

***ÉCOLE DOCTORALE Mathématiques, Informatique,
Physique Théorique et Ingénierie des Systèmes (MIPTIS)***

LABORATOIRE PRISME

THÈSE présentée par :
Bernard Robles

soutenue le : **18 décembre 2013**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline : **SCIENCES ET TECHNOLOGIES INDUSTRIELLES**

**Étude de la pertinence des paramètres stochastiques
sur des Modèles de Markov Cachés.**

THÈSE dirigée par :

Frédéric Kratz

Professeur des Universités, ENSI de Bourges

RAPPORTEURS :

Zohra Cherfi Boulanger

Professeur des Universités, Université de Technologie de
Compiègne

Abdessamad Kobi

Professeur des Universités, Université d'Angers

JURY :

Christian Olivier

Professeur des Universités, Université de Poitiers, Président
du jury

Zohra Cherfi Boulanger

Professeur des Universités, Université de Technologie de
Compiègne

Abdessamad Kobi

Professeur des Universités, Université d'Angers

Gilles Mourioux

Maître de Conférences, Université de Limoges

Yann Chamaillard

Professeur des Universités, Université d'Orléans

Frédéric Kratz

Professeur des Universités, ENSI de Bourges

Manuel Avila

Maître de Conférences, Université d'Orléans

Florent Duculty

Maître de Conférences, Université d'Orléans

Remerciements

Je tiens à remercier en tout premier lieu M. Frédéric Kratz d’avoir accepté, il y a quelques années de cela, que je débute une thèse sous sa direction. J’ai ainsi découvert le monde de la recherche au sein de l’équipe IRAuS du laboratoire PRISME.

Je remercie M^{me} Zohra Cherfi Boulanger et M. Abdessamad Kobi d’avoir accepté de rapporter sur mon travail de thèse. Leur intérêt pour ce travail et leurs remarques pertinentes m’ont poussé à aller plus loin dans certaines réflexions.

M. Christian Olivier, M. Gilles Mourioux et M. Yann Chamaillard d’avoir acceptés de faire parti de mon jury de thèse.

Merci à M. Manuel Avila et M. Florent Duculty qui ont co-dirigé ce travail de thèse, ils ont cadrés mon travail tout au long de ma thèse et ont su me diriger sur ce long chemin sinueux.

M. Stéphane Bégot pour les corrections à « quatre mains » que nous avons réalisé durant de longues heures.

M. Manuel Avila, M. Florent Duculty, M. Pascal Vrignat et M. Stéphane Bégot, pour leurs relectures et leurs corrections pertinentes.

M. Pascal Vrignat qui m’a transmis toutes les données récoltées sur le terrain ainsi que ces données de recherche.

J’ai découvert le plaisir d’enseigner en parallèle de mon travail d’informaticien. Je remercie M. Manuel Avila qui m’a fait confiance et m’a donné la responsabilité des divers cours magistraux en informatique, travaux dirigés et travaux pratiques que j’ai assuré durant ces années.

Merci aussi à tous les personnels de l’IUT de l’Indre qui m’ont encouragé pendant toutes ces années. Merci à Nicole Stride qui m’a montré comment motiver les étudiants de LPSAR pendant 4 heures de CM...

Merci à M. Jean-Christophe Bardet, directeur de l’IUT de l’Indre et M. Gérard Guillaume responsable administratif, qui m’ont autorisé à participer aux diverses conférences, colloques et réunions recherche, sur mon temps de travail.

Merci à Mme Pinkney pour son aide précieuse dans la rédaction de communications en

anglais.

Merci à Fanny, pour la correction des nombreuses « erreurs de frappe » sur ce manuscrit et aussi pour sa patience de m'avoir supporté à travailler durant de nombreux soirs et week-end.

Table des matières

Avant-propos	1
1 Principes généraux, objectifs et politique de maintenance industrielle	5
1.1 Introduction	7
1.2 Présentation de la maintenance dans le domaine industrielle	9
1.2.1 Définitions normatives	10
1.2.2 Politique de maintenance	10
1.2.3 Objectifs d'une politique de maintenance	12
1.2.4 Typologies des actions de maintenance	13
1.3 La Gestion de la Maintenance Assistée par Ordinateur : GMAO	14
1.4 Évolution des politiques de maintenance industrielle	16
1.5 Utilisation de Modèles de Markov Cachés dans le cadre d'une politique de maintenance industrielle	21
1.5.1 Anticiper une situation de panne	21
1.5.2 Utilisation des MMC dans la détection de panne	22
1.6 Conclusion et objectifs de cette étude	24
2 Approches classiques de mesures de pertinence dans la chaîne de modélisation	27
2.1 Introduction	29
2.2 Problématique générale	30
2.3 Analyse de sensibilité d'un modèle	30
2.3.1 Comment réaliser une analyse de sensibilité ?	31
2.3.2 Méthodes d'analyse de sensibilité	31
2.3.3 Les enjeux de la prise en compte des incertitudes	35
2.3.4 Incertitudes de conception	35
2.4 Méthodes de mesures de pertinence de modèles	38
2.4.1 Introduction	38
2.4.2 Concepts de base	39
2.4.3 Méthodes avec score d'intérêt	40
2.4.4 Méthodes par apprentissage	48
2.4.5 Méthodes statistiques	49
2.4.6 Comparaison de deux modèles par corrélation linéaire de Bravais-Pearson	52
2.5 Mesure du caractère stochastique d'un modèle	53
2.6 Discussion sur les méthodes de sélection de modèles	55
2.7 Conclusion	57

3	Évaluation de modèles par une approche markovienne	59
3.1	Introduction	61
3.2	Attentes théoriques	61
3.3	Approches classiques	62
3.3.1	Filtres particuliers	62
3.3.2	Filtre de Kalman étendu	62
3.3.3	Réseau de Petri stochastique	63
3.4	Approche markovienne	64
3.4.1	Chaînes de Markov	65
3.4.2	Modèle de Markov Caché discret	67
3.4.3	Complexité de l'évaluation d'un MMC	68
3.4.4	Topologies des modèles étudiés	69
3.4.5	Apprentissage	69
3.4.6	Symboles manquants	72
3.5	Modèle de synthèse	72
3.5.1	Caractère stochastique du modèle de synthèse	77
3.5.2	Processus d'évaluation	78
3.5.3	Processus de génération des symboles	81
3.5.4	Détermination de la loi statistique	81
3.6	Conclusion	86
4	Expérimentations et résultats	87
4.1	Résultats de simulation	89
4.1.1	Paramètres du modèle de simulation	89
4.1.2	Pertinence des observations de simulation	90
4.1.3	Évolution de la modélisation	92
4.1.4	Pertinence de l'architecture des modèles de simulation	98
4.1.5	Résultats avec les autres topologies	104
4.2	Résultats des études réelles	107
4.2.1	Présentation des environnements d'étude	107
4.2.2	Modélisation des processus industriels	107
4.2.3	Description des MMC utilisés	108
4.2.4	Pertinence des observations empiriques	110
4.2.5	Évolution de la modélisation	119
4.2.6	Confrontation du modèle de synthèse avec l'application réelle	125
4.2.7	Conclusion	125
4.3	Comparaison du modèle de synthèse avec le cas industriel	126
4.3.1	Ajustement du modèle de synthèse	126
4.3.2	Résultats après ajustement du modèle de synthèse	129
4.4	Conclusion	130
	Conclusion générale et perspectives	131

A Algorithmes	139
A.1 Apprentissage Interactive Dichotomizer 3 (ID3)	141
A.2 Compression de Lempel-Ziv-Welch	142
A.3 Itératif d'Espace de Version (IVSA)	143
A.4 Viterbi	145
A.5 Algorithme de calcul de l'entropie moyenne	146
B Principales lois de probabilités	149
B.1 Loi Normale (Laplace – Gauss)	149
B.2 Loi Uniforme	150
B.3 Loi Exponentielle	151
C Notions de mathématiques utilisées	153
Bibliographie	155
Communications	171
Liste des Figures	175
Liste des tableaux	178

Avant-propos

« Le but des modèles n'est pas de représenter les données mais de préciser les questions. »

Samuel Karlin, 1924–2007

« Un but majeur de la science est la compréhension et pas seulement la prédiction. »

Willard van Orman (La poursuite de la Vérité), 1908–2000

La notion de modèles de Markov cachés MMC¹ ou HMM², apparaît dans de nombreux domaines comme la théorie de l'information, la physique statistique, la compression de données, la bio-informatique (segmentation de gènes ou protéines afin de comparer des séquences d'ADN), la reconnaissance vocale, la maintenance industrielle, etc. Cette modélisation est une approche prometteuse dans certains domaines où l'on cherche à conceptualiser des données partiellement manquantes, comme la reconnaissance de caractères dans l'écriture manuscrite ou la reconnaissance d'empreintes digitales.

Nous proposons dans cette thèse des outils permettant de choisir ou d'optimiser des modélisations markoviennes utilisées notamment dans le cadre d'une politique de maintenance industrielle. Nous définirons alors des critères d'évaluation afin de mesurer la pertinence de MMC.

Pour rester compétitives tout en pérennisant leur productivité, les entreprises doivent maintenir leurs équipements de production dans un état de bon fonctionnement. En effet, un simple dysfonctionnement dans un processus de fabrication peut s'avérer catastrophique pour tout le reste de la chaîne de production et ainsi engendrer des pertes importantes de bénéfice pour l'entreprise, voire dans le pire des cas, des périodes de chômage technique pour les employés. Sous la pression de tels facteurs socio-économiques, de nombreuses entreprises doivent non seulement investir continuellement dans leur politique de maintenance industrielle mais également dans l'innovation technologique afin d'améliorer la qualité des interventions de maintenance tout en assurant un retour sur expérience.

Nos travaux s'inscrivent dans la continuité de ceux menés dans la thèse de Pascal Vri-gnat : « Génération d'indicateurs de maintenance par une approche semi-paramétrique et par une approche markovienne » soutenue le 14 octobre 2010 [177]. Dans ces précédents travaux, l'auteur s'est concentré sur l'estimation globale du niveau de dégradation d'un processus industriel. L'auteur a démontré qu'à partir d'observations empiriques, il était possible d'effectuer une modélisation de ces niveaux de dégradation.

Dans ce manuscrit, nous nous concentrerons sur l'étude d'un ensemble de modèles définis dans la thèse de Pascal Vri-gnat [177]. Nous étudierons en effet la pertinence de ces différentes architectures. Pour cela, nous parlerons de topologies, d'observations, de critères d'échantillonnage, d'algorithmes d'apprentissage et de décodage ou de critères de distributions des observations.

Ainsi, dans le cadre d'une politique de maintenance préventive, notre démarche tente d'améliorer l'aide à la décision des experts. Ce manuscrit s'articule autour d'une présentation des stratégies modernes de maintenance industrielle. Nous en définirons ensuite leurs objectifs principaux. Pour étayer notre propos, un état de l'art des différentes approches visant à étudier la pertinence des paramètres de modélisation sera présenté dans un premier temps. Nous exposerons par la suite les éléments de modèle(s) (architecture) les plus appropriés. Pour terminer, les résultats issus des données simulées seront comparés à ceux provenant de deux applications concrètes de maintenance industrielle.

1. Modèle de Markov Caché
2. Hidden Markov Model

La plupart des travaux publiés ont pour objectif d’optimiser la qualité du diagnostic suivant la stratégie de fiabilisation appliquée. Notre démarche se différencie des approches habituelles de la littérature : utilisation des *réseaux de Petri* (Labadi et al. [105]) ou des *filtres particuliers* (Li et al. [109]) ou encore des *réseaux Bayésiens* (Przytula et Thompson [133]). En effet, la démarche utilisée dans ces travaux consiste à modéliser les dégradations d’un processus en utilisant des MMC. Cet outil nous est apparu comme particulièrement adapté à l’identification de signatures de l’état d’un système. Afin d’être décorrélé d’a priori, un processus de synthèse a été créé pour produire des événements simulés associés à un niveau de dégradation connu. C’est en comparant le comportement de ce modèle à ceux définis empiriquement dans le cadre industriel que nous arriverons à présenter des préconisations concernant :

- la pertinence des observations collectées (voire à collecter) ;
- les critères d’échantillonnage de ces observations ;
- le nombre d’observations nécessaire et suffisant pour décrire le modèle de manière optimale ;
- la topologie d’un MMC ;
- l’usage d’algorithmes d’apprentissage et de décodage.

Les informations utilisées comme observations des modèles empiriques proviennent des activités de maintenance issues de divers secteurs de l’industrie. Le but de notre démarche est d’établir des indicateurs de niveaux de dégradation globale d’un processus quelconque. Ainsi, on peut espérer valider un modèle indépendant de caractéristiques trop liées à un processus industriel particulier.

Ceci est d’autant plus vrai qu’une telle stratégie peut s’appliquer à des domaines très différents. Il pourrait s’agir, par exemple, d’un routeur indiquant à l’ingénieur informatique la charge probable de son réseau. Lors d’une augmentation du niveau de collisions des paquets IP, cet outil lui permettrait alors d’alléger cette charge avant qu’un arrêt complet du système ne survienne.

A voir ces applications potentielles, on imagine bien que les modèles doivent se construire autour de caractéristiques globales et non particulières. En suivant cette démarche, nous pouvons espérer obtenir une modélisation beaucoup plus transposable à n’importe quel processus.

Dans un premier chapitre, nous détaillerons les différents aspects de la maintenance industrielle, ses objectifs principaux, ses politiques de mise en œuvre classiques ainsi que ses évolutions. Nous présenterons la problématique générale ainsi que les objectifs de cette étude. Comme nous l’avons déjà précisé, cette thèse se situe dans la continuité de la thèse de Pascal Vignat [177] sur la génération d’indicateurs de maintenance par une approche markovienne.

Nous présenterons dans un second chapitre, les différentes approches d'analyse de modèles. Nous définirons les différents types d'analyse de sensibilité dans la conception d'un modèle, les incertitudes dans la chaîne de modélisation ainsi que leurs impacts sur les modèles. Cette analyse devrait nous permettre de déterminer les caractéristiques des modèles les plus sensibles. Ensuite, nous réaliserons une étude sur les différentes approches d'évaluation de pertinence d'un modèle.

Dans un troisième chapitre, nous définirons nos attentes théoriques, en ce qui concerne la pertinence des observations, l'évolution de la modélisation et la pertinence de l'architecture des modèles étudiés. Ensuite, nous présenterons notre approche markovienne et ces différents aspects et propriétés, ainsi que d'autres approches pouvant répondre à notre problématique. Nous présenterons également la complexité du calcul de la séquence la plus probable d'états cachés ayant conduit à la production d'une séquence d'observations donnée. En effet, le nombre de chemins possibles pour générer une telle séquence est de l'ordre de N^T (N étant le nombre d'états d'un MMC et T la longueur d'une séquence d'observations). L'approche directe n'est pas acceptable sachant que pour notre cas $N = 4$ et $T = 1000$, le calcul demanderait alors approximativement 10^{600} opérations. Afin que notre modèle soit réactif, il est quand même préférable de mettre en place des alternatives au calcul direct. Enfin, le processus de synthèse sera introduit et ses caractéristiques seront comparées à celles des architectures empiriques mises en œuvre à partir de données industrielles.

Le quatrième chapitre sera consacré à l'analyse des résultats issus de notre modèle de simulation représentant le fonctionnement d'une GMAO industrielle. Nous présenterons dans un premier temps les caractéristiques de ce modèle, puis les résultats sur les trois points suivants : la pertinence des observations, la pertinence de l'architecture des différents modèles utilisés et l'évolution de la modélisation afin d'en améliorer son architecture. Nous utiliserons notamment le principe du maximum de vraisemblance pour évaluer la pertinence des paramètres d'un ensemble de MMC. Ce principe permet en général de déterminer l'appartenance d'un échantillon de n observations indépendantes à un ensemble de distributions données. Nous appliquerons la notion d'entropie de Shannon, un second critère communément utilisé dans la sélection de modèle. Ce critère est utilisé dans de nombreux domaines. Nous le trouvons dans le langage courant comme synonyme de « désorganisation ». En fonction de la minimalité ou de la maximalité de l'entropie, nous évaluerons la pertinence des séquences de symboles en terme de quantité d'information. Nous réaliserons ensuite deux études expérimentales, issues de cas concrets de processus industriels. Nous débuterons par une présentation de ces environnements expérimentaux puis nous présenterons nos résultats sur les trois points précédemment cités. Nous terminerons en donnant des perspectives d'améliorations possibles pour la modélisation.

Chapitre 1

Principes généraux, objectifs et politique de maintenance industrielle

Afin de mieux appréhender les objectifs de la maintenance industrielle, il est important de comprendre ses grands principes. C'est l'objet de ce premier chapitre.

Dans un premier temps, nous donnerons les définitions normatives de la maintenance. Nous détaillerons les deux types de maintenance couramment utilisés dans un contexte industriel. Leurs objectifs principaux ainsi que leurs politiques de mise en œuvre seront ensuite définis.

Dans un second temps, nous aborderons la gestion de la maintenance informatisée, implantée dans les divers secteurs économiques. Les différentes fonctionnalités de ces outils informatiques seront détaillées par la suite. Il est notable que ces outils sont souvent de fabrication « maison » et qu'ils sont devenus indispensables à la bonne pratique de la maintenance. Néanmoins, de nombreuses solutions professionnelles sont aujourd'hui disponibles sur le marché.

La littérature expose un nombre croissant de nouvelles stratégies de politiques de maintenance. Afin de mieux appréhender ces évolutions, nous tenterons de présenter ensuite les grands axes de ces nouvelles stratégies ainsi que leurs impacts sur les politiques de maintenances industrielles implémentées.

Depuis quelques années, des auteurs proposent l'utilisation de MMC pour la maintenance. Nous vous présenterons ces approches afin de comprendre la genèse de ces travaux débutés en 2007 [176].

Enfin, une présentation succincte des travaux de Pascal Vrignat [177] : « Génération d'indicateurs de maintenance par une approche semi-paramétrique et par une approche markovienne », viendra conclure ce chapitre, point de départ de notre contribution. En effet, ces indicateurs de dégradation d'un processus industriel quelconque doivent permettre à l'expert en maintenance d'améliorer la productivité des équipements dont il a la responsabilité. Sur ces bases ainsi établies, nous exposerons les objectifs de cette étude.

1.1 Introduction

La base de données ARIA¹, du Ministère de l'Écologie, répertorie toutes les informations sur les accidents technologiques. Sur les 7716 accidents recensés sur une vingtaine d'années, les experts ont mis en cause, pour 12% d'entre eux, les stratégies de maintenances, de réparations ou les protocoles de tests. Parmi eux, 30 % sont des accidents mortels ! Les circonstances de ces accidents sont les suivantes :

- sous estimation du risque industriel ;
- inadaptation de la mise en sécurité de certains équipements sensibles ;

1. Depuis 1992, elle répertorie les incidents ou accidents qui ont, ou auraient, pu porter atteinte à la santé ou la sécurité publique, l'agriculture, la nature et l'environnement. Site : <http://www.aria.developpement-durable.gouv.fr>

- insuffisance du dialogue entre les acteurs ;
- manque de contrôle et de suivi des chantiers.

Sur ces accidents mortels, 6,7% sont dûs à des opérations de « mise en service/arrêt/redémarrage » dont les causes sont :

- des manques d'étanchéité des équipements ;
- des mauvais positionnements de vannes ;
- des difficultés pour stabiliser des paramètres.

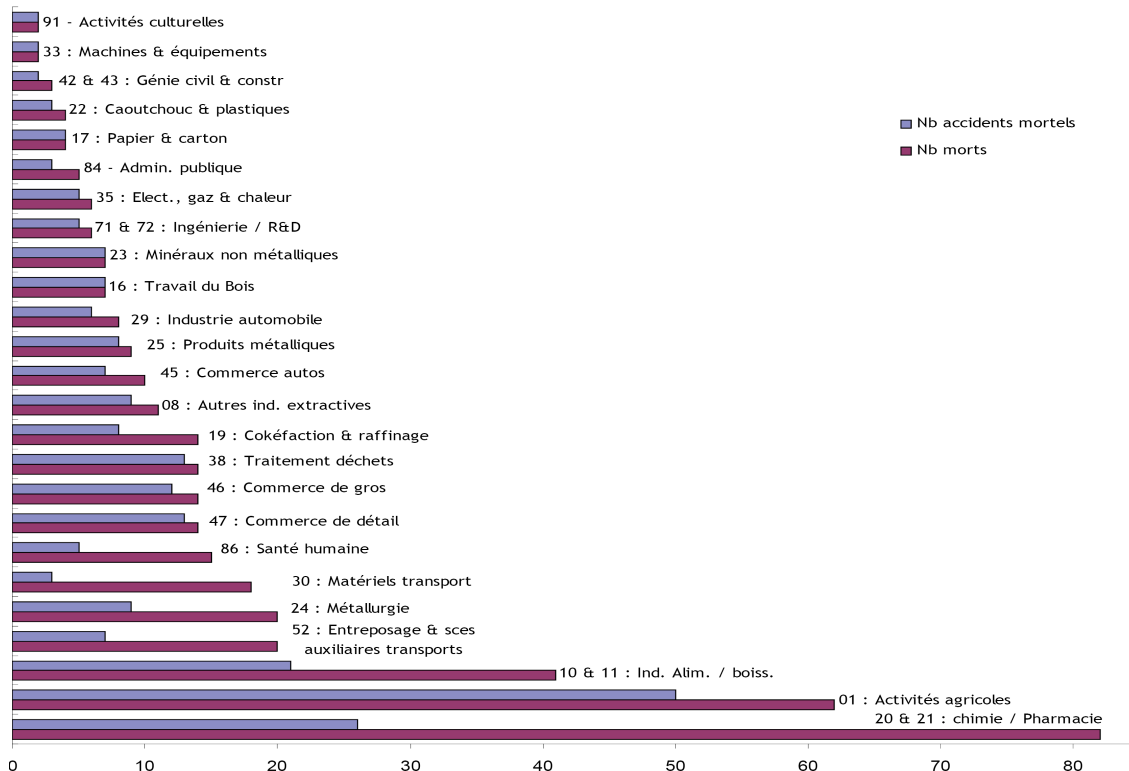


Fig 1.1 – Nombres d'accidents mortels et de victimes liés à un manque de suivi des équipements, pour les principales activités concernées.

Concrètement, les installations impliquées (voir Figure 1.1, p. 8) sont à l'origine de 254 accidents mortels et de 425 victimes entre 1992 et 2010. Il est regrettable de constater qu'un défaut ou un manque de suivi des équipements est trop souvent mis en cause par les experts.

Une meilleure gestion de la maintenance aurait-elle pu éviter de tels drames ?

Dans le même ordre d'idée, Geneviève Montaigu du journal Luxembourgeois « Le Quotidien » titrait récemment (mars 2012) : « Le parquet a requis hier une peine de 3 ans et demi contre le pilote, une peine de 2 ans contre l'ancien directeur du service technique et 18 mois contre deux autres responsables du **service maintenance**. La défense rappelle que les prévenus ne peuvent être condamnés pour la même faute dans un cas d'homicides involontaires ». Cet article concernant la tragédie de la Luxair en 2002 a fait jurisprudence en matière d'organisation de la maintenance. En effet, le technicien de maintenance n'ayant pas donné suite aux recommandations du constructeur alors qu'il avait les documents en sa possession, est en partie responsable de la mort de 10 personnes. Toute une chaîne de responsabilité est mise en cause, du pilote au personnel de maintenance. **Ainsi ce drame aurait pu être évité si les consignes de maintenance avaient été respectées !**

De tels drames ne sont pas inéluctables dans le domaine de la maintenance. Ainsi, certaines « bonnes » pratiques, que nous allons aborder par la suite, peuvent minimiser de tels risques encourus par les personnes.

Comme nous l'avons dit précédemment, la maintenance industrielle est en constante évolution depuis plusieurs années, elle est devenue un secteur stratégique dans les entreprises. En effet, outre les considérations humanistes, ces accidents ont un coût pour les entreprises concernées. Dans la course effrénée au rendement, le moindre dysfonctionnement du processus d'une chaîne de fabrication peut avoir des effets économiquement catastrophiques. Nous avons pu constater nous même que c'est le sentiment d'un bon nombre d'entreprises, notamment dans l'Indre, où une baisse de production pourrait engendrer une baisse de compétitivité entraînant des licenciements voire des fermetures d'usines.

La maintenance n'a plus aujourd'hui comme vocation unique de « simple réparation » de l'outil de production, car elle doit permettre de prévoir et d'éviter ces dysfonctionnements. Elle doit sans cesse s'adapter aux innovations technologiques tout comme aux nouveaux modes de gestion. Ces adaptations doivent aussi prendre en compte les nouvelles réglementations internationales sans accroître pour autant les coûts de production.

Il ne faut pas perdre de vue que l'objectif premier de la maintenance est de garantir la disponibilité optimale de l'outil de production d'une entreprise. Nous verrons dans les sections suivantes que cette idée est toujours la base de la maintenance mais qu'elle se décline suivant beaucoup d'autres aspects.

1.2 Présentation de la maintenance dans le domaine industrielle

De nos jours, les entreprises travaillent de plus en plus à flux tendu, i.e. « production à stock minimal ». Cette politique de fonctionnement a énormément fragilisé la chaîne de production. En effet, la moindre défaillance sur un chaînon du processus de fabrication peut impliquer un arrêt de toute la production. De ce fait, la fiabilisation de ces équipements est devenue un enjeu majeur. Suivant cette politique du « zéro stock », le « zéro défaut » est ainsi devenu une exigence vitale pour l'entreprise. La sécurité des personnes est devenue un enjeu fondamental en général, voire primordial pour certains secteurs comme l'aéronautique ou le nucléaire (en particulier les centrales nucléaires).

Les coûts liés à la maintenance font désormais partie intégrante de la politique budgétaire de l'entreprise. Son budget consacré à la maintenance doit être un juste équilibre entre compétitivité et qualité de la production. En effet, environ 10 % des effectifs industriels sont concernés par des tâches de maintenance et sont toujours en constante croissance malgré la conjoncture économique actuelle. Selon « l'Observatoire Réseau Maintenance[®] 2012 », les dépenses de maintenance ont progressé de + 1,9 % en 2011, par rapport à 2010 (les chiffres pour 2013 sont basés sur des estimations). Le secteur de la mécanique reste toujours dynamique mais ceux de l'automobile, du raffinage, de la fonderie ou de la papeterie ont minimisé ces coûts, en ajustant leur budget maintenance à la baisse ou en fermant des unités de production. Voir Tableau 1.1, p. 10.

	2005	2006	2007	2008	2009	2010	2011	2012	2013
Production en valeur (G€)	828	875	922	941	810	864	921	931	943
Dépenses maintenance/production (%)	2,7	2,5	2,4	2,3	2,6	2,4	2,3	2,3	2,3
Dépenses maintenance (G€)	22	22,1	22	21,6	20,6	20,9	21,2	21,3	21,4
Sous-traitance de la maintenance (%)	31,4	32,2	32,8	32,4	33,2	32,4	33,3	34,4	34,6
Maintenance sous-traitée (G€)	6,9	7,1	7,2	7	6,9	6,8	7,1	7,3	7,4

Tableau 1.1 – Évolution des valeurs et des grands ratios de la maintenance. Source : l'Observatoire Réseau Maintenance[®] 2012.

1.2.1 Définitions normatives

Une définition donnée par l'AFNOR² de la maintenance est la suivante : « Ensemble de toutes actions techniques, administratives et de management durant le cycle de vie d'un bien, destinées à le maintenir ou à le rétablir dans un état dans lequel il peut accomplir la fonction requise ».

Le « Petit Larousse 2010 » donne la définition suivante : « Ensemble de tout ce qui permet de maintenir ou de rétablir un bien dans un état spécifié ou en mesure d'assurer un service déterminé ».

En effet, le rôle premier de la maintenance est la réparation et le dépannage au moindre coût. C'est aussi le maintien des équipements en état de marche avec une sécurité de fonctionnement maximale. La maintenance doit se conjuguer avec la notion de maintenabilité. Les équipements de production doivent en effet pouvoir être entretenus tout au long de leur durée de vie (voir l'ouvrage de Roucoules et al. [143] sur le cycle de vie d'un produit).

1.2.2 Politique de maintenance

Le rôle du responsable de maintenance a bien évolué, au-delà de ses connaissances techniques. Il doit désormais trouver le juste équilibre entre la montée en puissance des exigences économiques imposées par les dirigeants et la disponibilité sans faille des équipements dont il est responsable. Le lecteur pourra se référer à la thèse de Castanier [37], qui réalise une bonne synthèse et donne un large éventail des politiques de maintenance.

2. Association Française de Normalisation

Avant le passage d'une politique corrective vers une politique de maintenance préventive, il faut tenir compte d'éventuelles rénovations futures. Il ne faudrait pas intégrer dans cette politique, un équipement qui serait mis au rebut dans un avenir proche.

Il faut ensuite évaluer les répercussions d'une panne sur le plan :

- économique et social : perte d'exploitation, chômage technique, image de l'entreprise ;
- environnemental : traitement de la pollution, conséquences néfastes sur l'environnement ;
- qualitatif : l'arrêt et le redémarrage ont une incidence sur la qualité des produits.

Si l'on ne tient pas compte des cas extrêmes où le risque est inacceptable, la politique de maintenance mise en œuvre devra tenir compte d'autres situations pour lesquelles le risque peut être tolérable, ou même totalement acceptable :

- pour un risque inacceptable, il faut supprimer totalement les causes de panne ;
- pour un risque tolérable, il faut choisir entre maintenance préventive systématique et maintenance conditionnelle (voir § 1.2.4, p. 13) ;
- et pour un risque acceptable, la maintenance préventive peut être inutile.

Une tendance actuelle, de plus en plus fréquente dans les grandes industries, est de confier la maintenance de niveau 1 et 2 (voir ci-dessous la classification de l'AFNOR) aux opérateurs de production afin de décharger le service de maintenance sur des tâches basiques et de responsabiliser les opérateurs au bon entretien de leurs outils de travail.

L'AFNOR distingue 5 niveaux de maintenance, suivant la complexité des opérations :

- 1^{er} niveau : réglage simple prévu par le constructeur ou le service de maintenance, au moyen d'éléments accessibles sans aucun démontage pour ouverture de l'équipement. Ces interventions peuvent être réalisées par l'utilisateur sans outillage particulier à partir des instructions d'utilisation.

