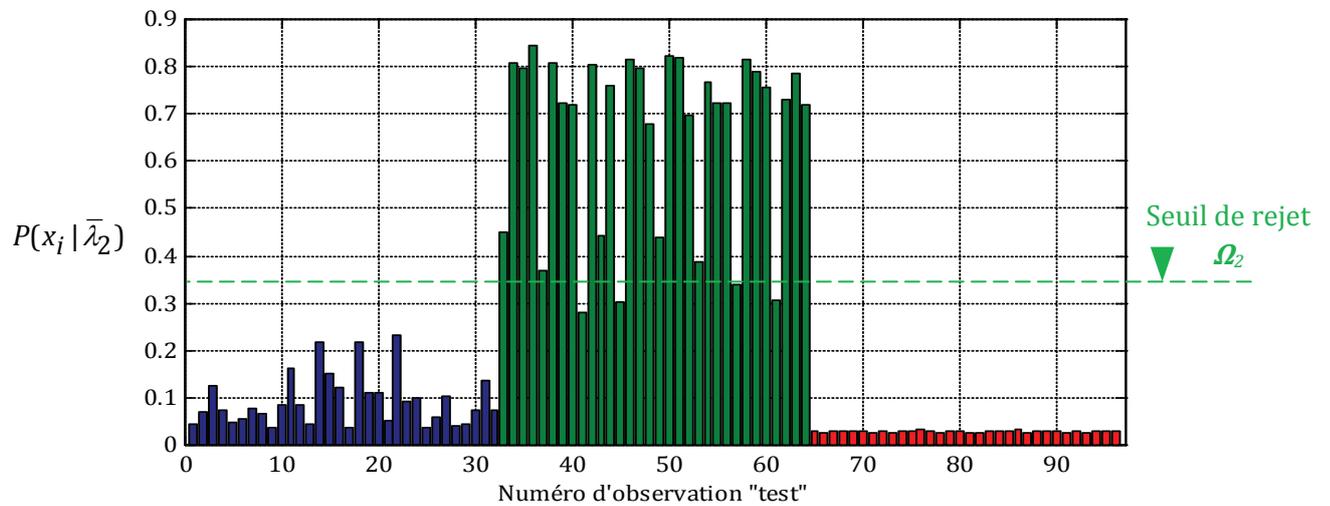
Les paramètres des 03 MMCs, après re-estimation, sont donnés comme suit :

MMC de la machine saine	MMC du défaut rotor (3 barres cassées)	MMC du défaut de roulements
$\bar{A}^1 = \begin{bmatrix} 0,9646 & 0 & 0,0354 \\ 0 & 0,0006 & 0,9994 \\ 0 & 0,0006 & 0,9994 \end{bmatrix}$	$\bar{A}^2 = \begin{bmatrix} 0,0256 & 0,9668 & 0,0076 \\ 0 & 0,9907 & 0,0093 \\ 0 & 0,9911 & 0,0089 \end{bmatrix}$	$\bar{A}^3 = \begin{bmatrix} 0,1792 & 0,0002 & 0,8206 \\ 0 & 0,0002 & 0,9998 \\ 0 & 0,0002 & 0,9998 \end{bmatrix}$
$\bar{\pi}^1 = [0,9978 \quad 0 \quad 0,0022]$	$\bar{\pi}^2 = [0,1216 \quad 0,8513 \quad 0,0271]$	$\bar{\pi}^3 = [0,1848 \quad 0,0001 \quad 0,8151]$

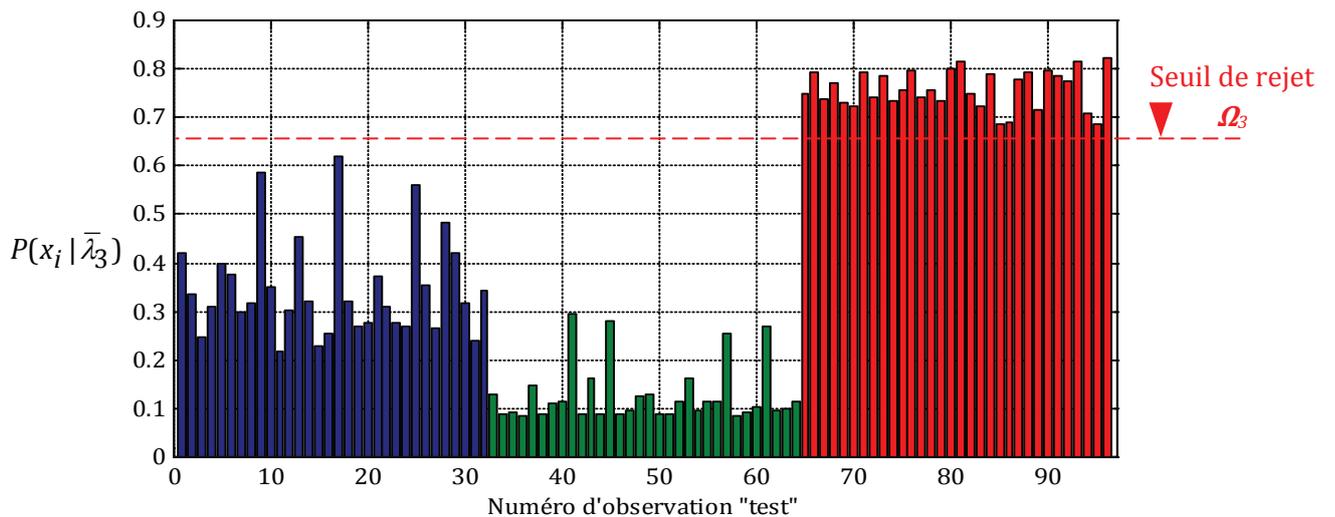
Les matrices de transition qui ont été identifiées pour chaque mode de fonctionnement indiquent que la probabilité $P(q_t = S_k | q_{t-1} = S_k, \bar{\lambda}_k) \approx 1$ pour $1 \leq k \leq 3$. Si on se réfère à la formule (2.53) qui a été définie dans le chapitre 2, cela indique une bonne séparation entre les classes ($\Omega_1, \Omega_2, \Omega_3$) (voir figure 4.28). Les figures 4.34 (a, b, c) montrent les probabilités générées par l'affectation de l'ensemble d'observations « test ».



(a) Probabilités générées par l'affectation des observations test à la classe Ω_1 étant donné $\bar{\lambda}_1$



(b) Probabilités générées par l'affectation des observations test à la classe Ω_2 étant donné $\bar{\lambda}_2$



(c) Probabilités générées par l'affectation des observations test à la classe Ω_3 étant donné $\bar{\lambda}_3$

Figure 4. 34. Classification de l'ensemble des observations test par modèles de Markov cachés
 Les probabilités générées par l'affectation des 96 observations test à la classe Ω_1 varient entre 0,23 et 0,78 avec un pourcentage de 29,1% d'observations test affectées avec une probabilité

supérieure à 0,58. Un seuil de similarité égal à 0,57 est appliqué sur les observations test afin de réduire les erreurs de classification. Ce seuil a pour but de classer que les observations les plus similaires à la classe Ω_1 ce qui permet au final de classer avec certitude 28/96 observations test à la classe Ω_1 .

Les probabilités générées par l'affectation des 96 observations test à la classe Ω_2 varient entre 0,0257 et 0,845 avec un pourcentage de 29,1% d'observations test affectées avec une probabilité supérieure à 0,38. Là encore, un seuil de similarité égal à 0,37 est appliqué sur les observations test dans le but de classer que les observations les plus similaires à la classe Ω_2 . Cela permet au final de classer avec certitude 28/96 observations test à la classe Ω_2 .

Les probabilités générées par l'affectation des 96 observations test à la classe Ω_3 varient entre 0,0829 et 0,824 avec un pourcentage de 33,3% d'observations test affectées avec une probabilité supérieure à 0,67. Un seuil de similarité égal à 0,66 est fixé pour la classe Ω_3 afin de réduire les erreurs de classification. Ce seuil permet au final de classer avec certitude 32/96 observations test à la classe Ω_3 .

La répartition des observations sur les classes obtenues par les MMCs est donnée comme suit, avec un $E_c=1,04\%$ et un taux de rejet de 8,32%.

Classement obtenu	Ω_1	28 0%	1 1,04%	0 0%
	Ω_2	0 0%	27 0%	0 0%
	Ω_3	0 0%	0 0%	32 0%
	Rejet	4 4,16%	4 4,16%	0 0%
		Ω_1	Ω_2	Ω_3
Classement réel				

4 observations test numérotées (14, 17, 18, 22) appartenant à la classe Ω_1 ainsi que 4 autres observations test numérotées (41, 45, 57, 61) appartenant à la classe Ω_2 sont rejetées en raison de leurs faibles probabilités d'appartenance aux classes ($\Omega_1, \Omega_2, \Omega_3$). Une observation numérotée (37) appartenant à la classe Ω_2 a été mal classée.

Il est possible d'améliorer la classification en se servant des classes obtenues comme base d'apprentissage pour les MMCs. L'apprentissage des MMCs, dans ce cas, prend en compte, en plus de l'ancienne base d'apprentissage, les observations test qui ont été classées. Cette stratégie permet d'obtenir une erreur de classification de 0 %. La répartition des observations, dans ce cas, est donnée comme suit, avec $E_c=0\%$ et un taux de rejet de 8,32%.

Classement obtenu	Ω_1	26 0%	0 0%	0 0%
	Ω_2	0 0%	30 0%	0 0%
	Ω_3	0 0%	0 0%	32 0%
	Rejet	6 6,25%	2 2,08%	0 0%
		Ω_1	Ω_2	Ω_3
Classement réel				

Conclusion

Au final, les deux méthodes de classification utilisées donnent quasiment les mêmes résultats dans le premier cas de figure (apparition d'un nouveau mode de fonctionnement). En ce qui concerne le deuxième cas de figure (connaissance limitée sur chacun des modes de fonctionnement), les modèles de Markov cachés se sont montrés plus efficaces que l'algorithme FA pour la classification des nouvelles observations avec 0% d'erreur de classification et 8,32% de rejet. Ces résultats sont d'autant plus satisfaisants que les modes de fonctionnement utilisés avaient une base de données limitée. Cette partie a montré l'efficacité de l'algorithme FA et des modèles de Markov cachés pour la classification des nouvelles observations et pour la détection des nouveaux modes de fonctionnement.

4.4. Pronostic des roulements

4.4.1. Présentation du banc de test pronostic

Les modes de fonctionnement qui vont servir à valider la procédure de pronostic représentent les différents stades de dégradation des roulements à rouleaux. Le but du pronostic dans ce cas-là est :

- De détecter l'imminence de chaque stade de dégradation ;
- D'estimer le temps restant avant l'apparition du prochain stade de dégradation en prédisant, à court et à long termes, l'évolution des paramètres de surveillance.

La dégradation des roulements a été réalisée sur un banc d'essai conçu par la société Rexnord Corp. Le banc d'essai est composé de quatre roulements à rouleaux de type ZA-2115 installés sur un arbre, comme indiqué dans la figure 4.35. Les caractéristiques de ces roulements sont détaillées dans le tableau 4.11.

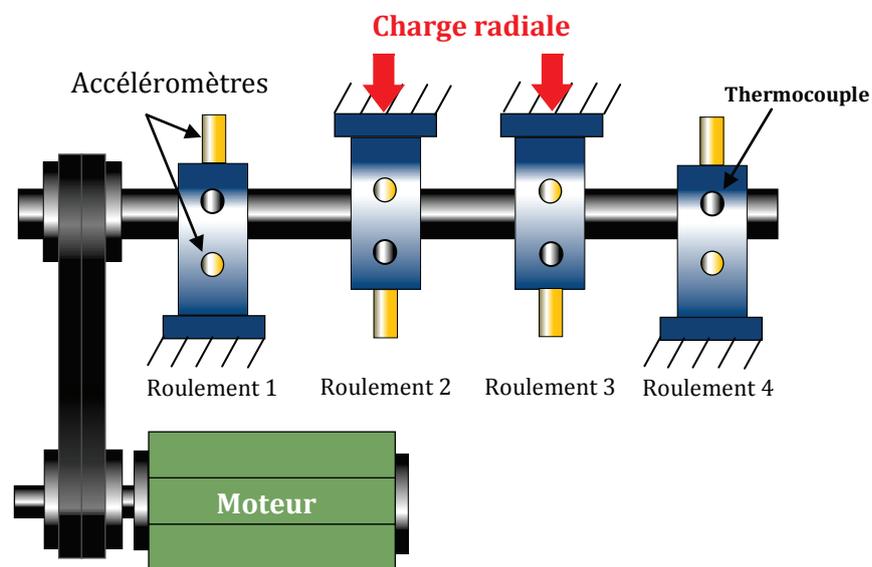


Figure 4. 35. Banc de vieillissement des roulements à rouleaux [Qiu_06]

Tableau 4. 11. Caractéristiques des roulements

Diamètre du rouleau (mm)	8,4
Nombre de rouleaux	16
Diamètre primitif du rouleau (mm)	71,62
Angle de contact (°)	15,17

La vitesse de rotation de l'arbre est de 2000 tr/min avec une charge radiale de 26689 Newton placée sur l'arbre et les roulements par un mécanisme à ressort. Un bouchon magnétique installé dans le tuyau de retour d'huile recueille les débris de l'huile comme preuve de la dégradation des roulements. Un accéléromètre du type PCB 353B33 a été installé sur l'axe horizontal (X) et l'axe vertical (Y) de chaque roulement. Les signaux vibratoires ont été enregistrés toutes les 10 minutes avec une fréquence d'échantillonnage fixée à 20 kHz. Ces signaux sont disponibles sur le site du centre d'excellence de pronostic [<http://ti.arc.nasa.gov/dash/pcoe/pronostic-data-repository/>].

La figure 4.36 montre le schéma de principe de l'approche utilisée pour la maintenance préventive conditionnelle (Chapitre 1, figure 1.3). Dans un premier temps, le procédé de pronostic commence par identifier les différents stades de dégradations liés aux contraintes imposées. Différents types de capteurs peuvent être utilisés afin de mesurer les signaux porteurs de ces dégradations. Nous utiliserons dans notre cas des capteurs de vibrations. Diverses techniques de traitement des signaux peuvent être appliquées à ces signaux pour extraire des paramètres sensibles à la dégradation. Nous utiliserons des paramètres extraits du domaine temporel afin de suivre la dégradation et des paramètres extraits du domaine fréquentiel afin d'identifier les défauts responsables de la dégradation. Une sélection de paramètres sera ensuite utilisée afin de ne garder que les paramètres permettant de mieux distinguer les modes de fonctionnement étudiés.

Dans la phase diagnostic, la classification d'une observation dans l'un des modes de fonctionnement permet d'identifier le stade de dégradation du roulement testé. Enfin, dans la phase pronostic, les informations fournies par le diagnostic vont servir à détecter l'imminence des stades de dégradation et à estimer le temps restant avant leur apparition. C'est là qu'intervient nos méthodes de pronostic.

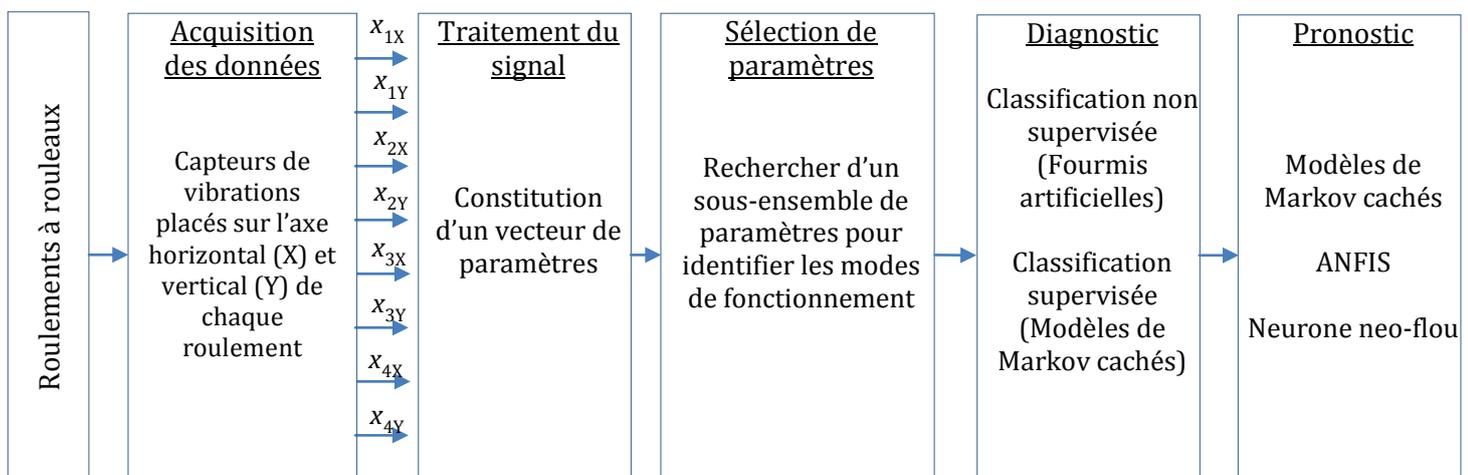


Figure 4. 36. Schéma de principe du processus de diagnostic et de pronostic

4.4.2. Validation des méthodes de pronostic

Un ensemble de données expérimentales, extrait des signaux vibratoires du roulement n°4, est utilisé pour évaluer la performance de nos méthodes de pronostic. Pour ce faire, nous supposons que les données historiques de dégradation des roulements (1, 2 et 3) sont disponibles. Cette base de données sera utilisée pour la suite afin d'identifier les M -inconnus stades de dégradation et identifier les paramètres des MMCs, du neurone neuro-flou et d'ANFIS.

4.4.2.1. Construction du vecteur de paramètres

Les techniques de traitement du signal, décrites dans le chapitre 1 (section 1.4), vont être appliquées afin d'extraire des indicateurs de dégradation à partir des signaux vibratoires des roulements (1, 2, 3 et 4). Nous allons dans un premier temps nous baser sur des paramètres statistiques extraits du domaine temporel. Ces paramètres vont nous servir afin d'identifier les différents stades de dégradation des roulements et servir comme paramètres de surveillance pour la procédure de pronostic. Cependant, ces paramètres-là ne permettent pas d'identifier le composant défectueux des roulements (bague intérieure, bague extérieure, rouleaux). Pour remédier à ce problème, nous compléterons la procédure de pronostic par des paramètres extraits du domaine fréquentiel en utilisant la formule (4.7) afin de prédire les défauts des roulements.

Les paramètres qui vont nous servir pour suivre la dégradation des roulements sont résumés dans le tableau 4.12.

Tableau 4. 12. Paramètres statistiques caractéristiques d'un signal vibratoire

<i>Énergie</i>		<i>Puissance</i>	
$E = \sum_{\ell=1}^N x_{\ell}^2$	γ_1	$P = \frac{1}{N} \sum_{\ell=1}^N x_{\ell}^2$	γ_5
<i>Amplitude crête</i>		<i>Valeur efficace</i>	
$S_{crête} = \sup_{1 \leq \ell \leq N} x_{\ell} $	γ_2	$S_{eff} (RMS) = \sqrt{\frac{1}{N} \sum_{\ell=1}^N (x_{\ell} - \bar{x})^2}$	γ_6
<i>Valeur moyenne</i>		<i>Kurtosis</i>	
$\bar{x} = \frac{1}{N} \sum_{\ell=1}^N x_{\ell}$	γ_3	$S_{kurt} = \frac{\frac{1}{N} \sum_{\ell=1}^N (x_{\ell} - \bar{x})^4}{S_{eff}^4}$	γ_7
<i>Facteur de crête</i>		<i>Facteur K</i>	
$F_c = \frac{S_{crête}}{S_{eff}}$	γ_4	$F_K = S_{crête} \cdot S_{eff} = F_c \cdot S_{eff}^2$	γ_8

avec :

- $x(t)$: signal analogique conditionné à analyser ;
- N : nombre de points de l'échantillon choisi pour représenter le signal ;
- f_{ech} : fréquence d'échantillonnage ;
- $x_{\ell} = x(t=\ell/f_{ech})$, avec ℓ nombre entier positif non nul inférieur ou égal à N .