Il s'agit ici de contrôler les paramètres nécessaires au bon fonctionnement de la machine comme les niveaux d'huile, les niveaux d'eau, les températures de fonctionnement et de réaliser un contrôle visuel et auditif du fonctionnement des organes sensibles.

- 2^{ème} niveau : dépannage par échange standard des éléments prévus à cet effet et d'opération mineure de maintenance préventive. Ces interventions peuvent être réalisées par un technicien habilité ou l'utilisateur de l'équipement dans la mesure où ils ont reçu une formation particulière.

Ce sont des opérations de remplacement, d'analyse, de réglages simples nécessitant éventuellement un outillage spécifique.

- 3^{ème} niveau : identification et diagnostic de panne suivis éventuellement d'échange de composant, de réglage et d'étalonnage général. Ces interventions peuvent être réalisées par un technicien spécialisé sur place ou dans un local de maintenance à l'aide de l'outillage prévu dans des instructions de maintenance.
Ces opérations de maintenance de type curative, concernent des réglages et des réparations mécaniques ou électriques mineures, comme le remplacement de capteurs ou de modules défaillants.
- 4^{ème} niveau : travaux importants de maintenance corrective ou préventive à l'exception de la rénovation et de la reconstruction. Ces interventions peuvent être réalisées par une équipe disposant d'un encadrement technique très spécialisé et des moyens importants adaptés à la nature de l'intervention.
Ce sont des opérations assez complexes comme par exemple le rebobinage d'un moteur électrique, la rectification de culasse, etc.
- 5^{ème} niveau : travaux de rénovation, de reconstruction ou de réparation importante confiés à un atelier central de maintenance ou une entreprise extérieure prestataire de service.
Ce niveau nécessite des moyens similaires à ceux utilisés en fabrication.

1.2.3 Objectifs d'une politique de maintenance

Dans un contexte de maintenance industrielle, le maintien en état de marche des équipements est devenu une exigence économique et sociale. L'entretien (activité opérationnelle de maintenance) peut-être correctif ou préventif. Dans le premier cas, les interventions de maintenance se font uniquement après qu'un défaut a été constaté ou subi. Cela peut conduire à des situations subies, en particulier pour les systèmes sensibles aux risques industriels (nous pouvons citer par exemple l'explosion récente sur une plate-forme pétrolière dans le golfe du Mexique, le 16 novembre 2012). Pour éviter que de tels cas ne se produisent, un entretien systématique peut être mis en œuvre.

Les principaux objectifs de la maintenance sont :

- assurer la sécurité humaine : dans certains cas, elle permet de limiter les interventions dangereuses qui comporteraient des risques pour les opérateurs de maintenance comme par exemple la manipulation de métal en fusion chez un métallurgiste. La maintenance doit minimiser les risques aux personnes sans augmenter pour autant les risques encourus par les personnels de maintenance ;
- optimiser le temps de fonctionnement de l'équipement : la planification des interventions de maintenance préventive sur une machine doit tenir compte du planning d'arrêt de cet équipement ;
- améliorer la fiabilité de l'équipement : le contrôle des réglages de la machine permet de conserver une qualité optimale des produits fabriqués tout en limitant les rebuts.

1.2.4 Typologies des actions de maintenance

Nous présentons ici les deux principaux types de maintenance utilisés dans l'industrie : la maintenance préventive et la maintenance corrective.

La maintenance préventive

Se prémunir d'une défaillance matérielle avant qu'elle n'arrive est devenu une règle d'or en matière de maintenance industrielle. La maintenance préventive doit être appliquée lorsque la panne potentielle a une incidence notable sur la sécurité en devenant donc inacceptable économiquement pour l'entreprise. Celle-ci permet d'améliorer la sécurité des personnes en minimisant les dépannages, d'augmenter le taux de productivité en optimisant les temps d'arrêt de production (i.e. d'assurer une continuité de service en planifiant des opérations de maintenance pendant l'arrêt de travail) et d'augmenter la durée de vie des équipements.

Selon l'AFNOR, c'est une « maintenance exécutée à des intervalles prédéterminés ou selon des critères prescrits. Cette stratégie de maintenance est destinée à réduire la probabilité de défaillance ou la dégradation du fonctionnement d'un bien ». Les définitions suivantes sont extraites de la norme NF EN 13306 X 60-319 [121].

La maintenance préventive est souvent subdivisée en trois principaux types (voir Figure 1.2, p. 15) :

- la maintenance préventive conditionnelle : « Maintenance préventive basée sur une surveillance du fonctionnement du bien et/ou des paramètres significatifs de ce fonctionnement intégrant les actions qui en découlent ». Elle est basée sur le franchissement d'un seuil de dégradation critique permettant de déclencher une opération de maintenance. Nous trouvons aussi dans la littérature un autre sous ensemble de la maintenance conditionnelle : la maintenance prévisionnelle (« Maintenance conditionnelle exécutée en suivant les prévisions extrapolées de l'analyse et de l'évaluation de paramètres significatifs de la dégradation du bien ». C'est une maintenance conditionnelle qui est basée sur le franchissement d'un seuil de dégradation critique permettant de déclencher une opération de maintenance) ;
- la maintenance préventive systématique : « Maintenance préventive exécutée à des intervalles de temps préétablis ou selon un nombre défini d'unités d'usage mais sans contrôle préalable de l'état du bien ». Cette maintenance préventive est effectuée selon un échancier établi à l'avance selon une stratégie complexe qui allie plusieurs facteurs comme le taux d'utilisation du matériel, la législation en vigueur (sécurité réglementée), etc. ;
- la maintenance préventive programmée : « Maintenance préventive exécutée selon un calendrier préétabli ou selon un nombre défini d'unités d'usage ».

La maintenance corrective

Cette norme [121] définit la maintenance corrective comme suit : « Maintenance exécutée après détection d'une panne et destinée à remettre un bien dans un état dans lequel il peut accomplir une fonction requise ».

Elle désigne l'élimination d'une avarie après une défaillance dans le fonctionnement d'une « entité »³ matérielle, par réparation ou remplacement de celle-ci. Le caractère temporel est parfois utilisé dans certaines politiques de maintenance :

- maintenance corrective immédiate, effectuée tout de suite après la panne ;
- maintenance corrective différée, retardée en fonction de la politique de maintenance.

La maintenance corrective est souvent subdivisée en deux types (voir Figure 1.2, p. 15) :

- la maintenance corrective palliative : « Action de maintenance corrective destinée à permettre à un bien d'accomplir provisoirement tout ou partie d'une fonction requise. Appelée couramment *dépannage*, la maintenance palliative est principalement constituée d'actions à caractère provisoire qui doivent être suivies d'actions curatives » [121]. Elle correspond au dépannage provisoire de l'équipement (parfois même, sans interprétation de la défaillance). Ce type de dépannage présentant un **caractère provisoire** devra être suivi d'activités curatives ;
- la maintenance corrective curative : elle a pour but la réparation d'un équipement, afin de lui permettre d'accomplir une fonction requise. Ces activités doivent présenter un caractère **permanent**.

Ce type de politique est destiné à rendre opérationnelle une machine qui est tombée en panne. Elle est peu efficace pour les équipements « vitaux » de production mais son application est bien adaptée à certains matériels peu coûteux.

1.3 La Gestion de la Maintenance Assistée par Ordinateur : GMAO

Pour rester compétitive, l'informatisation est devenue une nécessité dans l'industrie. L'implémentation d'un outil de GMAO est devenue un levier de performance indispensable pour les entreprises. Il s'intègre aujourd'hui avec une gestion globale appelée ERP⁴. Une large gamme de logiciels est disponible dans le commerce mais beaucoup d'entreprises ont développé leurs propres solutions en local afin de l'adapter plus particulièrement à leurs besoins. Ces outils permettent, entre autre, d'enregistrer toutes les observations, anomalies, défaillances ou opérations de maintenance. Toute personne autorisée peut ainsi signaler un problème ou une opération de maintenance effectuée (même banale). Ces données peuvent

3. jargon de la spécialité

4. Enterprise Resource Planning ou EAM Enterprise Asset Management

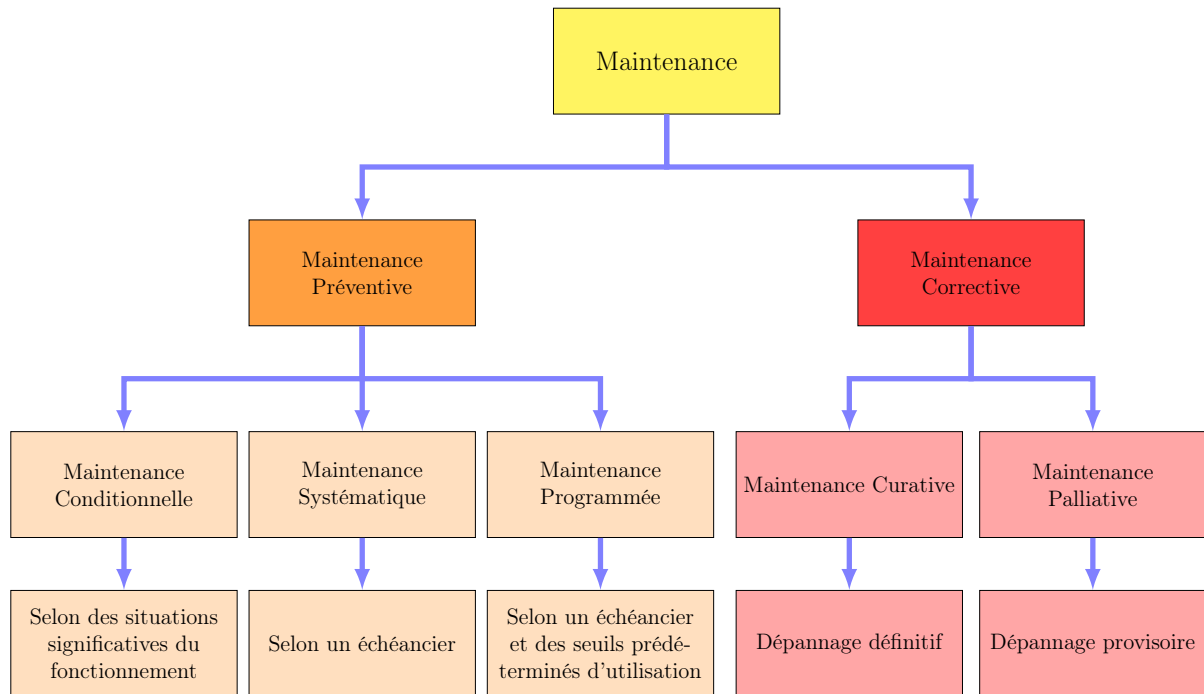


Fig 1.2 – Typologies des actions de maintenance (NF EN 13306 (indice de classement : X60319)) [121].

être ainsi prises en considération a posteriori, par l'expert. Ces logiciels possèdent de nombreuses fonctionnalités comme la planification des interventions, la gestion des stocks et même, le calcul en temps réel du KPI⁵. Certains logiciels intègrent des prévisions de dégradation. Sans faire une étude approfondie, de nombreuses entreprises n'utilisent pas d'outils de ce type. Nous donnons dans le Tableau 1.2, p. 16 les données issues de l'étude de Fumagalli et al. de 2009 [69] montrant l'utilisation de logiciels de GMAO dans les différents secteurs d'activités. Nous pouvons voir que le secteur automobile (19%) est particulièrement bien doté par rapport aux autres secteurs industriels.

Le Tableau 1.3, p. 16 montre pour exemple une base de données utilisée en GMAO dans une usine de production agroalimentaire.

Ces outils font désormais partie intégrante du système d'information de l'entreprise [101]. Ils se rapprochent même parfois d'un logiciel d'ERP possédant une base de données centralisée et intégrant toutes les fonctionnalités de gestion d'une entreprise (stock, commande, facturation, ressources humaines, etc.). Ces données vont servir de base de travail à notre étude. Une phase conséquente d'interprétation des données d'entreprises a été menée par Pascal Vrignat [177]. Le problème majeur consistait à codifier toutes les actions de maintenances, **observations**, sous forme de symboles (cf. colonne action « ACT. » (Tableau 1.3, p. 16)). L'alphabet ainsi créé est alors utilisable très simplement pour des traitements informatisés de calculs divers, de statistiques ou de modélisations.

5. Key Performance Indicator ou indicateur clé de performance qui mesure l'efficacité des mesures prises en amont pour atteindre un objectif fixé

SECTEUR	%
Automobile	19,10%
Chimie	8,40%
Logistique et transport	13,80%
Energie	0,60%
Alimentation, boisson	5,00%
Santé	20,90%
Tourisme, hotels	0,20%
Industrie mécanique	5,00%
Industries pharmaceutiques	0,40%
Services publics	6,00%
Papeterie	1,00%
Sidérurgie	2,50%
Textiles	1,90%
Immobiliers	9,30%
Autres	5,90%

Tableau 1.2 – Proportion d’utilisation des logiciels de « Gestion de la Maintenance Assistée par Ordinateur », dans les différents secteurs économiques.

Nom	Demande	Date estimée	Réalisation	N° OT	Lieu	Equipement	ACT.	TI
DUPOND	23/09/2008	13/01/2009	14/01/2009	M4291	FUSION - FOUR	ARCHE	VEP	4
DUPOND	06/10/2008	15/01/2009	15/01/2009	E1788	FINITION	BATIMENT STOCK	DEP	5,5
DUPOND	26/11/2008	05/01/2009	05/01/2009	MO936	FABRICATION	PRESSE	DEP	2,5
DUPOND	26/11/2008	06/12/2008	06/01/2009	MO940	FUSION - FOUR	ARCHE	AU	2
DUPOND	13/01/2009	13/01/2009	13/01/2009	M7	FABRICATION	REBRULEUSE 92	DEP	0,25
DUPOND	29/01/2009	29/01/2009	29/01/2009	E1792	FINITION	BURN OFF	DEP	3
DUPOND	26/02/2009	28/02/2009	27/02/2009	MO987	FABRICATION	REBRULEUSE	DEP	4
DUPOND	02/03/2009	05/03/2009	02/03/2009	MO984	FINITION	ETIQUETEUSE	TEP	1
DUPOND	04/03/2009	25/03/2009	17/03/2009	E1813	FINITION	CONVOYEUR CALCIN	DEP	5,5
DUPOND	02/04/2009	30/03/2009	08/04/2009	C2295	FABRICATION	PRESSE 93	VEP	4,5
DUPOND	15/04/2009	15/04/2009	15/04/2009	C2334	FUSION - FOUR	VIBRANT	RM	1
DUPOND	15/04/2009	24/04/2009	21/04/2009	C2334	FUSION - FOUR	VIBRANT	RM	2
DUPOND	15/04/2009	24/04/2009	22/04/2009	C2334	FUSION - FOUR	VIBRANT	RM	7
DUPOND	20/04/2009	21/04/2009	21/04/2009	C1318	EXTERIEUR	CENTRALE AZOTE	MOD	1,5
DUPOND	21/06/2009	21/06/2009	21/06/2009	M4839	FABRICATION	FOUR A MOULES	DEP	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tableau 1.3 – Exemple d’une base de données utilisée dans une Gestion de la Maintenance Assistée par Ordinateur.

1.4 Évolution des politiques de maintenance industrielle

Comme explicité au § 1.2.4, p. 14, la mise en œuvre d’une maintenance corrective ne permet pas d’éviter les défaillances et leurs conséquences. Une maintenance préventive destinée à limiter ces pannes est devenue indispensable. Un équilibre est alors à trouver afin de limiter les dépenses excessives ainsi que les temps d’indisponibilité parfois inutile, des entités impactées. Ainsi, la surveillance et la réparation ne sont plus les seules préoccupations du responsable de maintenance. Il doit désormais apporter à l’entreprise une stratégie différente basée sur un mélange subtilement dosé de « technicité » et « d’organisation ». Pour cela, il doit utiliser au plus juste les moyens techniques et financiers dont il dispose.

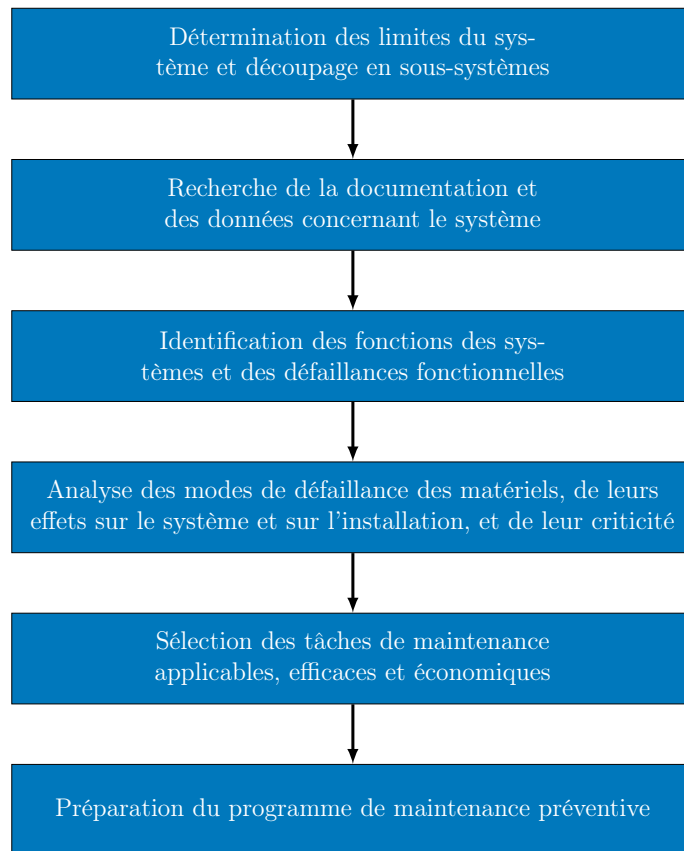


Fig 1.3 – Exemple d’optimisation de la Maintenance Basée sur la Fiabilité.

De nouvelles méthodes ont ainsi fait leur apparition comme la méthode MBF⁶ [194] ou RCM⁷ [120], qui s’intègre aux politiques de maintenance en vigueur en y rajoutant de nouvelles instructions de maintenance préventive. Cette méthode se base sur les conséquences des défaillances. Elle est représentée sur la Figure 1.3, p. 17. La méthode consiste, avant tout, à aider l’expert en maintenance dans son choix parmi toutes les tâches de maintenance disponibles. Par ailleurs, le caractère itératif de cette méthode permet de faire évoluer les programmes de maintenance en fonction du retour d’expérience. La méthode met à disposition un certain nombre d’outils pour l’utilisateur (listes génériques de fonctions pour les systèmes et les différents types de défaillance pour les matériels). Ce guide comporte des listes de tâches génériques utilisées en maintenance préventive. Il décrit les procédures à utiliser pour chaque type de défaillance de matériels génériques (vérins, pompes, etc.) en détaillant les préconisations d’utilisation recommandées de ces matériels (niveau de criticité, environnement particulier, conditions de fonctionnement).

Une autre méthode plus récente, décrite dans l’ouvrage d’Antoine Despujols de 2009 [58] : l’« Optimisation de la Maintenance par la Fiabilité » (OMF), mise en œuvre par EDF, est aujourd’hui utilisée dans d’autres secteurs industriels. Elle possède les mêmes spécificités

6. Maintenance Basée sur la Fiabilité

7. Reliability-Centered Maintenance

que la méthode basée sur la fiabilité explicitée ci-dessus. Cette méthode est basée sur trois principes :

- l'évaluation des risques ;
- l'analyse des retours d'expériences ;
- la mise en œuvre de tâches de maintenance élémentaires, suivant une certaine logique.

Il s'agit dans un premier temps de déterminer les objectifs visés : sécurité, sûreté de fonctionnement, taux de disponibilité, coûts de fonctionnement, etc. La méthode décrite Figure 1.4, p. 19, permet de déterminer les risques potentiels.

L'analyse des dysfonctionnements permet d'identifier les défaillances des équipements. L'analyse du retour d'expérience, nous apporte les données concernant les matériels et modes de défaillance critiques (AMDEC⁸), nécessaires pour l'établissement d'une stratégie de maintenance adéquate. Cette analyse donne une estimation de la fréquence d'apparition des défaillances, en utilisant les études probabilistes de sûreté [30], [106]. Ces procédures, fondées sur une analyse fonctionnelle poussée, indiquent à l'utilisateur les conduites à tenir en fonction des spécificités techniques des matériels ainsi que de leur durée de réparation. L'Optimisation de la Maintenance par la Fiabilité est notamment utilisée dans les centrales nucléaires pour son aspect **sûreté** [106].

D'autres techniques prometteuses comme le Soutien Logistique Intégré (SLI) [13] sont utilisées par Alstom, Aréva et EADS. Elle a pour but de réduire les coûts de fonctionnement et d'investissements d'un matériel, tout en maximisant l'efficacité opérationnelle pour l'utilisateur (sécurité, régularité, disponibilité). Une autre méthode comme la maintenance productive totale (TPM⁹) [72], consiste à assurer la maintenance tout en produisant, ou en pénalisant le moins possible la production. Tous les aspects de l'entreprise ainsi que tous ses acteurs sont pris en considération. Cette technique est surtout mise en place dans l'industrie automobile comme le « lean manufacturing »¹⁰ chez Toyota.

Le lecteur pourra se référer au livre d'Antoine Despujols [57], qui fait un état des lieux de différentes méthodes d'optimisation de la maintenance dans divers secteurs industriels.

Récemment publié dans l'article de Fumagalli et al. [69], la maintenance basée sur des conditions ou CBM¹¹ est considérée comme l'une des politiques les plus pertinentes en ce qui concerne l'amélioration de la gestion de la maintenance. Elle est très proche de la maintenance prévisionnelle ou prédictive. C'est une **maintenance conditionnelle** qui est basée sur le franchissement d'un seuil de dégradation critique permettant de déclencher une opération de maintenance. Couplée avec le système de GMAO ou d'ERP de l'entreprise,

8. Analyse des Modes de Défaillance, de leurs Effets et de leur Criticité

9. Total Productive Maintenance = maintenance productive totale

10. Amélioration continue et l'élimination des gaspillages de la gestion de la production

11. Condition Based Maintenance = maintenance basée sur des conditions

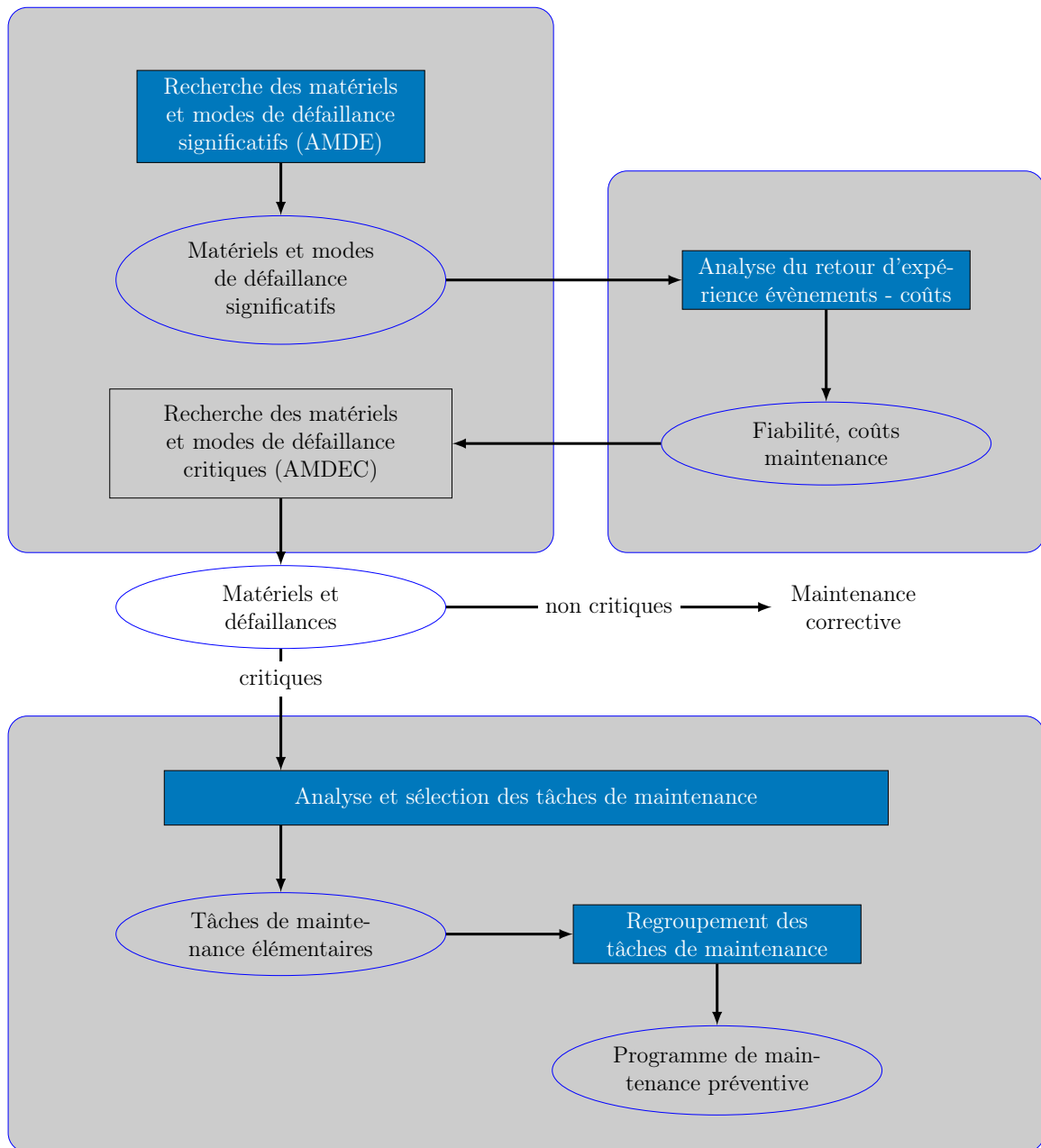


Fig 1.4 – Exemple d’Optimisation de la Maintenance par la Fiabilité.

elle utilise sa base de données centralisée. Elle permet de réduire le nombre de maintenances préventives planifiées inutiles et ainsi diminuer considérablement les coûts de la maintenance. Dans certains cas, un programme de surveillance peut éviter d’inutiles tâches de maintenance en prenant des mesures d’entretien uniquement lorsque il y a la preuve d’un comportement anormal de l’équipement.

La mise en œuvre d’une CBM [108] peut conduire à des avantages significatifs. En effet, contrairement à un entretien systématique, une CBM est réalisée en fonction de l’état de santé réel du matériel, estimé ou mesuré par des capteurs présents sur l’équipement. Les techniques

de surveillance les plus utilisées sont la mesure des vibrations, la thermographie¹² et la tribologie¹³. Ainsi, en utilisant des outils mathématiques adaptés (statistiques, modélisation, etc.), un état de panne peut-être alors anticipé par l'équipe de maintenance.

Selon Tobon-Mejia et al. [167], lors de la mise en place d'une CBM, le pronostic de panne est considéré comme l'un des principaux objectifs, car il permet d'estimer ce qu'on appelle la durée de vie utile restante (ou RUL¹⁴ en Anglais) avant la défaillance d'un système donné [92]. De nombreuses méthodes et outils peuvent être utilisés pour prédire la valeur de la RUL. Ces méthodes peuvent être classées selon trois groupes principaux [108], [80], à savoir les modèles prédictifs, ceux à prédiction basée sur les données et enfin ceux à prédiction « empirique » :

- le modèle prédictif, détermine la RUL des composants critiques à l'aide de modèles mathématiques ou physiques du phénomène de dégradation (fissure par fatigue, usure, corrosion, etc.). Ce modèle est basé sur des résultats de simulation ;
- les pronostics basés sur les données de tests réels, visent à modéliser la dégradation du système à l'aide des données fournies par les capteurs ;
- enfin, le pronostic « empirique », basé sur des retours d'expérience, exploite les données recueillies à partir de la machine pendant une longue période de temps pour estimer les paramètres des lois de fiabilité traditionnelles. Ils sont utilisés pour faire des extrapolations et des projections afin d'en estimer la RUL.

La pratique du CBM peut aider à améliorer la disponibilité, la fiabilité et la sécurité tout en réduisant les coûts de maintenance. Elle peut remplacer les traditionnelles politiques de maintenance impliquant les maintenances correctives et préventives.

Pour rester compétitif, les grandes entreprises doivent intégrer ces nouvelles stratégies dans leur organisation et ainsi investir des sommes importantes. L'objectif général de toutes ces nouvelles techniques de gestion de la maintenance est de **maximiser le temps de production**. Ces nouvelles techniques sont basées sur un historique de fonctionnement. La fiabilité de ces études repose tout de même en partie sur une saisie correcte des actions de maintenance (observations) et sur de bonnes pratiques des opérateurs de production et du service maintenance.

Quelques méthodes liées à notre problématique utilisant des MMC, proposées dans la littérature seront abordées dans la prochaine section. Celles-ci feront probablement l'objet d'outils opérationnels mis à disposition à l'expert.

12. « Technique permettant d'obtenir, au moyen d'un appareillage approprié, l'image thermique d'une scène observée dans un domaine spectral de l'infrarouge » selon l'AFNOR

13. Science qui étudie les phénomènes se produisant entre deux objets en contact (frottement, usure et lubrification, etc.)

14. Remaining Useful Life = durée de vie utile restante

1.5 Utilisation de Modèles de Markov Cachés dans le cadre d'une politique de maintenance industrielle

L'objectif de notre étude se situe au niveau de la maintenance préventive. Nous utilisons des outils mathématiques de modélisation comme les MMC, afin de réduire la probabilité de défaillance d'un équipement ou d'un service rendu. Dans cette section, nous situons nos travaux par rapport à l'estimation de la dégradation. Nous donnons ensuite quelques techniques proches de celles que nous allons présenter dans ce manuscrit.

Principe d'un MMC

Soit un système comportant un nombre d'états cachés, états qui évoluent au cours du temps. Si on peut observer périodiquement des symboles (observations) émis par ce système, alors il peut-être modélisé sous forme de MMC (voir Figure 1.5, p. 21). A chaque instant, l'évolution d'un état est déterminée à partir d'une distribution de probabilité, fixée au préalable, et de l'état présent.

La matrice de transition est de la forme :

$$\mathcal{P} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \quad (0 < \alpha, \beta \leq 1), \quad (1.1)$$

où $\mathcal{P}(i, j)$ donne la probabilité de transition de l'état i vers l'état j .

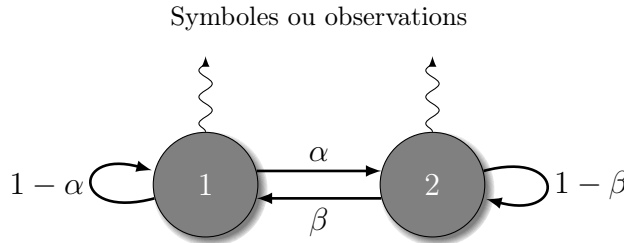


Fig 1.5 – Modèle de Markov Caché à deux états.

1.5.1 Anticiper une situation de panne

La définition d'une panne dans le secteur de la maintenance industrielle est l'arrêt accidentel et temporaire du fonctionnement d'une machine. Le diagnostic de panne qui consiste à isoler et trouver la cause probable de la panne [93], est une opération à effectuer a posteriori, i.e. après l'apparition du défaut. Le pronostic de panne vise quant à lui, à anticiper le moment de son apparition (Figure 1.6, p. 22) en s'aidant des observations passées du système (interventions sur le système, signaux émis par le système, températures, etc.). Nous retrouvons dans la littérature récente plusieurs travaux concernant la détection de panne en maintenance industrielle [118], [108].

Dans ce manuscrit, nous nous situons dans le domaine a priori i.e. à droite de la Figure 1.6, afin d'isoler l'information qui va nous permettre d'anticiper cette panne.

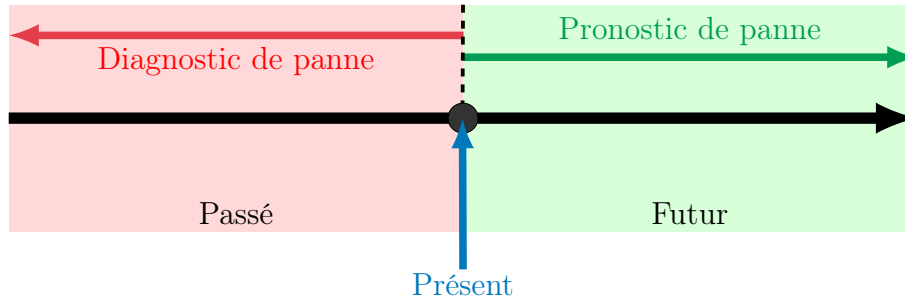


Fig 1.6 – Diagnostic vs Pronostic de panne.

Dans la littérature récente, de nombreuses études utilisent des MMC pour anticiper une défaillance d'un système. Dans la section suivante, quelques uns de ces travaux les plus pertinents seront résumés.

1.5.2 Utilisation des MMC dans la détection de panne

Pour améliorer l'efficacité de « l'apprentissage » pour MMC, Davis et al. [51] proposent une classe de nouvelles méthodes d'estimation, où la procédure de réestimation Baum-Welch [97] est gérée séparément sur plusieurs observations. Cette technique utilise l'ensemble des séquences d'observation, au lieu d'une seule séquence (une séquence se terminant après l'apparition d'un arrêt). Les paramètres qui maximisent la vraisemblance du modèle estimé sont ainsi déterminés. Nous utiliserons une partie de cette méthode dans ce manuscrit (§2.4.3, p. 42), dans le but d'évaluer les paramètres les plus pertinents sur nos modèles.

L'approche de Tobon-Mejia et al. [167], modèle prédictif fondé sur les données, utilise les MMC avec des mélanges de Gaussiennes (MoGHMM¹⁵) pour modéliser la dégradation et estimer la valeur de la RUL avant une éventuelle défaillance. Ce modèle contient le nombre d'états de « santé » ainsi que la durée de chaque état. Les paramètres temporels sont estimés par l'algorithme de Viterbi [175]. Les autres paramètres sont estimés par celui de Baum-Welch [55]. Nous utiliserons par la suite ces deux algorithmes pour tester l'apprentissage et le décodage des données de nos modèles.

Une autre étude sur le diagnostic à base de MMC [190], qui tente de trouver la meilleure estimation des états d'un système, arrive à quantifier les avantages d'une approche MMC (basée sur une distance de Hamming¹⁶).

Jian Zhou et Xiao-Ping Zhang [192] utilisent une **analyse de données séquentielles**, pour étendre le modèle classique (MMC) à un système continu¹⁷ en modélisant les densités d'observation comme un mélange de non-Gaussiennes. Ils ont développé des formules de réestimation pour les trois problèmes fondamentaux d'un MMC, à savoir le calcul de la vraisemblance, l'estimation de séquence d'états et l'apprentissage des paramètres du modèle. Afin d'obtenir une représentation paramétrique de la densité, ils appliquent une Analyse en

15. Mixture of Gaussians Hidden Markov Model

16. La distance de Hamming permet de quantifier la différence entre deux séquences de symboles

17. Modélise un processus continu

Composantes Indépendantes (ACI). Elle permet de séparer plusieurs des sources mélangées sur chacune des composantes du mélange.

De nombreuses approches basées sur l’analyse des données ont été développées en raison de leur grande fiabilité et du faible coût des capteurs. La plupart de ces études sont axées sur des méthodes d’estimation de densité de probabilité. On peut citer notamment celle de Serir et al. [151], qui utilise la théorie de l’évidence (voir § 2.3.2, p. 34) pour des données d’apprentissage limitées, en particulier les données d’états de défaillance, qui sont généralement coûteuses et difficiles à obtenir. Les paramètres estimés ne sont alors plus fiables. Ce nouvel outil basé sur des MMC (ou EvHMM¹⁸) semble très prometteur [138].

Une autre étude de prédiction de la RUL [168], aussi basée sur les données, utilise les méthodes temps-fréquence, et en particulier, **l’analyse par ondelettes** associée à un MMC. Cette technique utilise dans un premier temps, les données issues des capteurs. Celles-ci sont traitées pour extraire les caractéristiques sous la forme de coefficients permettant la décomposition par ondelettes. Les caractéristiques ainsi extraites, sont ensuite introduites dans les algorithmes d’apprentissage pour estimer les paramètres du MMC correspondant au mieux au phénomène de dégradation étudié. Le modèle ainsi obtenu est exploité dans un deuxième temps, afin d’évaluer l’état de santé de l’équipement et d’estimer sa durée de vie (RUL).

Une autre étude [35], utilise la **Décomposition Modale Empirique** (EMD). Elle consiste à décomposer un signal en un ensemble de fonctions, comme la décomposition en séries de Fourier ou une décomposition en ondelettes. La particularité de l’EMD réside dans le fait que la base de fonctions n’est pas donnée a priori mais est construite à partir des propriétés du signal. Associée à une modélisation par Markov caché, elle permet par exemple de détecter des défauts (fissures ou dents cassées) dans une boîte de vitesse.

Pour tenir compte de ces facteurs, Zhou Zhi-Jie et al. [193] ont développé un nouveau MMC basé sur la **théorie des fonctions de croyances** voir § 2.3.2, p. 34. Ce MMC est proposé pour prédire une défaillance cachée en temps réel, en tenant compte des influences des facteurs environnementaux. Dans le modèle proposé, le MMC est utilisé pour capturer les relations entre la panne cachée et les observations d’un système. Les fonctions de croyances sont utilisées pour modéliser les relations entre les facteurs environnementaux et les probabilités de transition entre les états cachés du système, y compris la défaillance cachée.

18. Evidential Hidden Markov Models

1.6 Conclusion et objectifs de cette étude

Conclusion

La maintenance industrielle n'est plus l'apanage des grosses entreprises industrielles. Elle est devenue un maillon essentiel dans la compétitivité des petites et moyennes entreprises. Malheureusement, nombreuses sont celles qui ont des systèmes de gestion de leur maintenance peu efficaces. Ceci met en évidence un manque de réflexion globale et durable :

- des actions de maintenance préventives réalisées par habitude ;
- de la planification des tâches de maintenance (qui n'est peu, voire jamais remise en question) ;
- de l'historisation des événements (description de la panne, pièces changées, temps d'intervention, etc.).

Nous venons de voir qu'il existe des solutions plus ou moins complexes à mettre en œuvre avec l'expert. Les travaux initiés par Pascal Vrignat dans sa thèse [177], sur une modélisation de la maintenance en utilisant des MMC ont montré qu'il était possible de modéliser la dégradation d'un processus industriel quelconque. A travers des cas concrets, il a pu donner aux experts une nouvelle approche de leur gestion de la maintenance préventive.

Objectifs de cette étude

Dans la continuité des précédents travaux présentés dans ce chapitre, notre objectif est d'étudier la pertinence des observations (ou symboles) utilisées dans la modélisation afin d'en optimiser leurs utilisations. La détermination de la pertinence de ces symboles consiste à évaluer la quantité d'informations qu'ils apportent, afin de privilégier ceux qui portent en eux une information de qualité. Pour cela, nous utiliserons un modèle de synthèse reproduisant des données de maintenance d'un processus industriel. Ainsi, sans connaissance a priori, nous étudierons la pertinence des données empiriques et des données issues de résultats de simulations.

Dans un deuxième temps, nous déterminerons la période d'échantillonnage des observations la plus efficace, par rapport aux architectures des modèles proposés dans [177], afin d'en faire évoluer la modélisation. Ensuite, nous essaierons de trouver un nombre minimal de données afin d'estimer de manière optimale les modèles, sans utiliser trop d'information. Cette « fenêtre glissante » de symboles composée d'une partie de l'historique des observations de maintenance, nous permettrait de faire une mise à jour régulière des modèles.

Enfin, nous analyserons les différentes architectures présentées dans [177] afin d'en confirmer ou d'en infirmer les choix théoriques. Nous étudierons ainsi la pertinence des topologies, des algorithmes d'apprentissage et de décodage ainsi que des lois statistiques utilisées dans la modélisation. Les données empiriques sont issues de GMAO (voir § 1.3, p. 14) d'entre-

prises, du secteur industriel de la région. Nous comparerons ensuite ces résultats avec ceux provenant de la simulation. Cette architecture la plus pertinente, permettra d'améliorer la conception du modèle de simulation.

Chapitre 2

Approches classiques de mesures de pertinence dans la chaîne de modélisation

2.1 Introduction

Afin de mieux appréhender les objectifs de notre travail sur l'évaluation des Modèles de Markov Cachés, il est important de resituer les méthodes de la littérature traitant de l'évaluation de modèles en général. Nous trouvons dans ce deuxième chapitre, les bonnes pratiques pour la qualification de nos modèles, un état de l'art des différentes approches d'analyse de modèles ainsi qu'une réflexion préliminaire sur leur adéquation aux MMC.

Nous définissons en premier lieu le terme « analyse de sensibilité » d'un modèle. Ces définitions sont suivies d'une présentation des méthodes recensées dans la littérature, concernant les mesures de pertinence d'un modèle, associées à cette analyse. Nous détaillons ainsi les méthodes avec score d'intérêt qui permettent d'évaluer et de comparer plusieurs modèles entre eux, de manière quantitative. Les méthodes statistiques dirigées par les données sont une alternative à ces méthodes avec score d'intérêt. Elles sont étudiées par la suite afin de nous permettre de comparer nos modèles « candidats », par une analyse quantitative. Dans le cadre de notre modèle de simulation de type « modèle stochastique », nous donnons quelques méthodes mesurant le caractère stochastique d'un modèle.



Fig 2.1 – Études de pertinences de Modèles de Markov Cachés.

Outre la présentation de ces différents critères d'évaluation, cette analyse bibliographique a pour ambition d'établir un point de départ pour le développement et la validation d'une méthode de sélection de modèles à base de MMC. Nous terminons ce chapitre par une discussion sur les méthodes de sélection d'un modèle et leurs intérêts. Les différents thèmes abordés pour chacune des méthodes sont illustrés par la Figure 2.1, p. 29.

2.2 Problématique générale

Cette étude a pour point de départ, les travaux de Pascal Vignat [177] sur la modélisation du niveau de dégradation d'un processus industriel quelconque. Dans sa thèse [177], Pascal Vignat propose d'utiliser plusieurs MMC comportant chacun quatre états cachés. Ces états S1, S2, S3 et S4, représentent les niveaux de dégradation. Ainsi, différentes architectures de modèles sont présentées : trois topologies sur les MMC, deux algorithmes d'apprentissage et deux distributions différentes (pour le modèle de simulation que nous présenterons par la suite). Nous étudions ces différentes architectures proposées afin de déterminer les éléments de ces modèles qui nous donnent les informations les plus intéressantes. C'est cette notion que nous appelons « pertinence de modèle ». L'étude est réalisée **sans connaissance a priori** sur les données. Les meilleures architectures pourront ainsi être utilisées dans le cadre d'une politique de maintenance préventive industrielle.

2.3 Analyse de sensibilité d'un modèle

Dans cette section, nous commençons par définir l'« analyse de sensibilité » (communément appelée SA pour « Sensitivity Analysis »). Nous présentons ensuite les méthodes d'évaluation de la sensibilité d'un modèle ainsi que celles du domaine connexe, concernant les incertitudes liées à sa conception. Nous discutons ensuite des enjeux de la prise en compte des incertitudes dans la chaîne de modélisation, puis nous définissons les différents types d'incertitudes liées à un modèle. A partir de ces analyses, nous pourrions ainsi déterminer les éléments du modèle qui influent sur sa pertinence.

- Que signifie *analyse* ?

Définition du « Petit Larousse 2010 » : « Analyse vient du grec *analysis*, signifiant décomposition. L'analyse est l'étude faite en vue de discerner les différentes parties d'un tout, de déterminer ou d'expliquer les rapports qu'elles entretiennent les unes avec les autres ».

- Que signifie *sensibilité* ?

Définition du « Petit Larousse 2010 » : « sensibilité vient du latin *sensibilis*, aptitude à réagir à des excitations externes ou internes ».

- Que signifie *analyse de sensibilité* ?

Nous parlons ici de l'analyse de sensibilité d'un modèle. Il s'agit de la réaction d'un modèle suite à un ensemble d'excitations externes ou internes.

Selon Rosen [142], l'analyse de sensibilité étudie les relations des flux d'informations entre les sorties et les entrées d'un modèle. L'analyse de sensibilité est quantifiée par des **indices de sensibilité** que nous allons détailler dans les sections suivantes. La valeur de l'indice est directement liée à l'importance de la variable i.e. **plus l'indice sera grand, plus la variable aura de l'importance**.

Le lecteur pourra se référer à l'excellent ouvrage de Saltelli et al. [145] ainsi que celui de Iooss [89], donnant un bon aperçu des méthodes de calcul des indices de sensibilité. Il distingue ainsi deux principaux types d'analyse de sensibilité :

- l'analyse de sensibilité dite « locale », qui est la variation de la réponse de sortie, par rapport à la variation d'une entrée (cette approche est souvent déterministe) ;
- l'analyse de sensibilité « globale », qui répond à la question : quelle est la contribution de l'incertitude des variables d'entrée (X_j), sur l'incertitude de la variable de sortie (Y) ?

2.3.1 Comment réaliser une analyse de sensibilité ?

Selon Saltelli et al. [145], l'analyse de sensibilité peut se décomposer en cinq étapes :

1. concevoir l'expérimentation (i.e. à quelle question le modèle doit-il répondre) et déterminer les facteurs d'entrées concernés par l'analyse ;
2. attribuer des fonctions probabilistes ou des plages de variation à chaque facteur d'entrées ;
3. concevoir un vecteur ou une matrice d'entrées adapté au domaine d'application ;
4. évaluer le modèle, définir ainsi une fonction de mesure ou distribution, pour la réponse étudiée ;
5. évaluer l'influence ou l'importance relative de chaque facteur d'entrées sur les variables de sortie.

La Figure 2.2, p. 32 illustre cette démarche d'analyse de sensibilité.

2.3.2 Méthodes d'analyse de sensibilité

Nous présentons ici les différentes méthodes issues de la littérature courante, permettant de réaliser une analyse de sensibilité d'un modèle. Nous discutons chacune de ces méthodes selon notre problématique d'analyse de sensibilité de MMC.

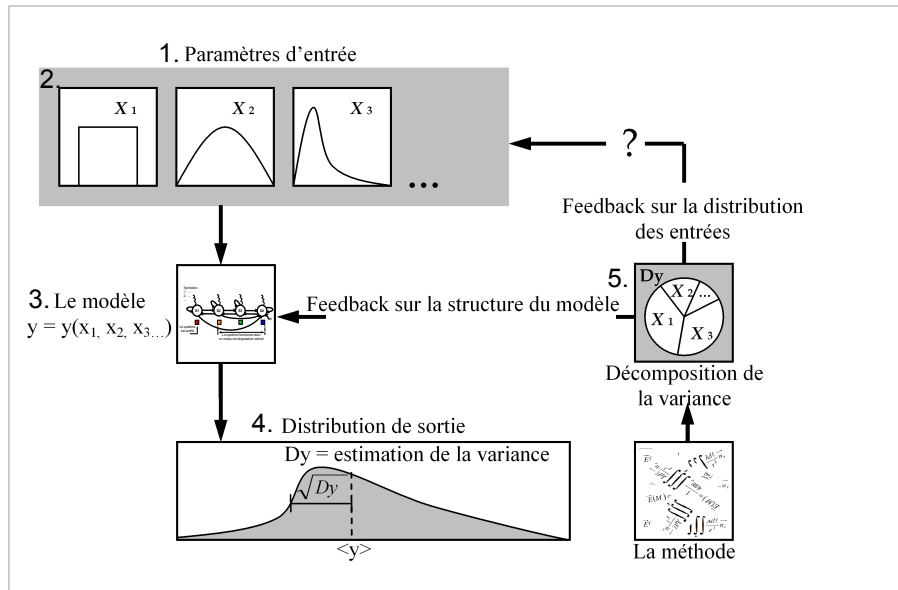


Fig 2.2 – Principe d'analyse de sensibilité d'un modèle [145].

Plans d'expériences, méthode de Morris

Les plans d'expériences permettent d'organiser au mieux les essais qui accompagnent une recherche scientifique ou des études industrielles [73]. Cette méthode s'appuie sur différentes notions :

- la notion d'espace expérimental (très utilisée en modélisation « B » [1]). Il s'agit de définir exactement le domaine d'application de chaque variable afin d'en étudier la valeur de sortie ou « réponse » mesurée ;
- la notion de surface de réponse, est formée par les réponses (sorties du modèle) associées aux domaines des variables ;
- la notion de modélisation mathématique : la simulation des valeurs de sortie permet d'étudier la réponse du modèle sans refaire d'expérimentations.

En 1991, Morris [117] a proposé une mesure de sensibilité pour identifier les variables les plus significatives d'un modèle. Cette mesure est surtout utilisée dans le cadre de modèles déterministes¹ ayant un nombre très élevé d'entrées et dont l'analyse mathématique s'avère assez compliquée. Cette méthode utilise des plans d'expériences basés sur la technique du **One factor At a Time (OAT)**, explicitée ci-dessous. L'analyse des données est basée sur l'observation des « effets élémentaires » sur un échantillon aléatoire et des modifications d'une sortie par rapport à une entrée particulière du modèle. Campolongo et al. [34] ont récemment amélioré la stratégie d'échantillonnage de Morris [117], permettant un coût de calcul

1. Les relations entre les variables sont strictement fonctionnelles. Aucune variable aléatoire n'intervient dans la modélisation

moindre. Cette dernière est employée pour évaluer la sensibilité d'un modèle de réaction chimique du sulfure de diméthyle (DMS), un gaz impliqué dans le changement climatique. Dans sa finalité, cette méthode permet d'identifier les éléments du modèle nécessitant d'être approfondis et ceux pouvant être simplifiés.

Nous n'utiliserons pas cette technique assez lourde à mettre en œuvre. Il nous faudrait en effet formaliser les domaines de fonctionnement de chaque variable et modéliser nos MMC sous forme de preuves mathématiques.

Les méthodes de « screening » (criblage)

Saltelli et al. [145] présentent des méthodes dites de « screening », qui consistent à analyser qualitativement l'importance des variables d'entrée sur la variabilité de la réponse du modèle. Elles permettent de hiérarchiser les variables d'entrée en fonction de leur influence sur la variabilité de la réponse. Elle permet de déterminer le poids de chaque variable, pour ensuite les classer par ordre d'importance [150].

Il existe d'autres techniques comme le **One factor At a Time (OAT)** dont le principe consiste à modifier chaque entrée du modèle étudié de +10% et -10% par rapport à leur valeur réelle. L'effet de ces petites variations est analysé en sortie du modèle. Cela nous donne un pourcentage de variation appelé « Indice de Sensibilité » qui nous donne les sorties les plus sensibles aux paramètres d'entrée. Le lecteur curieux trouvera quelques applications supplémentaires dans [169] et [4].

Borne de Cramér-Rao semi-paramétrique

Cette théorie fournit une borne inférieure pour la variance de la loi d'estimation [64]. Cette borne s'apparente à une mesure d'efficacité de la fonction à estimer. La borne de Cramér-Rao² [49], [161] ou FDCR, en l'honneur de Fréchet, Darmois, Cramér et Rao, est une inégalité indiquant que la variance de tout estimateur non biaisé est minorée par l'inverse de la matrice d'information de Fisher $\mathcal{I}(\theta)$:

$$\text{var}(\hat{\theta}) \geq \mathcal{I}(\theta)^{-1}, \quad (2.1)$$

$\hat{\theta}$ représente un estimateur non-biaisé, d'un paramètre θ inconnu mais déterministe et $f(x, \theta)$, la densité de probabilité des observations $X = (x_1, \dots, x_n)$. La matrice d'information de Fisher [171] est définie par :

$$\mathcal{I}(\theta) = \mathbb{E}_{\theta} \left[\left(\frac{\partial \log f(x, \theta)}{\partial \theta} \right)^2 \right]. \quad (2.2)$$

L'intérêt de cette borne est qu'elle fournit un indicateur d'efficacité de l'estimateur, lorsque l'on veut estimer des paramètres d'un modèle dans un environnement stochastique.

2. Historiquement, c'est Fréchet qui l'a découverte en premier : « Sur l'extension de certaines évaluations statistiques au cas de petits échantillons [68] »

Pour estimer les paramètres non déterministes d'un modèle, Hijazi et al. [82] utilisent la BCRB (Borne Cramér-Rao Bayésienne) qui permet d'analyser des performances optimales en terme d'Erreur Quadratique Moyenne³ d'un estimateur.

Dans notre problématique de comparaison de modèles, cette borne ne nous donnerait qu'une information de base sur l'efficacité de nos modèles. Cette technique est notamment utilisée par Cappe et al. [36]. Ils utilisent des chaînes de Markov à temps continu pour l'observation irrégulière de l'état d'une file d'attente.

Théorie de l'évidence ou des fonctions de croyances

Cette théorie introduite par Dempster [55] et Shafer [152] permet de modéliser l'incertitude de sources de données. Elle est ensuite enrichie par Smets [158] avec le modèle des croyances transférables TBM (Transferable Belief Model). L'approche de cette théorie à base de MMC est utilisée pour le diagnostic et la prédiction de défaillances dans le domaine de la maintenance de systèmes dynamiques [151] et [138]. Utilisés dans le cadre de la mise en place d'une CBM⁴ [167] (voir § 1.4, p. 16), les MMC fournissent une estimation de la durée de vie utile restante avant la défaillance d'un système donné. Cette théorie permet aussi de traiter les données incomplètes, imprécises, incertaines, redondantes, contradictoires [22] d'un système. Basée sur la théorie de l'évidence, Dezert et Smarandache l'ont implémentée de nouveau avec la théorie de « Dezert-Smarandache » (DSmT) [59], [157] et [165]. Celle-ci permet de traiter les imprécisions, les conflits et les informations incertaines des sources de données. Elle est notamment utilisée dans l'analyse du mouvement humain [139].

Il s'agit d'établir un degré de confiance appelé « distributions de masses d'évidence élémentaires » (basic belief assignment) aux sources d'informations (observations, mesures, etc.). En pratique, il y a généralement plusieurs types de sources d'informations. Toute la difficulté du processus décisionnel pour l'expert consiste donc à synthétiser toutes ces informations. C'est la notion de « fusion » d'informations [22] qui consiste à combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision.

Pour en savoir plus, sur le processus décisionnel, nous pouvons nous intéresser aux travaux d'Isabelle Bloch [21]. Elle fait un comparatif sur les trois principales approches de fusion de données : l'approche probabiliste, la théorie de l'évidence et la théorie des possibilités.

Cette approche, complexe à mettre en œuvre pour une orientation « industrielle », ne sera pas développée dans ce manuscrit.

Méthode d'estimation de Sobol

Très utilisé dans la littérature, l'indice de Sobol est une assez bonne estimation de la sensibilité d'un modèle. Son inconvénient majeur est sa complexité et son coût important en calcul, du fait de la décomposition multi-dimensionnelle de la variance.

Indice de Sobol [159] :

$$S_j = \frac{Var[\mathbb{E}(Y|X_j)]}{Var(Y)}, \quad (2.3)$$

3. Elle est utilisée pour comparer plusieurs estimateurs

4. Condition Based Maintenance = maintenance basée sur des conditions

$$0 \leq S_j \leq 1.$$

- $\mathbb{E}(Y|X_j)$ est la fonction de X_j qui approche le mieux Y (expected fonction) ;
- $Var[\mathbb{E}(Y|X_j)]$: fluctuation de la sortie si elle était fonction des X_j ;
- $Var(Y)$: pour la normalisation par la fluctuation totale.

Méthode GLUE (Generalized Likelihood Uncertainty Estimation)

La méthodologie GLUE [18] est fondée sur le concept « d'équifinalité » de modèles (i.e. atteindre le même objectif final à partir de modèles différents). L'idée simple d'une solution unique et optimale est abandonnée. Ce principe d'équifinalité provient de la connaissance imparfaite du modèle étudié. En analysant les paramètres et les variables, la méthode GLUE permet de sélectionner les meilleurs modèles en définissant un degré d'appartenance (i.e. dans quelle mesure le modèle se rapproche de la réalité). Ce degré d'appartenance est lui même estimé à l'aide du maximum de vraisemblance [119], [20].

Dans notre étude, nous nous inspirons de cette méthode pour déterminer les modèles les plus pertinents.

C'est aussi une méthode statistique permettant de quantifier l'incertitude des prédictions du modèle. L'importance de ces incertitudes sera abordée dans les sections suivantes.

2.3.3 Les enjeux de la prise en compte des incertitudes

L'analyse de sensibilité fut créée à l'origine pour traiter les problèmes d'incertitude des entrées d'un modèle [145]. « L'analyse d'incertitude consiste à propager les incertitudes des paramètres pris en compte dans la démarche sur le résultat final afin de lui affecter un intervalle d'incertitude ou de confiance » (source : AFSSET).

Le problème majeur lié aux différentes sources d'incertitude (étape B de la Figure 2.3, p. 36) concerne la fiabilité dans la prédiction du modèle.

Cette prise en compte dans un processus industriel permet selon Iooss [90] :

- l'optimisation de la sûreté, en délimitant qualitativement les défaillances du système ;
- l'optimisation conceptuelle, pour améliorer le système et le modèle.

2.3.4 Incertitudes de conception

Dans cette section, les différentes incertitudes liées à la conception d'un modèle vont être discutées. Les modèles décrivant des phénomènes physiques contiennent deux types d'incertitude : celle liée à la description physico-mathématique du système et celle liée aux données d'entrée. D'autres types d'incertitudes liés à d'autres types d'environnement seront évoqués. L'environnement stochastique qui nous intéresse plus particulièrement dans la modélisation sous forme de MMC sera abordé.

Incertitudes physico-mathématiques

Cette incertitude est liée au passage du phénomène physique au modèle mathématique. Elle est d'une part, liée à l'interprétation humaine du phénomène qui engendre des imperfec-

tions dans la conception du modèle. Elle est d'autre part, inhérente à l'Ingénierie Dirigée par les Modèles MDE (Model Driven Engineering). Le MDE propose, en effet, la définition d'un modèle en se limitant à un niveau d'abstraction fini. Cette limite engendre inévitablement des incertitudes de description. En effet, « Tous les modèles sont faux, mais certains sont utiles » [27]. Selon Burnham et al. [31], le choix d'un modèle est perçu comme un moyen d'approcher la réalité plutôt que de l'identifier complètement.

Incertitudes des données d'entrée

La seconde source d'incertitude est liée aux données d'entrée du modèle. Le lecteur pourra se référer à l'article de McKay et al. [114] qui présente des techniques pour évaluer les effets des incertitudes d'entrée sur un modèle. Ce sont des incertitudes stochastiques dues à la variabilité de l'environnement et des conditions de mesures (température, hygrométrie, etc.).

Incertitudes probabilistes

Dans les années 1970, les ingénieurs probabilistes ont commencé à prendre en compte les incertitudes apparaissant dans la modélisation des systèmes physiques et ont étudié l'impact de ces incertitudes sur la réponse des systèmes. Une étude récente d'EDF division R&D [52] et [53] sur les incertitudes industrielles propose une analyse pertinente des incertitudes, dans un contexte industriel. Les principales étapes de cette analyse sont résumées Figure 2.3, p. 36.

Tarentola et al. [166] ont récemment défini un nouvel indice de sensibilité basé sur la variance dite « total order » permettant de quantifier la propagation de l'incertitude d'un modèle d'entrée sur un modèle de sortie en tenant compte de l'interaction entre le modèle d'entrée et les autres paramètres du modèle.

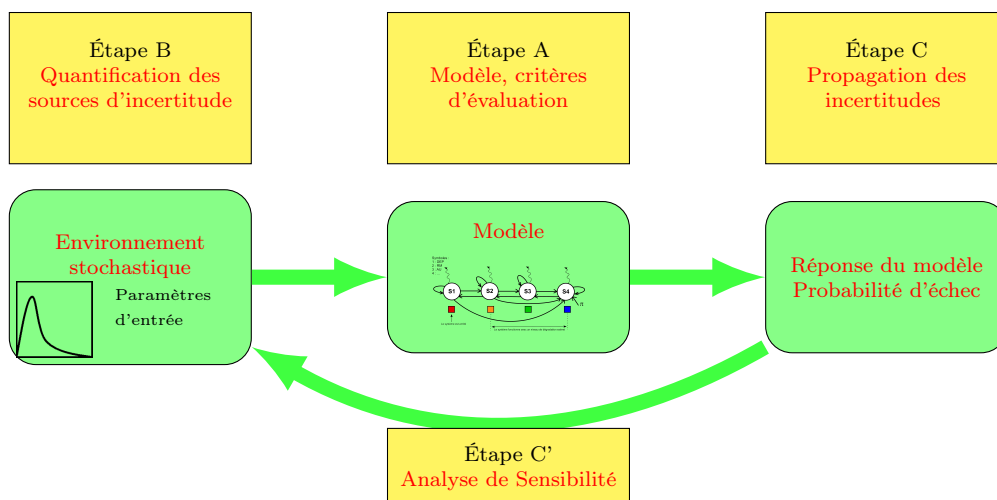


Fig 2.3 – schéma général d'analyse de l'incertitude probabiliste inspiré de [163].