Les signaux vibratoires sont extraits à partir d'accéléromètres installés sur l'axe horizontal (X) et l'axe vertical (Y) de chaque roulement. Afin d'évaluer l'importance de l'emplacement des accéléromètres, nous avons effectué une comparaison entre les signaux vibratoires extraits de

l'axe (X) et (Y). La figure 4.37 montre l'évolution de la dégradation des roulements à partir des paramètres de surveillance extraits des signaux vibratoires des accéléromètres placés sur l'axe (X) et (Y). Notons que les roulements (1, 2, 3 et 4) sont testés à 2000 tr/min avec une charge radiale de 26689 Newton pendant 35 jours de test. Le test s'arrête lorsque les débris accumulés qui adhèrent au bouchon magnétique dépassent un certain niveau et provoquent la fermeture d'un interrupteur électrique. La base de données dont nous disposons sur le vieillissement des roulements est constituée de 1944 observations, avec 486 observations par roulement et 14 observations par jour.

À partir de la figure 4.37, et quel que soit l'emplacement des accéléromètres, nous pouvons éliminer quelques paramètres non pertinents pour le suivi de la dégradation des roulements.

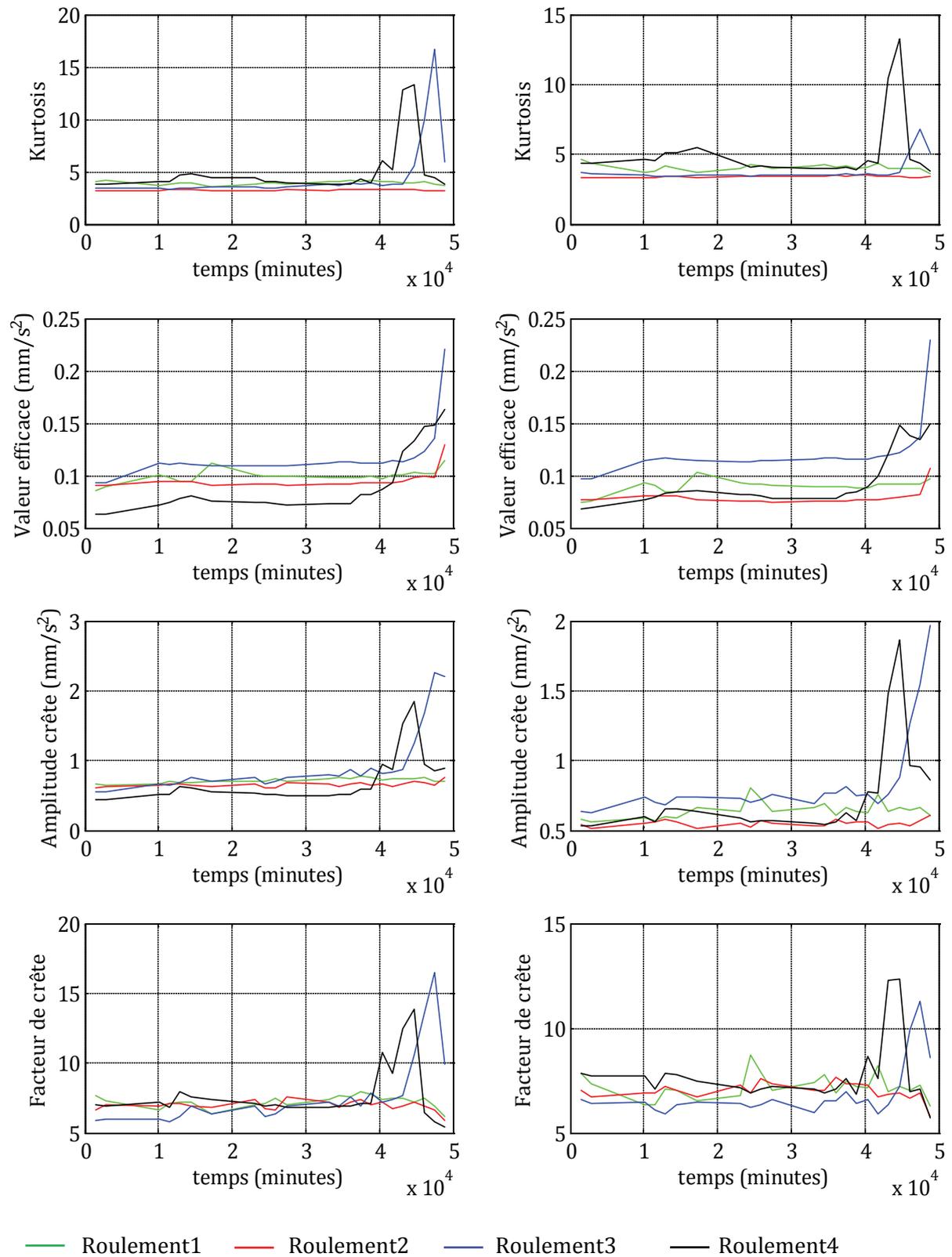
Les paramètres qui sont résumés dans le tableau 4.12 sont considérés comme étant de bons paramètres pour le suivi de la dégradation des roulements. La figure 4.37 montre l'évolution la valeur efficace, la puissance, l'énergie et le facteur K des signaux vibratoires recueillis durant la période de test. Après deux jours de test (2880 minutes), les premiers signes de vieillissement ont été observés dans les quatre roulements. Ces signes sont dus à un phénomène appelé «rodage» du roulement à rouleaux : le lubrifiant présent dans le roulement est progressivement évacué et les imperfections du métal disparaissent petit à petit. Après 30 jours de test (43200 minutes), les premiers signes de dégradation ont été observés dans le roulement n°4. Après cela, les valeurs du RMS, l'énergie et la puissance continuent à augmenter à mesure que la dégradation augmente. Nous pouvons observer que le facteur K n'est pas un bon indicateur pour le suivi de la dégradation des roulements que ce soit pour l'axe (X) ou pour l'axe (Y). Ceci est montré dans la figure 4.37, où le facteur K, pour le roulement n°4, commence à décroître à partir de 44640 minutes. On peut aussi donner la même remarque en ce qui concerne le Kurtosis, le facteur de crête et l'amplitude crête. On observe une décroissance de la valeur de ces paramètres pour les roulements 3 et 4 après 44640 et 47520 minutes de test.

La valeur de la moyenne est proche de zéro pendant les deux premiers jours. Ceci indique que les roulements (1, 2, 3 et 4) sont en bon état. Après deux jours de test (2880 minutes), les premiers signes de vieillissement ont été observés avec la décroissance de la valeur moyenne. Après cela, la valeur moyenne ne change pas de manière significative.

En ce qui concerne l'énergie du signal vibratoire, elle est considérée comme l'image de la puissance dans le temps. Donc, il nous suffit d'en choisir un des deux pour le suivi de la dégradation. On choisira la puissance pour la suite. Nous conservons finalement trois paramètres : la valeur moyenne, la puissance et le RMS. On peut aussi remarquer que ces trois paramètres donnent une meilleure représentation de la dégradation des roulements sur l'axe (X) que sur l'axe (Y). Le vecteur de paramètre qui servira pour construire l'espace de représentation sera donc composé de : $[\gamma_3, \gamma_5, \gamma_6]$.

Accéléromètre placé sur l'axe horizontal

Accéléromètre placé sur l'axe vertical



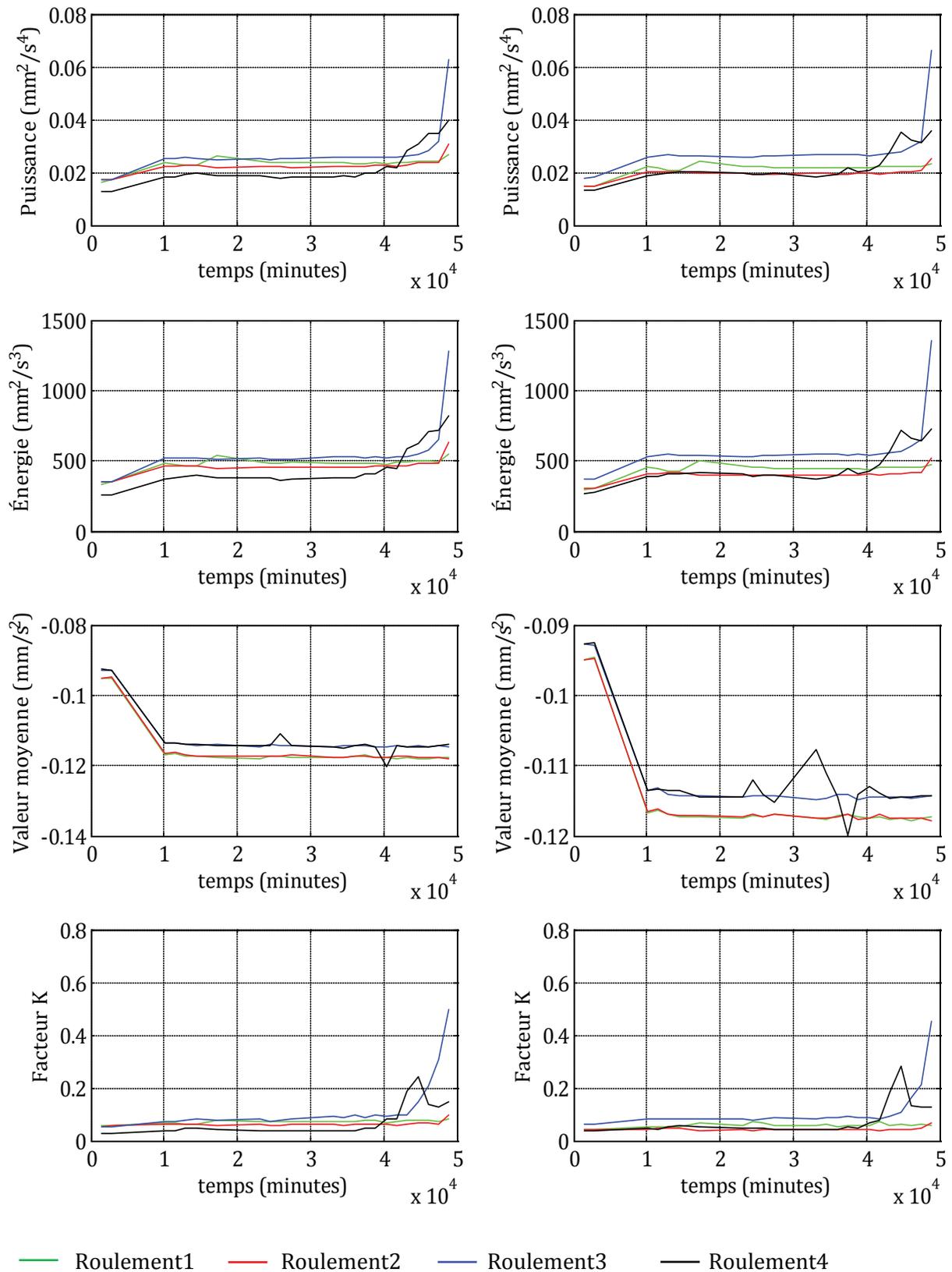


Figure 4. 37. Evolution de la dégradation des roulements (1, 2, 4 et 4) en fonction des paramètres statistiques

4.4.2.2. Identification des différents stades de dégradation

L'algorithme FA est utilisé afin d'identifier les M -inconnus stades de dégradation à partir de l'ensemble des données historiques des roulements (1, 2 et 3). Une analyse détaillée sur les différents paramètres de surveillance peut nous aider à choisir les plus pertinents. En se basant sur la figure 4.37, nous avons choisi le RMS, la valeur moyenne et la puissance du signal vibratoire. Le reste des paramètres ne sont pas sélectionnés car lorsque la dégradation augmente, la valeur de ces derniers décroît. Pour les paramètres sélectionnés ($\gamma_3, \gamma_5, \gamma_6$), plusieurs observations perturbées par le bruit peuvent apparaître. Ainsi, une classe qui regroupe les observations d'un même état de dégradation peut être représentée par une zone géométrique d'observations. Si les paramètres sont bien choisis, chaque stade de dégradation peut être représenté par une classe compacte regroupant les observations similaires dans l'espace de représentation comme illustré sur la figure 4.40.

- Etape 1

La figure 4.38 montre l'évolution du critère J en fonction des différentes valeurs de seuil de similarité (S_{sim}). Nous pouvons différencier à partir de cette figure deux zones bien distinctes qui indiquent l'évolution du critère J . Dans la première zone, S_{sim} décroît de 132 à 43. Cette décroissance correspond à un regroupement d'observations qui varie entre 24 classes pour $J=132$ et 9 classes pour $J=43$. La variance intra-classe dans ce cas est voisine de 0 ce qui donne une valeur élevée pour J . En revanche, plus le nombre de classes diminue, plus J diminue car la variance intra-classe augmente. Ceci nous place dans la deuxième zone avec un critère $J \in [5,8-43]$. Un seuil de similarité fixé à 0,4 génère un nombre de m -classes égale à 9, suffisant pour optimiser la classification.

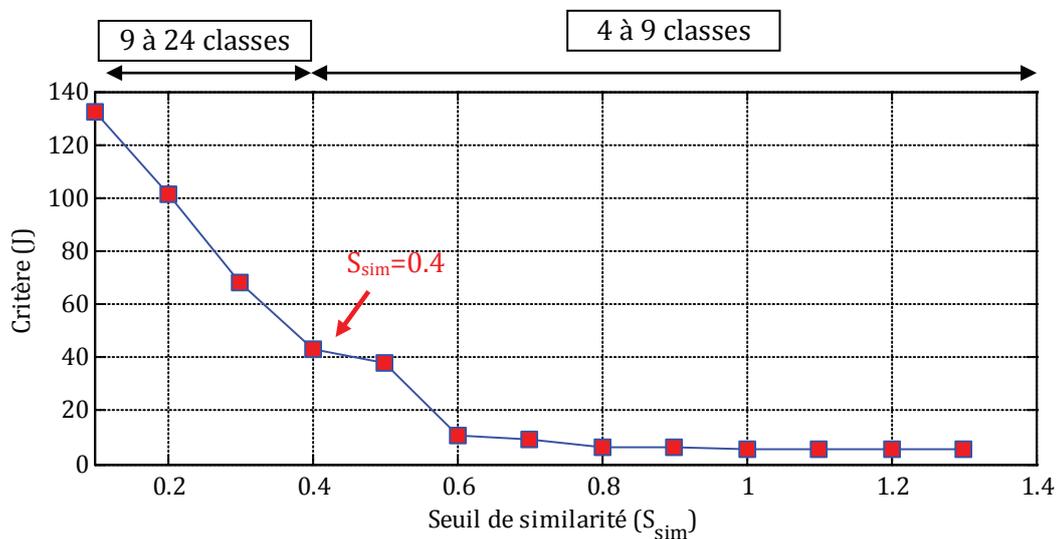


Figure 4. 38. Evolution du critère J en fonction du seuil de similarité

- Etape 2

La deuxième étape permet de rassembler les classes précédentes en un nombre de classes plus restreint. Le paramètre ε est fixé à 0,01 afin de limiter le poids de chaque classe. Le paramètre ξ est arbitrairement fixé à 1. Le critère J le plus élevé définit les meilleurs paramètres e et f , tel que : pour J avoisinant les 99 %, $e = -0,4$ et $f = 0,6$ (voir figure 4.39).

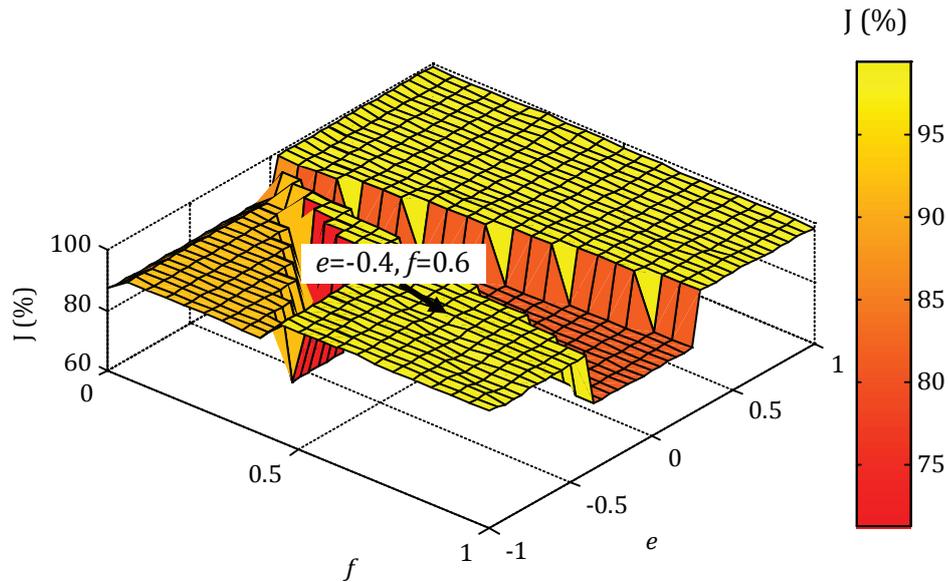


Figure 4.39. Détermination des paramètres e et f à partir du critère J

- Etape 3

La détermination des paramètres e et f permet à la deuxième étape de l'algorithme FA d'obtenir un deuxième ensemble de m' -classes, avec $m'=4$. Une analyse en composantes principales est effectuée afin de changer l'axe d'orientation des paramètres et permettre ainsi une meilleure visualisation des classes. La figure 4.40 montre la position des M -classes de l'ensemble test dans le sous - espace de dimension 3 après l'ACP.

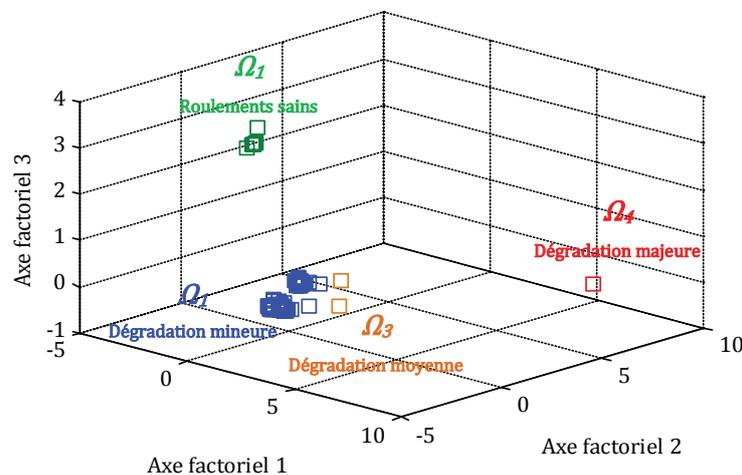


Figure 4.40. Position des classes obtenues par l'algorithme FA dans le sous - espace de dimension 3 après l'ACP

L'utilisation de l'algorithme K-means indique que la séparation de ces classes induit une diminution du critère J . On obtient au final un ensemble de 4 classes ($\Omega_1, \Omega_2, \Omega_3, \Omega_4$).

L'algorithme FA a détecté quatre classes ($M = 4$) ($\Omega_1, \Omega_2, \Omega_3, \Omega_4$) pour $e = -0,4$ et $f = 0,6$. La première classe (Ω_1) comprend des observations appartenant au roulement (1, 2 et 3) enregistrées au cours des deux premiers jours de test. Cette classe représente des roulements

en bon état de fonctionnement car les premiers signes de vieillissement ont été constatés qu'après deux jours de test. La dernière classe (Ω_4) comprend une observation du roulement 3 enregistrée le dernier jour de test. Cette classe correspond à un défaut sur la bague intérieure découvert dans le roulement 3 (voir figure 4.41). Cette classe représente des roulements avec une dégradation majeure. En plus de (Ω_1) et (Ω_4), l'algorithme FA a détecté deux autres classes, (Ω_2) et (Ω_3). La deuxième classe (Ω_2) comprend des observations appartenant aux roulements (1, 2 et 3) et enregistrées entre le 2^{ème} et le 33^{ème} jour de test. Les valeurs des paramètres de surveillance de cette base de données sont légèrement supérieures à ceux de la première classe. Ainsi, cette classe représente des roulements avec une dégradation mineure. La troisième classe (Ω_3) comprend les observations appartenant aux roulements 2 et 3 enregistrées entre le 33^{ème} et le 34^{ème} jour de test. Les valeurs des paramètres de surveillance sont légèrement supérieures à ceux de la deuxième classe. Cette classe représente des roulements avec une dégradation moyenne.

Les classes obtenues ($\Omega_1, \Omega_2, \Omega_3, \Omega_4$) représentent respectivement des roulements : sains, avec une dégradation mineure, dégradation moyenne et dégradation majeure.



Figure 4. 41. Déficit de bague interne dans le roulement 3

La base de données dont nous disposons sur le vieillissement des roulements est constituée de 1944 observations. 1458 observations sont utilisées comme base d'apprentissage (486 observations pour chacun des roulements 1, 2 et 3) et 486 observations du roulement 4, qui sont différentes et indépendantes de la base d'apprentissage, sont utilisées pour la validation des méthodes de pronostic. Les paramètres statistiques ($\gamma_3, \gamma_5, \gamma_6$) du roulement 1, 2 et 3 sont utilisés pour estimer les paramètres des MMCs, le neurone neo-flou et ANFIS.

4.4.2.3. Imminence des stades de dégradation du roulement 4

La première méthode de pronostic, basée sur l'identification du processus de dégradation du système étudié, nécessite une base de données sur plusieurs systèmes identiques soumis aux mêmes contraintes (vitesse, charge.. etc.). L'évolution de la dégradation de ces systèmes dans le temps nous permet de prédire l'imminence des différents stades de dégradation. Cependant, dans notre cas, nous observons que l'évolution de la dégradation dans chaque roulement est différente d'un roulement à un autre (voir figure 4.37). En effet, si on prend l'exemple des roulements (1), (2) et (3), on remarque qu'à la fin du test, seul le roulement (3) arrive au stade d'une dégradation majeure. Le roulement (2) s'arrête au stade d'une dégradation moyenne et le roulement (1) s'arrête au stade d'une dégradation mineure. Nous pouvons aussi remarquer à partir de la figure 4.37 que le roulement (4) vieillit plus prématurément par rapport aux autres roulements. Toutes ces différences font que nous ne disposons pas d'assez d'informations sur les différentes évolutions possibles de la dégradation de ce type de roulement. C'est d'ailleurs le

principal inconvénient de cette méthode. Pour pallier à ce problème, nous avons proposé les méthodes (2) et (3).

4.4.2.3.1. Méthode (2) : modèle de Markov évolutif

Comme indiqué dans le chapitre 3 (section 3.2.2), la deuxième méthode de pronostic (modèle de Markov évolutif) consiste à créer un MMC à partir d’une séquence d’observations qui évolue à chaque pas de temps. Cela revient à estimer le MMC ($\bar{\lambda}_t$) à chaque pas de temps. Il faut pour cela définir les différents modes de fonctionnement. Dans notre cas, nous avons des : roulements sains (S_1), roulements avec dégradation mineure (S_2), roulements avec dégradation moyenne (S_3) et roulements avec dégradation majeure (S_4). La séquence d’observations utilisée pour estimer les paramètres du MMC peut être formulée comme suit : $X = \{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$, avec x_{t+1} inconnue. La longueur de la séquence d’observations est fixée à 3. Les mesures sont effectuées avec un intervalle égal à 98 minutes et équivalent à un pas de mesure (r) égal à 1.

Résultats

La figure 4.42 montre la détection de l’imminence du premier stade de dégradation après 7707 minutes de fonctionnement alors que l’apparition du premier stade de dégradation est diagnostiquée qu’après 9176 minutes de fonctionnement, soit environ 24 heures avant l’apparition de la dégradation mineure du roulement 4.

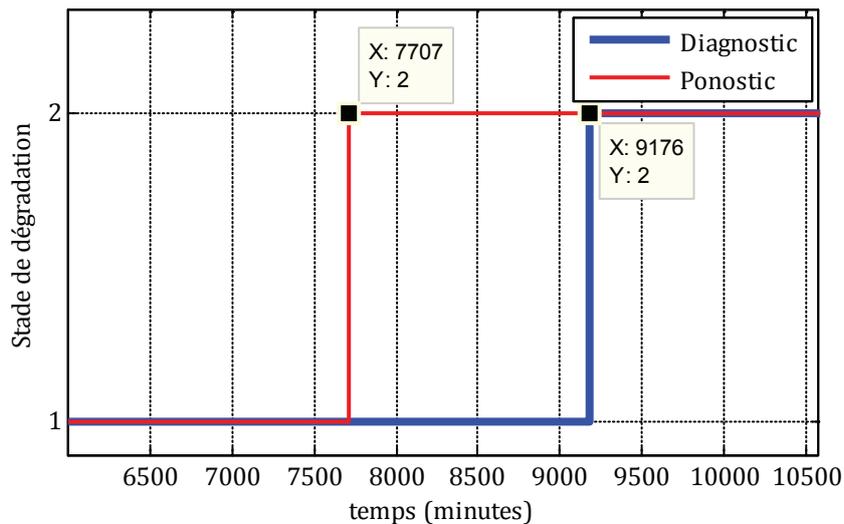


Figure 4. 42. Imminence du premier stade de dégradation du roulement 4

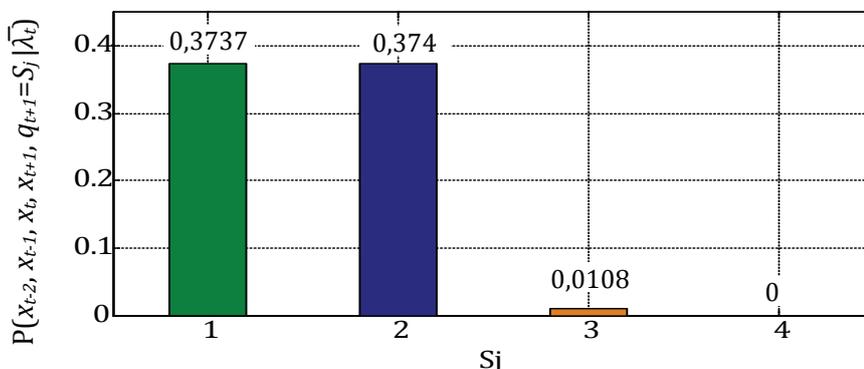


Figure 4. 43. Probabilité qu’une séquence d’observations arrive à un fonctionnement (S_j) à $t = 7707$ minutes

En se basant sur la formule (3.13), la probabilité que la séquence d'observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ arrive à un fonctionnement sain à l'instant $(t+1)$ est de 0,3737 (voir figure 4.43). Cette probabilité est inférieure à celle des roulements avec une dégradation mineure (0,374). Cela permet de détecter l'imminence de la dégradation mineure du roulement (4) à l'instant $t = 7707$ minutes.

La figure 4.44 montre la détection de l'imminence du deuxième stade de dégradation après 42766 minutes de fonctionnement alors que l'apparition du deuxième stade de dégradation est diagnostiquée qu'après 42860 minutes de fonctionnement, soit environ 1h34 minutes avant l'apparition de la dégradation moyenne du roulement 4.

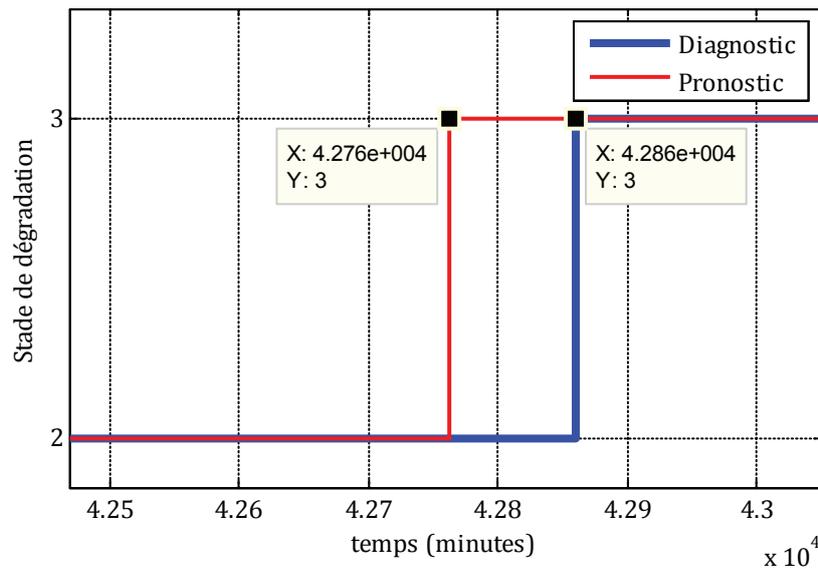


Figure 4. 44. Imminence du deuxième stade de dégradation

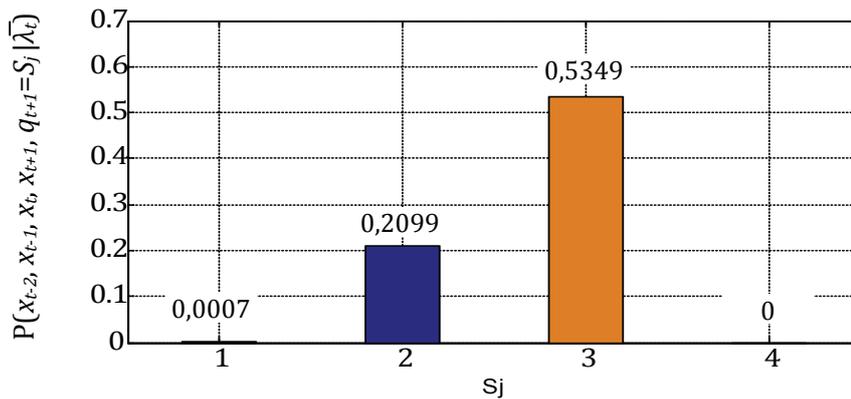


Figure 4. 45. Probabilité qu'une séquence d'observations arrive à un fonctionnement (S_j) à $t = 42766$ minutes

En se basant sur la formule (3.13), la probabilité que la séquence d'observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ arrive à une dégradation mineure à l'instant $(t+1)$ est de 0,2099 (voir figure 4.45). Cette probabilité est inférieure à celle des roulements avec une dégradation moyenne (0,5349). Cela permet de détecter l'imminence de la dégradation moyenne du roulement (4) à l'instant $t = 42766$ minutes.

La figure 4.46 montre la détection de l'imminence du troisième stade de dégradation après 50110 minutes de fonctionnement alors que l'apparition du troisième stade de dégradation est diagnostiquée qu'après 50300 minutes de fonctionnement, soit environ 3h10 minutes avant l'apparition de la dégradation majeure du roulement 4.

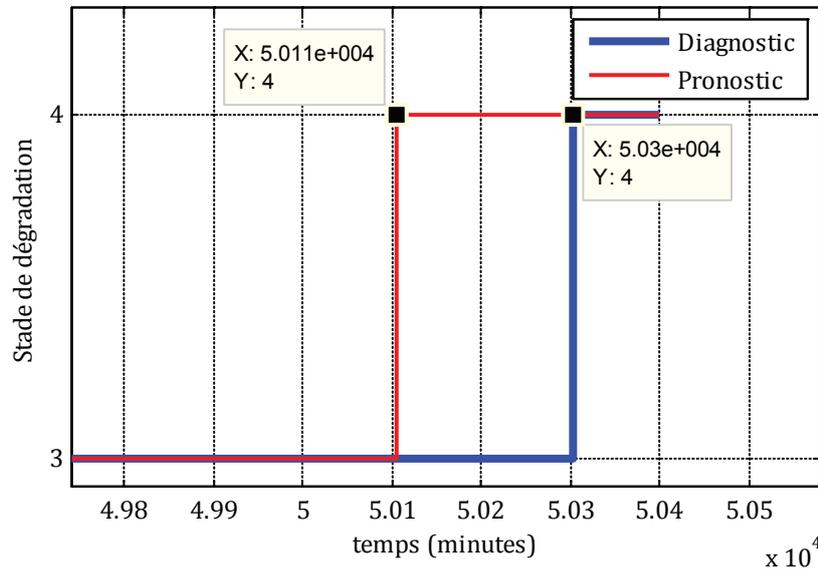


Figure 4. 46. Imminence du troisième stade de dégradation

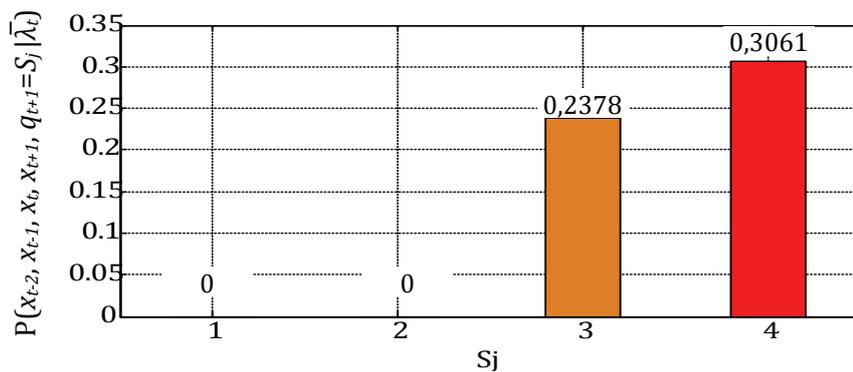


Figure 4. 47. Probabilité qu'une séquence d'observations arrive à un fonctionnement (S_j) à t = 50110 minutes

La probabilité que la séquence d'observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ arrive à une dégradation moyenne à l'instant (t+1) est de 0,2378 (voir figure 4.47). Cette probabilité est inférieure à celle des roulements avec une dégradation majeure (0,3061). Cela permet de détecter l'imminence de la dégradation majeure du roulement (4) à l'instant $t = 50110$ minutes.

Ces premiers résultats montrent que la méthode (2) de pronostic parvient à détecter l'imminence des stades de dégradation avec un intervalle de temps pouvant atteindre 24 heures pour une évolution assez lente de la dégradation et 1h34 voir 3h10 minutes pour une évolution très rapide de la dégradation du roulement 4.

4.4.2.3.2. Méthode (3) : Modèles de Markov cachés

En ce qui concerne la méthode (3), les modes de fonctionnement restent les mêmes que ceux de la méthode (2) : roulements sains (S_1), roulements avec dégradation mineure (S_2), roulements avec dégradation moyenne (S_3) et roulements avec dégradation majeure (S_4). À partir de chaque classe, une séquence d'observations de longueur ($T=3$) est extraite. Chaque observation est décrite par les paramètres statistiques ($\gamma_3, \gamma_5, \gamma_6$). Les séquences d'observations utilisées pour estimer les paramètres des MMCs peuvent être formulées comme suit : $X^k = \{x_0^k, x_1^k, \dots, x_T^k\}$ pour $1 \leq k \leq 4$ et $T = 3$ (Chapitre 2, section 2.3.2.3.2). La longueur T correspond à un tirage successif de $T+1$ observations de chacune des classes.

Les paramètres initiaux de chaque MMC (matrice de transition $A^{(k)}$ et vecteur d'état initial $\pi^{(k)}$), à partir desquels nous commencerons l'apprentissage, sont définis comme suit :

$$A^{(1,2,3,4)} = \begin{matrix} & \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 & \Omega_4 \end{matrix} \\ \begin{matrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \\ \Omega_4 \end{matrix} & \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0,5 & 0,5 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}, \quad \pi^{(1,2,3,4)} = \begin{matrix} \Omega_1 & \Omega_2 & \Omega_3 & \Omega_4 \\ \left(\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \right) \end{matrix} \quad (4.30)$$

Notons que les probabilités de transition $a_{43}, a_{42}, a_{41}, a_{32}, a_{31}, a_{21}$ sont fixées à zéros. La classe Ω_1 correspond à des roulements sains et $\Omega_2, \Omega_3, \Omega_4$ correspondent à des dégradations progressives des roulements. Dans ce cas, il est normal de fixer à zéros $a_{43}, a_{42}, a_{41}, a_{32}, a_{31}, a_{21}$ car nous ne prenons pas en compte le fait qu'un roulement dégradé puisse revenir à un fonctionnement normal.

Les paramètres des 03 MMCs, après re-estimation, sont donnés comme suit :

MMC des roulements sains	MMC des roulements avec dégradation mineure
$\bar{A}^1 = \begin{pmatrix} 0,9949 & 0,0051 & 0 & 0 \\ 0 & 0,9878 & 0,0122 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ $\bar{\pi}^1 = (0,9998 \quad 0,0002 \quad 0 \quad 0)$	$\bar{A}^2 = \begin{pmatrix} 0,1076 & 0,8924 & 0 & 0 \\ 0 & 0,9999 & 0,0001 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ $\bar{\pi}^2 = (0,0443 \quad 0,9557 \quad 0 \quad 0)$

MMC des roulements avec dégradation moyenne MMC des roulements avec dégradation majeure

$\bar{A}^3 = \begin{pmatrix} 0,0075 & 0,0253 & 0,9672 & 0 \\ 0 & 0,0216 & 0,9784 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ $\bar{\pi}^3 = (0,0015 \quad 0,0147 \quad 0,9838 \quad 0)$	$\bar{A}^4 = \begin{pmatrix} 0,0013 & 0,0011 & 0,0066 & 0,9910 \\ 0 & 0,0014 & 0,0100 & 0,9886 \\ 0 & 0 & 0,0069 & 0,9931 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ $\bar{\pi}^4 = (0,0001 \quad 0,0003 \quad 0,0043 \quad 0,9953)$
--	---

Résultats

La figure 4.48 montre la détection de l'imminence du premier stade de dégradation après 5553 minutes de fonctionnement alors que l'apparition du premier stade de dégradation est

diagnostiquée qu’après 9176 minutes de fonctionnement, soit environ 2 jours 12h23min avant l’apparition de la dégradation mineure du roulement 4.

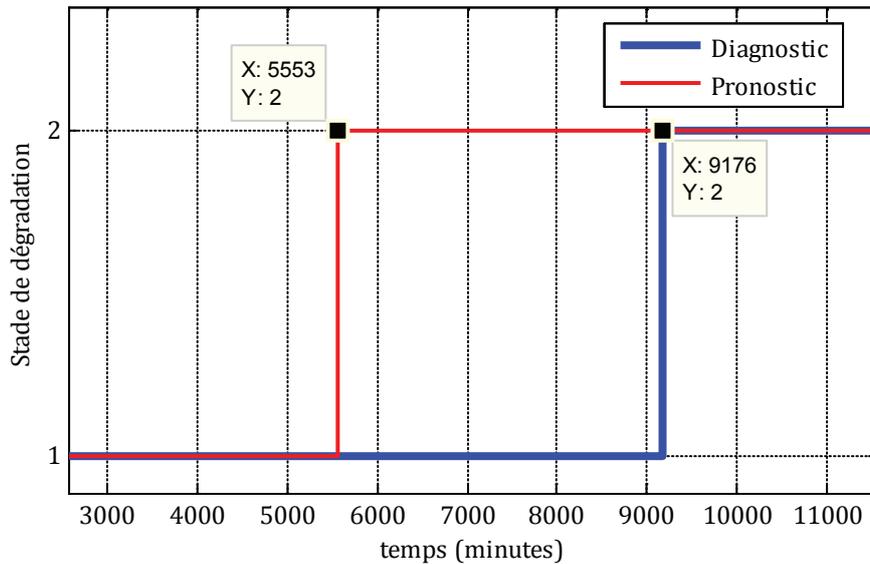


Figure 4. 48. Imminence du premier stade de dégradation

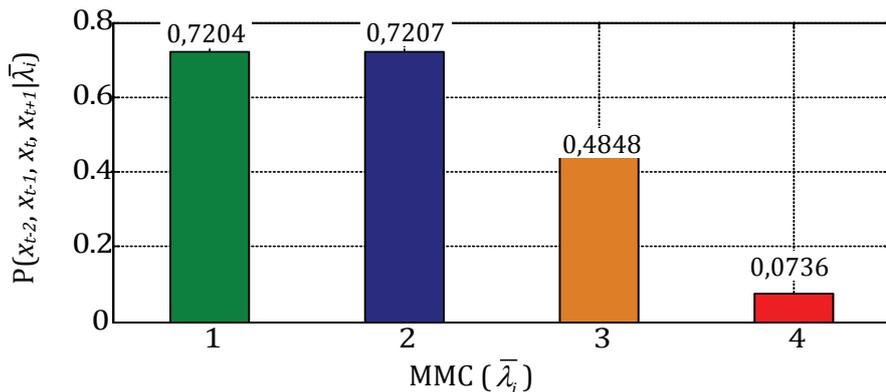


Figure 4. 49. Probabilité de générer une séquence d’observations étant donné les MMCs ($\bar{\lambda}_t$) à $t = 5553$ minutes

La probabilité d’émettre une observation en étant dans un fonctionnement sain à l’instant $t = 5553$ minutes est $b_1(x_t) = 0,92$ (Chapitre 2, formule (2.72)). La probabilité d’émettre une observation en étant dans un fonctionnement sain à l’instant $(t+1)$ est inconnue. Dans ce cas, le coefficient C_{sj} que nous avons introduit, (Chapitre 3, section 3.2.2), prend la même valeur que la probabilité $b_1(x_t)$ à l’instant $(t+1)$ pour un fonctionnement sain ($C_{sj} = b_1(x_t) = 0,92$) ($j=1$) et un coefficient de 1 ($C_{sj}=1$) pour les prochains stades de dégradation ($j=2, 3, 4$). La figure 4.49 montre les probabilités de générer la séquence d’observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ à l’instant $t=5553$ minutes étant donné les MMCs. Comme nous pouvons le voir, pour $C_{sj}=0,92$, la probabilité de générer la séquence observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ à l’instant $t=5553$ minutes, étant donné un fonctionnement sain des roulements, est égale à 0,7204. Cette probabilité est inférieure à celle des roulements avec une dégradation mineure (0,7207). Cela permet de détecter l’imminence de la dégradation mineure du roulement (4) à l’instant $t = 5553$ minutes.