- Étape A : spécification du problème : définir le modèle (ou la séquence de sous-modèles dans le cadre de systèmes complexes) et ses critères d'incertitude permettant une éva-

luation du système à étudier ;

- Étape B : quantification des sources d'incertitude i.e. identifier les paramètres connus ou inconnus. L'introduction d'un processus stochastique est parfois nécessaire dans certains cas d'évolution de variables dans le temps. Cette étape peut-être aussi assimilée à la phase d'identification de paramètres, ou du calage/qualification du modèle ;
- Étape C : la « propagation des incertitudes » : calcul de la mesure d'incertitude au travers du modèle i.e. caractériser la réponse stochastique du modèle, en rapport avec les critères d'évaluation définis à l'étape A ;
- Étape C' : « Analyse de sensibilité » ou comment réagit le modèle suite à un ensemble d'excitations externes ou internes.

Incertainces d'apprentissage

Prenons par exemple l'apprentissage par « Version Space » (Mitchell [116]). L'avantage est : plus on apprend, plus on « colle » aux données. Mais le risque est, qu'à la fin, tous les exemples sont appris par cœur, y compris le bruit (les informations contenues dans l'exemple ne servant pas dans l'étude). C'est le phénomène de sur-apprentissage. Pour éviter ce piège, le critère d'arrêt de l'apprentissage doit être défini. Soit $erreur_{app}(h)$ l'erreur commise par l'hypothèse h sur les données d'apprentissage et $erreur_D(h)$ l'erreur commise par h sur la distribution totale des données.

Définition : on dira que h sur-apprend les données d'apprentissage s'il existe une hypothèse h' telle que :

$$erreur_{app}(h) < erreur_{app}(h'), \quad (2.4)$$

et

$$erreur_D(h) > erreur_D(h'). \quad (2.5)$$

Mesure d'erreur stochastique de base

Pour une série de n mesures $x_1, x_2, \dots, x_i, \dots, x_n$, la valeur moyenne ou moyenne arithmétique est définie par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (2.6)$$

et la variance empirique d'échantillon par :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (2.7)$$

On montre alors que S^2 est un estimateur biaisé de σ^2 (variance de la population dont est extrait l'échantillon). De manière à avoir un estimateur non biaisé, on construit

$$S^{*2} = \frac{n}{n-1} S^2. \quad (2.8)$$

(Car $\mathbb{E}(S^2) = \frac{n-1}{n} \sigma^2$ que l'on retrouve dans l'équation (2.8)).

Si nous calculons la variance sur $\bar{x}$:

$$Var(\bar{x}) = \frac{\sigma^2}{n}. \quad (2.9)$$

L'incertitude sur la moyenne de Pibouleau [129] :

$$\Delta \bar{x} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{1}{n \cdot (n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2.10)$$

Si l'écart type est fini, l'incertitude de Pibouleau [129] converge inexorablement vers zéro lorsque n tend vers l'infini [76].

Propagation des incertitudes

Dans tout modèle, les variables d'entrée issues de mesures ou d'estimations, sont soumises à des transformations mathématiques. La propagation des incertitudes d'un modèle est le calcul ou l'estimation de l'incertitude induite par cette transformation.

Au travers de l'article [134], Pugol et al. proposent d'évaluer des critères d'« incertitude », des critères de « fiabilité des performances » et des critères de « robustesse des performances », ce dernier étant justement sensible aux incertitudes. La méthode utilise un critère de « *quantile* » pour définir des intervalles de confiance. Pour le quantile q_α , un intervalle de confiance est donné par :

$$P(q_\alpha \in [L_1, L_2]) = \sum_{i=L_1}^{L_2-1} C_n^i \cdot \alpha^i \cdot (1-\alpha)^{n-i}. \quad (2.11)$$

L'estimation des incertitudes dans notre problématique, permettra de trouver les architectures des modèles les plus pertinentes, sous la forme d'intervalle de confiance.

2.4 Méthodes de mesures de pertinence de modèles

Dans ce paragraphe, une étude bibliographique des principales méthodes de mesures de pertinence de modèles utilisées dans la littérature est réalisée. Nous terminons par une discussion sur ces différentes méthodes d'évaluation (§ 2.6, p. 55) en expliquant notre choix sur les méthodes utilisées.

2.4.1 Introduction

Selon Stoica et Babu [162], la problématique de la sélection d'un modèle est basée sur la minimisation d'un critère de pénalité. Les premiers critères qui apparaissent dans la littérature sont l'*AIC* : l'Akaike Information Criterion [3] [38] [153] [32], le *BIC* : *Bayesian*

Information Criterion [148] [102], le *MDL* : *Minimum Description Length* [140] [50] puis le C_p de Mallows [112]. De nombreux travaux théoriques ont été réalisés sur leurs propriétés statistiques afin de les adapter à des modèles spécifiques. Prenons comme exemple l' AIC_c , qui est une version corrigée du critère AIC d'Hurvich [86] [54] et [32]. Nous trouvons également Sugiura [164] avec le $c - AIC$ pour les petites tailles d'échantillons par rapport au nombre de paramètres à estimer, l' $AICR$ [141] [95] pour une régression d'erreurs non gaussiennes, le $QAIC$ [32] et le $c - QAIC$ [155] pour les données *sur-dispersées*. De nouveaux outils statistiques apparaissent comme par exemple la technique *PAC-Bayésienne*, pour l'estimation et la prédiction de modèles de grandes dimensions [74]. Elle propose ainsi un oracle pour l'estimation et la reconstruction de modèle.

2.4.2 Concepts de base

Nous commençons par présenter les éléments intervenant dans les différentes méthodes de mesures de pertinence de modèles que nous allons développer.

L'échantillon

Les échantillons ou observations peuvent être divisés en trois types [61] :

- catégorie : les observations appartiennent à un nombre limité de catégories qui n'ont pas d'échelle évidente ;
- discrètes : il y a une échelle réelle mais les valeurs de l'échantillon ne sont pas toutes possibles (par exemple : nombre d'espèces dans un échantillon) ;
- continues : où toute valeur est théoriquement possible, seulement limitée par l'appareil de mesure (longueur par exemple).

Les MMC que nous étudions sont de type discret i.e. le champ d'application des valeurs des observations n'utilise pas une échelle continue.

Test d'hypothèse

Un test statistique d'hypothèse est défini par Steinebach [160] comme une méthode de prise de décisions statistiques utilisant des données expérimentales. Le concept est assez simple : lorsque l'on effectue un test statistique, on teste la probabilité qu'une hypothèse est correcte. Si cette probabilité est faible, alors l'hypothèse est considérée comme fausse et elle est rejetée en faveur de l'hypothèse la plus intéressante (souvent notée H_1). L'hypothèse nulle (souvent notée H_0) est l'hypothèse où rien ne se passe. Par exemple : pour un échantillon donné, si l'on rejette l'hypothèse nulle (H_0 = l'échantillon est issue d'une population de même distributions) alors nous pouvons accepter l'hypothèse la plus intéressante (H_1 = l'échantillon est issue d'une population de distributions différentes). Il n'y a pas de test statistique sans hypothèse.

P-value

La p-value permet de définir un seuil au delà duquel le postula d'une égalité ou d'une inégalité entre deux données d'un modèle est vrai. Si cette p-value est inférieure à la valeur du seuil préalablement définie, on rejette ce postula. Il est habituel de prendre une valeur de 5% comme niveau critique.

Échantillonnage

Selon Steinebach et al. [160], les observations doivent être recueillies d'une certaine façon. Ce processus d'acquisition de données est appelé échantillonnage. Bien qu'il existe de nombreuses méthodes pouvant être utilisées pour un échantillonnage, il y a néanmoins quelques règles générales : une des plus évidentes est qu'un grand nombre d'observations est généralement mieux qu'un petit nombre. Il faut ensuite équilibrer l'échantillonnage i.e. prendre un même nombre d'observations pour chaque séquence étudiée. La plupart des tests statistiques supposent que les échantillons sont pris au hasard.

Statistiques

Les statistiques permettent en général d'obtenir un résultat simple, issu de la manipulation d'observations complexes. Selon [61], il y a plusieurs catégories de statistiques :

- les statistiques descriptives : très simples à mettre en œuvre, elles permettent de résumer des ensembles de données, elles ne doivent pas être négligées ;
- tests de différence : permettent de faire la différence entre 2 groupes d'échantillons. L'hypothèse nulle est donc ici : « le groupe 1 et le groupe 2 ne sont pas différents » ;
- tests de relation : permettent de répondre à la question : A est-il associé à B ? Ou encore avec l'hypothèse nulle : A n'est pas associé à B. Ce test est divisible en test de *corrélation* et test de *régression*, selon le type d'hypothèse à l'étude. La *corrélation* est un test pour mesurer le degré auquel un ensemble d'observations varie en fonction de l'autre : elle ne veut pas dire qu'il n'y a aucune relation de cause à effet. Le test de *régression* est utilisé pour ajuster une relation entre deux variables, cette relation étant prédite à partir de l'autre ;
- tests d'investigation de données : contrairement aux tests précédents, ces tests n'ont pas besoin d'hypothèse. Dans le cas de groupes trop nombreux de population, il est préférable d'utiliser une technique multivariée afin de faire apparaître des relations entre groupes d'échantillons.

2.4.3 Méthodes avec score d'intérêt

Maintenant que les concepts de base sont établis, les méthodes de sélection de modèles avec score d'intérêt vont pouvoir être présentées.

Mesure de l'entropie de Shannon

Selon Bouquet [25], l'entropie de Shannon est une fonction mathématique qui permet de mesurer la quantité d'informations contenue dans une source d'informations. Cette source peut être un texte écrit dans une langue donnée, un signal électrique ou encore un fichier informatique quelconque, etc.

Wolsztynski [187] propose de minimiser l'entropie des variables aléatoires de ces modèles, dans le cas discret, afin d'en estimer les événements rares. C'est effectivement cette différence principale, ainsi que les autres critères de mesure de dispersion classiques que nous allons utiliser.

Définition de l'entropie

L'entropie de Shannon est définie dans [48] et [83] comme suit :

$$H(S) = - \sum_{i=1}^n P_i \log P_i, \quad (2.12)$$

P_i est la probabilité moyenne de voir apparaître le symbole i dans S .

Propriétés formelles de l'entropie de Shannon

Deux propriétés nous intéressent dans la théorie de l'entropie :

- *la minimalité* : une distribution « pure » (un seul symbole) a une entropie nulle : si un seul symbole est représenté, l'impureté est nulle. On a donc une première condition de minimalité : l'entropie d'une distribution est nulle si cette distribution est pure.

$$\begin{aligned} H([, , , 1,]) &= - \sum_{i=1}^5 P_i \log P_i \\ &= -(0. \log 0 + 0. \log 0 + 0. \log 0 + 1. \log 1 + 0. \log 0) \\ &= 0. \end{aligned}$$

- *la maximalité* : à l'inverse, une distribution mélangée est « impure » : la valeur de l'entropie, lorsque plusieurs symboles sont représentés, doit donc être élevée. C'est la condition de maximalité : l'entropie d'une distribution est maximale lorsque tous les symboles sont représentés dans les mêmes proportions. Pour de plus amples informations, le lecteur se référera à Beirlant et al. [15] qui présentent un état de l'art concernant les méthodes d'estimation de l'entropie et leurs propriétés.

Nous utiliserons les deux propriétés précédentes dans ce manuscrit. Dans notre cas d'étude, nous aurons à déterminer les observations les plus discriminantes de notre modèle en utilisant la propriété de minimalité. La maximalité sera utilisée pour trouver les états du système où le maximum d'observations est représenté.

Principes de l'entropie maximale

Les deux principes de maximisation d'entropie de Jaynes [94] et [2] sont les suivants :

- principe d'assignation des probabilités à une distribution lorsque nous ne disposons pas d'informations complètes sur elle ;
- de toutes les distributions de probabilité qui satisfont les contraintes, choisir celle qui présente l'entropie maximale au sens de Shannon [154].

Chandrasekaran [43] utilise ce 2^{ème} principe pour la sélection de modèles, ainsi que [5] pour construire des modèles de plus en plus précis simplement en rajoutant de l'information. Notre démarche consiste à comparer la moyenne des entropies des différents modèles. La valeur d'entropie moyenne serait alors maximale pour les modèles les plus pertinents.

Notion de filtre entropique

Les valeurs extrêmes de l'entropie peuvent perturber le calcul de l'entropie moyenne d'un modèle. Il est donc préférable d'éliminer ces valeurs pour ce calcul. Cette démarche est notamment utilisée au travers des algorithmes d'apprentissage ID3 (voir annexe A.1, p. 141) [135], [25] et C4.5 [136] lors de la création d'arbre de décision en éliminant récursivement l'attribut ayant une entropie nulle. Afin d'évaluer au plus juste notre modèle, nous proposons donc d'éliminer les symboles totalement discriminés d'entropie nulle et les symboles à large spectre où l'entropie est maximale. Cette opération d'élimination sera nommée par la suite : « **filtre entropique** ». Nous utiliserons cette notion pour déterminer les observations inutiles dans les bases de données, ainsi que celles nous apportant une quantité importante d'informations en terme d'entropie.

Maximum de vraisemblance

Pour un modèle statistique P_μ donné, et étant donnée la séquence d'observations X , la probabilité de son apparition suivant P peut être mesurée par $f(X, \mu)$ qui représente la densité de X où μ apparaît. Puisque μ est inconnue, il semble alors naturel de favoriser les valeurs de μ pour lesquelles $f(X, \mu)$ est élevée : c'est la notion de la *vraisemblance* de μ pour l'observation X , sous la condition d'indépendance des observations.

Définition 2.4.3.1 *Expression de la vraisemblance V :*

$$V(x_1, \dots, x_n; \mu) = \prod_{i=1}^n f(x_i; \mu), \quad (2.13)$$

μ est l'espérance mathématique des x_i .

Définition 2.4.3.2 *Le maximum de vraisemblance :*

$$\hat{V}(x_1, \dots, x_n; \mu). \quad (2.14)$$

Une transformation strictement croissante ne change pas un maximum. Le maximum de vraisemblance peut donc aussi s'écrire :

$$\log(V(x_1, \dots, x_n; \mu)). \quad (2.15)$$

Donc

$$\log(V(x_1, \dots, x_n; \mu)) = \sum_{i=1}^n \log(f(x_i; \mu)). \quad (2.16)$$

Définition 2.4.3.3 Pour un échantillon discret :

$$f(X; \mu) = P_\mu(X = x_i), \quad (2.17)$$

$P_\mu(X = x_i)$ représente la probabilité discrète où μ apparaît.

Définition 2.4.3.4 Maximum de vraisemblance pour un échantillon discret $P_\mu(x_i)$ qui représente la probabilité discrète où μ apparaît :

$$\log(V(x_1, \dots, x_n; \mu)) = \sum_{i=1}^n \log(P_\mu(x_i)). \quad (2.18)$$

En pratique, on maximise le logarithme de la fonction de vraisemblance pour comparer plusieurs modèles.

Pour notre problématique de MMC où les états sont cachés, la fonction de vraisemblance ne possède pas d'expression analytique simple. Pour résoudre le problème de recherche de maxima, nous utilisons l'algorithme Expectation-Maximization (voir § 2.4.3, p. 47). Plus précisément, nous emprunterons les données de cet algorithme calculé dans la thèse de Vri-gnat [177], pour faire des estimations de lois de survie. Ainsi, ce principe de maximum de vraisemblance sera nécessaire dans la validation des architectures des modèles les plus pertinents. Ce critère nous aidera à déterminer la meilleure topologie, le meilleur algorithme d'apprentissage, la meilleure distribution entre autre.

Critères d'Akaike, de Bayes et d'Hannan-Quinn

D'après Ash [7], l'entropie d'une variable aléatoire est une mesure de régularité. Nous pouvons aisément étendre cette notion à un modèle possédant plusieurs variables aléatoires. Dans la littérature, le critère d'Akaike (*AIC*) [153] est souvent associé à un autre critère connu, appelé critère d'information de Bayes (*BIC*) [45], [102]. Dans son rapport de recherche, Lebarbier [107] décrit avec précision les hypothèses nécessaires à son application.

Définition 2.4.3.5 *AIC* (*Akaike Information Criterion*) :

$$AIC = -2 \cdot \ln V + 2k, \quad (2.19)$$

où k est le nombre de paramètres, $2k$ représente la pénalité, V est la vraisemblance.

Le modèle à retenir est celui qui montre l' AIC le plus faible. L' AIC utilise le principe du maximum de vraisemblance (équation 2.18). Il pénalise les modèles comportant trop de variables, et évite le sur-apprentissage. Hurvich et Tsai, [87] préconisent d'utiliser l' AIC corrigé lorsque le nombre de paramètres k est grand par rapport au nombre d'observations n ($\frac{n}{k} < 40$).

$$AIC_c = AIC + \frac{2.k.(k+1)}{n-k-1}. \quad (2.20)$$

Ce critère est souvent présenté avec celui de Schwarz : le BIC , qui pénalise davantage le sur-paramétrage. Le critère BIC a été introduit dans Schwarz [148] et Kapetanios [102]. La différence entre les deux critères concerne le terme de correction.

Définition 2.4.3.6 *BIC (Bayesian Information Criterion) ou Schwarz Criterion :*

$$BIC = -2. \ln V + k. \ln n, \quad (2.21)$$

où k est le nombre de paramètres libres du modèle de Markov, n est le nombre de données, $k. \ln n$ est le terme de pénalité [10].

Comparaison des deux critères : choisir entre ces deux critères revient à choisir entre un modèle prédictif et un modèle explicatif [107]. Il permet de vérifier la validité d'un modèle mais surtout de comparer plusieurs modèles entre eux.

Définition 2.4.3.7 φ ou HQC (Hannan-Quinn information Criterion) est défini dans [78] et [39] par :

$$\varphi = -2. \ln V + k. \ln(\ln n), \quad (2.22)$$

où k est le nombre de paramètres libres du modèle de Markov, n est le nombre de données, $k. \ln(\ln n)$ représente la pénalité du critère.

Ce critère apparaît comme un compromis entre AIC et BIC . Par contre, Yu et al. [191] montrent que le temps de calcul est très élevé pour les grands ensembles de données, ce qui peut poser des problèmes pour le choix d'un modèle parmi un ensemble de modèles candidats.

Critère de Kullback-Leibler

L' AIC est relié à l'information de Kullback-Leibler [104]. Pour le cas de deux distributions de probabilités discrètes P et Q , la divergence ou information de Kullback-Leibler est :

$$D_{KL}(P \parallel Q) = \sum_{i=1}^n P_i. \log_2 \frac{P_i}{Q_i}. \quad (2.23)$$

Aussi appelé entropie relative, ce critère permet de mesurer l'écart entre deux distributions de probabilités. L'objectif dans notre cas est de mesurer l'écart ou divergence d'un modèle par rapport à un autre. Ce critère nous permettra d'établir un classement entre nos modèles candidats.

Mesure du MDL

En 1978, Jorma Rissanen [140], a développé l'idée de minimiser l'erreur de développement sur la conception d'un modèle, en pénalisant celui-ci en fonction de la longueur de sa description. A cette époque, la seule autre méthode qui a réussi à empêcher le sur-apprentissage par la pénalisation était le critère d'information d'Akaike (AIC) vu précédemment. La mesure choisie pour évaluer ce travail de modélisation est une mesure de type « Minimum Description Length (MDL) ». La mesure MDL est une formalisation du principe du rasoir d'Occam⁵ dans laquelle la meilleure hypothèse pour un ensemble de données est celle qui conduit à la plus grande compression des données. Le principe est de choisir le modèle qui donne la meilleure compression en tenant compte de l'erreur, c'est à dire de ce qui n'est pas expliqué par le modèle ; soit la minimisation de la « longueur » du modèle + la « longueur » des erreurs. La « longueur » se mesure en bits.

C_p de Mallows

Selon Joshi [96], le critère de choix C_p est une méthode de choix du modèle de régression en fonction du nombre de variables explicatives entrant dans le cas d'une régression linéaire. Un faible C_p correspond à un meilleur modèle dans le sens qu'il représente la somme des résidus (SC résiduelle) la plus faible et la moins pénalisée par le nombre de paramètres entrant dans le modèle. Ce critère permet d'optimiser l'erreur d'apprentissage du modèle. C'est la différence entre l'erreur de généralisation (en espérance sur les données) et l'erreur d'apprentissage (celle du modèle choisi). Le C_p de Mallows permet d'estimer cette erreur sans utiliser les données elles mêmes, mais en utilisant de l'information sur la richesse du modèle et le nombre de données n .

$$C_p = \frac{(SC \text{ Résiduelle})_p}{\sigma^2} - (n - 2p), \quad (2.24)$$

p : le nombre de paramètres du modèle,

n : le nombre d'observations,

σ^2 : la variance,

$(SC \text{ Résiduelle})_p$: SCR du modèle à p paramètres.

- si le modèle est juste bien : $SCR = (n - p)\sigma^2$ alors : $C_p = p$,
- si le modèle est sous-paramétré : $SCR > (n - p)\sigma^2$ alors : $C_p > p$,
- si le modèle est sur-paramétré : $SCR \approx (n - p_{optimal})\sigma^2$ alors : $C_p = p + 2\alpha > p = p_{utilisé}$.

Information mutuelle

Dans le cas continu :

$$I(X, Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy. \quad (2.25)$$

5. « C'est en vain que l'on fait avec plusieurs ce que l'on peut faire avec un petit nombre »

Dans le cas discret :

$$I(X, Y) = \sum_{i,j} P(X = i, Y = j) \log \frac{P(X = i, Y = j)}{P(X = i) P(Y = j)}. \quad (2.26)$$

Propriété 2.4.3.1 $I(X, Y) = 0$ ssi X et Y sont des variables aléatoires indépendantes.

Dans son étude sur l'estimation rapide de modèles de Markov semi-continus discriminants, Linares et al. [110] nous donnent une méthode d'estimation de SHMM⁶ par maximisation de l'information mutuelle. Le principe général est de minimiser le risque d'erreur en maximisant l'écart de vraisemblance (voir § 2.4.3, p. 42) entre la bonne transcription et les mauvaises.

Dans ce manuscrit, nous utilisons cette information mutuelle pour trouver d'éventuelles dépendances entre les observations puis entre les états et symboles.

Nouveaux critères : AIC3, AICu, AICp et CAIC

Selon Pigeau [130], le critère AIC est obtenu en se basant sur la théorie classique du test des hypothèses. Ce critère n'est donc pas théoriquement pertinent pour déterminer le nombre de composantes dans un modèle. Pour résoudre ce problème, une variante de AIC, appelée AIC3, a été proposée par Bozdogan [28].

La pénalisation avec les paramètres libres du modèle est ici plus forte que pour le critère AIC.

- AICu : le critère AICu McQuarrie [115] est utilisé pour la sélection d'un modèle d'ordre infini ;
- AICp : Vandewalle [172] propose un nouveau critère AICp, qui cherche à évaluer les performances en prédiction d'un « modèle génératif appris ». Ce critère est composé de la vraisemblance pénalisée par une quantité nouvelle qui s'interprète comme la dimension prédictive du modèle considéré. Les modèles génératifs consistent en une modélisation de la vraie distribution de probabilité p par une loi paramétrique ;
- CAIC : enfin, le CAIC est un coefficient important, car il tient compte à la fois du degré d'ajustement du modèle et du nombre de degrés de liberté. Ceci permet d'estimer quel modèle semble le plus approprié, i.e. lequel devrait avoir les plus petites valeurs de CAIC (Bentler [16]).

Critère ICOMP

Le critère ICOMP (Informational Complexity Criterion) (Bozdogan [29]) présente la particularité d'utiliser une mesure de complexité non linéaire. Celle-ci se base sur la matrice d'information de Fisher et le nombre de paramètres du modèle. Le calcul de ce critère est néanmoins beaucoup moins évident que celui des critères précédents : la matrice de Fisher est difficile à obtenir.

6. Semi Continuous Hidden Markov Models

Critères de classification

Ces critères permettent de sélectionner des modèles avec des composantes distinctes et des observations bien regroupées. Pigeau [130] présente un état de l'art de ces critères pour la sélection de modèle pour la classification. Ce sont les critères suivants : NEC (Normalized Entropy Criterion), PC (Partition Coefficient), MIR (Minimum Information Ratio) et LP (logarithme de la probabilité de la partition). Le critère de BEC (Bayesian Entropy Criterion) quant à lui, est utilisé en classification supervisée [24].

Critère ICL

Le critère ICL (Integrated Completed Likelihood) est proposé par Biernacki [19]. Ici, la vraisemblance est pénalisée par la complexité du modèle et le critère de classifiabilité E . Il est défini à partir du critère BIC :

$$ICL = \frac{1}{2}BIC + E, \quad (2.27)$$

où E est un critère de classifiabilité.

Détection des valeurs marginales

Dans la même problématique, Low-Kam et al. [111] nous présentent une approche basée sur des modèles de Markov d'ordre variable afin de mesurer les similarités entre objets séquentiels. Cette mesure permet entre autre la détection de valeur aberrante (une observation qui s'écarte tellement des autres qu'elle est susceptible d'avoir été générée par un mécanisme différent). Dans le cas de notre étude, nous pourrions avoir une estimation des séquences atypiques des modèles de Markov.

Ceci permettra ensuite de déterminer les « mauvais modèles » en utilisant la procédure simplifiée de Ueda [170] et Kadota et al. [99] :

$$U = \frac{1}{2}AIC = n \cdot \log(\sigma) + \sqrt{2} \cdot s \cdot \frac{\log(n!)}{n}, \quad (2.28)$$

où s est le nombre de valeurs aberrantes candidats.

Algorithme Expectation-Maximization (EM)

L'algorithme EM, proposé par Dempster et al. [56] est une classe d'algorithmes qui permet de trouver le maximum de vraisemblance des paramètres de modèles probabilistes lorsque le modèle dépend de variables non observables comme dans notre problématique de MMC.

Voici le principe en deux étapes :

- étape E : évaluation de l'espérance, où l'on calcule l'espérance de la vraisemblance en tenant compte des dernières variables observées ;
- étape M : maximisation où l'on estime le maximum de vraisemblance des paramètres en maximisant la vraisemblance trouvée à l'étape E.

On utilise ensuite les paramètres trouvés à l'étape M comme point de départ d'une nouvelle phase d'évaluation de l'espérance, etc.

L'algorithme EM, bien que très performant et souvent simple à mettre en œuvre, pose quand même parfois quelques problèmes qui ont donné lieu à des développements complémentaires. Parmi ceux-ci, nous évoquerons un développement appelé GEM (Generalized EM) [56] qui permet de simplifier le problème de l'étape maximisation. Un autre, appelé CEM (Classification EM) [42] permet de prendre en compte l'aspect classification lors de l'estimation. Un dernier, SEM (Stochastic EM) [40] et [41] dont l'objectif est de réduire le risque de tomber dans un optimum local de vraisemblance.

Récemment amélioré, Huda et al. [85] proposent un algorithme hybride basé sur une méta-heuristique SAS (Simulated Annealing Stochastic). Cette étape stochastique supplémentaire qui reformule le processus d'estimation du MMC permet d'empêcher la convergence vers un maximum local. Cet algorithme basé sur des MMC, donne une meilleure précision en reconnaissance de la parole.

2.4.4 Méthodes par apprentissage

Version space, graphe de généralisation/spécialisation

La « version space » ou « espace de versions » est le fruit des travaux de Mitchell [116], [84]. Par l'utilisation d'exemples positifs et négatifs d'observations, on réalise une « version space »⁷ qui permet de créer un modèle de description de ce concept. Nous réalisons ainsi un « graphe de généralisation/spécialisation » à l'aide de l'algorithme IVSA⁸ tiré de [75] (voir annexe A.3, p. 143). L'algorithme « d'élimination » construit de manière incrémentale l'espace de versions à partir d'un espace d'hypothèses H et d'un ensemble d'instances I . Les exemples sont ajoutés un par un. Chaque exemple réduit éventuellement l'espace des versions en supprimant les hypothèses qui sont incompatibles avec l'exemple. Pour chaque nouvel exemple ou séquence d'observations, l'algorithme met à jour la frontière entre sa généralisation ou sa spécialisation.

Certaines conditions sont nécessaires au bon fonctionnement du mécanisme :

- existence d'un nombre fixe d'attributs pour définir les exemples présentés au système ;
- le modèle de description d'un exemple doit être une combinaison des valeurs de ces attributs ;
- en aucune façon, les exemples positifs et négatifs ne doivent être erronés.

Avec les techniques précédentes (apprentissage par analyse des différences), le modèle de description était modifié à chaque fois qu'un nouvel exemple était présenté au système. Pour éviter une mauvaise modification, chaque nouvel exemple ne doit être que légèrement différent du modèle. Ainsi, on évite de faire évoluer le modèle vers des interprétations incertaines. Avec la *version space*, on explore chaque interprétation possible, et ce, tant qu'elle

7. Espace de versions

8. Iterated Version Space Algorithm

reste viable.

Une *version space* est une représentation qui enregistre toutes les informations utiles des exemples fournis au système sans conserver un seul de ces exemples. Une *version space* est une représentation dans laquelle :

- il y a un arbre de spécialisation et un arbre de généralisation ;
- chaque noeud correspond à un modèle de description ;
- la racine de l'arbre de généralisation est un modèle qui accepte tous les exemples ;
- les liens entre les noeuds dénotent des relations de généralisation et de spécialisation.

L'idée de l'apprentissage par *version space* est que la généralisation des modèles spécifiques et la spécialisation des modèles généraux conduisent à un modèle assurément correct qui accepte tous les exemples positifs présentés au système et rejette tous ceux qui sont négatifs. Dans ce manuscrit, nous utilisons cette méthode pour tenter de construire un modèle qui accepte les différentes instances i.e. les chaînes de Markov de différents ordres. La construction de tels arbres va nous permettre de détecter les futures pannes du système.