Les figures 4.50 et 4.52 montrent la détection de l'imminence du second et du troisième stade de dégradation après 42667 et 49910 minutes de fonctionnement alors que l'apparition du second stade de dégradation est diagnostiquée qu'après 42860 minutes de fonctionnement, soit environ 3^h13^{min} avant l'apparition de la dégradation moyenne du roulement 4. L'apparition du troisième stade de dégradation est diagnostiquée qu'après 50300 minutes de fonctionnement, soit environ 6^h30^{min} avant l'apparition de la dégradation majeure du roulement 4.

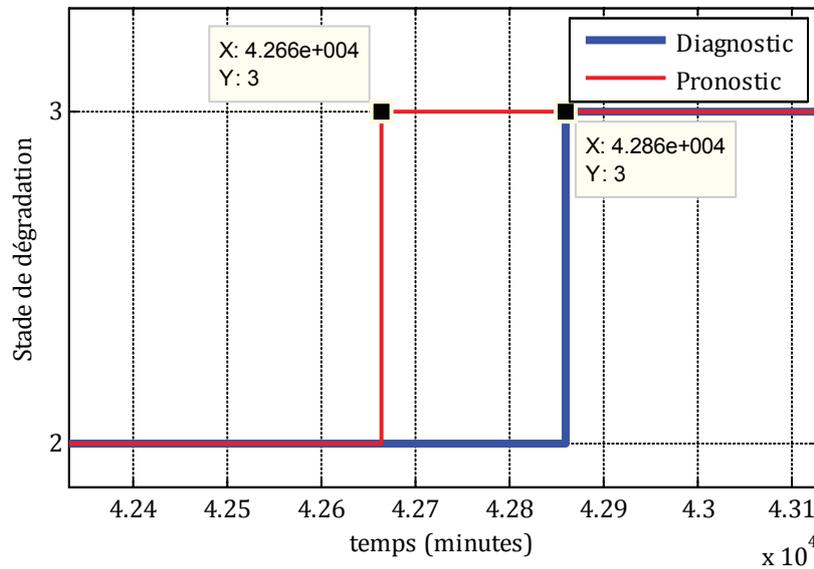


Figure 4. 50. Imminence du deuxième stade de dégradation

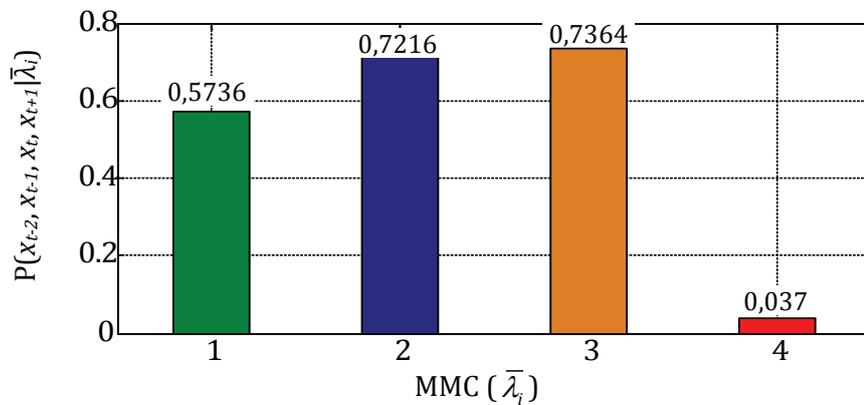


Figure 4. 51. Probabilité de générer une séquence d'observations étant donné les MMCs ($\bar{\lambda}_j$) à $t = 42667$ minutes

La probabilité d'émettre une observation en étant dans le premier stade de dégradation à l'instant $t = 42667$ minutes est $b_2(x_t) = 0,92$. La probabilité d'émettre une observation en étant dans le premier stade de dégradation à l'instant $t+1$ est inconnue. Le coefficient C_{sj} , dans ce cas, prend la même valeur que la probabilité $b_2(x_t)$ à l'instant $(t+1)$ pour le premier stade de dégradation ($C_{sj} = b_2(x_t) = 0,92$) ($j = 2$) et un coefficient de 1 ($C_{sj} = 1$) pour les prochains stades de dégradation ($j = 3, 4$). La figure 4.51 montre les probabilités de générer la séquence d'observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ à l'instant $t = 42667$ minutes étant donné les MMCs ($\bar{\lambda}_j$). Comme nous pouvons le voir, pour $C_{sj} = 0,92$, la probabilité de générer la séquence observations

$\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ à l'instant $t = 42667$ minutes, étant donné le premier stade dégradation, est égale à 0,7216. Cette probabilité est inférieure à celle des roulements avec une dégradation moyenne (0,7364). Cela permet de détecter l'imminence du second stade de dégradation du roulement (4) à l'instant $t = 42667$ minutes.

Pour la troisième prédiction, la probabilité d'émettre une observation en étant dans le second stade de dégradation à l'instant $t=49910$ minutes est $b_3(x_t)=0,83$. La probabilité d'émettre une observation en étant dans le second stade de dégradation à l'instant $(t+1)$ est inconnue. Le coefficient C_{sj} , dans ce cas, prend la même valeur que la probabilité $b_3(x_t)$ à l'instant $(t+1)$ pour le second stade de dégradation ($C_{sj}=b_3(x_t)=0,83$) ($j=3$) et un coefficient de 1 ($C_{sj}=1$) pour le prochain stade de dégradation ($j=4$). La figure 4.53 montre les probabilités de générer la séquence d'observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ à l'instant $t=49910$ minutes, étant donné les MMCs $(\bar{\lambda}_i)$. Comme nous pouvons le voir, pour $C_{sj}=0,83$, la probabilité de générer la séquence observations $\{x_{t-2}, x_{t-1}, x_t, x_{t+1}\}$ à l'instant $t=49910$ minutes, étant donné le second stade dégradation, est égale à 0,4779. Cette probabilité est inférieure à celle des roulements avec une dégradation majeure (0,4886). Cela permet de détecter l'imminence du troisième stade de dégradation du roulement (4) à l'instant $t = 49910$ minutes.

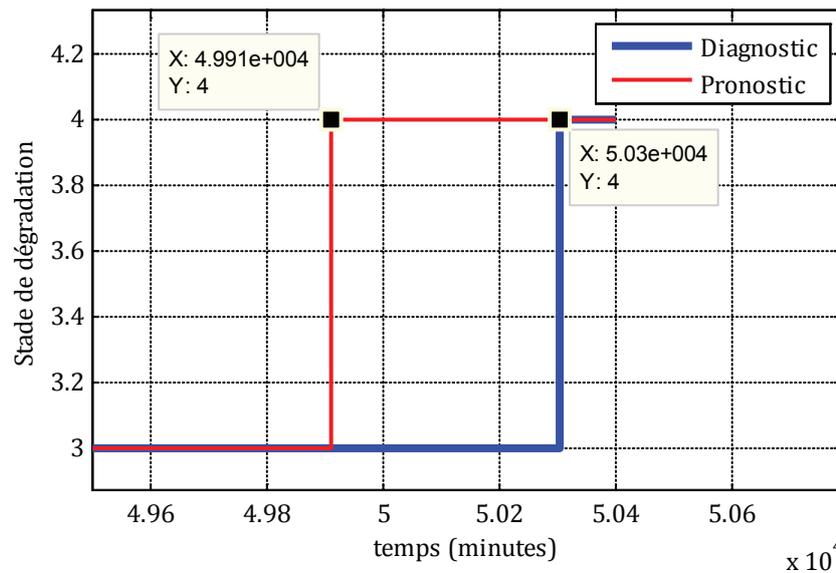


Figure 4. 52. Imminence du troisième stade de dégradation

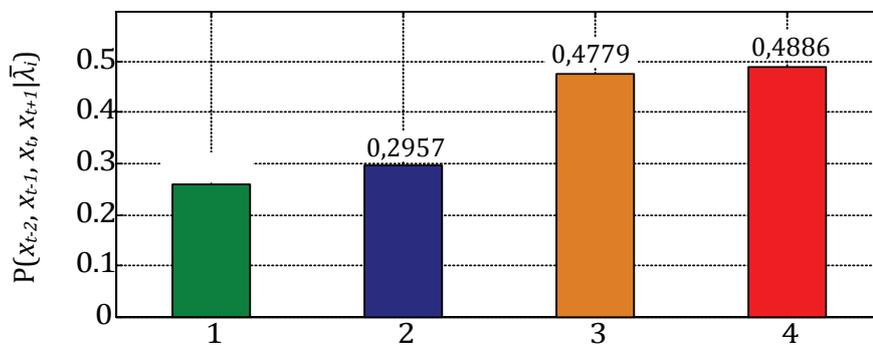


Figure 4. 53. Probabilité de générer une séquence d'observations étant donné les MMCs $(\bar{\lambda}_i)$ à $t = 49910$ minutes

Ces résultats montrent que la méthode (3) de pronostic parvient à détecter l'imminence des stades de dégradation avec un intervalle de temps pouvant atteindre 2 jours 12h pour une évolution assez lente de la dégradation et 3h13 voir 6h30 minutes pour une évolution très rapide de la dégradation du roulement 4.

4.4.2.4. Prédiction des défauts du roulement 4

Les fréquences caractéristiques des défauts des roulements peuvent être obtenues par le biais de la formule (4.7). Connaissant la vitesse de rotation des roulements et leurs caractéristiques, nous pouvons en déduire les fréquences des défauts. Elles sont données comme suit :

- Défaut bague externe : 236,4 Hz ;
- Défaut de bague interne : 296,8 Hz ;
- Défaut de rouleaux : 280,8 Hz.

Une analyse fréquentielle (voir figure 4.54) a été réalisée aux instants correspondant aux prédictions de l'imminence du deuxième et du troisième stade de dégradation données par la méthode (3) : $t=42667$ minutes et $t=49910$ minutes, respectivement. On observe que l'amplitude du pic à la fréquence de la bague externe (236.4 Hz) a augmenté à $t=42667$ minutes. Nous pouvons en déduire que le deuxième stade de dégradation est dû à un défaut de la bague externe du roulement 4 (voir figure 4.55). On observe aussi que l'amplitude du pic à la fréquence des rouleaux (280,8 Hz) a augmenté à $t=49910$ minutes. Il en est de même pour la fréquence caractéristique de la bague interne. Nous pouvons en déduire que le troisième stade de dégradation est dû à un défaut de la bague externe, interne, et des rouleaux du roulement 4 (voir figure 4.56).

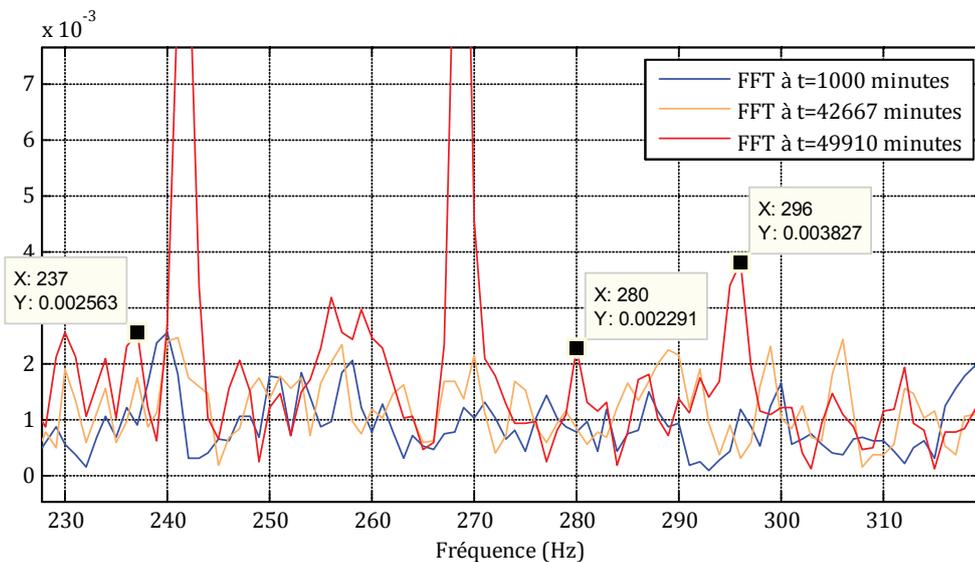


Figure 4. 54. FFT du signal vibratoire du roulement 4 à $t=1000, 42667, 49910$ minutes

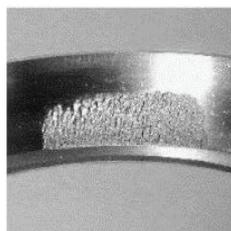


Figure 4. 55. Défaut de bague externe dans le roulement 4

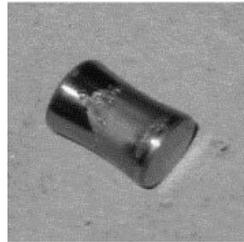


Figure 4. 56. Défaut de rouleau dans le roulement 4

4.4.2.5. Estimation du temps restant avant le prochain stade de dégradation du roulement 4

Le système neuro flou ANFIS et le neurone neo-flou que nous avons présentés dans le chapitre 3 vont nous servir pour estimer le temps restant avant l'apparition du prochain stade de dégradation. Pour cela, nous allons prédire l'évolution des trois paramètres que nous avons sélectionnés ($\gamma_3, \gamma_5, \gamma_6$) entre l'instant (t), où l'on détecte l'imminence du prochain stade de dégradation, et l'instant ($t+p$), où l'on diagnostique l'apparition de ce stade de dégradation. Nous utiliserons les prédictions données par la méthode (3) (section 4.4.2.3.2) afin de juger de l'efficacité du neurone neo-flou et d'ANFIS. Le choix de la méthode (3) est justifié car qu'il permet d'obtenir des prédictions plus précoces par rapport à la méthode (2) ce qui peut être un avantage pour mettre au point une stratégie de maintenance efficace.

Comme mentionné dans le chapitre 3, section 3.3, la prédiction de x_{t+p} notée \hat{x}_{t+p} peut être obtenue par deux approches :

a) *Approche (1) : prédiction multi-pas*

La première approche consiste à prédire jusqu'à un horizon (p) la valeur de l'observation \hat{x}_{t+p} à partir des estimations des prédictions précédentes (Chapitre 3, section 3.3). Cette approche est définie par la formule (3.20) du chapitre 3. La valeur de p n'étant pas connue, elle sera déterminée en effectuant un diagnostic sur les p -valeurs prédites. Le temps restant avant l'apparition du prochain stade de dégradation sera alors déterminé à partir de l'instant (t), où l'imminence du prochain stade de dégradation est détectée, et l'instant ($t+p$), où l'apparition du prochain stade de dégradation est diagnostiquée à partir des p -valeurs prédites. Le tableau 4.13 résume les différents paramètres qui doivent être fixés pour ANFIS et le neurone neo-flou.

Tableau 4. 13. Configuration d'ANFIS et du neurone neo-flou

ANFIS	
Nombre d'entrées (longueur de la série temporelle) (n)	3
Le pas de la série temporelle (r)	1
Nombre de fonctions d'appartenance	2
Type de fonction d'appartenance	Gaussienne
T- norme	Produit
Type d'inférence	Linéaire Sugeno
Apprentissage	
Nombre d'observations	1944
Nombre d'itérations	100
Ensemble d'apprentissage / test	1458/486
Algorithme d'apprentissage	LES+ Descente Gradient

Neurone neo-flou	
Nombre d'entrées (longueur de la série temporelle) (n)	3
Le pas de la série temporelle (r)	1
Nombre de fonctions d'appartenance	10 à 16
Type de fonction d'appartenance	Triangulaire
Apprentissage	
Nombre d'observations	1944
Nombre d'itérations	100
Ensemble d'apprentissage / test	1458/486
Algorithme d'apprentissage	Descente Gradient

Résultats

Les résultats donnés par la méthode (3) nous indiquent les instants (t) à partir des quels nous commencerons le pronostic. Les instants (t) correspondent à l'imminence des trois stades de dégradations et sont donnés comme suit :

- Dégradation mineure $\rightarrow t= 5553$ minutes ;
- Dégradation moyenne $\rightarrow t= 42667$ minutes ;
- Dégradation majeure $\rightarrow t= 49910$ minutes.

- Dégradation mineure $\rightarrow t= 5553$ minutes

Etant donné que le roulement (4) débute par un fonctionnement sain, il s'agit maintenant de prédire la dégradation mineure. Les figures 4.57, 4.59 et 4.61 montrent l'évolution de la RMSE (Root Mean Square Error) pour chacun des paramètres de surveillance ($\gamma_3, \gamma_5, \gamma_6$).

Puissance du signal vibratoire γ_3

Le système neuro-flou ANFIS possède un ensemble de trois entrées $\{x_t, x_{t-1}, x_{t-2}\}$ sous forme d'une série temporelle. Le modèle de prédiction ANFIS donne des résultats satisfaisants pour des prévisions à court terme (jusqu'à 6000 minutes), les erreurs obtenues augmentent entre 6×10^{-5} et $8,83 \times 10^{-4}$. À moyen terme (jusqu'à 7500 minutes), les erreurs obtenues augmentent entre $8,83 \times 10^{-4}$ et 0,0109, contribuant d'une manière insatisfaisante à la dégradation de la précision et de l'exactitude du pronostic (voir figure 4.58).

Le neurone neo-flou, comme ANFIS, possèdent un ensemble de trois entrées $\{x_t, x_{t-1}, x_{t-2}\}$ sous forme d'une série temporelle. L'erreur varie cette fois-ci entre $8,26 \times 10^{-5}$ et 8×10^{-4} pour une prédiction à court terme et entre $1,83 \times 10^{-4}$ et 0,00375 pour une prédiction à moyen et long termes. Cette architecture donne des résultats satisfaisants pour des prévisions à court et à long termes (voir Figure 4. 58).

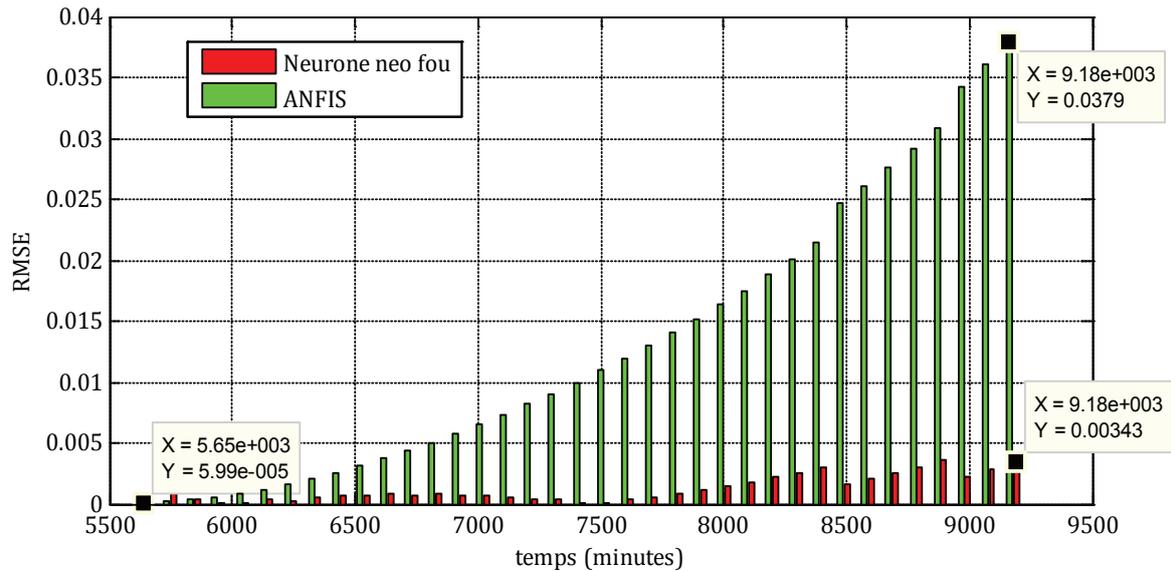


Figure 4. 57. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps

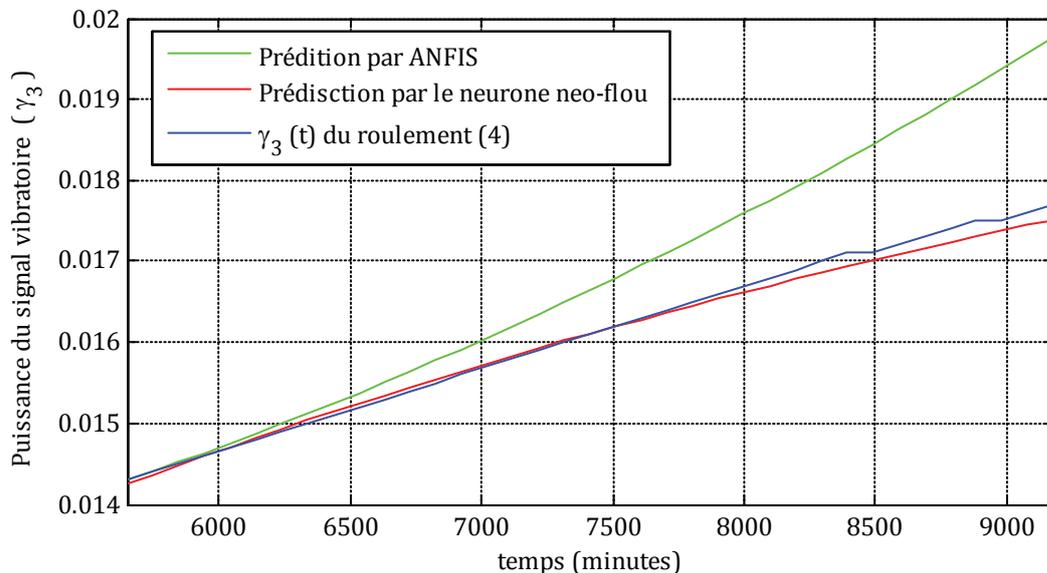


Figure 4. 58. Prédiction de la puissance du signal vibratoire par ANFIS et le neurone neo-flou

Valeur efficace γ_5

Le système neuro-flou ANFIS avec trois entrées donne des résultats satisfaisants pour des prévisions à court et à moyen termes (jusqu'à 7500 minutes), les erreurs obtenues augmentent entre 5×10^{-4} et $4,42 \times 10^{-3}$. À long terme (jusqu'à 9175 minutes), les erreurs obtenues augmentent jusqu'à $9,54 \times 10^{-3}$, contribuant d'une manière satisfaisante à une bonne prédiction du processus de dégradation du roulement 4 (voir figure 4.60).

Les résultats obtenus par le neurone neo-flou avec trois entrées varient cette fois-ci entre $6,94 \times 10^{-5}$ et $3,98 \times 10^{-4}$ pour une prédiction à court et à moyen termes. À long terme, les erreurs obtenues augmentent jusqu'à 0,0446, contribuant d'une manière insatisfaisante à la dégradation de la précision et de l'exactitude du pronostic (voir figure 4.60). Cependant, cette architecture donne des résultats satisfaisants pour des prévisions à court terme et même à moyen terme.

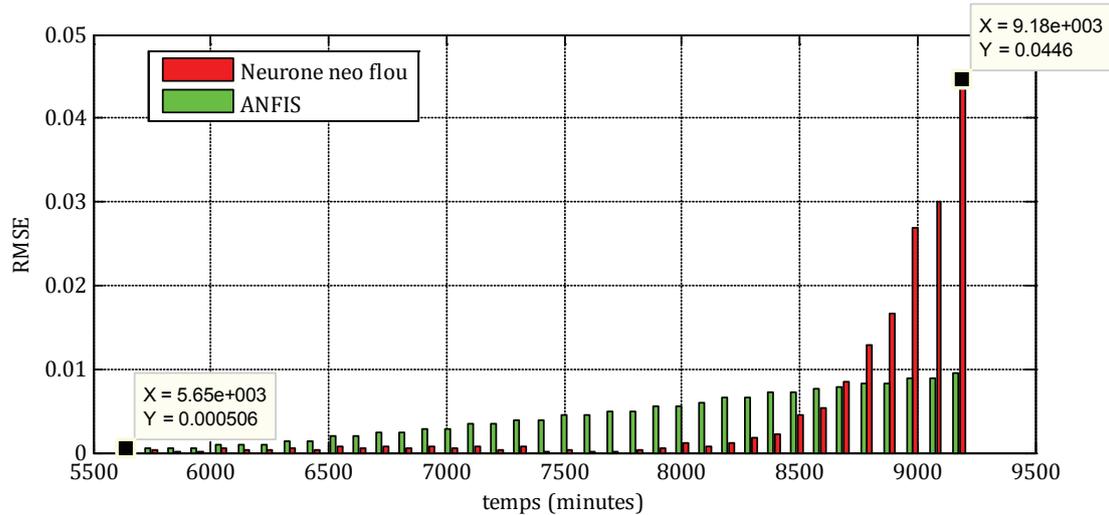


Figure 4. 59. Evolution de la RMSE de la valeur efficace du signal vibratoire en fonction du temps

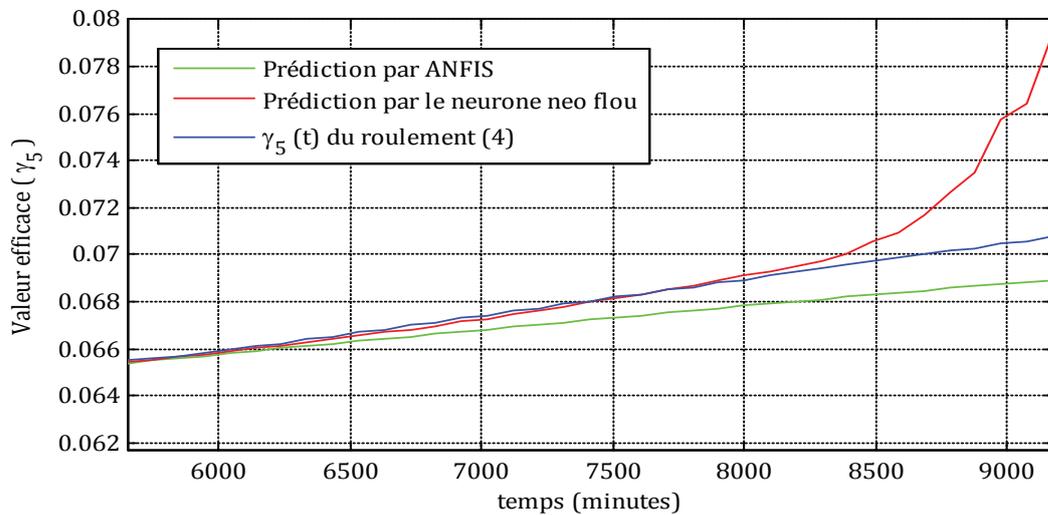


Figure 4. 60. Prédiction de la valeur efficace du signal vibratoire par ANFIS et le neurone neo-flou

Valeur moyenne γ_6

Le système neuro-flou ANFIS avec trois entrées donne des résultats satisfaisants pour des prévisions à court terme. Pour une prédiction à moyen et long termes, l'augmentation de la RMSE indique une mauvaise prédiction du processus de dégradation du roulement (4) (voir figures 4.61 et 4.62). Les résultats obtenus par le neurone neo-flou avec trois entrées sont moins précis que ceux obtenus par ANFIS que ce soit à court ou à long termes.

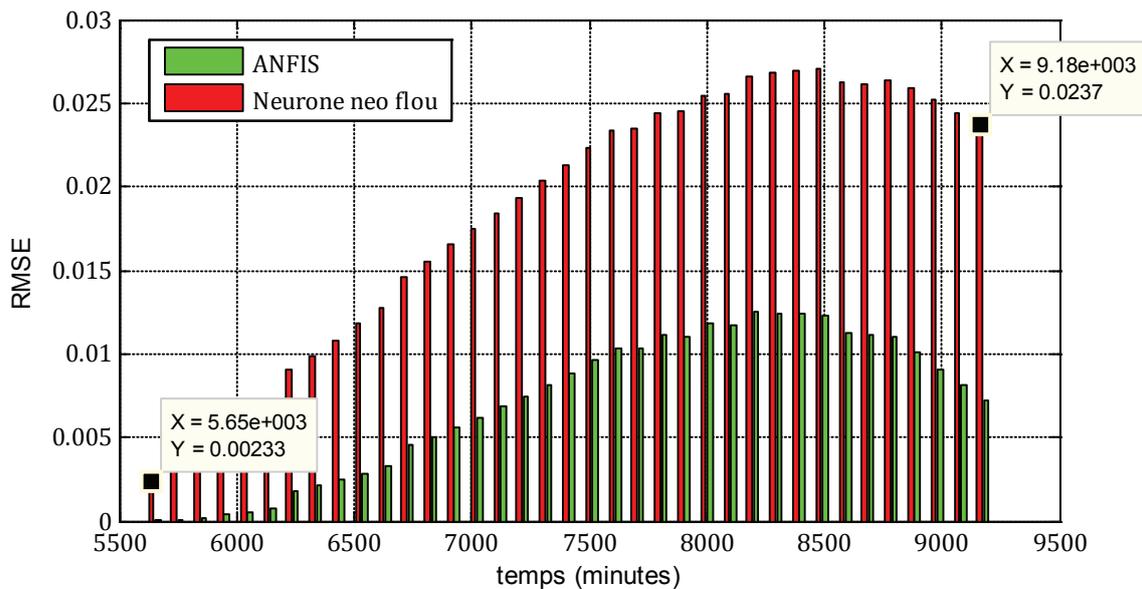


Figure 4. 61. Evolution de la RMSE de la valeur moyenne du signal vibratoire en fonction du temps

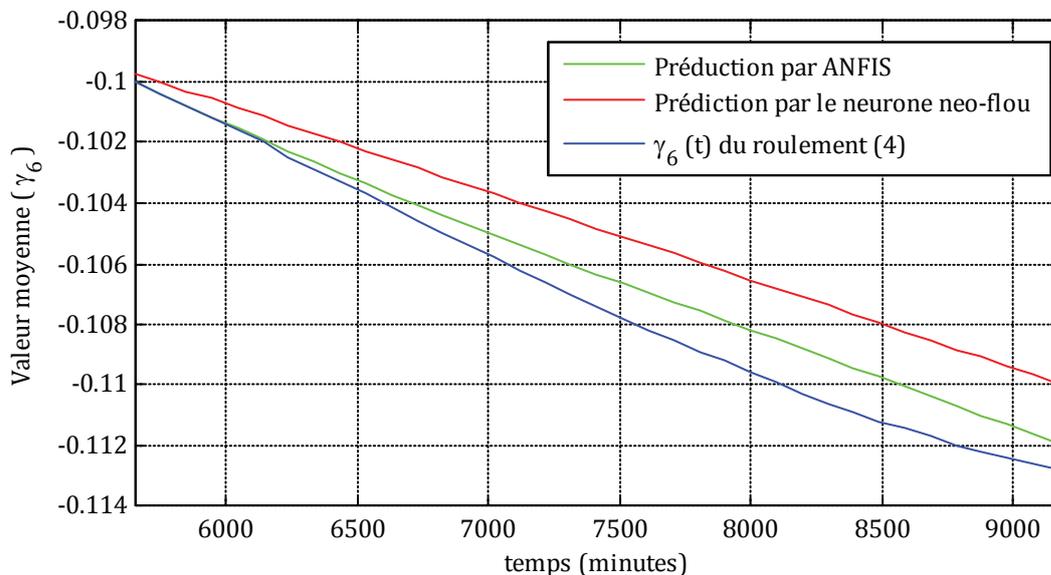


Figure 4. 62. Prédiction de la valeur moyenne du signal vibratoire par ANFIS et le neurone neo-flou

La classification des observations en fonction des trois paramètres statistiques (γ_3 , γ_5 , γ_6), prédites par le neurone neo-flou, a permis d'estimer l'apparition du premier stade de dégradation après 9274 minutes de fonctionnement. Le premier stade de dégradation étant diagnostiqué après 9176 minutes, cela permet d'en déduire que l'approche (1) combinée avec le système neo-flou permet d'avoir un excellent résultat de prédiction. Nous pouvons aussi donner la même remarque au système neuro-flou ANFIS qui estime l'apparition du premier stade de dégradation après 9372 minutes de fonctionnement.

- Dégradation moyenne \rightarrow $t=42667$ minutes

Le premier stade de dégradation étant diagnostiqué, il s'agit maintenant de prédire la dégradation moyenne. Les figures 4.63, 4.65 et 4.67 montrent l'évolution de la RMSE pour chacun des paramètres de surveillance (γ_3 , γ_5 , γ_6).

Puissance du signal vibratoire γ_3

Le système neuro-flou ANFIS avec trois entrées donne des résultats satisfaisants pour des prévisions à court terme, les erreurs obtenues ne dépassent pas les 0,0014. Les erreurs obtenues par le neurone neo-flou sont légèrement supérieures à ceux d'ANFIS et ne dépassent pas les 0,0029.

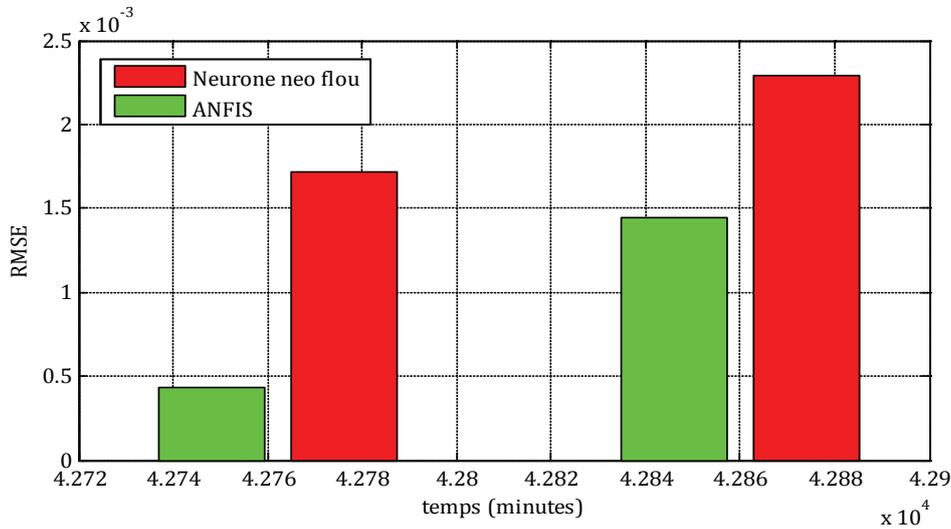


Figure 4. 63. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps

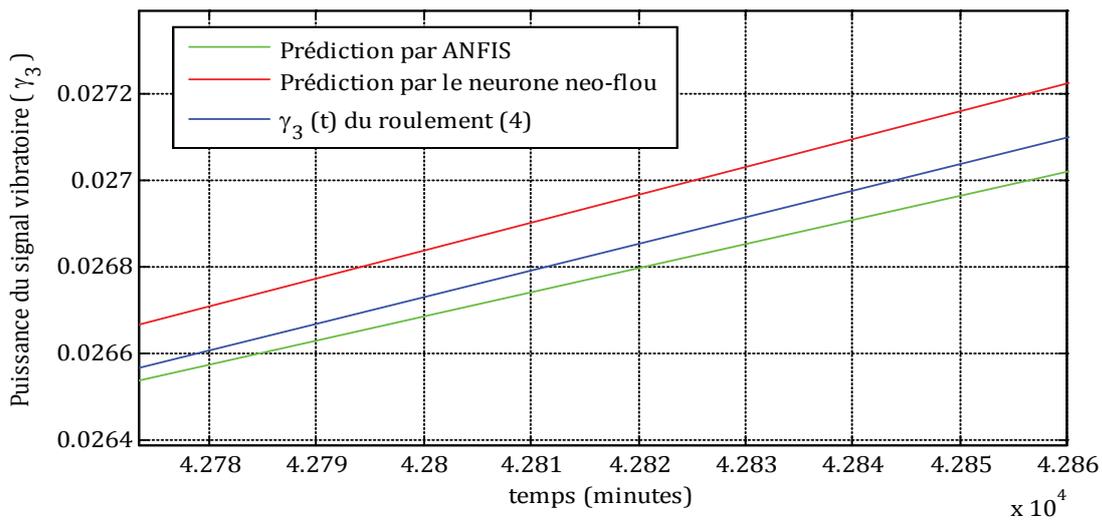


Figure 4. 64. Prédiction de la puissance du signal vibratoire par ANFIS et le neurone neo-flou

Valeur efficace et moyenne du signal vibratoire (γ_5, γ_6).

Nous pouvons aussi donner les mêmes remarques en ce qui concerne la prédiction de l'évolution de la valeur efficace et de la moyenne du signal vibratoire par ANFIS et le neurone neo-flou. En effet, les erreurs de prédiction pour les deux paramètres ne dépassent pas les 0,0016 pour le neurone neo-flou et 0,00083 pour ANFIS.