Arbre de décision

La construction d'arbre de décision utilise la notion d'attribut le plus discriminant. Les données issues de GMAO industrielle nous donnent les tables de contingence à étudier. Nous pouvons alors appliquer plusieurs critères comme l'entropie maximale (voir § 2.4.3, p. 42). Nous utilisons aussi cette technique, qui va nous permettre de faire une classification sur les attributs (ou observations dans notre cas), les plus pertinents.

2.4.5 Méthodes statistiques

Kleijnen [103] disait avec humour : « si la statistique mathématique est utilisée, les statistiques appropriées doivent être utilisées ! ». Selon Easterling et al. [62], l'objectif fondamental de l'analyse des données statistiques est d'extraire et de transmettre « ce que les données ont à dire au sujet de questions diverses ». Dans le cadre de la validation de modèle de simulation, Sargent [146] fait la distinction entre trois types de modèles différents :

- aucune donnée réelle : utilisation de la théorie statistique sur la conception de plan d'expériences (voir § 2.3.2, p. 32) ;
- données d'entrée uniquement : utilisation de tests statistiques d'hypothèses ;
- données d'entrée et de sortie : réalisation d'une analyse de sensibilité (voir § 2.3, p. 30).

Dans notre étude, nous nous situons dans le deuxième cas. Nous étudions uniquement les données d'entrée. Les données de sortie que nous ne connaissons pas, correspondent aux états

cachés du MMC.

Méthodes de Monte-Carlo

Ces méthodes permettent de calculer une quantité inconnue en utilisant une suite de nombres aléatoires. Son intérêt est la convergence plus rapide vers la solution qu'une exploration dit « systématique ». Elles permettent de quantifier des variables en utilisant des outils statistiques. Selon Pardoux [126], cette méthode consiste dans un premier temps à mettre sous la forme d'une espérance, la quantité que l'on cherche à calculer. La deuxième étape consiste à faire une simulation de la variable aléatoire et à calculer $\mathbb{E}(X)$, où X est une variable aléatoire. Pour calculer $\mathbb{E}(X)$, il faut savoir simuler des variables aléatoires indépendantes $X_1, \dots, X_n$, ayant toutes la loi de X . Pour finir, il faut approcher la valeur de $\mathbb{E}(X)$ par :

$$\mathbb{E}(X) \approx \frac{1}{n}(X_1 + \dots + X_n). \quad (2.29)$$

Le théorème de la loi forte des grands nombres permet de justifier la convergence de la méthode et le théorème de la limite centrale précise la vitesse de convergence. Pour plus de détails, le lecteur pourra se référer à la « loi forte des grands nombres, théorème de la limite centrale et méthode de Monte-Carlo » [126].

Test de Kolmogorov-Smirnov

Ce test permet de comparer les distributions de deux échantillons [60], [181], [182], [91] et [125]. C'est un test non paramétrique. Il consiste à comparer la distribution des fréquences d'une variable observée avec la distribution théorique que cette variable aurait si elle était distribuée normalement. On cherche alors l'écart entre la distribution théorique et la distribution observée. On veut tester l'hypothèse nulle H_0 : « il n'y a pas de différence entre les deux échantillons ». Ce test repose sur le fait que si les fonctions de répartition théoriques sont égales, les différences entre les fonctions de répartition empiriques sont faibles.

Pour le lecteur curieux, la preuve du théorème de Kolmogorov-Smirnov est décrite dans [17].

La fonction de répartition empirique d'un échantillon $X_1, \dots, X_n$ est définie par (2.30) :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i \leq x} \quad (2.30)$$

avec $\delta_{X_i \leq x} = \begin{cases} 1 & \text{si } X_i \leq x, \\ 0 & \text{sinon.} \end{cases}$

Avec l'hypothèse nulle H_0 : les deux échantillons suivent la même loi.

La distance de Kolmogorov-Smirnov est définie en (2.31) :

$$D_n = \sup_x |F_{1n}(x) - F_{2n}(x)|. \quad (2.31)$$

Nous utilisons ce test afin d'évaluer la différence entre la fonction de répartition empirique étudiée et les lois de distributions testées.

Test d'Aspin-Welch

Un autre test d'adéquation qui prends en compte les moyennes des deux échantillons. L'hypothèse nulle H_0 est la même que celle du test de Kolmogorov-Smirnov (« il n'y a pas de différence entre les deux échantillons »).

Le test d'Aspin-Welch [184], [185], [79] est défini par :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \quad (2.32)$$

$$\sigma^2 = \frac{n_1 \sigma_1^2 + n_2 \sigma_2^2}{n_1 + n_2 - 2}, \quad (2.33)$$

- $\bar{x}_i$: moyenne de l'échantillon,
- σ : la variance des deux échantillons,
- σ_i : les variances des échantillons,
- n_i : la taille de l'échantillon,
- H_0 : les deux échantillons suivent la même loi.

avec ν , le nombre de degrés de liberté est estimé en utilisant l'équation de Welch-Satterthwaite :

$$\nu = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^2}{\frac{\sigma_1^4}{n_1^2 \cdot (n_1 - 1)} + \frac{\sigma_2^4}{n_2^2 \cdot (n_2 - 1)}}. \quad (2.34)$$

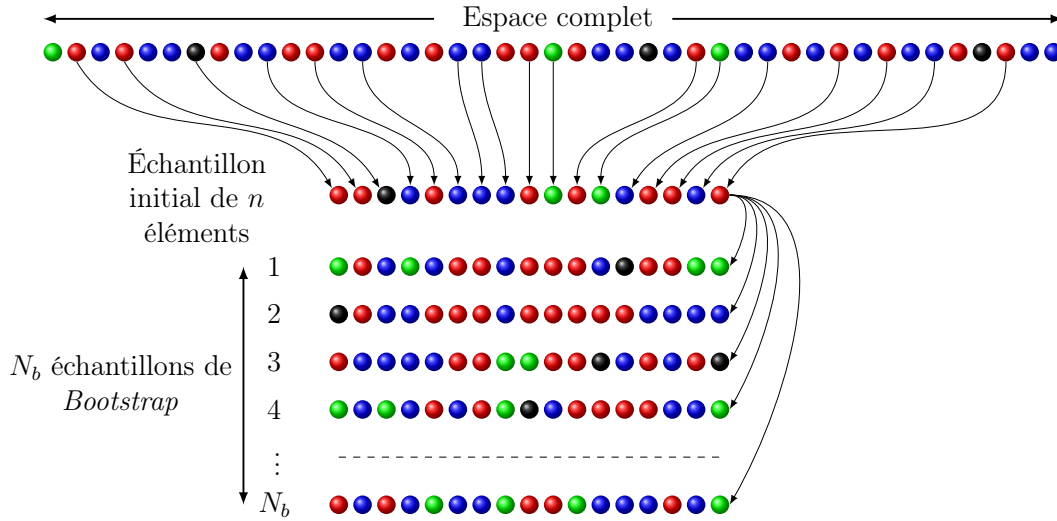
Nous utilisons également ce test statistique afin d'évaluer la différence entre la fonction de répartition empirique et les lois de distributions.

Bootstrap

Le *Bootstrap* est une méthode issue des recherches de Bradley Efron [63] à la fin des années 70. Son but est d'estimer la précision d'un intervalle de confiance ou d'une distribution. Elle consiste à re-échantillonner un échantillon de taille limitée, sans rajouter de nouvelles données. Il permet d'obtenir des informations sur les incertitudes statistiques liées à un échantillon de taille limitée.

Théorème 2.4.1 *Lorsque n tend vers l'infini, la distribution des valeurs moyennes calculées à partir des échantillons de bootstrap est égale à la distribution des valeurs moyennes obtenues à partir de tous les échantillons avec des n éléments qui peuvent être construits à partir de l'espace complet. Ainsi, la largeur de la distribution donne une évaluation de la qualité de l'échantillon (Bradley Efron [63]).*

Pour étudier une population donnée (« espace complet »), le *Bootstrap* (voir théorème 2.4.1), consiste à extraire un premier échantillon représentatif de cette population puis d'échantillonner de nouveau ce premier échantillon (sans recours à de nouvelles observations) : c'est le ré-échantillonnage. Cela permet de conserver toujours le même nombre de données. Après un certain nombre d'itérations (formule 2.35), résultant de la méthode de Monte-Carlo (voir


 Fig 2.4 – Méthode du *Bootstrap*.

§ 2.4.5), nous obtenons une statistique finale en faisant la moyenne des différentes statistiques obtenues. Cette méthode permet de remplacer les difficultés mathématiques par d'importants calculs. Le lecteur peut visualiser cette méthode Figure 2.4, p. 52 pour une meilleure compréhension.

En pratique, un nombre d'itérations minimum est nécessaire pour l'estimation des variances ou des intervalles de confiance de statistique des paramètres du modèle.

- $N_b \simeq 1000$: pour estimer l'erreur-standard,
- $N_b \simeq 5000$: pour l'évaluation d'intervalles de confiance.

$$\text{Nombre maximal d'itérations pour un } n\text{-échantillon : } N_{max} = \frac{(2n-1)!}{n!(n-1)!}. \quad (2.35)$$

2.4.6 Comparaison de deux modèles par corrélation linéaire de Bravais-Pearson

Contrairement à un test de statistiques comme celui de Kolmogorov-Smirnov, ce coefficient de corrélation nous permettra de comparer les distributions des données empiriques avec celles issues de la simulation. Nous voulons ici comparer les distributions des symboles et essayer de se rapprocher le plus possible de la réalité. Le but est de comparer deux courbes entre elles, afin de regarder la différence entre chacun des points. Pour cela, nous utilisons le coefficient de corrélation linéaire de Bravais-Pearson [6]. Pour deux séries de données de même taille $X(x_1, \dots, x_n)$ et $Y(y_1, \dots, y_n)$, le coefficient de corrélation linéaire liant ces deux séries est défini par :

$$\rho_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2.36)$$

Les deux courbes ne sont pas corrélées si $\rho_{x,y}$ est proche de 0. Les deux courbes sont d'autant plus corrélées entre elles que $\rho_{x,y}$ est proche de -1 ou de 1.

La covariance entre x et y est définie par :

$$\sigma_{xy}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y}). \quad (2.37)$$

Cette technique de comparaison nous permettra de quantifier les différences entre les distributions empiriques et simulées. Cela pourra nous donner, dans un deuxième temps, d'éventuelles améliorations à apporter à notre modèle de synthèse.

2.5 Mesure du caractère stochastique d'un modèle

Le but de cette section est de quantifier les aspects stochastiques des modèles étudiés. Nous devons, en effet, vérifier que les modèles étudiés sont de type stochastique. Pour vérifier ce caractère, le modèle doit comporter des séquences considérées comme aléatoires. Ainsi, pour éviter d'utiliser des modèles biaisés, nous devons vérifier le caractère aléatoire des données du modèle. Nous pouvons citer le guide du NIST⁹, qui a publié le guide [144]. C'est une suite de 15 tests statistiques qui ont été développés afin d'estimer le caractère aléatoire de séquences binaires. Ces tests sont très utilisés en cryptographie, ils permettent de définir le caractère prédictible ou pas d'un générateur de nombres aléatoires ou pseudo-aléatoires.

Nous vérifions que le générateur d'états du modèle de synthèse est « suffisamment » aléatoire en utilisant le test en fréquence § 2.5, p. 53.

Nous décrivons succinctement quelques uns de ces tests les plus pertinents :

Test de Fréquence (Monobit)

Le but de ce test est de déterminer si les nombres de uns et de zéros d'une séquence donnée, sont approximativement les mêmes qu'une réelle séquence aléatoire. Tous les tests suivants dépendent du passage de ce test.

- Soit n , la longueur de la chaîne de bits à tester,
- ϵ est la séquence de bits à tester,
- $X_i = 2\epsilon - 1 = \pm 1$,
- $S_n = X_1 + X_2 + \dots + X_n$,

9. National Institute of Standards and Technology

- $s_{obs} = \frac{|S_n|}{\sqrt{2}}$,
- $\text{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^{+\infty} e^{-u^2} du$: la fonction d'erreur complémentaire dite de Gauss,
- $\text{p-value} = \text{erfc}\left(\frac{s_{obs}}{\sqrt{2}}\right)$.

Si la séquence est aléatoire, les +1 et −1 auront tendance à s'annuler mutuellement et la statistique tendra vers 0. L'hypothèse nulle (H_0) considère que la séquence est *aléatoire*. Le test accepte H_0 si $\text{p-value} \geq 0,01$. La séquence est considérée comme non aléatoire si $\text{p-value} < 0,01$.

Test de Maurer – « test statistique universel »

Une séquence significativement compressible est considérée comme non-aléatoire. Ce test mesure simplement la longueur d'une séquence compressée. La compression utilise l'algorithme de Lempel-Ziv-Welch [186] (voir annexe A.2, p. 142).

Test de complexité linéaire

Ce test étudie la longueur d'un LFSR (Linear Feedback Shift Register). C'est un registre à décalage à rétroaction linéaire, c'est-à-dire que le bit entrant est le résultat d'un OU exclusif entre plusieurs bits du M-bloc. La séquence est considérée aléatoire si celle-ci est suffisamment complexe.

Test de l'entropie approximative

Ce test utilise l'entropie définie au § 2.4.3, p. 41, pour comparer la fréquence d'apparition de blocs de longueurs adjacentes m et $m + 1$. Cette fréquence doit être bien évidemment conforme à celle d'une séquence aléatoire.

Test des sommes cumulées

Le but du test est de déterminer si la somme cumulée de toutes les séquences partielles (équation 2.38) est proche de zéro. Comme pour le *Test de Fréquence*, les 0 de la séquence sont transformés en −1.

- ϵ est la séquence de bits à tester,
- $X_i = 2\epsilon - 1 = \pm 1$,

$$S_n = X_1 + X_2 + \dots + X_n. \quad (2.38)$$

Test « Random Excursions »

Ce test examine toutes les séquences partielles S_1 à S_n (équation 2.38), et détermine le nombre de cycles particuliers i.e. lorsque la somme cumulée des séquences partielles s'annule (les 0 de la séquence sont transformés en −1 : $X_i = 2\epsilon - 1 = \pm 1$). Il compare ce nombre

à celui attendu pour une séquence aléatoire. Ce test est en fait une série de huit tests (et conclusions) pour chacun des états : $-4, -3, -2, -1$ et $+1, +2, +3, +4$.

Test « Random Excursions Variant »

Le test est identique au précédent, avec une série de dix-huit tests (et conclusions) pour les états suivants : $-9, -8, -7, -6, -5, -4, -3, -2, -1$ et $+1, +2, +3, +4, +5, +6, +7, +8, +9$.

Vérification des tests précédents

Pour vérifier la performance des tests précédents, le NIST utilise le test de Kolmogorov-Smirnov d'uniformité (voir § 2.4.5, p. 50). La conformité est alors étudiée à partir des p-values obtenues à partir des séquences étudiées dans [144].

2.6 Discussion sur les méthodes de sélection de modèles

Il n'existe pas de critère universellement meilleur. En pratique, seule une parfaite connaissance du milieu à analyser permet de donner un sens à la notion de supériorité d'un critère sur un autre. D'après Olivier et al. [123], le principe du maximum de vraisemblance conduit en général à sur-paramétriser le modèle pour avoir de bons résultats. Une pénalisation du terme de vraisemblance peut pallier cet inconvénient. Le critère de type log-vraisemblance pénalisé le plus célèbre est AIC Akaike (1973) [3]. Même s'il n'est pas totalement satisfaisant, il améliore le principe du maximum de vraisemblance mais conduit aussi à une sur-paramétrisation. D'autres critères désormais classiques, BIC et HQC, assurent une meilleure estimation en pénalisant justement le sur-dimensionnement du modèle. Enfin, le critère de comparaison de Kullback-Leibler [174] mesure la dissimilarité entre 2 modèles.

Une grande attention doit être portée au choix du critère de sélection du modèle. Il doit être conditionné par l'objectif de l'analyse et de la connaissance des données. Olivier et al. [122] ont remarqué que les critères comme BIC et AIC étaient utilisés indifféremment, quel que soit le problème posé, alors qu'ils n'ont pas le même objectif.

Les données dont nous disposons dans ce manuscrit sont essentiellement basées sur des probabilités. Notre démarche de sélection se fera ainsi sur des modèles probabilistes. Si nous nous référons à la littérature purement mathématique, celle-ci recommande l'utilisation d'une méthode Bayésienne. En effet, l'inférence bayésienne est parfaitement maîtrisée dans ce domaine. Mais selon Burnham [32], presque tous les arguments en faveur de l'utilisation de BIC plutôt que AIC, avec des données réelles, manquent de pertinence ! Ceci contribue davantage à la confusion qu'à la compréhension de l'utilisation de ces critères. Il y a des contextes clairement définis permettant le choix de la bonne méthode. Il apparaît pour certains qu'il vaut mieux choisir BIC car c'est une méthode Bayésienne. La différence doit se faire dans la distribution a priori définissant le modèle [32]. La comparaison des 2 méthodes doit être basée selon [32], sur la comparaison des mesures de leur performance dans des conditions réalistes d'application. Celle-ci est fonction du nombre de paramètres du modèle. De plus, une augmentation de ce nombre de paramètres permet de réduire le biais d'un modèle mais induit en contre partie une augmentation de la variance [131].

Consistance et efficacité

Certains critères sont basés sur le principe du maximum de vraisemblance (voir § 2.4.3, p. 42) tels que AIC [3] et BIC [148]. Nous utiliserons par la suite certains de ces critères comme AIC, BIC et HQC. Le critère de BIC est connu comme consistant, i.e. si le vrai modèle est contenu dans l'ensemble des modèles candidats, alors la probabilité de choisir le vrai modèle est proche de 1. D'autre part, le critère d'AIC est optimal pour les cas à la fois paramétriques et non paramétriques. Ce critère donne de bons résultats dans l'estimation d'une fonction de régression par exemple. Selon Claeskens et al. [46], BIC et HQC sont des critères fortement consistants. AIC, AICc et le C_p de Mallows sont efficaces (un critère est efficace lorsque l'erreur de prédiction attendue est proche de l'erreur de modélisation) [47]. Cependant, nous ne pouvons pas combiner la consistance de BIC avec l'efficacité d'AIC. En effet, Yang [189] montre que si le vrai modèle est inclus dans l'ensemble des candidats alors les points forts des 2 précédents critères, mentionnés ci-dessus, ne peuvent être partagés. Autrement dit, pour être consistant, les critères de sélection de modèles doivent se comporter de manière optimale par rapport à l'AIC, en terme d'erreur quadratique moyenne. Selon [98], BIC a été conçu pour trouver le modèle le plus probable. Par contre AIC est meilleur lorsque les modèles candidats sont peu probables, il minimise la distance de Kullback-Leibler (voir § 2.4.3, p. 44) qui permet aussi de comparer des modèles entre eux.

Sur-apprentissage

Les données d'apprentissage sont déterminantes dans la construction d'un modèle. Wooldroffe [188] étudie le caractère de sur-apprentissage des critères comme l'AIC ou le C_p de Mallows. Il montre l'influence du paramètre k (pour $k \rightarrow +\infty$) pour la loi de probabilité « Arc Sinus ». Par exemple pour $k \rightarrow +\infty$, la probabilité de sur-apprentissage est de 0,946 alors que la probabilité d'identifier correctement le vrai modèle est de 0,712 (voir Tableau 2.1, p. 56)

k	Probabilité de sur-apprentissage	Probabilité d'identifier le vrai modèle
5	0,571	0,736
10	0,718	0,791
$+\infty$	0,946	0,712

Tableau 2.1 – Caractère de sur-apprentissage de l'AIC (distribution Arc Sinus [188]).

Cet exemple illustre le fait que ce critère de sélection de modèle minimise le score si le nombre de paramètres est trop important.

2.7 Conclusion

Des approches telles que l'analyse de sensibilité d'un modèle, les méthodes avec score d'intérêt ou statistiques, nous offrent de nombreuses possibilités pour trouver les architectures des modèles les plus pertinentes. De nombreuses méthodes utilisées en mathématiques appliqués ne seront pas développées ici à cause de leur complexité. Nous utiliserons les méthodes décrites ci-dessus pour tenter de mesurer la pertinence des architectures de MMC proposées dans Vrignat [177]. De plus, afin de pouvoir comparer les données empiriques avec des données issues de simulation, nous allons mettre en place un modèle de synthèse. Nous allons alors vérifier dans un premier temps son caractère stochastique (avec un test du NIST). Ensuite, nous allons déterminer les éléments du modèle les plus pertinents à l'aide des méthodes précédentes qui ont déjà été utilisées avec une problématique similaire. Enfin, dans un but d'améliorer les performances de notre modèle, nous proposerons des améliorations en les quantifiant à l'aide de la corrélation linéaire de Bravais-Pearson.

Chapitre 3

Évaluation de modèles par une approche markovienne

Nous débutons ce chapitre par une description des attentes théoriques de cette étude. Nous présentons ensuite quelques approches classiques issues de la littérature qui pourraient être utilisées dans notre démarche. Nous détaillons ensuite les différents aspects liés à notre problématique, notre choix de l'approche markovienne et les différentes architectures de modèle que nous utilisons (topologies, algorithmes d'apprentissage, algorithmes de décodage). Nous introduisons ensuite le modèle de synthèse utilisé pour cette étude, afin d'obtenir des données simulées proches d'un processus industriel. Ce modèle établit ainsi des relations cohérentes entre les observations de sortie d'un processus industriel et l'état « caché » de celui-ci. Comme préconisé dans les travaux précédents [177], nous utilisons une modélisation markovienne à quatre états (S1, S2, S3 et S4). S4 est l'état de fonctionnement optimal et S1 est l'état critique où le processus est arrêté. Enfin, nous présenterons les lois statistiques utilisées dans la production des symboles à l'aide du Principe de Maximum d'Entropie (PME).

3.1 Introduction

Certaines entreprises de la région ont mis en place des systèmes de GMAO. Des données de maintenance sur des processus industriels ont ainsi pu être collectées. Ces dernières nous servent de base pour notre étude empirique. Les comparaisons avec les données issues d'un modèle de synthèse nous permettent d'évaluer les différentes architectures des modèles utilisés. Nous présentons ici les méthodes et donnons les résultats dans le chapitre suivant. Tous les calculs sont menés sur deux fronts : données empiriques et données de simulation.

3.2 Attentes théoriques

Notre étude est basée sur des analyses comportementales de modèles en fonction des données d'entrée (informations de maintenance). Mais, il est évident, que toutes ces observations (informations d'entrée) n'ont pas le même poids dans la construction du modèle. De ce fait, dans un premier temps, une analyse de la nature des données nous permet d'étudier la **pertinence des symboles** qui alimentent le modèle (le symbole étant la nature de l'observation). Donc, cela revient à répondre à la question : les données apportent-elles de l'information au modèle ? Dans un deuxième temps, nous faisons **évoluer la modélisation** en proposant différentes améliorations à apporter afin que le modèle soit plus efficace. Enfin, nous étudierons la **pertinence de l'architecture des modèles** par différents critères de sélection.

Nous pouvons décomposer nos attentes théoriques selon les axes suivants :

- (a) **Pertinences des observations** sans connaissance a priori sur les résultats :
 - quels sont les symboles les plus pertinents ? Ils proviennent d'une GMAO industrielle,
 - quelles sont les instances qui n'apportent aucune ou trop peu d'information au modèle ?
 - peut-on retrouver l'état d'un système ou d'un processus, à partir d'une séquence d'instances particulières ?
- (b) **Évolution de la modélisation** :

- quel est le meilleur échantillonnage des observations pour obtenir une réactivité efficace des modèles ?
- quel est le nombre minimal de données nécessaire au modèle pour qu’il ait une information suffisante ? Est-il possible d’établir une fenêtre optimale de symboles, de manière à pouvoir réévaluer le modèle en fonction de cette fenêtre « glissante » ? Le modèle se fonderait donc sur un historique limité.

(c) **Pertinence de l’architecture des modèles :**

- quelle est la meilleure topologie parmi celles étudiées ?
- quel est le meilleur algorithme d’apprentissage et de décodage ?
- quelle est la meilleure distribution des observations ?

3.3 Approches classiques

L’estimation de données non connues ou cachées n’est pas l’apanage des MMC. D’autres techniques peuvent estimer des états cachés d’un système et modéliser des niveaux de dégradation d’un processus. Nous présentons dans cette section d’autres approches pouvant répondre à cette problématique. En effet, les filtres particuliers [33] (voir § 3.3.1, p. 62) fondés sur la théorie des filtres de Kalman, ainsi que les réseaux de Petri stochastiques [124] (voir § 3.3.3, p. 63) permettent d’avoir des propriétés proches de celles de la théorie markovienne.

3.3.1 Filtres particuliers

L’objectif des filtres particuliers [109] est l’estimation de modèles fondés sur la simulation. Il s’agit ici d’estimer une séquence de paramètres cachés en se basant uniquement sur des observations.

Soit x_k , les paramètres cachés et y_k , les paramètres bayésiens estimés de x_k . Les filtres particuliers estiment la distribution de filtrage $p(x_k|y_0, y_1, \dots, y_k)$ au lieu d’utiliser les probabilités jointes a posteriori $p(x_0, x_1, \dots, x_k|y_0, y_1, \dots, y_k)$.

Ils constituent une alternative aux filtres de Kalman étendus (voir § 3.3.2, p. 62) et peuvent être plus rapides que les méthodes de Monte-Carlo par chaînes de Markov. Ils donnent une estimation proche de la valeur optimale en utilisant la méthode des réseaux Bayésiens. Nous n’utilisons pas cette technique, trop complexe à mettre en œuvre.

3.3.2 Filtre de Kalman étendu

Le filtre de Kalman [100], [11], [128], et [71] permet d’estimer l’état d’un système à partir d’une information a priori sur l’évolution de cet état et de mesures réelles. Il est notamment utilisé dans l’identification paramétrique d’un modèle.

Le principe est d’utiliser une estimation de l’état précédent pour prédire l’état présent, un peu comme l’hypothèse de Markov (voir § 3.4.2, p. 68). Une deuxième phase consiste à corriger la prédiction en utilisant l’observation de l’instant présent. Ceci a pour but de corriger l’estimation.

Prenons par exemple un modèle d'état discret représenté par l'équation d'état suivante :

$$\begin{cases} x(t_{k+1}) &= A_k x(t_k) + B_k u(t_k) + v(t_k) & \text{(équation d'état)} \\ y(t_k) &= C_k x(t_k) + \varepsilon(t_k) & \text{(équation de mesure).} \end{cases} \quad (3.1)$$

où :

- $k \geq 0$ représente les instants successifs du temps, $t_k = kT_e$ où T_e est la période d'échantillonnage ;
- $x(t_k) \in \mathbb{R}^n$ est le vecteur d'état du système ;
- $u(t_k) \in \mathbb{R}^m$ est le vecteur des entrées déterministes ;
- $v(t_k) \in \mathbb{R}^p$ est le vecteur bruit sur les entrées ;
- $y(t_k) \in \mathbb{R}^q$ est le vecteur des mesures ;
- $\varepsilon(t_k) \in \mathbb{R}^q$ est le vecteur des signaux stochastiques (erreur de mesure).

Les hypothèses suivantes doivent être vérifiées :

- les matrices A_k , B_k , C_k ainsi que l'entrée $u(t_k)$ sont déterministes ;
- les bruits de mesure ainsi que les bruits sur les entrées sont supposés de moyenne nulle, non corrélés entre eux $\forall t_k$;
- l'état initial x_0 d'espérance $\overline{x_0}$, de matrice de covariance P_0 est indépendant du bruit d'état et de mesure.

soit pour l'initialisation :

$$E \left(\begin{bmatrix} x(t_0) \\ v(t_i) \\ \varepsilon(t_i) \end{bmatrix} \right) = \begin{bmatrix} \overline{x_0} \\ 0 \\ 0 \end{bmatrix}, \quad \left(\begin{bmatrix} x(t_0) \\ v(t_i) \\ \varepsilon(t_i) \end{bmatrix} [x(t_0) \quad v(t_j) \quad \varepsilon(t_j)]^T \right) = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q_i \delta_{ij} & 0 \\ 0 & 0 & R_i \delta_{ij} \end{bmatrix}. \quad (3.2)$$

- E représente l'espérance mathématique ;
- δ_{ij} est le symbole de Kroknecker¹ ;
- P_0 , Q_i et R_i sont des matrices symétriques définies positives.

La méthode consiste à prédire l'état suivant en minimisant la variance de l'erreur d'estimation en deux phases :

- la phase de prédiction : utilise l'état précédent pour estimer l'état courant ;
- la phase de correction : les observations de l'instant courant permettent d'affiner l'état courant.

3.3.3 Réseau de Petri stochastique

Selon Labadi et al. [105], un réseau de Petri est un modèle mathématique utilisé dans la représentation de systèmes dynamiques. Initiés en 1962 dans la thèse de doctorat de Carl Adam Petri, les réseaux de Petri suscitent toujours autant d'intérêt de nos jours dans le domaine de la recherche. Malheureusement, peu d'applications industrielles voient le jour par

1. $\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$

manque d'une normalisation internationale contrairement au GRAFCET², qui est d'ailleurs dérivé du principe des réseaux de Petri. Les réseaux de Petri permettent de représenter le fonctionnement d'un automate. Ils peuvent ainsi modéliser un processus industriel quelconque.

Nous illustrons la démarche sur la Figure 3.1, p. 64. Il y a les étapes ou places, auxquelles sont associées des actions. Les transitions entre étapes sont liées par des conditions de transition et les liaisons orientées entre les étapes et les transitions. Les **Jetons** correspondent aux nombres de ressources disponibles.

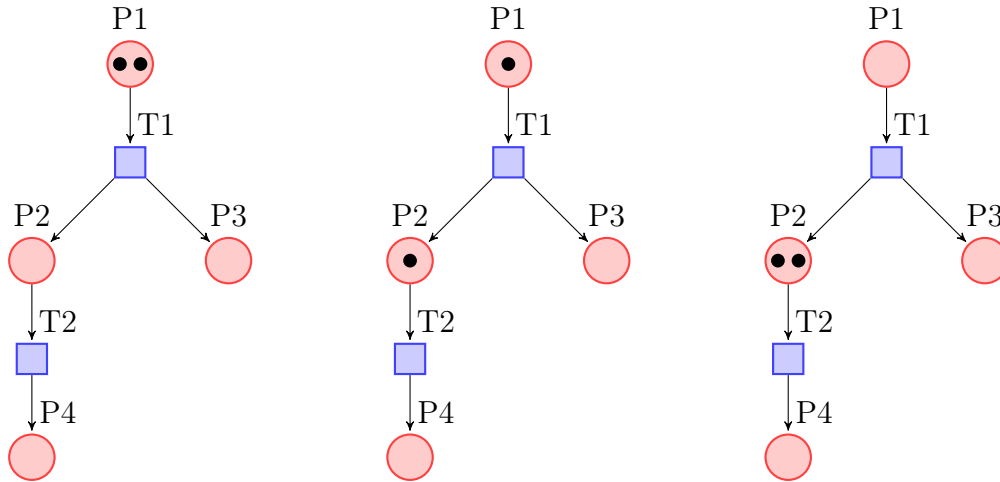


Fig 3.1 – Exemple d'un réseau de Petri à 4 places, 2 transitions et 2 jetons.