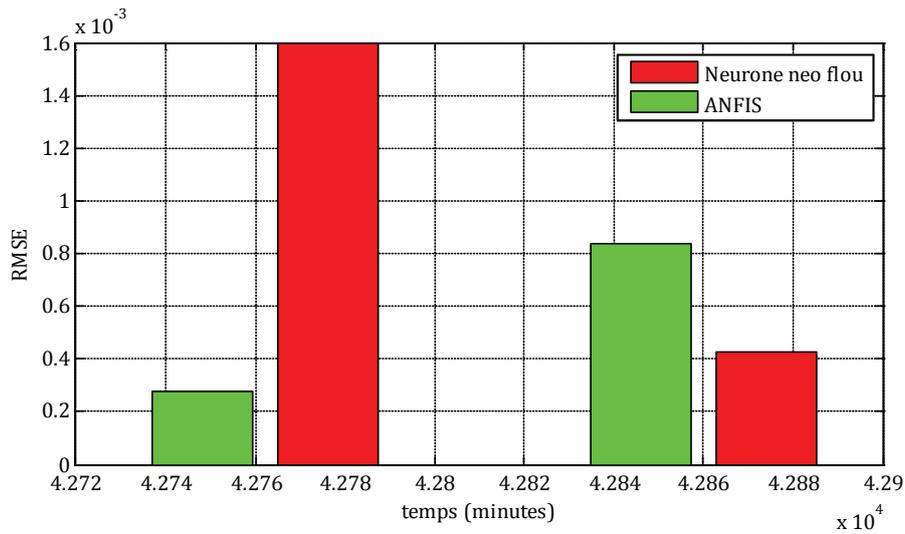


Figure 4. 65. Evolution de la RMSE de la valeur efficace du signal vibratoire en fonction du temps

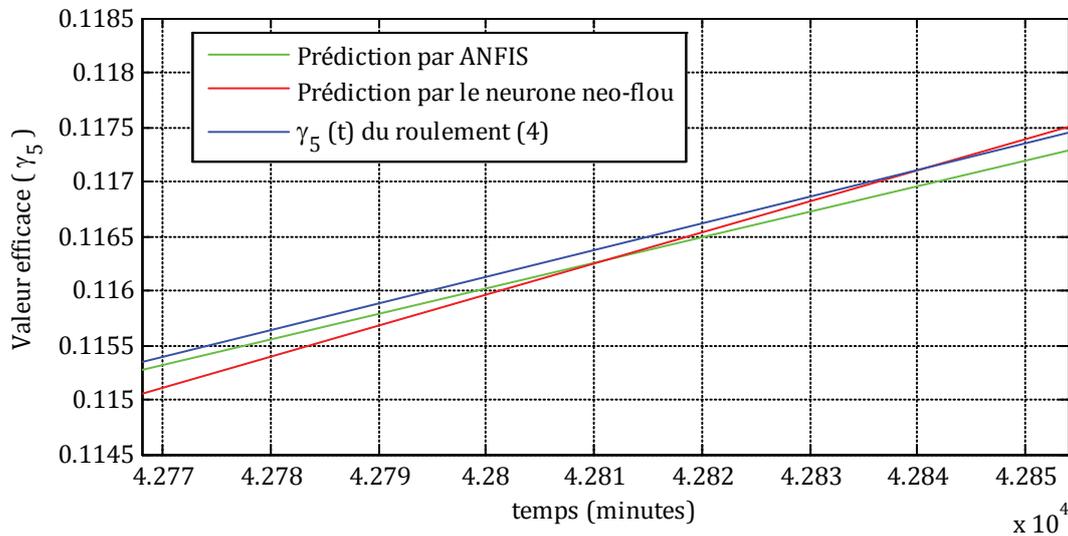


Figure 4. 66. Prédiction de la valeur efficace du signal vibratoire par ANFIS et le neurone neo-flou

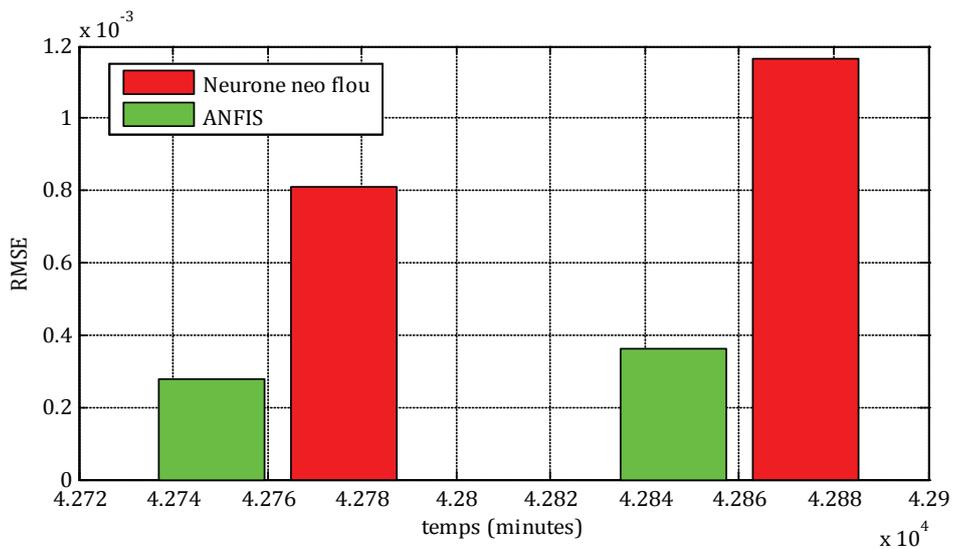


Figure 4. 67. Evolution de la RMSE de la valeur moyenne du signal vibratoire en fonction du temps

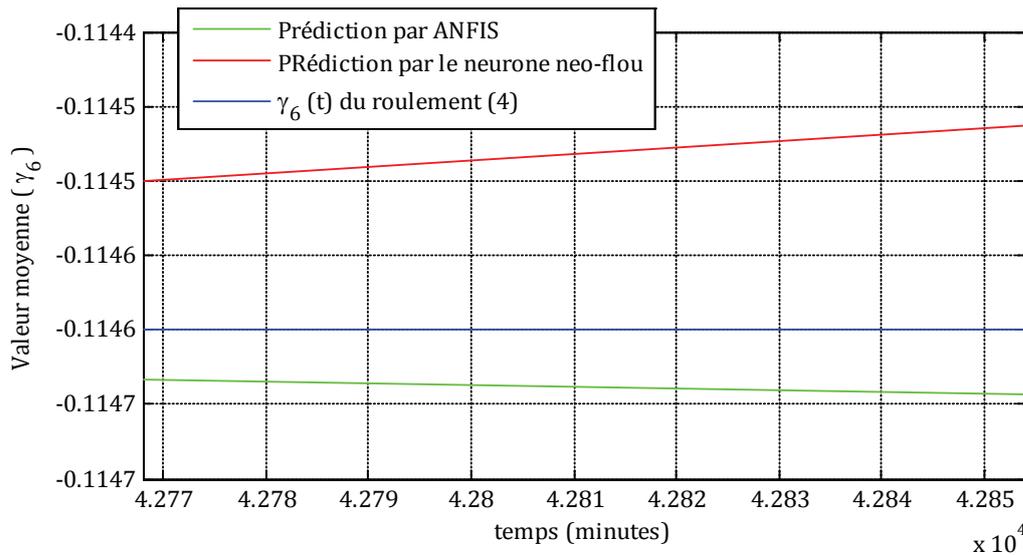


Figure 4. 68. Prédiction de la valeur moyenne du signal vibratoire par ANFIS et le neurone neo-flou

La classification des observations en fonction des trois paramètres statistiques ($\gamma_3, \gamma_5, \gamma_6$), prédites par le neurone neo-flou, a permis d’estimer l’apparition du deuxième stade de dégradation après 42860 minutes de fonctionnement. Le deuxième stade de dégradation étant diagnostiqué après 42860 minutes, cela permet d’en déduire que l’approche (1) combinée avec le neurone neo-flou permet d’avoir un excellent résultat de prédiction. Il en est de même en ce qui concerne le système neuro-flou ANFIS qui estime l’apparition du deuxième stade de dégradation après 42860 minutes de fonctionnement.

- Dégradation majeure → $t= 49910$ minutes

Le deuxième stade de dégradation étant diagnostiqué, il s’agit maintenant de prédire la dégradation majeure. Les figures 4.69, 4.71 et 4.73 montrent l’évolution de la RMSE pour chacun des paramètres de surveillance ($\gamma_3, \gamma_5, \gamma_6$).

Comme dans le cas précédent, et pour les trois paramètres de surveillance ($\gamma_3, \gamma_5, \gamma_6$), la prédiction à court terme est très satisfaisante que ce soit pour ANFIS ou pour le neurone neo-flou. Les erreurs de prédiction pour ANFIS ne dépassent pas les 0,0012. Pour le neurone neo-flou, les erreurs de prédiction ne dépassent pas les 0,00034.

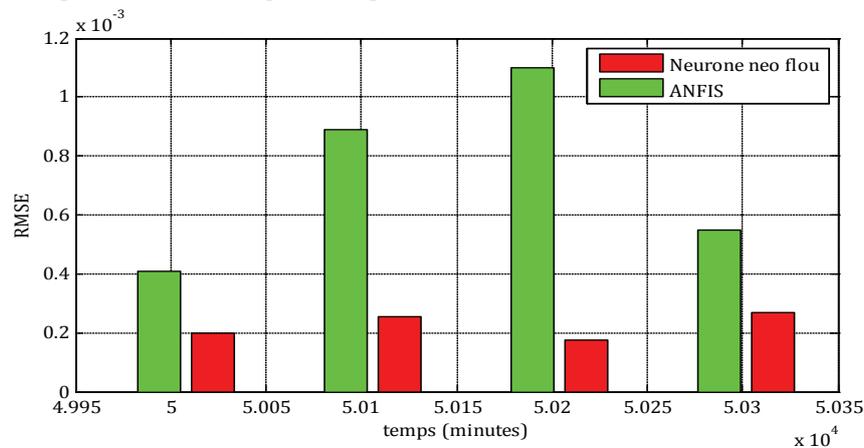


Figure 4. 69. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps

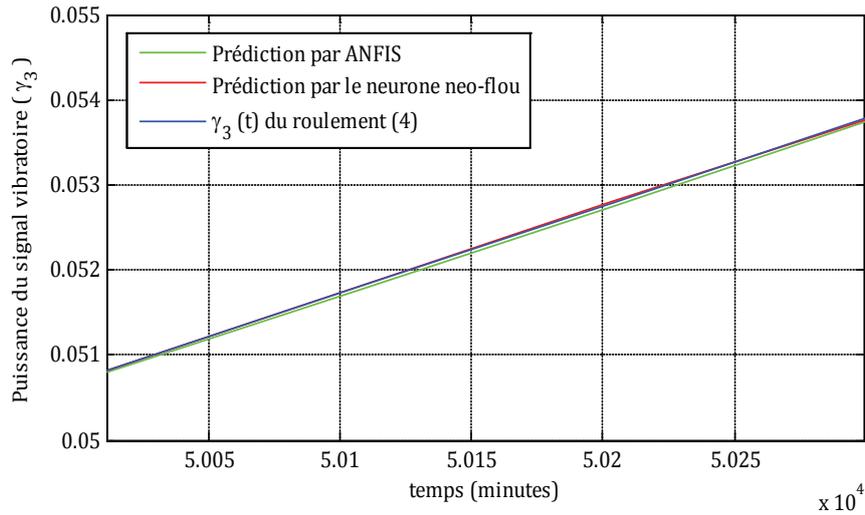


Figure 4. 70. Prédiction de la puissance du signal vibratoire par ANFIS et le neurone neo-flou

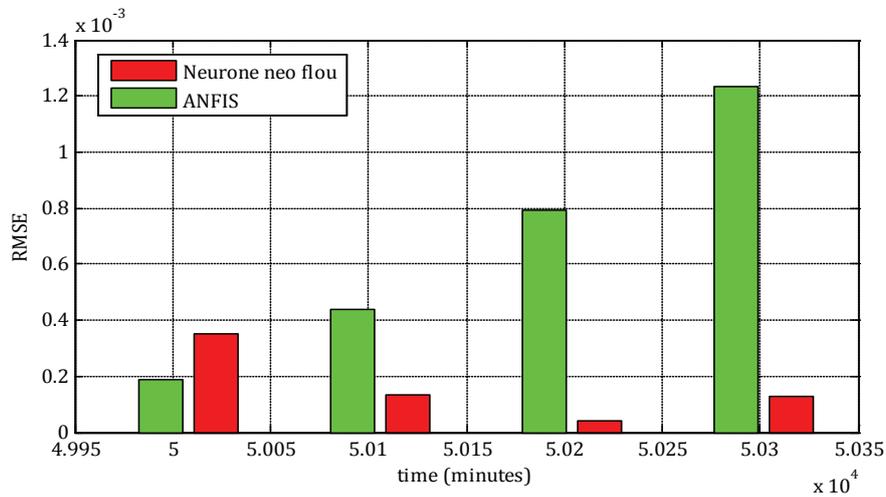


Figure 4. 71. Evolution de la RMSE de la valeur efficace du signal vibratoire en fonction du temps

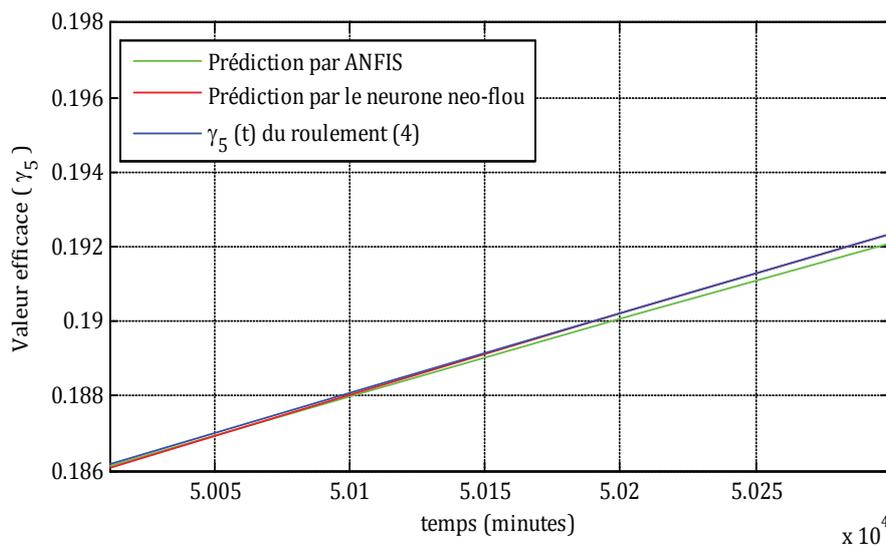


Figure 4. 72. Prédiction de la valeur efficace du signal vibratoire par ANFIS et le neurone neo-flou

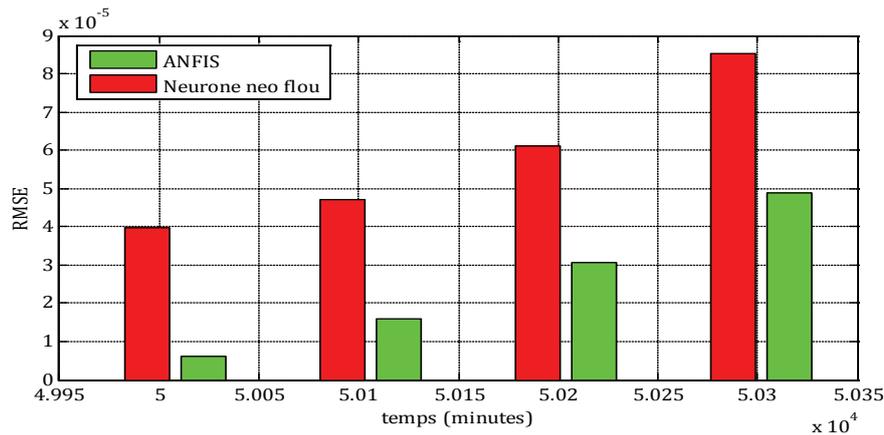


Figure 4. 73. Evolution de la RMSE de la valeur moyenne du signal vibratoire en fonction du temps

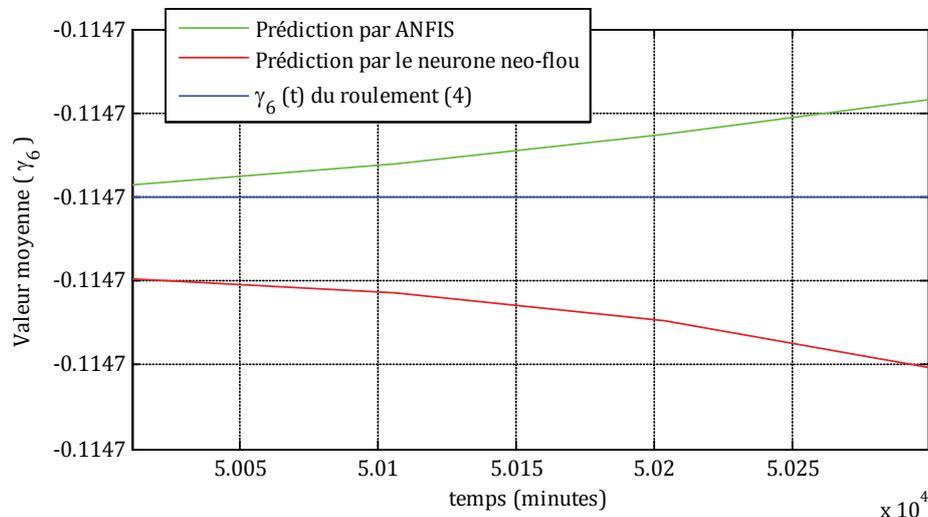


Figure 4. 74. Prédiction de la valeur moyenne du signal vibratoire par ANFIS et le neurone neo-flou

La classification des observations en fonction des trois paramètres statistiques (γ_3 , γ_5 , γ_6), prédites par le neurone neo-flou, a permis d'estimer l'apparition du troisième stade de dégradation après 50302 minutes de fonctionnement. Le troisième stade de dégradation étant diagnostiqué après 50302 minutes, cela permet d'en déduire que l'approche (1) combinée avec le neurone neo-flou permet d'avoir un excellent résultat de prédiction. Il en est de même en ce qui concerne le système neuro-flou ANFIS qui estime l'apparition du troisième stade de dégradation après 50302 minutes de fonctionnement.

b) Approche (2) : prédiction à un pas dans le futur

L'approche 2 consiste à prédire la valeur de l'observation \hat{x}_{t+p} à partir de la série temporelle d'observations $\{x_t, x_{t-r}, x_{t-2r}, \dots, x_{t-(n-1)r}\}$ jusqu'à l'instant (t). Les paramètres de réglage pour ANFIS et le neurone neo-flou sont les mêmes que ceux définis dans le tableau 4.13 avec $n=3$ et $r=1$ où r désigne le pas de prédiction et $(n-1)$ définit le nombre de pas précédents. Nous nous appuyerons sur les résultats donnés par la méthode (3) (voir section 4.4.2.3.2) pour débiter le pronostic.

Les tableaux suivants correspondent à la prédiction de la puissance moyenne du signal vibratoire à plusieurs horizons de prédiction ($p=1, 2, 10, 20, 30, 40$).

- Dégradation mineure → $t= 5553$ minutes

Etant donné que le roulement (4) débute par un fonctionnement sain, il s'agit maintenant de prédire la dégradation mineure. Le tableau 4.14 montre l'évolution de la RMSE de la puissance moyenne du signal vibratoire (γ_3) en fonction des horizons de prédiction (p) pour ANFIS et le neurone neo-flou.

Tableau 4. 14. RMSE de γ_3 à plusieurs horizons de prédiction (p)

Horizon de prédiction (p)	$t+p$ (minutes)	RMSE ANFIS (02 fonctions d'appartenance)	RMSE Neurone neo-flou (10 fonctions d'appartenance)
1	5651	6×10^{-5}	8×10^{-4}
2	5749	0,0057	0,0033
10	6533	0,1812	0,0123
20	7513	0,4426	0,0159
30	8493	1,1043	0,027
40	9473	1,0716	0,056
Temps de simulation		47 secondes	8 secondes

- Dégradation moyenne → $t= 42667$ minutes

Le premier stade de dégradation étant diagnostiqué, il s'agit maintenant de prédire la dégradation moyenne. Le tableau 4.15 montre l'évolution de la RMSE de la puissance moyenne du signal vibratoire (γ_3) en fonction des horizons de prédiction (p) pour ANFIS et le neurone neo-flou.

Tableau 4. 15. RMSE de γ_3 à plusieurs horizons de prédiction (p)

Horizon de prédiction (p)	$t+p$ (minutes)	RMSE ANFIS (02 fonctions d'appartenance)	RMSE Neurone neo-flou (10 fonctions d'appartenance)
1	42765	4.29×10^{-4}	1.72×10^{-3}
2	42863	0,0282	0,013
Temps de simulation		47 secondes	8 secondes

- Dégradation majeure → $t= 49910$ minutes

Le deuxième stade de dégradation étant diagnostiqué, il s'agit maintenant de prédire la dégradation majeure. Le tableau 4.16 montre l'évolution de la RMSE de la puissance moyenne du signal vibratoire (γ_3) en fonction des horizons de prédiction (p) pour ANFIS et le neurone neo-flou.

Tableau 4. 16. RMSE de γ_3 à plusieurs horizons de prédiction (p)

Horizon de prédiction (p)	$t+p$ (minutes)	RMSE ANFIS (02 fonctions d'appartenance)	RMSE Neurone neo-flou (10 fonctions d'appartenance)
1	50008	4×10^{-4}	2×10^{-4}
2	50106	0,0242	0,001
3	50204	0,088	0,002
4	50302	0,2435	0,003
Temps de simulation		47 secondes	8 secondes

Comme nous pouvons le voir, les prédictions à $p=1$ pour les trois stades de dégradation sont les mêmes que ceux de l'approche (1). Cependant, plus l'horizon de prédiction augmente, plus la valeur de la RMSE augmente. Ceci est démontré par la figure 4.75 qui indique que la RMSE obtenue par l'approche (1) et (2) du neurone neo-flou à partir de l'instant $t= 5553$ minutes.

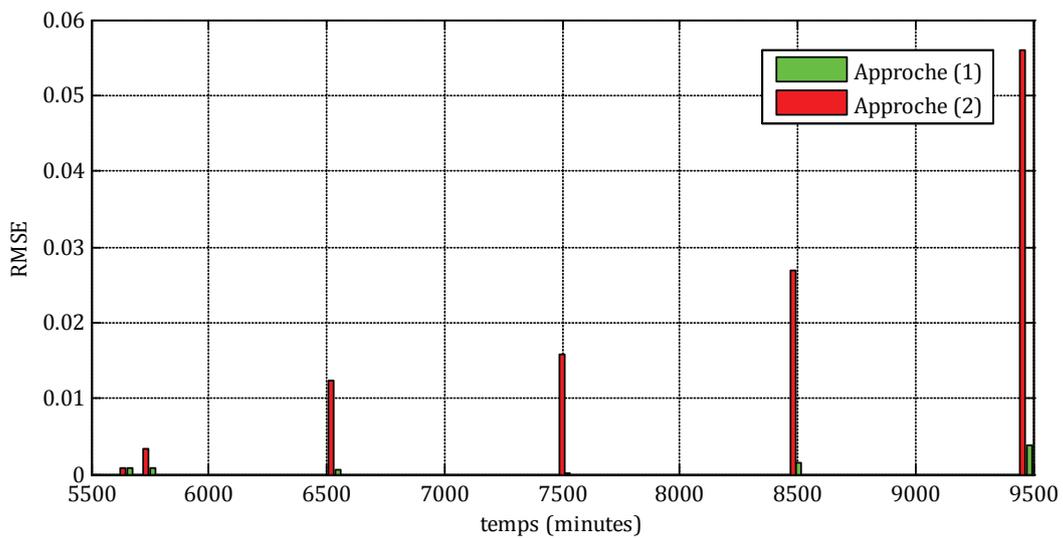


Figure 4. 75. Evolution de la RMSE de la puissance du signal vibratoire en fonction du temps

Ceci nous amène à conclure que l'approche (1) est la mieux adapté pour la prédiction à court et à long termes.

4.5. Conclusion

Les résultats précédents ont mis en évidence les performances des méthodes proposées pour le diagnostic des différents états de la machine asynchrone et pour le pronostic des roulements à rouleaux. Les performances de ces méthodes dépendent fortement du choix des indicateurs de dégradation et de défaut (autrement dit du vecteur de paramètres).

L'algorithme OSBS, pour la sélection de paramètres pertinents, a donné de très bons résultats. En effet, le diagnostic des observations définies par le vecteur de paramètres construit avec cet algorithme s'est montré efficace dans deux cas de figure : apparition d'un nouveau mode de fonctionnement et connaissance limitée sur chacun des modes de fonctionnement.

Les méthodes de diagnostic développées : classification non supervisée par l'algorithme fourmis artificielles et classification supervisée par modèles de Markov cachés, permettent de détecter l'apparition d'un nouveau mode de fonctionnement dans la machine asynchrone. Elles permettent aussi de donner de très bons résultats de diagnostic même avec une connaissance limitée sur chacun des modes de fonctionnement. L'algorithme fourmis artificielles analyse les observations inconnues et inexplorées afin de mettre en évidence les classes regroupant les observations similaires. Cela permet de détecter l'apparition d'un ou plusieurs modes de fonctionnement qui n'étaient pas présents dans la base d'apprentissage. La classification par modèles de Markov cachés permet d'associer un degré d'appartenance (sous forme d'une probabilité) de l'observation à une ou plusieurs classes. Ainsi, il est possible de définir un indice de fiabilité à l'affectation réalisée mais aussi de détecter l'apparition de nouveaux modes de fonctionnement.

Les méthodes de diagnostic développées peuvent être étendues au pronostic. Pour cela, les modèles de Markov cachés, le système neuro-flou ANFIS et le neurone neo-flou ont été appliqués au pronostic des roulements à rouleaux. Dans un premier temps, Les modèles de Markov cachés sont utilisés pour détecter l'imminence des différents stades de dégradation présents dans les roulements. Dans un deuxième temps, ANFIS et le neurone neo-flou sont utilisés pour estimer le temps restant avant le prochain stade de dégradation en prédisant, à court et à long termes, l'évolution des paramètres de surveillance. Dans les deux cas, les résultats de l'estimation et de la prédiction présentent une bonne précision.

Conclusion générale

Les travaux présentés dans cette thèse traitent de l'intégration du diagnostic et du pronostic de pannes dans la sûreté de fonctionnement des systèmes d'entraînement électriques. Le diagnostic permet de déterminer en ligne les composants du système qui sont en faute et donc qui doivent être réparés. Le pronostic permet d'anticiper les défauts en fournissant les informations sur l'état futur du système à partir desquels des actions de maintenance peuvent être envisagées.

Pour débiter l'étude (**Chapitre 1**), nous avons établi un état de l'art sur les différentes stratégies de maintenance utilisées pour assurer la fiabilité et la sûreté de fonctionnement des systèmes industriels. Le but de cette étude était de positionner le diagnostic et le pronostic au sein de la maintenance préventive conditionnelle. Ce travail préliminaire a permis de mettre en évidence deux grandes catégories d'approches. La première approche utilise un modèle mathématique basé sur les résidus pour détecter et suivre la dégradation du système. Cette approche est bien adaptée lorsqu'un modèle analytique précis peut être construit. Cependant, dans la réalité, la complexité du système est telle qu'extraire un modèle est souvent difficile à mettre en œuvre ou dans certains cas impossible à extraire. La deuxième approche consiste à analyser les signaux de mesure pour extraire un vecteur de paramètres indicateur d'état de santé. Ce vecteur de paramètres est ensuite utilisé dans la reconnaissance des formes (RdF) pour détecter des changements dans les données de mesure via des méthodes de classification supervisée et non supervisée. Ces méthodes s'appuient sur un vecteur de paramètres caractéristique des défauts du système. À des fins de diagnostic et de pronostic, des modes de fonctionnement ont été définis pour représenter le comportement du système. Dans le cas du diagnostic, ces modes de fonctionnement peuvent représenter les différents types de défauts qu'un système peut rencontrer durant son fonctionnement. Ces modes peuvent aussi représenter un seul type de défaut mais avec différents niveaux de charge. Dans le cas du pronostic, les modes de fonctionnement représentent les différents stades de dégradation par lesquels le système passera systématiquement.

La deuxième partie (**Chapitre 2**) du travail de thèse a pour but de montrer comment utiliser la reconnaissance des formes pour le diagnostic des défauts. Cette méthodologie est générique, et donc applicable à tout système complexe (véhicule automobile, être humain...). La mise en place d'une méthode de diagnostic nécessite de définir un espace de représentation regroupant les paramètres les plus pertinents vis-à-vis des modes de fonctionnement connus. Différents techniques peuvent être utilisées pour effectuer le choix d'un sous-ensemble de paramètres pertinents à partir d'un ensemble initial de paramètres. Nous avons proposé l'algorithme OSBS (Optimized Sequential Backward Selection) qui a l'avantage de prendre en compte d'éventuelles combinaisons entre les paramètres pour améliorer la sélection. L'espace de représentation permet d'identifier une structure de classes caractérisant géométriquement les différents modes de fonctionnement de l'ensemble d'apprentissage par des méthodes de classification. Nous avons mis au point deux méthodes de classification.

La première méthode repose sur l'algorithme des fourmis artificielles. Cet algorithme permet de mettre en évidence les classes regroupant les observations similaires. Cela améliore la

classification et permet la détection de nouveaux modes de fonctionnement qui n'étaient pas présents dans la base d'apprentissage.

La deuxième méthode repose sur un modèle probabiliste capable d'associer un degré d'appartenance qui remplace les critères classiques de distance lors de la classification d'une observation. Cette méthode, basée sur les modèles de Markov cachés, définit un indice de fiabilité capable de rejeter les observations mais aussi de détecter l'apparition de nouveaux modes de fonctionnement. Ces deux méthodes ont été mises au point pour répondre à une problématique qui est la disponibilité de mesures correspondant à chaque mode de fonctionnement du système.

Nous avons établi dans la troisième partie de la thèse (**Chapitre 3**) un couplage original entre le diagnostic et le pronostic en définissant le pronostic comme une extension du diagnostic. Le pronostic consiste à fournir des informations sur l'état futur du système à partir d'informations fournies par le diagnostic. Cinq méthodes de pronostic ont été proposées. Les trois premières sont des méthodes probabilistes basées sur les modèles de Markov cachés. L'objectif de ces méthodes est de détecter l'imminence de l'état futur du système en se basant sur une loi qui calcule la probabilité de l'apparition de l'état futur du système. Les deux méthodes restantes sont basées sur le système neuro-flou ANFIS et le neurone neo-flou. Ces deux méthodes permettent de prédire l'évolution des paramètres de surveillance à court et à long termes et, de ce fait, permettent d'estimer le temps restant avant l'apparition d'un défaut.

La dernière partie de la thèse (**Chapitre 4**) consiste à appliquer les méthodes de diagnostic et de pronostic sur une machine asynchrone et sur des roulements à rouleaux. Au niveau de la machine asynchrone, sont considérés des défauts créés au rotor ainsi qu'aux roulements. Les défauts étudiés sont les cassures de barres au rotor et l'usure (vieillesse naturelle) des roulements pour la machine. La machine est alimentée à vide et à plusieurs niveaux de charge par un onduleur de tension. Les mesures accessibles en direct sur la machine sont les courants et tensions ainsi que la vitesse. Deux cas de figure sont étudiés dans cette partie : la première concerne le diagnostic d'une machine asynchrone avec apparition d'un nouveau mode de fonctionnement ; la seconde concerne le diagnostic d'une machine asynchrone avec une connaissance limitée sur chacun de ses modes de fonctionnement. Un espace de représentation est défini à partir d'un vecteur initial de paramètres regroupant un ensemble de paramètres susceptibles d'indiquer la présence d'un défaut. Le nombre élevé de paramètres dans ce vecteur conduit à une dimension trop importante et éventuellement à une mauvaise définition des observations. De ce fait, l'algorithme OSBS est utilisé pour rechercher les paramètres les plus pertinents et pour réduire le temps de calcul. Pour pouvoir détecter la présence d'un défaut sur la machine, nous avons appliqué l'algorithme fournis artificielles et les modèles de Markov cachés. Ces deux méthodes donnent de très bons résultats de diagnostic pour les modes de fonctionnement connus et inconnus. L'option de rejet dans la classification joue un rôle prépondérant dans la détection des nouveaux modes de fonctionnement. Les observations rejetées peuvent alors être intégrées dans l'ensemble d'apprentissage afin d'enrichir la base de données du système de diagnostic.

Au niveau des roulements à rouleaux, la dégradation a été réalisée sur un banc d'essai conçu par la société Rexnord Corp. Les mesures accessibles en direct sur les roulements sont les signaux vibratoires. Un espace de représentation est défini à partir d'un vecteur initial de paramètres regroupant un ensemble de paramètres susceptibles d'indiquer la dégradation des roulements. Pour pouvoir identifier les différents stades de dégradation des roulements, nous

avons appliqué l'algorithme fourmis artificielles. Pour pouvoir détecter l'imminence d'un stade de dégradation sur un roulement, nous avons appliqué deux méthodes de pronostic : le modèle de Markov évolutif et les modèles de Markov cachés. Ces deux méthodes donnent de très bons résultats de pronostic. Toutefois, elles ne permettent pas d'identifier le défaut responsable de la dégradation. Une amélioration est apportée en ajoutant une analyse fréquentielle sur les signaux vibratoires. Ainsi, une prédiction sur le prochain stade de dégradation est réalisée ainsi qu'une indication sur le défaut responsable de la dégradation. Pour pouvoir estimer le temps restant avant l'apparition du prochain stade de dégradation, nous avons appliqué deux autres méthodes de pronostic : le système neuro-flou ANFIS et le neurone neo-flou. Ces deux méthodes permettent de prédire l'évolution des paramètres de surveillance dans le temps, à court et à long termes. Enfin, ces deux méthodes donnent de très bons résultats pour la prédiction de l'évolution de la dégradation des roulements ainsi que sur le temps restant avant leur apparition.

En conclusion, ce travail a permis de monter l'efficacité des méthodes de diagnostic et de pronostic sur la machine asynchrone ainsi que sur les roulements à rouleaux. Au final, une approche globale de diagnostic et de pronostic d'un système d'entraînement électrique a été réalisée.

Les travaux réalisés dans le cadre de cette thèse pourraient donner lieu à des études supplémentaires dont nous présentons quelques points qui nous semblent importants. Ils sont résumés comme suit :

- Maitriser l'erreur de prédiction du système neuro flou ANFIS et celui du neurone neo-flou. En effet, cela reste toujours un défi. Il faudrait mener des études afin d'améliorer les algorithmes d'apprentissage des paramètres d'ANFIS et du neurone neo-flou.
- Utiliser la fusion de données pour améliorer le vecteur de paramètres nécessaire au diagnostic et au pronostic. Un processus de fusion de données vise à l'association, la combinaison, l'intégration et le mélange de données fournies par au moins deux ensembles de données issues d'une même source au cours du temps ou de plusieurs sources, éventuellement au cours du temps. Le but d'un processus de fusion de données serait de fournir un autre ensemble de données de meilleure qualité que si l'on avait utilisé chaque ensemble initial séparément.
- Etendre les méthodes de diagnostic proposées à l'identification des défaillances sur d'autres systèmes d'entraînement électriques notamment les machines synchrones ainsi qu'aux systèmes électromécaniques.
- Prédire la durée de vie en introduisant une prise en compte de la durée d'un état (les modèles semi-markov cachés).
- Etendre le diagnostic au régime transitoire (variation de vitesse, ...).

Bibliographie

- [Albrecht_86] P. F. Albrecht, J. C. Appiarius, and D. K. Sharma, "Assessment of the reliability of motors in utility applications-Updated," *IEEE Transactions on Energy Conversion*, vol. 1, pp. 39-46, 1986.
- [Angeli_04] C. Angeli and A. Chatzinikolaou, "On-line fault detection techniques for technical systems: A survey," *International journal of computer science & application*, vol. 1, pp.12-30, 2004.
- [Babu_08] Babu. B.P, Srinivas J.V.S, Vikranth. B, Premchnad. P, "Fault diagnosis in multi-level inverter system using adaptive back propagation neural network," *IEEE Conference, INDICON 2008*, vol.2, pp.494-498, 11-13 Dec. 2008.
- [Baruah_05] P. Baruah and R. B. Chinnam, "HMMs for diagnostics and prognostics in machining processes," *International Journal of Production Research*, vol.43, no.6, pp.1275-1293, 2005.
- [Bask_79] C. Baskiotis, J. Raymond, and A. Rault, "Parameter identification and discriminant analysis for jet engine mechanical state diagnosis," *IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, vol.18, pp. 648-650, 1979.
- [Beard_71] R. V. Beard "Failure Accommodation in Linear System Through Self Reorganization" PhD thesis, Massachusetts Institute of Technology, 1971.
- [Becraft_93] W.R. Becraft and P.L. Lee, "An integrated neural network/expert system approach for fault diagnosis," *Computers & Chemical Engineering*, vol.17, issue.10, pp.1001-1014, 1993.
- [Benke_12] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, "Fault prognostic of bearings by using support vector data description," in *Proc. IEEE-PHM*, 2012, pp.1-7.
- [Boryczka_09] Urszula Boryczka, "Finding groups in data: Cluster analysis with ants," *Applied Soft Computing*, vol1, pp.61-70, 2009.
- [Bradley_98] P.S. Bradley and O.L, "Mangasarian. Feature selection via concave minimization and support vector machines," In *Machine Learning Proceedings of the Fifteenth International Conference ICML98*, pp. 82-90, 1998.
- [Burg_95] Lee. Burgess, Terry. Shimbel, "What is the prognosis on your maintenance program," *Engineering & Mining Journal*, May.1995.
- [Casimir(I)_03] Casimir. R, Boutleux. E, Clerc, G, "Fault diagnosis in an induction motor by pattern recognition methods," *Diagnostics for Electric Machines, Power Electronics and Drives*, pp.294-299, Aug. 2003
- [Casimir(II)_03] Casimir. R, Boutleux. E, Clerc. G, Chappuis. F, "Broken bars detection in an induction motor by pattern recognition," *Power Tech Conference Proceedings*, vol.2, June. 2003

- [Casimir(III)_03] R. Casimir, *Diagnostic des défauts des machines asynchrones par reconnaissance des formes*, Thèse de l'école Centrale de Lyon, CEGELY, 2003.
- [Chang_94] C. H. Chang, "A fast method for determining electrical and mechanical qualities of induction motors using a PC-aided detector," *IEEE transactions on Instrumentation and Measurement*, Vol. 43, issue 4, pp. 624-629, Aug. 1994.
- [Chen_12] K. Chen, X. Li, F. Wang, T. Wang, and C. Wu, "Bearing fault diagnosis using Wavelet analysis," in *Proc. IEEE Int. QRRMSE*, 2012, pp. 699-702.
- [Chen_11] C. Chen, B. Zhang, G. Vachtsevanos, and M. Orchard, "Machine condition prediction based on adaptive neuro fuzzy and high-order particle filtering," *IEEE Trans. Ind. Electron.*, vol. 58, no. 9, pp. 4353–4364, Sept. 2011.
- [Chinnam_04] R.B. Chinnam and P. Baruah, "A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems," *Materials and Product Technology*, vol.20, pp.166–179, 2004.
- [Chiu_94] S. Chiu, "Fuzzy Model Identification based on cluster estimation," *Journal of Intelligent Fuzzy Systems*, vol.2, pp.267–278, 1994.
- [Chow_84] E. Chow and A.S. Willsky, "Analytical redundancy and the design of robust failure detection systems," *Automatic Control, IEEE Transactions on*, vol.29, pp.603–614, 1984.
- [Ding_08] Steven X. Ding. *Model-Based Fault Diagnosis Techniques Design Schemes, Algorithms and Tools*. Springer, 2008.
- [Did_04] Didier. G, *Modélisation et diagnostic de la machine asynchrone en présence de défaillances*, Thèse de doctorat, Université Nancy 1, France, 2004.
- [Doak_92] J. Doak, *An Evaluation of Feature Selection Methods and Their Application to Computer Security*, University of California, Computer Science, 1992.
- [Dorigo_97] M. Dorigo and L.M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *Evolutionary Computation, IEEE Transactions on*, vol1, pp.53–66, 1997.
- [Dubuisson_93] Bernard Dubuisson and Mylène Masson, "A statistical decision rule with incomplete knowledge about classes," *Pattern Recognition*, vol.26, pp.155 – 165, 1993.
- [Fisher_36] R. A. FISHER, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol.7, pp.179–188, 1936.
- [Fla_93] Flandrin .P, *Temps - Fréquence, Traité des nouvelles technologies*, Hermès, 1993.
- [Forgy_65] E. W. Forgy, "Cluster analysis of multivariate data: efficiency vs interpretability of classifications," *Biometrics*, vol.21, pp.768–769, 1965.
- [Frank_94a] P. M. Frank, "On-line fault detection in uncertain nonlinear systems using diagnostic observers- a survey," *Int. Journal of Systems Science*, vol.25, pp.2129–2154, 1994.
- [Frank_96] P. M. Frank, "Analytical and Qualitative Model-based Fault Diagnosis - a survey and some new results," *European Journal of Control*, vol.2, pp.6–28, 1996.
- [Frosini_10] L. Frosini, and E. Bassi, "Stator current and motor efficiency as indicators for different types of bearing faults in induction motors," *IEEE Trans. Ind. Electron.*, vol. 57, no. 1, pp. 244 –251, Jan. 2010.