Ce réseau évolue après chaque transition. Des jetons sont redéployés au niveau des entrées et des sorties des transitions.

Notre problématique est assez proche de celle des réseaux de Petri. En effet, l'automate représenté par un réseau de Petri peut aussi bien se représenter sous forme d'un automate de Markov. Dans ce manuscrit, nous n'avons pas choisi cette méthode car les réseaux de Petri ne supportent pas la couche « états cachés » comme peut le faire l'approche markovienne. Peu d'applications industrielles voient le jour.

3.4 Approche markovienne

L'approche markovienne est très utilisée, dans les études probabilistes de processus à comportements dynamiques. C'est une approche « analytique par états » i.e. qui est fondée sur l'identification des états d'un processus quelconque et l'analyse de l'évolution au cours du temps de ses états. Elle est couramment utilisée en fiabilité, pour les calculs de disponibilité.

En choisissant cette approche, nous bénéficions d'un certain nombre de propriétés :

- celles d'un processus markovien qui imposent un découpage temporel régulier [113]. Ce découpage concerne évidemment l'échantillonnage des données, qu'elles soient empiriques

2. GRAPhe Fonctionnel de Commande Etape Transition

ou simulées ;

- des chaînes de Markov pour des processus à temps discrets (selon Foata [66]). Connaissant l'état présent, la prédiction du futur n'est pas rendue plus précise par la connaissance des états supplémentaires concernant leur passé ;
- les états inconnus du système à modéliser correspondent aux états cachés d'un MMC ;
- le système à modéliser doit émettre des observations ou symboles.

Des MMC permettent alors, la modélisation de processus industriels notamment utilisés pour la gestion de la maintenance. Les symboles émis correspondent à l'état inconnu du processus.

3.4.1 Chaînes de Markov

Les chaînes de Markov sont utilisées dans l'étude de suites de variables aléatoires. Il y a les chaînes de Markov dont l'ensemble des états est continu et celles dont l'ensemble E des états est discret ($E \in \mathbb{N}$). Nous étudions dans cette thèse uniquement le cas discret. Le lecteur pourra se référer à l'article de Baier et al. [12] pour plus de détails sur les chaînes de Markov.

Définitions d'une chaîne de Markov

Soit X_n , $n \geq 0$ une suite de variables aléatoires à valeurs dans l'ensemble des états E . Cette suite est une chaîne de Markov d'ordre 1, si pour tout $n \geq 1$ et toute suite $(i_0, \dots, i_{n-1}, i, j)$ d'éléments de E , pour laquelle la probabilité $P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i)$ est strictement positive, on a la relation suivante [66] :

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i). \quad (3.3)$$

Dans l'évolution au cours du temps, l'état du processus à l'instant $n + 1$ ne dépend que de celui à l'instant n , mais non de ses états antérieurs. Le processus est dit *sans mémoire* ou *non héréditaire*.

Pour une chaîne de Markov d'ordre supérieur à 1 (ordre m), l'état du processus à l'instant $n + 1$ ne dépend que des m états précédents. Il a une petite mémoire.

Définition 3.4.1.1 *La chaîne de Markov est dite homogène (dans le temps), si la probabilité précédente ne dépend pas de n :*

$$p_{i,j} = P(X_{n+1} = j | X_n = i) \quad (n \geq 0). \quad (3.4)$$

Cette probabilité est appelée « probabilité de passage » de l'état i à l'état j , en une transition.

Définition 3.4.1.2 Soit $\mathcal{A}$ la matrice suivante :

$$\mathcal{A} = \begin{pmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \cdots \\ a_{1,0} & a_{1,1} & a_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (3.5)$$

Les coefficients sont les probabilités de transition $p_{i,j}$ et forment la matrice de passage (ou de transition) de la chaîne de Markov. C'est une matrice finie ou dénombrable, suivant que l'ensemble des états est fini ou dénombrable.

Exemple d'une chaîne de Markov à deux états

Les chaînes de Markov sont intuitivement très simples à définir. En effet, un système comportant un certain nombre d'états qui changent au cours du temps discret peut-être modélisé sous forme d'une chaîne de Markov. Pour chaque changement d'état, celui-ci est déterminé à partir d'une distribution de probabilités fixée au préalable et ne dépendant que de l'état présent. Nous montrons, ci-dessous, l'exemple d'une chaîne de Markov à deux états. En excluant le cas trivial de la matrice unité, la matrice de transition est de la forme :

$$\mathcal{P} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \quad (0 < \alpha, \beta \leq 1). \quad (3.6)$$

Le graphe associé est donné Figure 3.2.

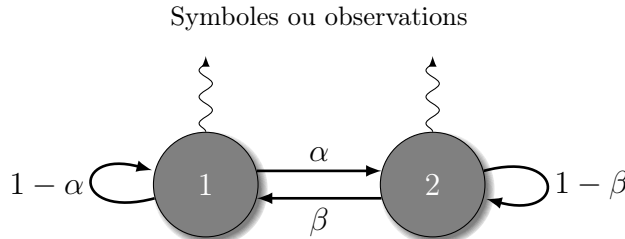


Fig 3.2 – Chaîne de Markov à deux états.

Exemple du modèle de diffusion d'Ehrenfest

Un autre exemple, un peu plus complexe est le *modèle des urnes* [66] et [147]. C'est un modèle stochastique introduit en 1907 par les époux Ehrenfest. Ce modèle est aussi appelé le « dog-flea model » (modèle des chiens et des puces) :

Deux urnes A et B contiennent, à elles deux, a boules, numérotées de 1 à a . A chaque instant, on choisit un nombre de 1 à a , avec une probabilité de $1/a$. Si ce nombre est i , on change d'urne la boule numérotée i .

L'ensemble des états est l'ensemble $E = \{0, 1, \dots, a\}$. Le processus est dit être dans l'état j si l'urne A contient j boules. Dans ces conditions, si le processus est dans l'état 0 (l'urne A est vide), la probabilité est égale à 1 lorsqu'il passe dans l'état 1 (respectivement l'état $(a - 1)$). La matrice de transition est donc donnée par :

$$\mathcal{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1/a & 0 & (a-1)/a & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2/a & 0 & (a-2)/a & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (a-1)/a & 0 & 1/a \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}. \quad (3.7)$$

et le graphe associé est donné Figure 3.3, p. 67.

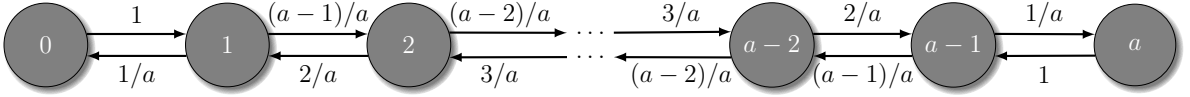


Fig 3.3 – Chaîne de Markov : modèle de diffusion d'Ehrenfest.

Pour $n \geq 0$, désignons par X_n le nombre de boules dans l'urne A à l'instant n . Si $X_0 = a$, alors le processus (X_n) ($n \geq 0$) décrit la « diffusion » d'un gaz de A vers B .

3.4.2 Modèle de Markov Caché discret

Un Modèle de Markov Caché discret est composé d'un processus stochastique caché modélisé par une chaîne de Markov et un processus stochastique observé dépendant des états du processus caché [9]. Selon Rabiner [137] et Fox [67], c'est un automate à états cachés qui est constitué d'une variable non observable. Celle-ci représente l'état du système à modéliser. Seule la variable de sortie est observable. Cela nous permet d'avoir une séquence d'observations en sortie de l'automate; à partir de maintenant, nous parlerons simplement de symboles représentant ces observations.

Définitions formelles d'un Modèle de Markov Caché à observations discrètes

- soit N , le nombre d'états cachés possibles et $S = \{S_1, S_2, \dots, S_N\}$, l'ensemble des valeurs possibles de cette variable. On notera q_t , la valeur de cette variable à l'instant t ;
- le processus ainsi modélisé, doit répondre à l'hypothèse markovienne : l'état à un instant t ne dépend que de l'état à l'instant $t - 1$;
- soit K , le nombre total de symboles d'observations et nous notons $\mathbb{V} = \{v_1, v_2, \dots, v_K\}$, l'ensemble des K symboles émissibles par le système. On notera v_t , la valeur de cette variable à l'instant t . Soit $V = (V_1, \dots, V_T)$, un T-uplet de valeurs aléatoires définies sur $\mathbb{V}$;
- soit $A = \{a_{ij}\}$, la distribution de probabilité de la transition d'état avec :

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \quad 1 \leq i, j \leq N. \quad (3.8)$$

- soit $B = \{b_j(k)\}$, la distribution de probabilité des observations v_k à l'état S_j , avec :

$$b_j(k) = P(V_t = v_k | q_t = S_j) \quad 1 \leq j \leq N \quad 1 \leq k \leq K, \quad (3.9)$$

- soit $\pi = \{\pi_i\}$, la distribution des probabilités initiales, avec :

$$\pi = P(q_1 = S_i) \quad 1 \leq i \leq N. \quad (3.10)$$

- le Modèle de Markov Caché sera noté :

$$\lambda = (A, B, \pi). \quad (3.11)$$

Les relations de dépendance entre les différentes variables aléatoires d'un MMC sont représentées sur la Figure 3.4, p. 68 [8].

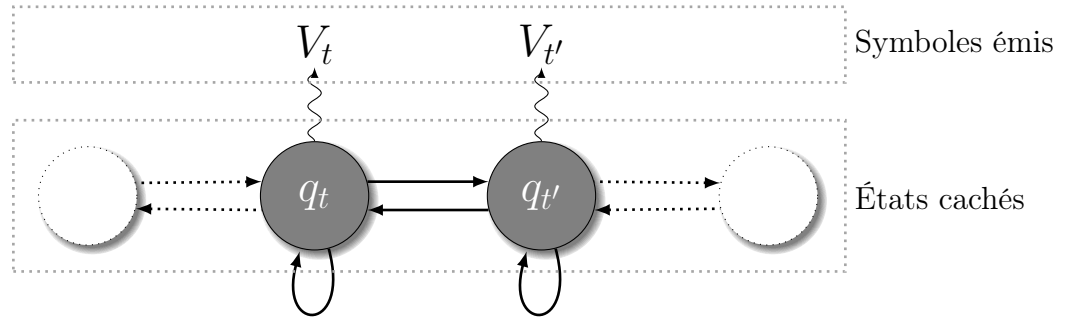


Fig 3.4 – Relations de dépendance entre les variables aléatoires d'un MMC. Pour chaque état q_t à un instant t , il y a émission d'un symbole V_t pris dans l'ensemble $\mathbb{V}$.

Hypothèse de Markov

La prédiction de l'état futur n'est pas rendue plus précise par connaissance supplémentaire d'information a priori i.e. toute l'information utile pour la prédiction du futur est contenue dans l'état présent du processus :

$$P(X_{n+1} = j | X_0, X_1, \dots, X_n = i) = P(X_{n+1} = j | X_n = i). \quad (3.12)$$

3.4.3 Complexité de l'évaluation d'un MMC

Comment trouver la séquence la plus probable d'états cachés ayant conduit à la production d'une séquence d'observations données ? Le nombre de chemins possibles pour générer une telle séquence est de l'ordre de N^T (N étant le nombre d'états d'un MMC et T la longueur d'une séquence d'observations). L'approche directe n'est pas acceptable sachant que pour notre cas $N = 4$ et $T = 1000$, le calcul demanderait alors approximativement 10^{600} opérations. Pour un ordinateur cadencé à 1GHz et en supposant qu'un calcul est égal à une opération du processeur, la solution sera trouvée dans :

$$\frac{10^{600}}{10^9} / (3600) / 365 = 10^{400} \text{ années}. \quad (3.13)$$

C'est pourquoi des alternatives doivent être mises en place. Certains algorithmes comme celui de « forward-backward » [70] appelé aussi algorithme Baum-Welch, avec une complexité de l'ordre de N^2T , ainsi que celui de « Viterbi » [81] ayant une complexité de TN^T , permettent de pallier à ce problème. Ils réduisent ainsi le temps de calcul.

3.4.4 Topologies des modèles étudiés

Nous reprenons pour notre étude, les trois topologies étudiées dans [177]. Nous allons étudier leurs comportements au travers des algorithmes d'apprentissage que nous présentons § 3.4.5, p. 69.

- topologie 1 : cet automate de Markov illustre toutes les transitions possibles entre tous les états S_1 , S_2 , S_3 et S_4 (voir Figure 3.5(a), p. 71) ;
- topologie 2 : cette topologie est moins permissive que la précédente. Pour passer de l'état S_4 (le système fonctionne et tout va bien) à l'état S_1 (le système est à l'arrêt en panne), il faut obligatoirement passer par S_3 et S_2 . L'objectif du modèle étant de réduire autant que possible le temps de séjour en S_2 avant S_1 , afin que cet état soit un indicateur pertinent de l'arrivée imminente de la panne (voir Figure 3.5(b), p. 71) ;
- topologie 3 : nous retrouvons ici la topologie 2, à une différence près : l'état S_1 est un état « aspirant » i.e. cette topologie autorise moins de marge de manœuvre pour aller vers l'état S_1 , lors de la phase d'apprentissage. Le passage de S_1 à S_4 est le redémarrage après un arrêt (voir Figure 3.5(c), p. 71).

Les états S_2 à S_4 sont des états où le processus modélisé fonctionne. S_1 est un état d'arrêt du système. Pour un système productif, cette situation d'arrêt doit être minimisée. Dans notre étude, cet état doit être prédit au plus juste.

Les symboles émis représentent des interventions de maintenance (voir la codification Tableau 3.1, p. 70).

3.4.5 Apprentissage

Algorithmes d'apprentissage et de décodage

Pour réaliser l'apprentissage des différents modèles, nous utilisons les deux algorithmes suivants :

- apprentissage Baum-Welch [14] et [88], décodé par Variables Forward [137] :
estimation du modèle de façon itérative $\eta = (A, B, \pi)$,
avec une séquence d'observations $O = \{o_1, o_2, \dots, o_T\} \in \mathbb{V}^T$,

$$\text{Maximiser } \rightarrow P(V = O|\eta). \quad (3.14)$$

- apprentissage Segmental K-means [97], décodé par Viterbi [175] et [23] :

$$\text{Optimiser la probabilité } \rightarrow P(O, S = Q^*|\eta), \quad (3.15)$$

Etat du processus		
MARCHE		
ARRET		
N°Obs.	Symboles	Nature des interventions
1	DEP	(Dépannage / arrêt de la production)
2	RM	(Réglage Machine)
3	AU	(Autre)
4	OBS	(Observation)
5	TEP	(Travaux Entretien Préventif)
6	SEC	(Sécurité)
7	RAN	(Remise A Niveau / planifié)
8	NET	(Nettoyage Machine)
9	VEP	(Visite Entretien Préventif)
10	RAS	(pas d'intervention)

Tableau 3.1 – Codification symbolique des interventions de maintenance.

Q^* : séquence d'états cachés qui a le plus probablement engendré la séquence telle que calculée par l'algorithme de Viterbi (voir annexe A.4, p. 145).

Pour décoder les informations et ainsi retrouver les données de sortie de nos modèles, issues de la phase d'apprentissage, nous utilisons les deux algorithmes de décodage suivants :

- décodage algorithme Variables Forward :

$$\alpha_t(j) = P(o_1, o_2, \dots, o_t, Q_t = s_j | \eta). \quad (3.16)$$

α_t est la probabilité d'être dans chaque état à l'instant t .

- décodage algorithme Viterbi :

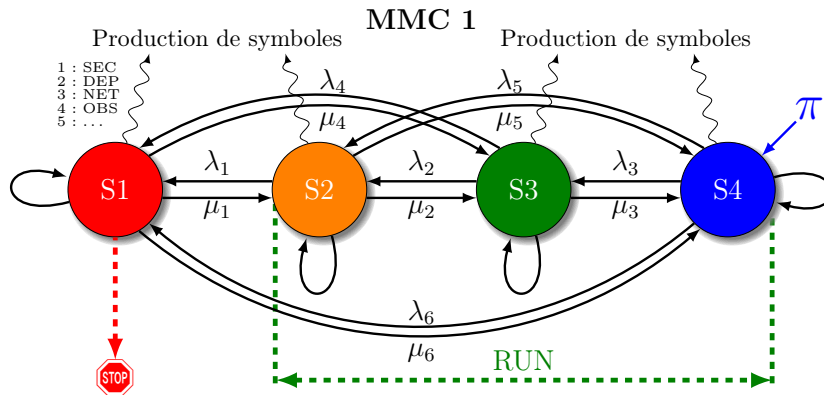
$$\delta_t(j) = \max_{(q_1, \dots, q_{t-1} \in \mathbb{S}^{t-1})} \{P(S_1 = q_1, \dots, S_{t-1} = q_{t-1}, S_t = s_j, V_1 = o_1, \dots, V_t = o_t | \eta)\}. \quad (3.17)$$

δ_t est la probabilité d'être dans chaque état à l'instant t .

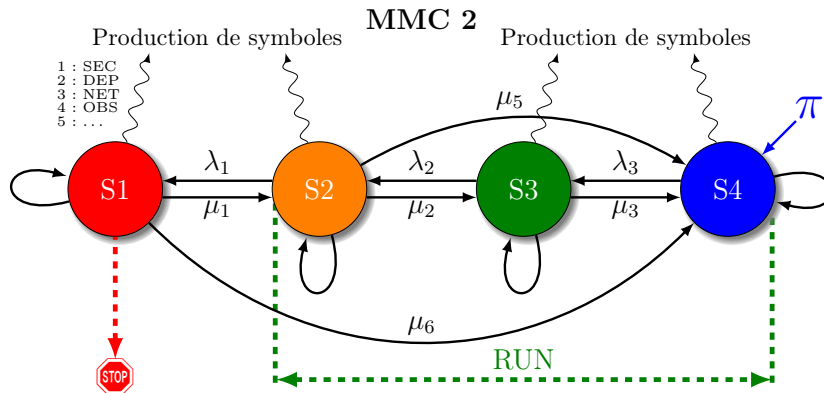
Arbres de décision par chaînes de Markov

Principalement utilisés à l'ordre 1, nous allons ici essayer de construire des arbres de décision (aussi appelés arbres de généralisations/spécialisations) en utilisant des chaînes de Markov d'ordres 1, 2 et 3, à partir des données. Notre but est d'essayer de trouver une relation de cause à effet entre les différents symboles étudiés. Cette notion [77] d'arbres de décision est empruntée à la problématique de la classification.

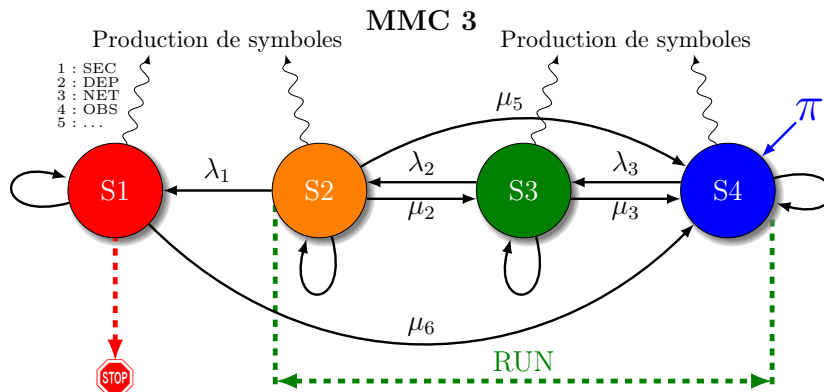
Pour notre cas d'étude et pour une chaîne de Markov d'ordre n , le symbole considéré à l'instant t dépend des n symboles précédents. Ainsi, à partir d'une chaîne de Markov d'ordre n , nous avons réalisé des arbres de décisions. Les observations (instances) constituent les données de l'apprentissage. Le but de cette démarche est d'essayer de trouver l'état d'un système, à partir d'une séquence d'instances particulières.



(a) Topologie 1



(b) Topologie 2



(c) Topologie 3

Fig 3.5 – Modèles de Markov Cachés, topologies à 4 états. Les λ_k, μ_k sont des a_{ij} illustrant les transitions S_i vers S_j . Les λ_k détériorent l'état et les μ_k améliorent l'état. La matrice d'initialisation π pointe obligatoirement sur l'état S4 puisque nous supposons démarrer toujours dans l'état optimal (S4 est l'état optimal, S1 est l'état du processus arrêté).

Espace de versions par chaînes de Markov

La technique des « espaces de versions » (voir § 2.4.4, p. 48) permet de déterminer, dans un ensemble d'hypothèses, lesquelles peuvent correspondre à un ensemble d'exemples pris dans un tableau de contingence correspondant à l'espace des données. Cela nous permettra de déterminer la correspondance d'un état particulier avec une séquence de symboles émis. Nous présentons nos résultats dans les paragraphes § 4.2.4, p. 116 et § 4.2.4, p. 115.

3.4.6 Symboles manquants

Les paramètres manquants d'un modèle de Markov peuvent être interprétés comme suit : le modèle établi est relativement satisfaisant à un certain niveau d'abstraction (voir § 2.3.4, p. 35). En effet, nous avons établi un certain nombre de variables sur nos modèles. Ces variables ont pour intérêt de décrire au mieux les processus physiques étudiés. Certaines données pourraient alors être rajoutées au modèle, par exemple certaines observations peuvent être anodines mais pourraient engendrer des modifications notables en sorties de nos modèles. Nous parlons alors de données incomplètes qui correspondent à un certain niveau d'abstraction considéré. Il est classique alors d'appliquer des méthodes d'estimations comme le maximum de vraisemblance (voir § 2.4.3, p. 42) [44].

3.5 Modèle de synthèse

Nous allons maintenant définir notre modèle de simulation. Nous répondons dans un premier temps aux questions suivantes :

- pourquoi utiliser un modèle de synthèse ?
- à quoi va-t-il nous servir ?

Nous avons créé un modèle de synthèse afin d'obtenir des données se rapprochant le plus possible du cas réel. En effet, ces données simulées vont nous permettre d'éprouver les architectures des modèles étudiés, les différentes topologies, les algorithmes d'apprentissage et de décodage ainsi que les deux distributions choisies dans l'émission de symboles. Ce modèle de synthèse va nous servir de référence pour valider le cas réel.

Nous présentons ici les différentes étapes de construction du modèle de synthèse (un modèle = une topologie, un algorithme d'apprentissage et un algorithme de décodage) :

- une topologie,
- une matrice de transition A ,
- une matrice de distribution des observations B et une matrice d'initialisation π .

Nous discutons dans un premier temps, le choix du *modèle de référence* qui nous a servi à construire notre modèle de synthèse. Connaissant les probabilités de transition du couple (*Symboles*, *Etats*), nous avons testé les différentes topologies étudiées (voir Figure 3.5, p.

71) sur des algorithmes d'apprentissage et de décodage. A la fin de ces essais comparatifs, nous avons conclu que la topologie 2 (voir Figure 3.6, p. 73) était la plus pertinente, en terme de détection de panne. Elle caractérise au mieux les activités de maintenance.

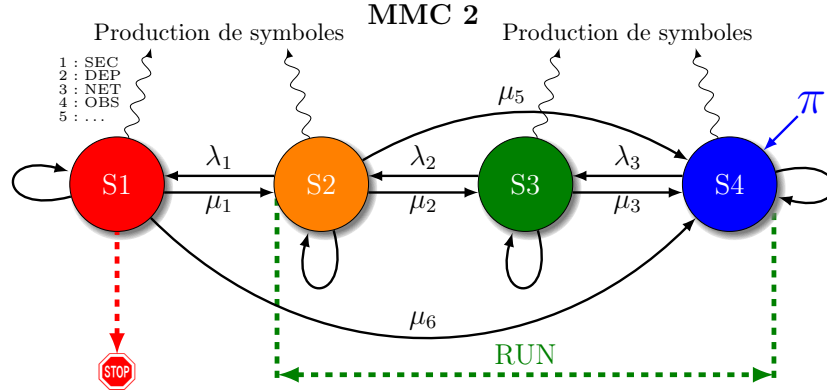


Fig 3.6 – Modèle de Markov Caché à 4 états, topologie 2, référence du modèle de synthèse.

La topologie d'un MMC dépend uniquement des éléments non nuls de la matrice de transition notée $A = \{a_{ij}\}$. Un tel modèle est dit à « connectivité totale » lorsque sa matrice de transition ne comporte aucun élément nul (voir Figure 3.5(a), p. 71). Le modèle de synthèse génère des émissions de symboles suivant les deux distributions suivantes :

- la distribution uniforme : cette loi permet de modéliser des variables aléatoires uniformément réparties sur un intervalle, comme le montre la Figure 3.8(a), p. 76. La variable aléatoire peut prendre n valeurs équiprobables $\{x_1, x_2, \dots, x_n\}$. La probabilité est : $P(x_i)_{i \in [1, n]} = 1/n$. Cette loi est peu représentative d'un système réel, car tous les symboles ont le même poids. Nous donnons quelques détails supplémentaires sur cette loi en annexe B.2, p. 150.
- la distribution normale (Laplace – Gauss) : cette loi permet de modéliser de nombreuses répartitions de variables aléatoires. La loi normale est définie par la densité de probabilité $\varphi : \mathbb{R} \rightarrow \mathbb{R}^+$, d'espérance μ et d'écart type σ :

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad (3.18)$$

Cette loi peut être une bonne représentation d'un système réel, car les symboles ont des poids différents. Nous donnons un exemple pratique utilisant cette loi en annexe B.1, p. 149.

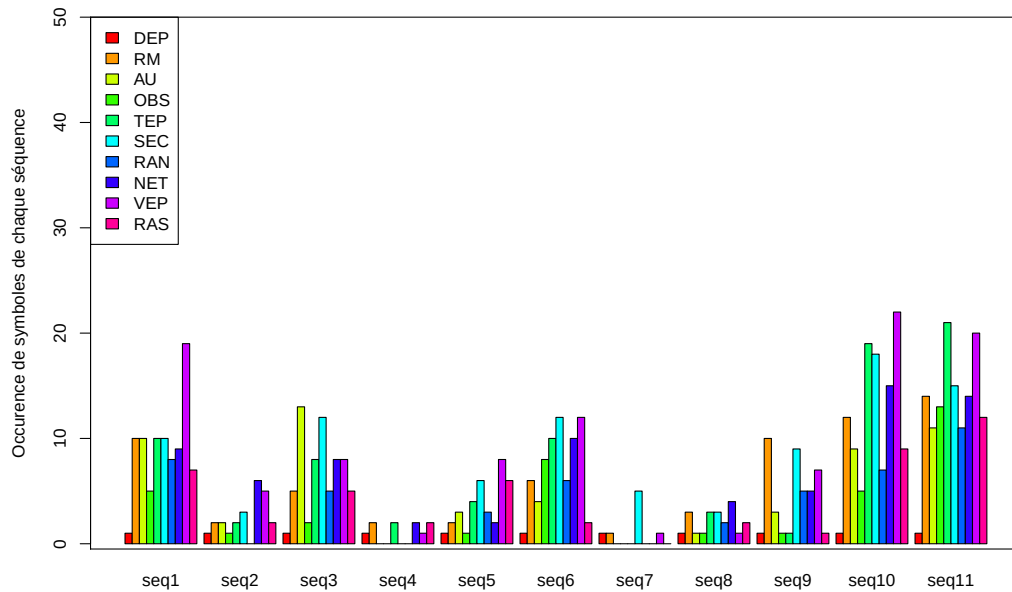
La matrice A (probabilité de passage entre les états) est le résultat de calculs issus de l'algorithme d'apprentissage utilisé pour le modèle considéré (topologie désirée a priori). Dans notre situation, ces résultats sont fournis par le modèle de synthèse avec ses caractéristiques topologiques et reflétant une distribution des symboles suivant une loi normale (Figure 3.8(b), p. 76).

Nous avons lancé une première simulation afin de produire 1000 observations. La matrice A est initialement définie comme suit :

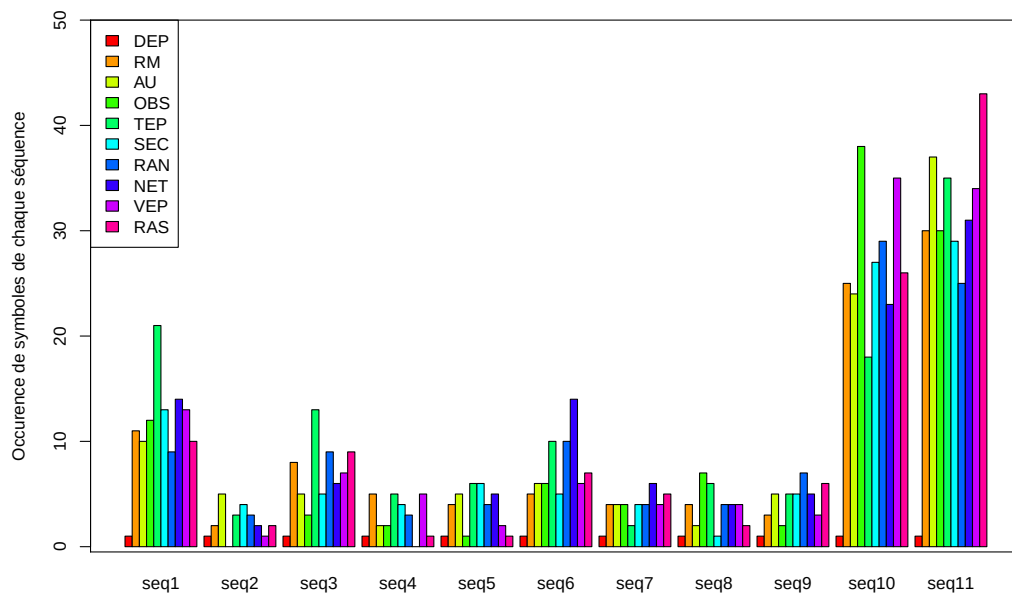
$$A = \begin{pmatrix} 0.500 & 0.250 & 0 & 0.250 \\ 0.100 & 0.070 & 0.500 & 0.330 \\ 0 & 0.005 & 0.495 & 0.500 \\ 0 & 0 & 0.001 & 0.999 \end{pmatrix}. \quad (3.19)$$

Cette matrice de transition est déterminée soit empiriquement pour un apprentissage supervisé, soit aléatoirement. Cette matrice est estimée de nouveau, une fois injectée dans les algorithmes d'apprentissage.

Ces 1000 données simulées correspondent aux observations que l'on trouve couramment dans une base de données de GMAO industrielle. Nous avons ainsi construit 11 séquences de symboles. A chaque séquence (enchaînement de symboles émis) correspond une signature qui reflète l'occurrence de chaque symbole dans la dite séquence. Ceci est illustré par les Figures 3.7(a) et 3.7(b), qui montrent la distribution de chaque symbole lors de la séquence (l'ordre des symboles n'est pas représentatif de l'ordre de la séquence. Remarque : chaque séquence est ponctuée par le symbole DEP (situation de dépannage cf. Tableau 4.9, p. 108), où le processus est arrêté). Ces séquences produisent des symboles stochastiques, issus de la loi **normale** et de la loi **uniforme**, voir Figures 3.8(a) et 3.8(b).

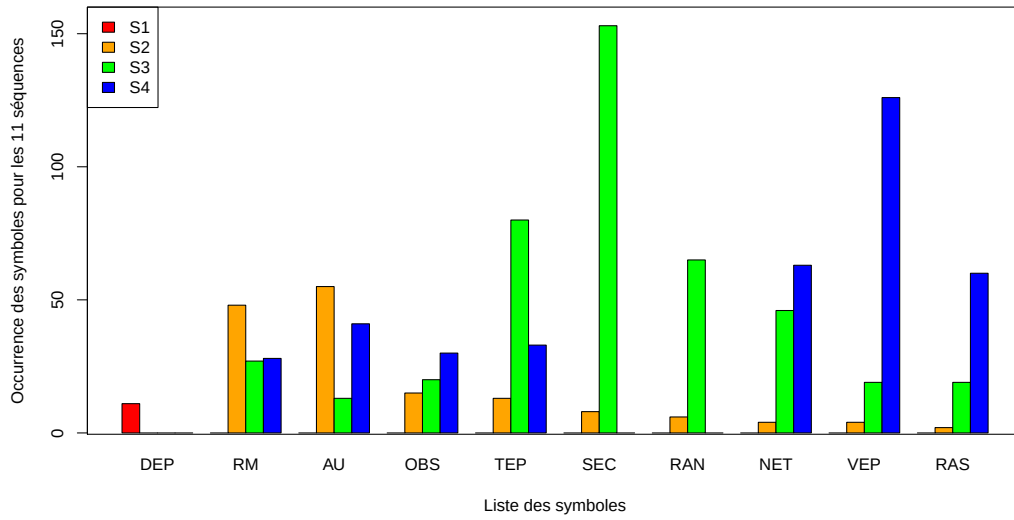


(a) Distribution normale

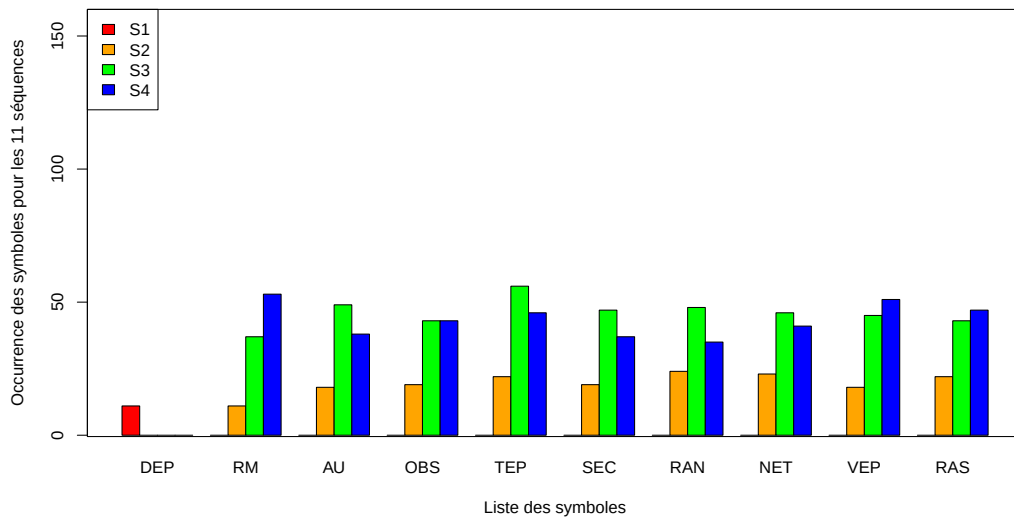


(b) Distribution uniforme

Fig 3.7 – Séquences V des T observations du modèle de synthèse. La Figure 3.7(a) représente la distribution normale et la Figure 3.7(b) représente la distribution uniforme. Sur l'abscisse, nous pouvons voir la liste des 11 séquences d'observations émises par le modèle de synthèse. Pour chaque séquence, on trouve les 10 symboles (DEP, RM, AU, OBS, TEP, SEC, RAN, NET, VEP et RAS). Ces figures illustrent la distribution de chaque séquence pour chaque symbole. Nous remarquons que chaque séquence se termine par le symbole d'arrêt DEP.



(a) Distribution normale



(b) Distribution uniforme

Fig 3.8 – Distribution des symboles par état du modèle de synthèse. La Figure 3.8(a) représente la distribution normale et la Figure 3.8(b) représente la distribution uniforme. Sur l'abscisse, nous pouvons voir la liste des 10 symboles. Pour chaque symbole, on trouve les 4 états (S1, S2, S3 et S4). Ces figures illustrent la distribution de chaque symbole pour chaque état. Nous remarquons que le premier symbole n'est émis que par l'état S1 (symbole d'arrêt pour l'état de non fonctionnement). Nous remarquons aussi que l'état S1 n'émet aucun des 9 autres symboles.

Le modèle de synthèse permet de produire des séquences de symboles Tableau 3.2, p. 77 à partir d'une topologie Figure 3.6, p. 73. Ce modèle de synthèse permet de produire des séquences de couples (Etats, Observations) Figure 3.9, p. 77 complètement définies qui serviront de référence (vérités terrains) pour comparer les résultats obtenus pour les différents tests effectués par la suite.

TEP	TEP	SEC	TEP	TEP	RAS	RAS	DEP	AU	...
-----	-----	-----	-----	-----	-----	-----	-----	----	-----

Tableau 3.2 – Séquence d'un message issue des données de maintenance.

Ces symboles vont donc pouvoir ensuite être implémentés dans les trois topologies étudiées par l'intermédiaire des algorithmes d'apprentissage Baum-Welch (décodage Variables Forward) et Segmental K-means (décodé par Viterbi). Finalement, nous obtenons des couples (Etats, Observations) pour chaque sortie d'automates. Nous évaluons ainsi la pertinence et l'incertitude de chaque topologie.

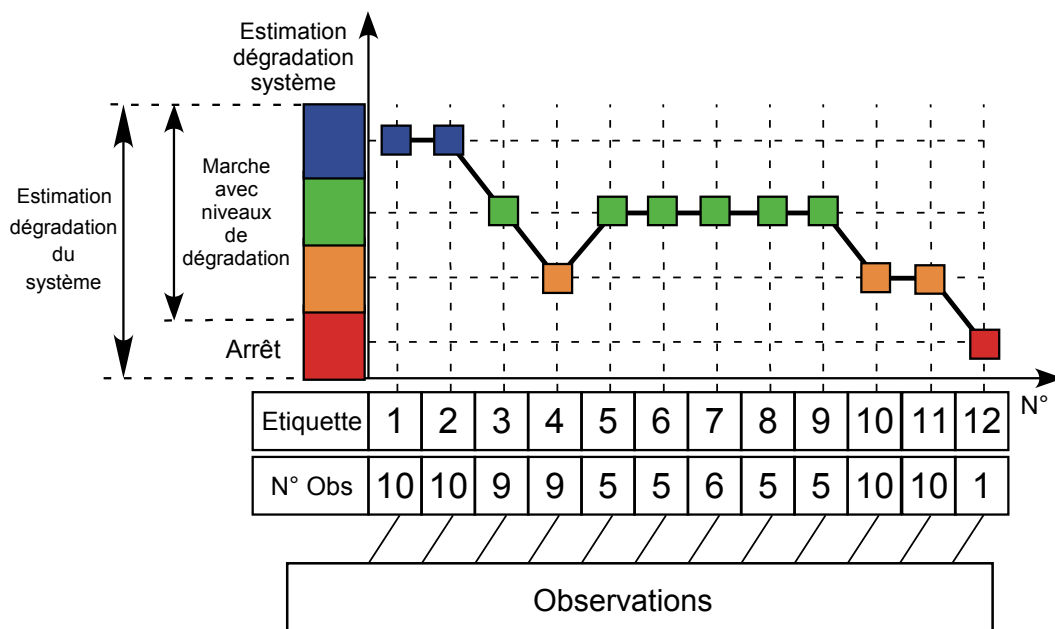


Fig 3.9 – Dégradation d'un processus.

3.5.1 Caractère stochastique du modèle de synthèse

Le Guide [144] du NIST est une suite de 15 tests statistiques qui ont été développés afin d'estimer le caractère aléatoire de séquences binaires. NIST a vérifié la performance de ces tests en utilisant le test statistique de Kolmogorov-Smirnov. Le but de ces tests est de déterminer si les nombres de uns et de zéros dans une séquence sont approximativement les mêmes, tels qu'ils seraient prévus pour une séquence véritablement aléatoire. Nous utilisons le **Test de Fréquence (Monobit)** § 2.5, p. 53, de NIST afin de valider le caractère stochastique du modèle de synthèse.

La séquence est considérée comme *aléatoire* pour $p\text{-value} \geq 0,01$, et non aléatoire pour $p\text{-value} < 0,01$. Nous vérifions que les séquences Figure 3.8(a) et 3.8(b) sont issues d'un générateur aléatoire : résultats Tableau 3.3, p. 78. Les symboles des topologies 1 et 3 sont générés à partir du modèle de référence utilisant la topologie 2. Ces topologies caractérisent un contexte réel de dégradation et de réparation.

Test NIST	p-value	
Topologie du modèle de synthèse	Loi uniforme	Loi normale
Topologie 1	0,47	0,06
Topologie 2	0,30	0,02
Topologie 3	0,47	0,06

Tableau 3.3 – p-value des états générée par le modèle de synthèse.

Les résultats obtenus, Tableau 3.3, p. 78, indiquent que pour tous les modèles, les $p\text{-value} \geq 0,01$. Les séquences obtenues par le générateur sont donc considérées comme *aléatoires*. La différence entre les lois uniformes et lois normales se fonde sur un constat empirique.

3.5.2 Processus d'évaluation

Nous réalisons l'évaluation des différentes topologies vues en Figure 3.5, p. 71 en utilisant un modèle de synthèse. Ce modèle de synthèse nous permet de reproduire les conditions réelles d'un processus industriel étant soumis à des « perturbations » : pannes, réparations, maintenances préventives, etc. (voir la distribution des observations sur la Figure 3.11(a), p. 80). Corrélativement, nous produisons des séquences d'observations (ou symboles) selon les lois de deux différentes distributions (normale et uniforme). Nous injectons ces symboles dans nos trois topologies étudiées au travers de deux algorithmes d'apprentissage :

- apprentissage Baum-Welch, décodé par Variables Forward (voir un exemple de distribution des observations Figure 3.11(b), p. 80),
- apprentissage Segmental K-means, décodé par Viterbi (voir un exemple de distribution des observations Figure 3.11(c), p. 80).

Nous réalisons ainsi des mesures de pertinence et d'incertitude sur les nouvelles observations obtenues pour nos trois topologies étudiées. Le processus complet est résumé sur la Figure 3.10, p. 79.

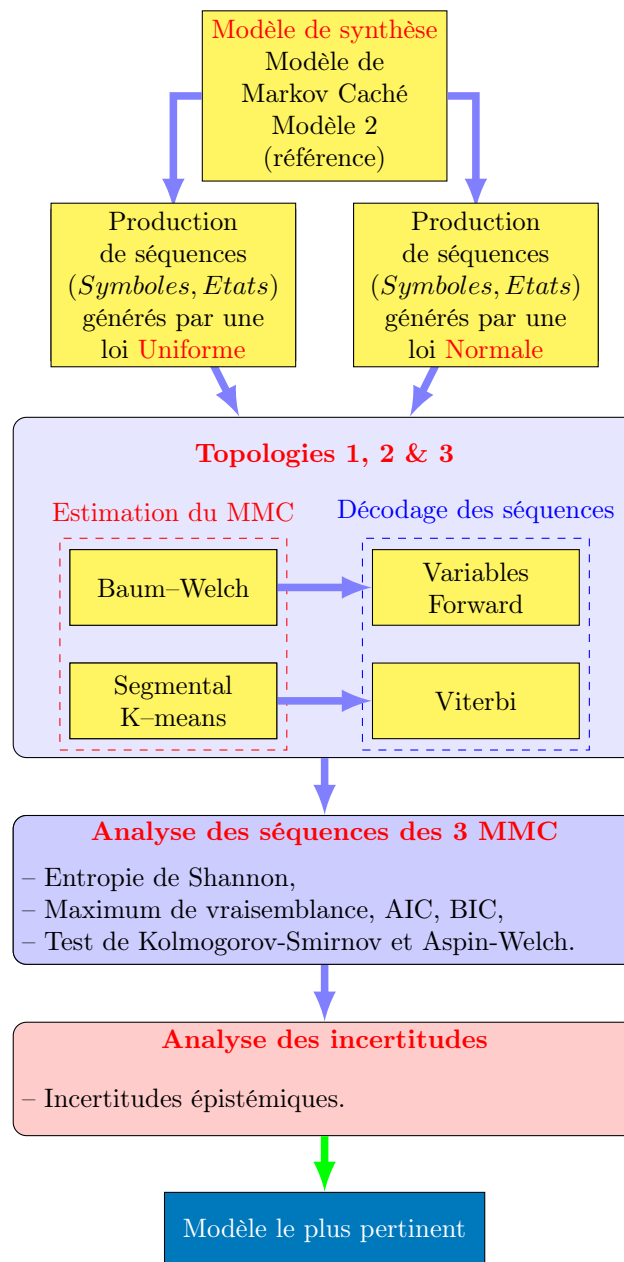


Fig 3.10 – Étapes d'évaluation de l'architecture des modèles, à l'aide d'un modèle de synthèse. Le modèle de synthèse utilise la topologie 2 pour générer des 2-uplet (*Symboles, Etats*) en utilisant les distributions (uniforme et normale). Nous injectons alors ces signatures dans les 3 topologies étudiées. Nous utilisons les 2 algorithmes d'apprentissage et de décodage pour obtenir de nouvelles séquences que nous analysons pour en évaluer la pertinence. Nous essayons ainsi de trouver la meilleure architecture des modèles.

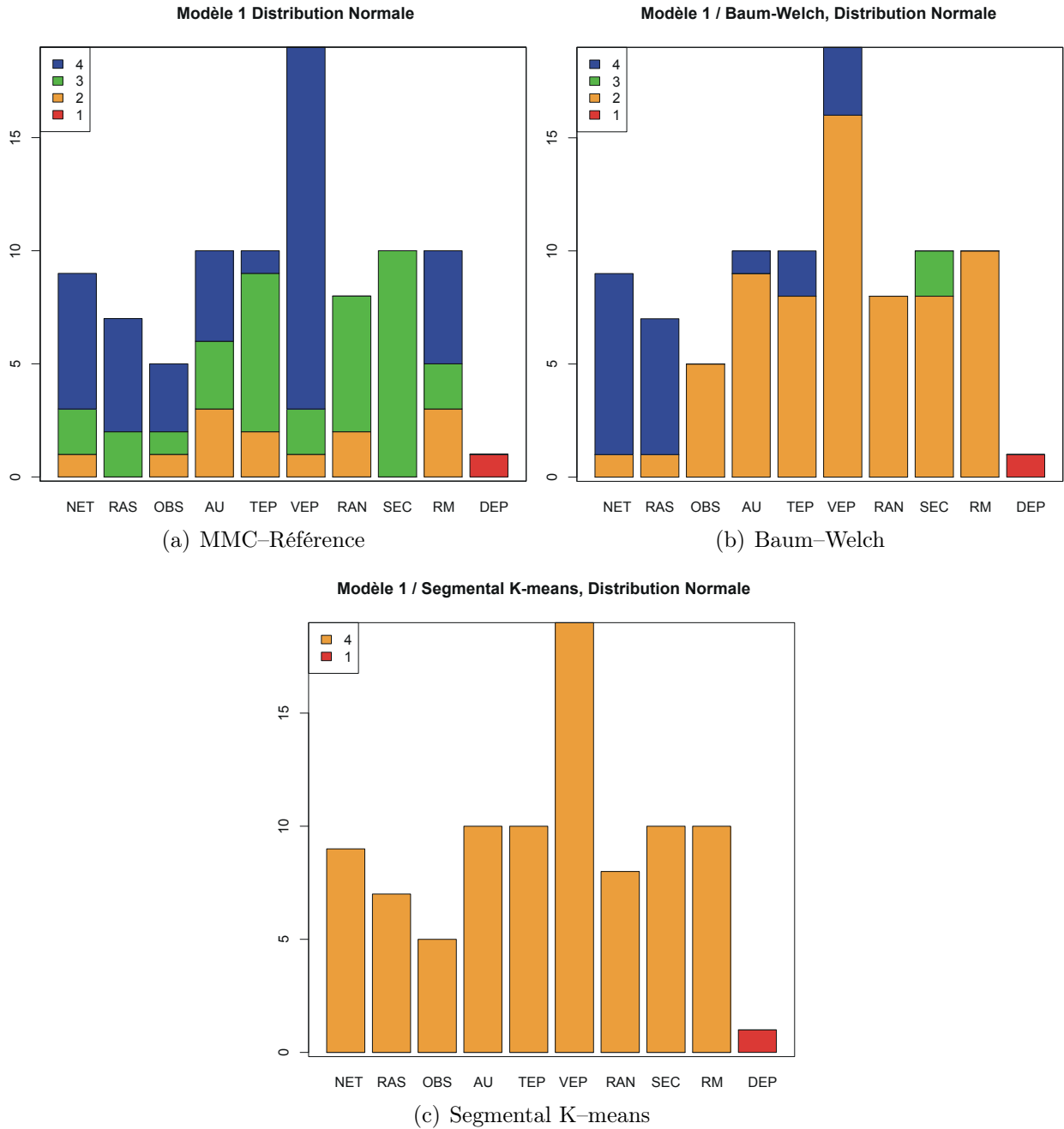


Fig 3.11 – Distribution des symboles par état. La Figure 3.11(a) représente la distribution normale des observations émises par le modèle de synthèse. La Figure 3.11(b) représente la distribution des observations après apprentissage Baum-Welch et décodage par Variables Forward. La Figure 3.11(c) représente la distribution des observations après apprentissage Segmental K-means et décodage par Viterbi.

3.5.3 Processus de génération des symboles

Le modèle de synthèse produit des 2-uplets (*Symboles*, *Etats*) de manière aléatoire pour les trois topologies étudiées. La distribution des symboles est donnée par une loi normale : (*Symb_N*, *Etats_N_i*) et une loi uniforme : (*Symb_U*, *Etats_U_i*) (les différentes topologies sont représentées par « *i* »). Ces différentes séquences sont injectées dans les deux algorithmes d'apprentissage et de décodage. Nous obtenons alors 12 séries de symboles supplémentaires : (*Symb_U*, *Etats_U_iBW*) pour l'algorithme Baum-Welch, décodage par Variables Forward et (*Symb_U*, *Etats_U_iSK*) pour Segmental K-means, décodage par Viterbi. Nous résumons le processus de génération des symboles Figure 3.12, p. 81.

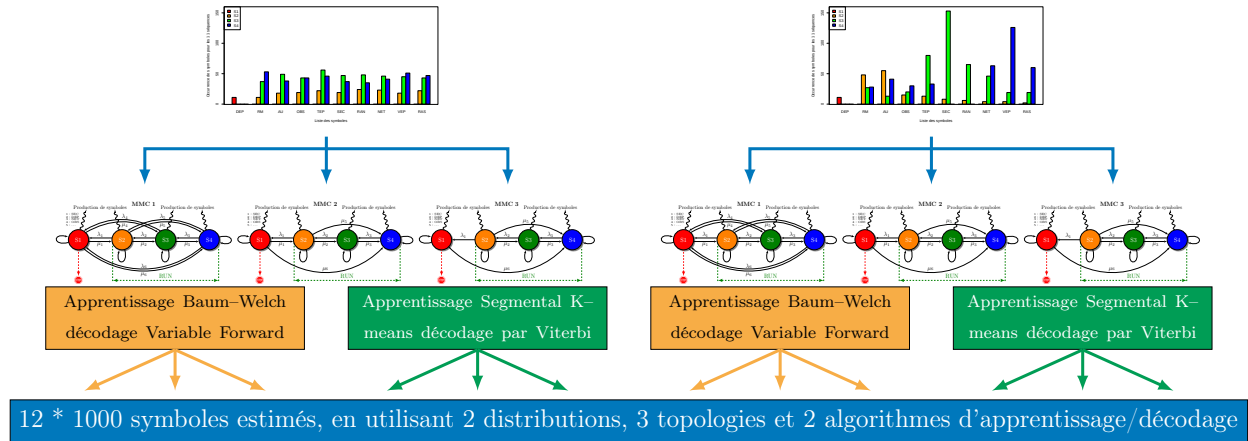


Fig 3.12 – Production des symboles. Cela nous permet de tester différentes distributions de symboles, différentes topologies et différents algorithmes d'apprentissage et de décodage.

3.5.4 Détermination de la loi statistique

Nous voulons ici confirmer nos choix dans la construction du modèle de synthèse : topologies, représentation des distributions des observations, algorithmes d'apprentissage et de « restitution » i.e. de décodage. Pour cela, nous allons construire des modèles les plus simples possibles, basés sur le Principe du Maximum d'Entropie (PME), en utilisant les données issues du modèle de synthèse.

Principe du Maximum d'Entropie (PME) : construction de modèles

A partir d'une distribution issue du modèle de synthèse, nous allons essayer de retrouver les densités des lois de probabilités utilisées sur les observations. En effet, le théorème de Glivenko-Cantelli [156] exprime qu'une loi de probabilité peut être révélée par la connaissance d'un grand échantillon de cette loi. Nos modèles du PME seront construits à partir de 1000 données stochastiques après implémentation dans les topologies et algorithmes d'apprentissage / décodage étudiés. Nous construisons de nouveaux modèles utilisant le PME après avoir injecté ces observations dans nos algorithmes « d'apprentissage et de décodage ». La loi du Maximum d'Entropie de chaque modèle sera obtenue à partir des moments successifs d'ordre K fini [173].

Loi de maximum d'entropie obtenue à partir des moments successifs : Observations : x^K

Niveaux de contraintes empiriques :

$$\overline{x^K} = \frac{1}{1000} \sum_{k=1}^K x_i^k. \quad (3.20)$$

Multiplicateurs :

$$m_k, k \in \{1, \dots, K\}. \quad (3.21)$$

États de Gibbs :

$$f^*(x) = \frac{1}{Z} \exp \left(\sum_{k=1}^K \alpha_k x^k \right), x \in [1, 8]. \quad (3.22)$$

Équations du Principe du Maximum d'Entropie :

$$\frac{1}{Z} \int_0^{1000} x^l \exp \left(\sum_{k=1}^K \alpha_k x^k \right) dx = \overline{x^l}, l \in \{1, \dots, K\}. \quad (3.23)$$

Constante de normalisation :

$$Z = \int_0^{1000} \exp \left(\sum_{k=1}^K \alpha_k x^k \right) dx. \quad (3.24)$$

Pour un ensemble d'observations $\{o_i\}$ dont l'espérance est définie par :

$$E(o_i) = a_i, i \in \{1, \dots, K\}, \quad (3.25)$$

les a_i étant les niveaux de **contraintes** intervenant dans le calcul des moments d'ordre K .

Exemple

Prenons le cas d'une variable X dans $\mathbb{R}$. Nous voulons utiliser les trois premiers moments :

$$\mathcal{M} = \{(x, a_1), (x^2, a_2), (x^3, a_3)\}, a_i \in \mathbb{R}. \quad (3.26)$$

Il existe un état de Gibbs de la forme suivante :

$$f(x) = \frac{1}{Z} \exp(\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3), \alpha_i \in \mathbb{R}, \quad (3.27)$$

c'est la loi d'entropie maximum pour les **multiplicateurs** α_i . Le système d'équations non linéaires suivant correspond aux équations du PME :

$$\begin{cases} a_1 = \frac{1}{Z} \int_x x \exp(\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3) dx \\ a_2 = \frac{1}{Z} \int_x x^2 \exp(\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3) dx \\ a_3 = \frac{1}{Z} \int_x x^3 \exp(\alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3) dx. \end{cases} \quad (3.28)$$

Z est la constante de normalisation définie comme suit :

$$Z = \int_x \exp \left(\sum_{k=1}^K \alpha_k x^k \right) dx. \quad (3.29)$$

Nous calculons les moments successifs sur les données du modèle de synthèse. Nous utilisons le logiciel SCILAB pour résoudre les équations non linéaires du PME. Les algorithmes de calcul basés sur la méthode hybride de Powell [65] et [132]. Pour optimiser ce calcul, nous évaluons la matrice Jacobienne (voir § C, p. 153). L'algorithme s'arrête lorsqu'il a atteint un niveau de précision prédéfini.

Nous réaliserons ensuite un test d'adéquation (Kolmogorov-Smirnov par exemple), afin d'évaluer les différences entre les fréquences empiriques et les fréquences estimées par ces différents modèles.

Résultats attendus

Différents critères de pertinence nous ont déjà fourni des résultats en adéquation avec ceux de [179] : l'algorithme d'apprentissage Baum-Welch / décodage Variables Forward ainsi que la distribution normale sur les observations. Avec cette méthode, nous espérons trouver une nouvelle adéquation entre les données issues du modèle de synthèse et les modèles issus du Principe du Maximum d'Entropie.

Exemple de résolution sous Scilab [149] en utilisant l'algorithme de Powell [132] :

$def f(z = f8(l), z = [(x^8) \exp(\lambda_1 x^1 + \lambda_2 x^2 + \lambda_3 x^3 + \lambda_4 x^4 + \lambda_5 x^5 + \lambda_6 x^6 + \lambda_7 x^7 + \lambda_8 x^8) - a(1), \dots]$

Méthode des moindres carrés par Newton

Cette méthode permet de comparer des données expérimentales à un modèle mathématique. Pour notre cas, il nous servira à comparer les séquences d'observations du MMC synthétique avec celles d'autres modélisations mathématiques.

Dans notre exemple, on suppose que deux variables physiques x et y sont liées entre elles par une relation exponentielle, du type :

$$y = a \cdot \exp(b \cdot x). \quad (3.30)$$

Nous voulons déterminer a et b de manière à minimiser :

$$E(a, b) = \sum_{i=1}^p (y_i - a \cdot \exp(b \cdot x_i))^2. \quad (3.31)$$

Solutions trouvées à l'aide de Scilab (voir graphes 3.13(a), p. 85, 3.13(b), p. 85 et 3.13(c), p. 85).

Modèle de la distribution des symboles pour l'état S2 :

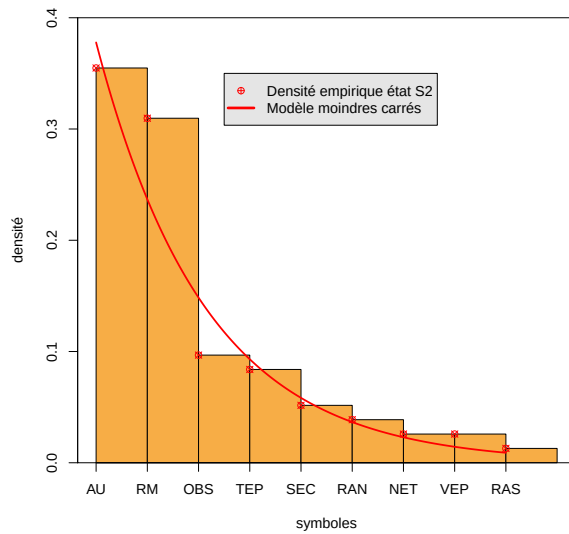
$$y = 0,6027 \cdot \exp(-0,4670 \cdot x). \quad (3.32)$$

Modèle de la distribution des symboles pour l'état S3 :

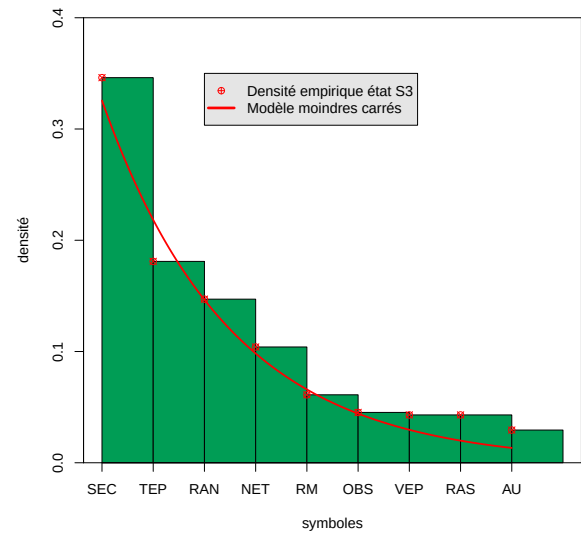
$$y = 0,4853 \cdot \exp(-0,3993 \cdot x). \quad (3.33)$$

Modèle de la distribution des symboles pour l'état S4 :

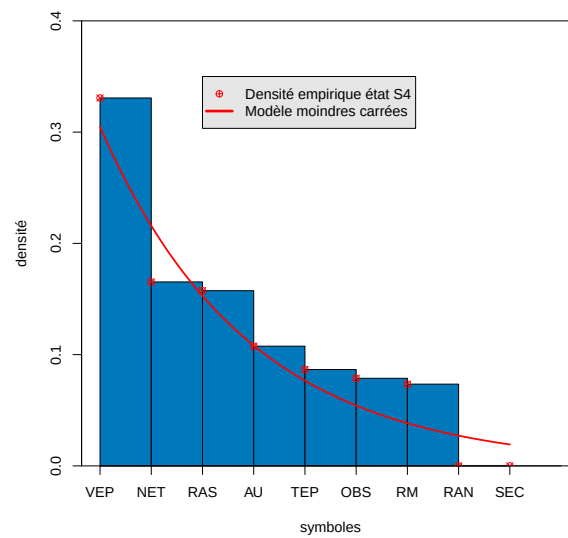
$$y = 0,4304 \cdot \exp(-0,3452 \cdot x). \quad (3.34)$$



(a) État S2



(b) État S3



(c) État S4

Fig 3.13 – Modélisation Newton de la densité des états.