- [Gouriveau_07] Rafael Gouriveau, Mohamed El Koujok, and Nouredine Zerhouni, "Spécification d'un système neuro-flou de prédiction de défaillances à moyen terme," Nîmes, France, 2007.
- [Helle_06] A. Helle, "*Prognostics for Industrial Machinery Availability*," *Maintenance, Condition Monitoring and Diagnostics*, 2006.
- [Hoskins_91] J. C. Hoskins, K. M. Kaliyur, and David M. Himmelblau, "Fault diagnosis in complex chemical plants using artificial neural networks," *AIChE Journal*, vol.37, pp.137–141, 1991.
- [Holroyd_05] Trevor J. Holroyd, "The application of ae in condition monitoring," In *Proceedings of International Conference on Condition Monitoring*, Cambridge, 2005.
- [Hountalas_00] Dimitrios.T Hountalas, "Prediction of marine diesel engine performance under fault conditions," *Applied Thermal Engineering*, vol.20, pp.1753 – 1783, 2000.
- [Hoyland_94] Arnljot Høyland and Marvin Rausand, *System Reliability Theory: Models and Statistical Methods*, Wiley, 1994.
- [Hu_64] M.J.C. Hu, *Application of the Adaline System to Weather Forecasting*, Department of Electrical Engineering, Stanford University, 1964.
- [Ibrahim_08] A. Ibrahim, M. El Badaoui, F. Guillet, and F. Bonnardot, "A New Bearing Fault Detection Method in Induction Machines Based on Instantaneous Power Factor," *IEEE Trans. Ind. Electron*, vol.55, pp.4252 -4259, Dec. 2008.
- [Immo_09] F. Immovilli, M. Cocconcelli, A. Bellini, and R. Rubini, "Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals," *IEEE Trans. Ind. Electron*, vol. 56, pp.4710 – 4717, Nov. 2009.
- [Immo_13] F. Immovilli, C. Bianchini, M. Cocconcelli, A. Bellini, and R. Rubini, "Bearing Fault Model for Induction Motor with Externally Induced Vibration," *IEEE Trans. Ind. Electron*, vol.60, pp.3408-3418, Aug. 2013.
- [Isermann_84] Rolf Isermann, "Process fault detection based on modeling and estimation methods - a survey," *Automatica*, vol.20, pp.387–404, July 1984.
- [Jang_93] J.S.R. Jang, "Anfis: adaptive-network-based fuzzy inference system," *Systems, Man and Cybernetics, IEEE Transactions on*, vol23, pp.665–685, 1993.
- [Jelinek_98] Frederick Jelinek, *Statistical Methods for Speech Recognition*, The MIT Press, January. 1998.
- [John_94] George H. John, Ron Kohavi, and Karl Pflieger, "Irrelevant features and the subset selection problem," In *Machine Learning*, pp.121–129, 1994.
- [Juang_08] Chia-Feng Juang, Chun-Ming Lu, Chiang Lo, and Chi-Yen Wang, "Ant colony optimization algorithm for fuzzy controller design and its fpga implementation," *Industrial Electronics, IEEE Transactions on*, vol.55, pp.1453–1462, 2008.
- [Jain_99] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Vol.*31, pp.264–323, September.1999.
- [Kar_06] Chinmaya Kar and A.R. Mohanty, "Monitoring gear vibrations through motor current signature analysis and wavelet transform," *Mechanical Systems and Signal Processing*, vol.20, pp.158– 187, 2006.
- [Kim_97] Daijin Kim and Chulhyun Kim, "Forecasting time series with genetic fuzzy

- predictor ensemble," *Fuzzy Systems, IEEE Transactions on*, vol.5, pp.523–535, 1997.
- [Kohavi_97] Ron Kohavi and George H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol.97, pp.273–324, 1997.
- [Kresta_91] James V. Kresta, John F. Macgregor, and Thomas E. Marlin, "Multivariate statistical monitoring of process operating performance," *The Canadian Journal of Chemical Engineering*, vol.69, pp.35–47, February.1991.
- [Krogh_93] Anders Krogh, Michael Brown, I. S Mian, Kimmen Sjolander, and David Haussler, "Hidden markov models in computational biology: Applications to protein modeling," *Technical report*, Santa Cruz, CA, USA, 1993.
- [Kudo_00] M. Kudo, J. Sklansky, "Comparison of algorithms to select features for pattern classifiers". *Pattern Recognition*, vol.33, pp.25-41, 2000.
- [Laerhoven_01] Kristof Van Laerhoven, Kofi Aidoo, and Steven Lowette, "Real-time analysis of data from many sensors with neural networks," In *ISWC*, pp.115–123, 2001.
- [Lebaroud_08] A. Lebaroud and G. Clerc, "Classification of induction machine faults by optimal time-frequency representations," *Industrial Electronics, IEEE Transactions on*, vol.55, pp.4290–4298, 2008.
- [Lebold_03] M. Lebold, K. Reichard, and D. Boylan, "Utilizing dcom in an open system architecture framework for machinery monitoring and diagnostics," In *IEEE Aerospace Conference*, vol.3, pp.1227–1236, 2003.
- [Liu_09] Jianbo Liu, Dragan Djurdjanovic, Kenneth A. Marko, and Jun Ni, "A divide and conquer approach to anomaly detection, localization and diagnosis," *Mechanical Systems and Signal Processing*, vol.23, pp.2488 – 2499, 2009.
- [Li_99] N Tandon and A Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, vol.32, pp.469 – 480, 1999.
- [Li_00] B. Li, M.Y. Chow, Y. Tipsuwan, and J. Hung, "Neural network based motor rolling bearing fault diagnosis," *IEEE Trans. Ind. Electron*, vol.47, pp.1060 – 1069, Oct. 2000.
- [Liu_10] Jiao-Min Liu, Jian-Li Zhao, Li Li, Ya-Ning Wang, "Mechanical fault diagnosis for L-V circuit breakers based on energy spectrum entropy of wavelet packet and Naive Bayesian classifier," *Machine Learning and Cybernetics (ICMLC)*, vol.3, pp.1059,1064, July. 2010.
- [Loparo_00] K.A. Loparo, M.L. Adams, W. Lin, M.F. Abdel-Magied, and N. Afshari, "Fault detection and diagnosis of rotating machinery," *IEEE Trans. Ind. Electron*, vol.47, pp.1005 -1014, Oct. 2000.
- [MacGregor_95] J. F. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol.3, pp.403–414, March. 1995.
- [Mackey_77] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol.197, pp.287–289, July.1977.
- [Medjaher_12] Kamal Medjaher, Fatih Camci, and Nouredine Zerhouni, "Feature extraction and evaluation for Health Assessment and Failure prognostics," In *First European Conference of the Prognostics and Health Management Society*, pp. 111–116, 2012.
- [Mahamad_10] A. K. Mahamad, S. Saon, and T. Hiyama, "Predicting remaining useful life of rotating machinery based artificial neural network," *Computers &*

- Mathematics with Applications*, vol. 60, pp. 1078–1087, Aug. 2010.
- [Mba_06] David Mba and Raj B K N Rao, "Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines; bearings, pumps, gearboxes, engines and rotating structures," *The Shock and vibration digest*, vol.38, pp.3–16, 2006.
- [Murthy_03] D N Prabhakar Murthy, *Case Studies in Reliability and Maintenance*, vol.20, Wiley Series in Probability and Statistics, 2003.
- [Nid_11] <http://membres.multimania.fr/fourmiliere/vue%20BC.jpg>
- [OCAK_03] H OCAK, *Fault Detection, Diagnosis and Prognosis of Rolling Element Bearings: Frequency Domain Methods and Hidden Markov Modeling*, PhD thesis, Case Western Reserve University, 2003.
- [Ondel(II)_06] O. Ondel, *Diagnostic par reconnaissance des formes: application à un ensemble convertisseur-machine asynchrone*, Thèse de l'école centrale de Lyon, CEGELY, 2006.
- [Ondel(I)_06] Ondel. O, Boutleux. E, Clerc. G, "A method to detect broken bars in induction machine using pattern recognition techniques," *Industry Applications, IEEE Transactions on*, vol.42, pp.916,923, 2006.
- [Paclík_06] P. Paclík, J. Novovičková, and R. P W Duin, "Building road-sign classifiers using a trainable similarity measure," *Intelligent Transportation Systems, IEEE Transactions on*, vol.7, pp.309–321, 2006.
- [Patton_94] R.J. Patton and J. Chen, "Review of parity space approaches to fault diagnosis for aerospace systems," *Journal of Guidance, Control and Dynamics*, vol.17, pp.278–285, March.1994.
- [Patton_97] R.J. Patton and J. Chen, "Observer-based fault detection and isolation: Robustness and applications," *Control Engineering Practice*, vol.5, pp.671–682, 1997.
- [Prieto_13] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing faults detection by a novel condition monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron*, vol. 60, pp.3398 –3407, Aug. 2013.
- [Pusey_07] PUSEY Henry, "Turbo machinery condition monitoring and failure prognosis," *Sound & vibration*, vol.41, pp.10–15, 2007.
- [Qiu_06] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *Journal of Sound and Vibration*, vol.289, pp.1066– 1090, Feb. 2006.
- [Rabiner_89] Lawrence R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," In *Proceedings of the IEEE*, pp.257–286, 1989.
- [Roemer_00] M.J. Roemer and G.J. Kacprzyński, "Advanced diagnostics and prognostics for gas turbine engine risk assessment," In *Aerospace Conference Proceedings*, vol.6, pp.345 –353, 2000.
- [Rich_87] Steven H. Rich, "Model-based reasoning in diagnostic expert systems for chemical process plants," *Computers & Chemical Engineering*, vol.11, pp.111–122, 1987.
- [Reng_00] Raghunathan Rengaswamy and Venkat Venkatasubramanian, "A fast training neural network and its updation for incipient fault detection and

- diagnosis," *Computers & Chemical Engineering*, vol.24, pp.431– 437, 2000.
- [Repo_85] I. C. Report, "Report of large motor reliability survey of industrial and commercial installation, Part I and Part II," *IEEE Transactions on Industry Applications*, vol. 21, pp. 853-872, 1985.
- [Roemer_05] M. Roemer, C. Byington, G. Kacprzyński, G. Vachtsevanos, "An Overview of Selected Prognostic Technologies with Reference to an Integrated PHM Architecture," In *IEEE Aerospace Conf*, 2005.
- [Sam_08] B. Samanta and C. Nataraj, "Prognostics of machine condition using soft computing," *Robotics and Computer-Integrated Manufacturing*, vol.24, pp.816 – 823, 2008.
- [Schoen_95] R. R. Schoen, T. G. Habetler, F. Kamran, R. G. Bartheld, "Motor bearing damage detection using stator current monitoring," *IEEE Transaction on Industry Applications*, vol.31, pp.1274–1279, 1995.
- [Serya_10] O.R. Seryasat, M.A. Shoorehdeli, F. Honarvar, and A. Rahmani, "Multi-fault diagnosis of ball bearing based on features extracted from time-domain and multi-class support vector machine (MSVM)," in *Proc. IEEE SMC*, 2010, pp. 4300-4303.
- [Soura_00] Sourabh Dash, Venkat Venkatasubramanian, "Challenges in the industrial applications of fault diagnostic systems," *Computers & Chemical Engineering*, Volume 24, Issues 2–7, 15 July 2000, pp 785-791.
- [Soualhi_11] Soualhi. A, Clerc. G, Razik. H, Ondel. O, "Detection of induction motor faults by an improved artificial ant clustering," *IECON 2011*, pp.3446-3451, Nov. 2011.
- [Stack_04] J. Stack, T. Habetler, and R. Harley, "Bearing fault detection via autoregressive stator current modeling," *IEEE Trans. Ind. Appl*, vol. 40, pp. 740 – 747, May-June. 2004.
- [Starr_97] A.G. Starr, "A structured approach to the selection of condition based maintenance," In *Factory 2000 - The Technology Exploitation Process*, pp.131 –138, April. 1997.
- [Strangas_11] E. Strangas, S. Aviyente, J. Neely, and S. Zaidi, "Improving the reliability of electrical drives through failure prognosis," in *Proc. IEEE SDEMPED*, 2011, pp. 172 –178.
- [Svozil_97] Daniel Svozil, Vladimír Kvasnicka, and Jiří Pospichal. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol.39, pp.43 – 62, 1997.
- [Takagi_85] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *Systems, Man and Cybernetics, IEEE Transactions on*, vol.15, pp.116–132, 1985.
- [Tandon_99] N. Tandon and A. Choudhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," *Tribology International*, vol.32, pp.469 – 480, 1999.
- [Thurston_01] M.G. Thurston, "An open standard for web-based condition-based maintenance systems," In *IEEE Systems Readiness Technology Conference*, pp.401 –415, 2001.
- [Tob_11] D.A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "Hidden markov models for failure diagnostic and prognostic," In *Prognostics and System Health Management Conference, 2011*, pp.1–8.

- [Tsoumas_05] I. Tsoumas, E. Mitronikas, G. Georgoulas, and A. Safacas, "A comparative study of induction motor current signature analysis techniques for mechanical faults detection," In *Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 1–6, Sept. 2005.
- [Uchino_97] Eiji Uchino and Takeshi Yamakawa, "Soft computing based signal prediction, restoration, and filtering," In *Intelligent Hybrid Systems*, pp.331–351, 1997.
- [Ungar_90] L.H. Ungar, B.A. Powell, and S.N. Kamens, "Adaptive networks for fault diagnosis and process control," *Computers & Chemical Engineering*, vol.14, pp. 561–572, 1990.
- [Ven(III)_03] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, and Kewen Yin, "A review of process fault detection and diagnosis: Part iii: Process history based methods," *Computers & Chemical Engineering*, pp.327–346, 2003.
- [Ven(II)_03] Venkatasubramanian Venkat, Rengaswamy R, and Kavuri S.N, "A review of process fault detection and diagnosis - part ii: Qualitative models and search strategies," *Computers and Chemical Engineering*, vol.27, pp.313–326, 2003.
- [Ven(I)_03] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, and Surya N. Kavuri, "A review of process fault detection and diagnosis: Part i: Quantitative model-based methods," *Computers & Chemical Engineering*, vol.27, pp.293 – 311, 2003.
- [Vacht_06] M. Roemer, A. Hess, G.J. Vachtsevanos, F.L. Lewis and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, 2006.
- [Vacht_01] G. Vachtsevanos and P. Wang, "Fault prognosis using dynamic wavelet neural networks," In *IEEE Systems Readiness Technology Conference*, pp.857 –870, 2001.
- [Wise_96] Barry M. Wise and Neal B. Gallagher, "The process chemometrics approach to process monitoring and fault detection," *Journal of Process Control*, vol.6, pp.329 – 348, 1996.
- [Walter_04] T. J. Walter and H. Lee, "Development of a smart wireless sensor for predicting bearing remaining useful life," In *Proceedings of the 58th Meeting of the society for machinery failure prevention technology*, 2004.
- [Wang_96] W.J. Wang and P.D. McFadden, "Application of wavelets to gearbox vibration signals for fault detection," *Journal of Sound and Vibration*, vol.192, pp.927 – 939, 1996.
- [Wang_07] Shiaokai Wang, William Pentney, Ana-Maria Popescu, Tanzeem Choudhury, and Matthai Philipose, "Common sense based joint training of human activity recognizers," In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp.2237–2242, 2007.
- [Wang_02] Hongzhou Wang, "A survey of maintenance policies of deteriorating systems," *European Journal of Operational Research*, pp.469–489, 2002.
- [Wang_04] Jingen Wang, Lin Shang, Shifu Chen, Yanfei Wang, "Application of fuzzy classification by evolutionary neural network in incipient fault detection of power transformer," *IEEE International Joint Conference on Neural Networks*, vol.3, pp.2279-2283, July. 2004.
- [Watanabe_89] Kajiro Watanabe, Ichiro Matsuura, Masahiro Abe, Makoto Kubota, and D. M.

- Himmelblau, "Incipient fault diagnosis of chemical processes via artificial neural networks," *AICHE Journal*, vol.35, pp.1803–1812, 1989.
- [Widrow_85] Bernard Widrow and Samuel D. Stearns, *Adaptive signal processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1985.
- [Wun_90] J. Wunnenberg, *Observer Based Fault Detection in Dynamic Systems*, PhD thesis, VDI-Fortschrittsber., 1990.
- [Xue_09] Zhihong Xue, Xiaoyun Sun, Yongchun Liang, "Application of Data Mining Technology Based on FRS and SVM for Fault Identification of Power Transformer," *Artificial Intelligence and Computational Intelligence*, vol.2, pp.452-455, Nov. 2009.
- [Yamak_99] T. Yamakawa and T. Miki, "Analog implementation of neo-fuzzy neuron and its on-board learning," In *Computational Intelligence and Applications*, pp. 144–149, 1999.
- [Yamak_92] T. Yamakawa, E. Uchino, T. Miki, and H. Kusanagi, "A neo fuzzy neuron and its applications to system identification and prediction of the system behaviour," In *Proc. 2nd Int. Conf. Fuzzy Logic Neural Networks*, pp.477–483, 1992.
- [Yam_01] R. C. M. Yam, P. W. Tse, L. Li, and P. Tu, "Intelligent predictive decision support system for condition-based maintenance," *The International Journal of Advanced Manufacturing Technology*, vol.17, pp.383–391, 2001.
- [YANG_04] Q YANG, *Model-Based and Data Driven Fault Diagnosis Methods With Applications to Process Monitoring*, PhD thesis, Electrical Engineering and Computer Sciences, Case Western Reserve University, Ohio, USA, 2004.
- [Yazidi_11] A. Yazidi, H. Henao, G.A. Capolino, F. Betin, and F. Filippetti, "A web-based remote laboratory for monitoring and diagnosis of ac electrical machines," *IEEE Trans. Ind. Electron*, vol. 58, pp.4950 –4959, Oct. 2011.
- [Ypma_99] Alexander Ypma, David M. J. Tax, and Robert P. W. Duin, "Robust machine fault detection with independent component analysis and support vector data description," In *IEEE Workshop on Networks for Signal Processing*, pp.67–76, 1999.
- [Ypma_02] Alexander Ypma, Er Ypma, and Tom Heskes, "Categorization of web pages and user clustering with mixtures of hidden markov models," In *proceeding of WEBKDD 2002*, pp.31–43, 2002.
- [Zaidi_11] S. Zaidi, S. Aviyente, M. Salman, K.K. Shin, and E. Strangas, "Prognosis of gear failures in dc starter motors using hidden markov models," *IEEE Trans. Ind. Electron*, vol. 58, pp. 1695 –1706, May. 2011.
- [Zhan_10] Huaqun Zhan, "Application of rough set and support vector machine in fault diagnosis of power electronic circuit," *Information Management and Engineering (ICIME)*, pp.289,292, April. 2010.
- [Zhang_11] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. Orchard, and G. Vachtsevanos, "A probabilistic fault detection approach: Application to bearing fault detection," *IEEE Trans. Ind. Electron*, vol. 58, pp. 2011–2018, May. 2011.
- [Zhen_09] Yang Zhengyou, Peng Tao, Li Jianbao, Yang Huibin, Jiang Haiyan, "Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Transform and Support Vector Machine," *Measuring Technology and Mechatronics Automation*, vol.1, pp.650-653, April. 2009.

[ISO13381]

ISO. Condition monitoring and diagnostics of machines - prognostics - Part 1: General guidelines. Int. Standard ISO13381-1, 2004.