3.6 Conclusion

Notre approche permet de modéliser des niveaux de dégradation d'un processus quelconque. Nous bénéficions ainsi de toutes les propriétés relatives à la théorie de Markov :

- le processus de Markov qui impose un échantillonnage régulier ;
- les chaînes de Markov pour des processus à temps discrets ;
- les MMC pour la modélisation d'un processus sous la forme d'automates.

Nous avons ainsi étudié trois modèles d'automates différents. Ils modélisent sur quatre niveaux les états de dégradation d'un processus (S1, S2, S3 et S4). Leurs topologies diffèrent uniquement sur les transitions permises entre états (forçage de la topologie) ; le choix de ces transitions étant déjà étudié dans Vrignat [177].

Nous avons décomposé nos attentes théoriques selon trois axes principaux : étudier la pertinence des observations ou symboles sans connaissance a priori sur les résultats, faire évoluer la modélisation (i.e. déterminer les éléments des modèles qui vont nous permettre d'optimiser nos MMC) et enfin déterminer la pertinence des architectures des modèles. Cette dernière nous donnera des indicateurs sur la meilleure topologie, le meilleur algorithme d'apprentissage et de décodage et la meilleure distribution des observations synthétiques.

Nous avons introduit également un modèle de simulation que nous appelons modèle de synthèse. Nous avons montré le comportement stochastique de ce modèle de synthèse. Celui-ci nous a permis de générer des séquences d'observations parfaitement définies entre niveau de dégradation et symboles observés. Le choix de la topologie utilisée pour la production des symboles ainsi que la matrice de transition A ont été démontrés dans Vrignat [177]. Nous présentons nos résultats dans le chapitre suivant.

Chapitre 4

Expérimentations et résultats

Nous donnons dans ce chapitre, les résultats issus de notre modèle de simulation représentant le fonctionnement d'une GMAO industrielle. Nous présentons dans un premier temps les caractéristiques de ce modèle, puis des résultats sur les trois points suivants : la pertinence des observations, la pertinence de l'architecture des différents modèles utilisés et l'évolution de la modélisation afin d'en améliorer son architecture. Nous réalisons ensuite deux études expérimentales, issues de cas concrets de processus industriel. Nous débutons par une présentation de ces environnements expérimentaux. Nous présentons ensuite nos résultats sur les trois points précédemment cités. Nous terminons en donnant des perspectives d'améliorations possibles pour la modélisation.

4.1 Résultats de simulation

4.1.1 Paramètres du modèle de simulation

Le modèle de simulation nous permet d'obtenir des données se rapprochant le plus possible de la réalité (voir § 3.5, p. 72).

Nous utilisons la topologie 2 (Figure 4.1, p. 89) comme modèle de référence (voir discussion § 3.5, p. 72).

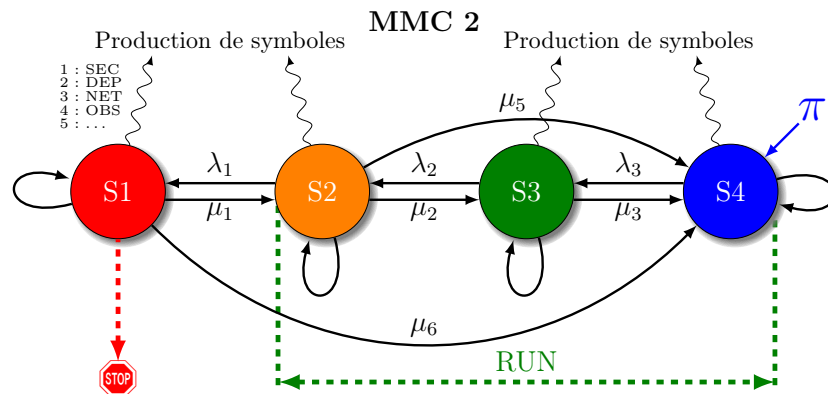


Fig 4.1 – Modèle de Markov caché à 4 états, topologie 2, référence du modèle de synthèse.

Nombre d'états

Nous avons choisi d'utiliser un modèle à 4 états notés S1, S2, S3 et S4 sur la Figure 4.1, p. 89. Nous retrouvons ces 4 niveaux d'alerte dans d'autres champs disciplinaires comme le plan canicule ou vigipirate.

Des tests ont été effectués dans [177] avec 5 états et plus. L'algorithme de calcul effectue alors un étalement des probabilités et le modèle ne permet plus d'avoir des états significatifs. La décision de l'expert s'avèrerait alors plus délicate à trancher. A l'inverse, avec 3 états, l'état médian serait trop rapidement atteint ne permettant pas ainsi à l'expert de réagir efficacement.

Le sens que nous pouvons attribuer à chacun des états est le suivant :

- S4 : « neuf ou quasi neuf »,
- S3 : « usé »,
- S2 : « très usé »,
- S1 : « inutilisable ».

Nombre de symboles

Les symboles correspondent aux observations liées au système à modéliser i.e. ce que l'on veut ou que l'on peut observer. Il faut que ces observations soient significatives *a priori*. Nous avons choisi ici un alphabet de 10 symboles (les symboles émis représentent des interventions de maintenance, voir la codification Tableau 3.1, p. 70). Ce nombre est déterminé en fonction du processus que l'on veut modéliser et dépend de la politique de maintenance de l'entreprise (une autre étude de cas nous a amené à considérer plus de 20 symboles).

Topologie de la chaîne de Markov

Nous utilisons les lois normale et uniforme pour l'émission des symboles. Ces deux distributions sont décrites dans le § 3.5, p. 72. La relation symbole / état est déterminée par la matrice de passage B définie dans Vrignat [177].

Le passage d'un état à un autre se réalise suivant une loi exponentielle (voir Annexe B.3, p. 151). Cette loi est très utilisée en fiabilité dont le taux de défaillance λ est constant. Elle décrit la vie des matériels qui subissent des défaillances brutales.

La densité de probabilité d'une loi exponentielle a pour expression :

$$\varphi(t; \lambda) = \begin{cases} \lambda e^{-\lambda t} & , t \geq 0, \\ 0 & , t < 0. \end{cases} \quad (4.1)$$

La fonction fiabilité s'écrit :

$$R(t) = e^{-\lambda t}. \quad (4.2)$$

4.1.2 Pertinence des observations de simulation

Dans cette section, nous voulons déterminer les symboles ou les séquences de symboles les plus pertinents, à partir de séquences d'observations issues du modèle de synthèse. Nous utilisons pour cela les différents principes de l'entropie de Shannon (voir § 2.4.3, p. 41).

Les entropies de chaque symbole (chaîne de Markov d'ordre 1) sont calculées par rapport aux états S1, S2, S3 et S4 pour les deux distributions du modèle de synthèse. Nous établissons ainsi les entropies pour chaque symbole, en fonction de tous les états du modèle.

Nous illustrons dans le Tableau 4.1, p. 91, l'entropie du modèle de synthèse avec la loi normale. Les résultats du modèle de synthèse avec la loi uniforme sont donnés dans le Tableau 4.2, p. 91.

Le Tableau 4.1 montre que le symbole « TEP » obtient une entropie maximale de **1,129**. Ceci prouve que ce symbole est le mieux réparti parmi les états du système. Nous remarquons aussi que l'entropie du symbole « DEP » est nulle. Ce symbole est donc totalement discriminé

Symboles	S1	S2	S3	S4	Total	Entropie
AU		18	49	38	105	1,0277
DEP	19				19	0,000
NET		23	46	41	110	1,077
OBS		19	43	43	105	1,036
RAN		24	48	35	107	1,064
RAS		22	43	47	112	1,082
RM		11	37	53	101	0,952
SEC		19	47	37	103	1,025
TEP		22	56	46	124	1,129
VEP		18	45	51	114	1,068

Tableau 4.1 – Mesure de l’entropie, données issues du modèle de synthèse, loi normale.

Symboles	S1	S2	S3	S4	Total	Entropie
AU				105	105	0,347
DEP	19				19	0,000
NET		1		109	110	0,686
OBS		1		104	105	0,677
RAN		1		106	107	0,681
RAS				112	112	0,359
RM		1		100	101	0,670
SEC		4		99	103	0,864
TEP			1	123	124	0,377
VEP		2		112	114	0,824

Tableau 4.2 – Mesure de l’entropie, données issues du modèle de synthèse, loi uniforme.

i.e. l’état S1 correspond toujours à l’émission de ce symbole (« DEP ») et uniquement ce symbole.

Pour la distribution uniforme, le Tableau 4.2 montre que le symbole « SEC » obtient une entropie maximale de **0,864**. Nous trouvons également une entropie nulle pour le symbole « DEP ».

Nous utilisons par la suite la notion de filtre entropique vu au § 2.4.3, p. 42 pour éliminer les symboles ayant une entropie nulle et ceux ayant une entropie maximale. Nous éliminons ainsi les deux symboles « DEP » et « TEP » pour la distribution normale puis « DEP » et « SEC » pour la distribution uniforme.

Une fois les symboles précédents éliminés du processus d’évaluation, nous établissons un classement des symboles les plus pertinents par ordre d’entropie décroissante. Ainsi, si nous classons les symboles du plus pertinent au moins pertinent, nous obtenons les séquences suivantes :

- RAS, NET, VEP, RAN, OBS, AU, SEC, RM pour la distribution normale,
- VEP, NET, RAN, OBS, RM, TEP, RAS, AU pour la distribution uniforme.

Ces séquences sont obtenues en utilisant l'algorithme d'élimination successive des symboles (voir §A.5, p. 146).

4.1.3 Évolution de la modélisation

Fenêtre glissante

Sans connaissance a priori, nous avons évalué différentes signatures par mesure de l'entropie de Shannon. Cette signature est considérée comme une chaîne de Markov cachée du 1^{er} ordre. L'entropie de Shannon nous a permis d'évaluer la pertinence des observations. Nous avons trouvé dans cette étude un maximum d'entropie pour des données simulées issues de deux distributions différentes. La première valeur maximale est de 88 symboles pour la distribution uniforme et 152 symboles pour la distribution normale. Les résultats Figure 4.2, p. 93 montrent aussi une valeur asymptotique de l'entropie. Cela laisse supposer qu'à partir d'un certain nombre d'échantillons observés, l'apport d'information par les symboles suivant devient très faible voire nulle. Le modèle ainsi établi avec un nombre fini de symboles, pourrait être remis à jour en fonction des nouvelles activités de maintenance. Pour une maintenance prédictive, nous pourrions ainsi établir une « fenêtre glissante » contenant au moins ce nombre de symboles trouvé précédemment. Nous constatons également de manière empirique que les deux courbes ont un écart constant au delà de 200 symboles.

Le processus d'évaluation de l'entropie utilise les spécifications suivantes (voir Figure 4.3, p. 94) :

- 1000 2-uplets ($Symb_U, Etat_U$), pour la distribution uniforme (Figure 3.8(a)),
- 1000 2-uplets ($Symb_N, Etat_N$), pour la distribution normale (Figure 3.8(b)).

Le modèle de synthèse génère 1000 2-uplets ($Symboles, Etats$) en utilisant la distribution uniforme ou normale ($(Symb_U, Etat_U)$ ou $(Symb_N, Etat_N)$). 12 séquences de 1000 2-uplets sont ainsi générées. Les 2-uplets sont utilisés dans le processus d'évaluation afin de déterminer le nombre minimal de données de modélisation, en utilisant l'entropie de Shannon. Chaque séquence se termine par un arrêt du processus (symbole DEP).

Évolution de l'entropie pour le modèle de synthèse

L'évolution de l'entropie de Shannon à partir des données issues du modèle de synthèse nous donne les résultats présentés sur la Figure 4.2, p. 93. Nous observons un maximum d'entropie pour 152 symboles avec la distribution normale et 88 symboles avec la distribution uniforme. Nous voyons également que la courbe est asymptotique au-delà de 200 symboles.

Algorithmes d'apprentissage

La courbe en noire, Figure 4.4, p. 94, correspond à l'évolution de l'entropie de Shannon sur le modèle de synthèse en utilisant la topologie 2. Nous avons trouvé les mêmes genres de courbes pour les topologies 1 et 3. Les courbes continues sont obtenues à partir des données issues de la loi normale et celles en pointillés, à partir de la loi uniforme.

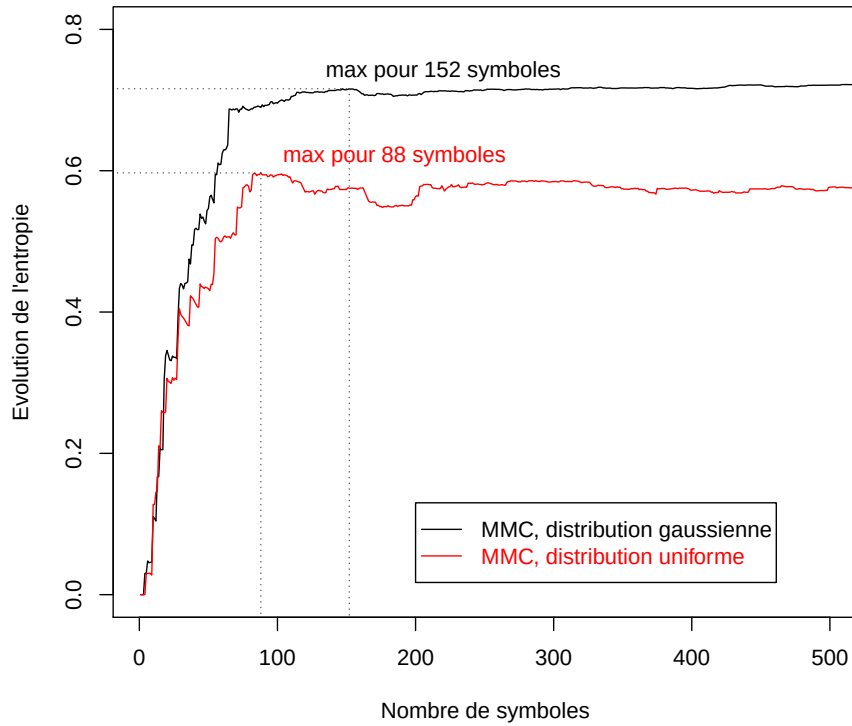


Fig 4.2 – Nombre minimal de données par entropie de Shannon. Les données sont issues du modèle de synthèse.

Les courbes rouges décrivent l'évolution de l'entropie de Shannon du modèle de synthèse, après implémentation dans l'algorithme d'apprentissage Baum-Welch décodé par Variables Forward.

Les courbes bleues décrivent l'évolution de l'entropie de Shannon du modèle de synthèse par l'algorithme d'apprentissage Segmental K-means décodé par Viterbi.

Nous observons ainsi que l'algorithme d'apprentissage Baum-Welch décodé par Variables Forward est plus efficace que celui de Segmental K-means décodé par Viterbi car son entropie est supérieure.

Analyse sur une fenêtre glissante

Le but de cette partie était de trouver une quantité minimale de données pour estimer correctement un modèle. En d'autres termes : existe-t-il une valeur limite (L) de l'entropie vers laquelle celle-ci converge ? Une telle convergence nous permettra de conclure qu'un nombre **fini** de symboles permettra une modélisation optimale. Cette limite est définie par la limite de la fonction d'entropie H lorsque S tend vers $+\infty$:

$$\lim_{S \rightarrow +\infty} H(S) = L, \quad (4.3)$$

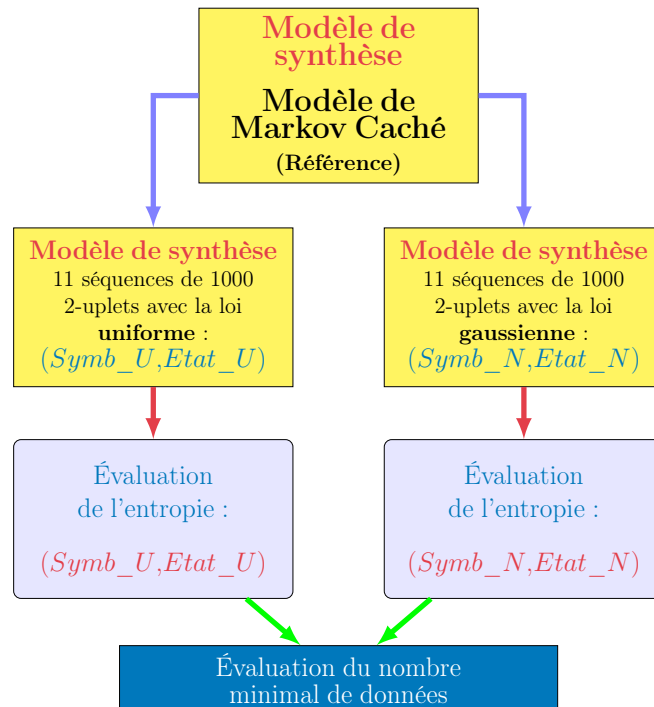


Fig 4.3 – Évaluation du nombre minimal de données.

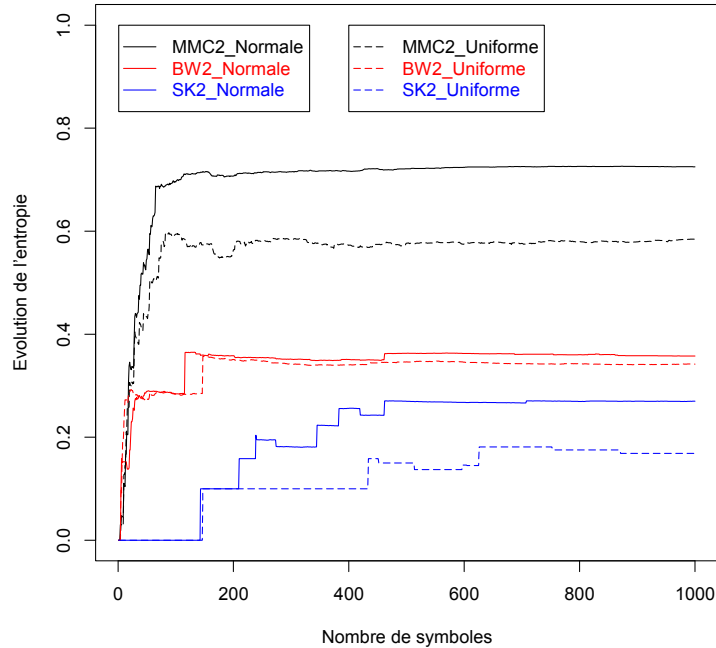


Fig 4.4 – Nombre minimal de données par entropie de Shannon en utilisant les 2 algorithmes d'apprentissage.

S est le nombre de symboles de la séquence étudiée.

Le processus d'évaluation du nombre minimal de données est donné Figure 4.5, p. 95.

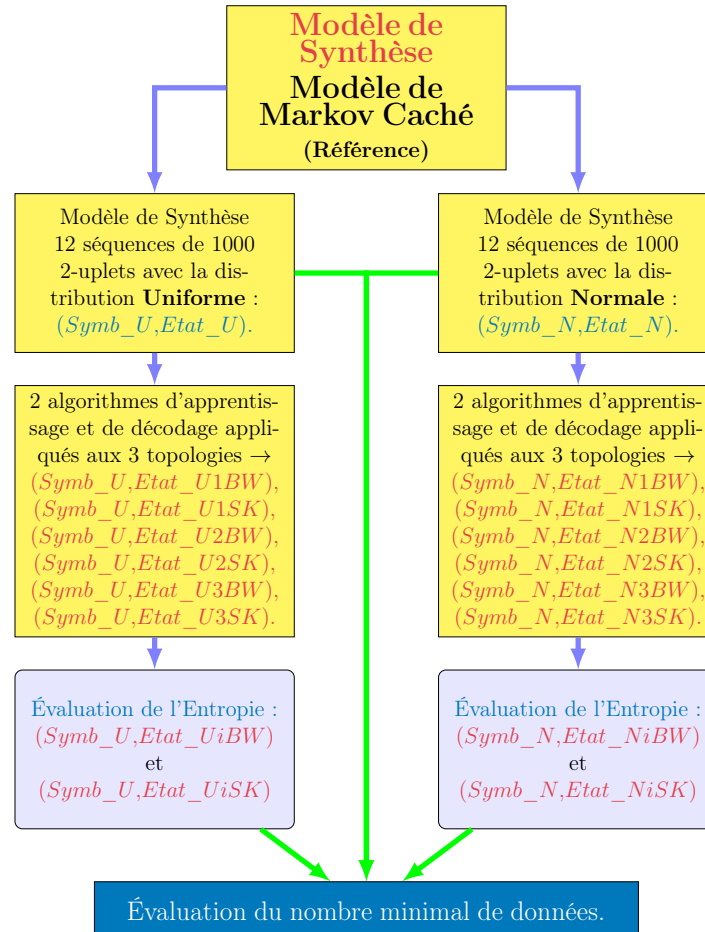


Fig 4.5 – Évaluation du nombre minimal de données pour une utilisation optimale des algorithmes d'apprentissage.

Ce nombre optimal de symboles ou d'observations permettra à l'expert en maintenance de réajuster périodiquement ses modèles, voir Figure 4.6, p. 96. Par exemple : avec un échantillonnage à la journée, nous pouvons supposer que la mise à jour du modèle existant, à partir des nouveaux événements, pourrait se faire mensuellement ou bi-mensuellement (30 à 60 nouveaux événements). Ce délai pourrait être réajusté lors d'un changement de recette de fabrication sollicitant de manière différente le processus.

Essais avec une fenêtre glissante normée

Pour tenter de valider la fenêtre normée précédemment trouvée, nous avons injecté des séquences de 200 symboles dans notre modèle. Nous avons ensuite réévalué ces symboles en utilisant l'algorithme **Baum-Welch** avec un décodage **variable Forward**. Nous étudions ainsi de nouveau l'entropie, en utilisant cette quantité minimale de symboles pour les données issues du modèle de synthèse :

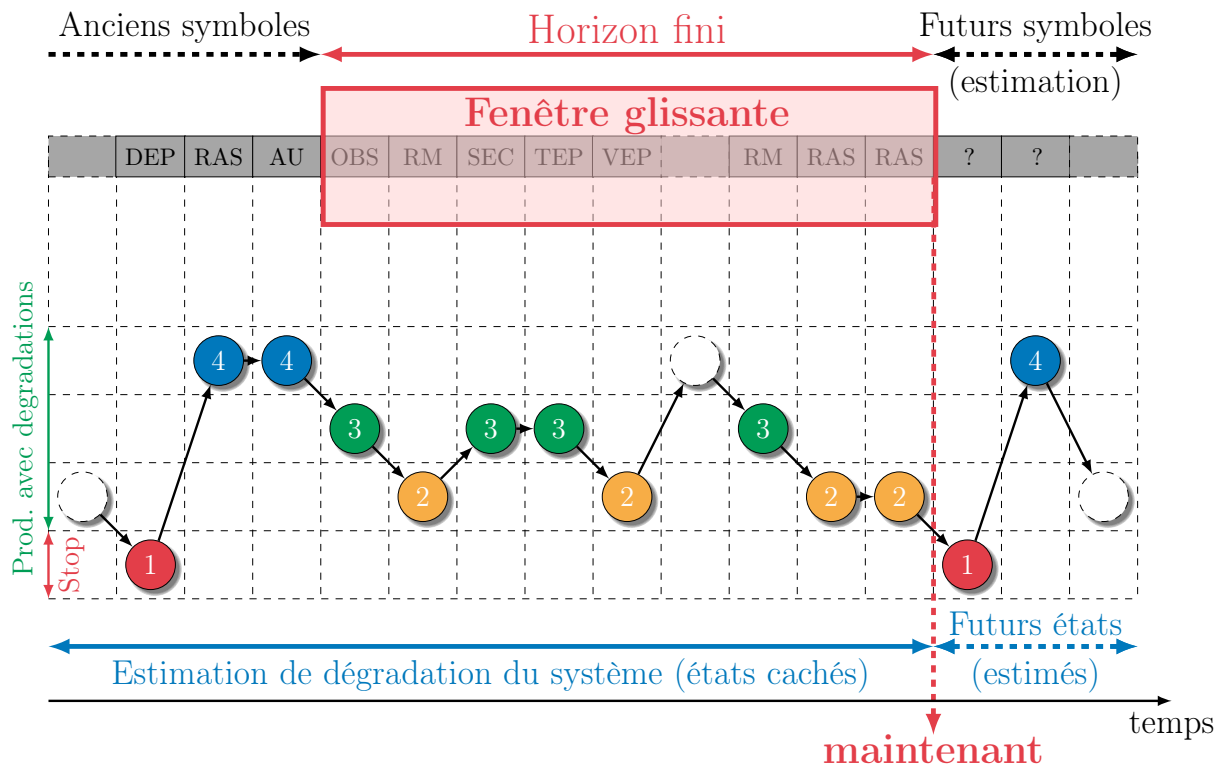


Fig 4.6 – La fenêtre glissante en rouge contient un nombre minimal de symboles. L’entropie est maximale pour ce nombre de symboles.

- les résultats sont donnés Figure 4.7, p. 97. Une fois la phase de transition terminée, l’entropie (en noire) reste inférieure à celle du modèle de synthèse (en rouge). Nous observons également qu’une fois l’apprentissage de la première séquence terminée, l’entropie rentre dans une phase « asymptotique », identique à la progression de l’entropie du modèle de synthèse.

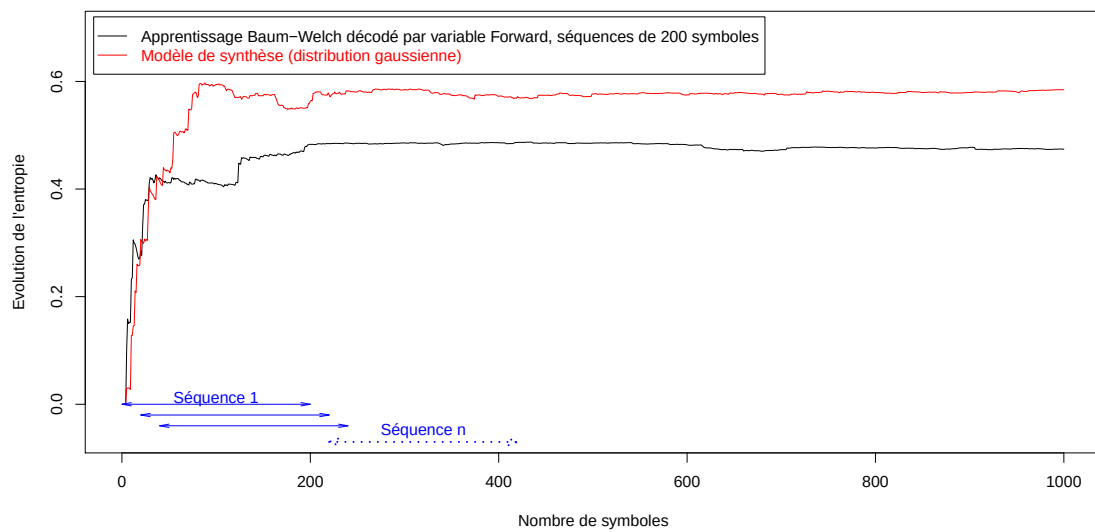


Fig 4.7 – Évolution de l'entropie évaluée avec une fenêtre glissante normée (200 symboles). Les données d'apprentissage sont issues du modèle de synthèse en utilisant la distribution normale.

Discussion

La fenêtre glissante (voir Figure 4.6) contient le nombre minimal de symboles trouvés à partir de la Figure 4.2, p. 93, où l'entropie est maximale. Tous les symboles à l'intérieur de cette fenêtre permettraient de réaliser régulièrement une estimation des modèles. Ainsi, dans une phase d'exploitation des indicateurs de niveaux de dégradation, les actions engagées vont modifier le comportement du système (notre but est de « repousser » la panne). Les événements qui ont permis d'apprendre les modèles ne seront donc plus à l'image du nouveau comportement. Nous cherchons alors à quel moment sera-t-il pertinent de réajuster les modèles.

4.1.4 Pertinence de l'architecture des modèles de simulation

Dans cette section, nous allons évaluer la pertinence de la modélisation. Le terme *architecture* désigne l'ensemble des éléments qui composent le modèle i.e. la topologie, l'algorithme d'apprentissage et la loi statistique.

Entropie de Shannon

Sans connaissance a priori, nous calculons les entropies des différentes architectures étudiées. L'entropie de la **Topologie 2**, MMC 2 (voir Figure 4.17(b), p. 109) est significativement plus élevée que celles des autres topologies avec la **distribution normale** et avec l'algorithme d'apprentissage **Baum-Welch** décodage **variable Forward** (voir Figure 4.8, p. 99). Les résultats quantitatifs sont donnés dans le Tableau 4.3, p 98.

Topologie	Estimations - loi Uniforme	Entropie	Avec filtre Entropique
1	Baum-Welch / Variables Forward	0.484	0.510
1	Segmental K-means / Viterbi	0.480	0.486
2	Baum-Welch / Variables Forward	1.215	1.326
2	Segmental K-means / Viterbi	1.216	1.225
3	Baum-Welch / Variables Forward	0.923	0.990
3	Segmental K-means / Viterbi	0.934	0.939
Estimations - loi Normale			
1	Baum-Welch / Variables Forward	0.774	0.933
1	Segmental K-means / Viterbi	0.517	0.554
2	Baum-Welch / Variables Forward	1.218	1.361
2	Segmental K-means / Viterbi	1.221	1.252
3	Baum-Welch / Variables Forward	0.774	0.882
3	Segmental K-means / Viterbi	1.258	1.208

Tableau 4.3 – Entropie moyenne pour les algorithmes d'apprentissage et décodage.

Maximum de vraisemblance

Les résultats du Tableau 4.4, p. 99 montrent un minimum pour la **topologie 2** avec la **distribution normale**.

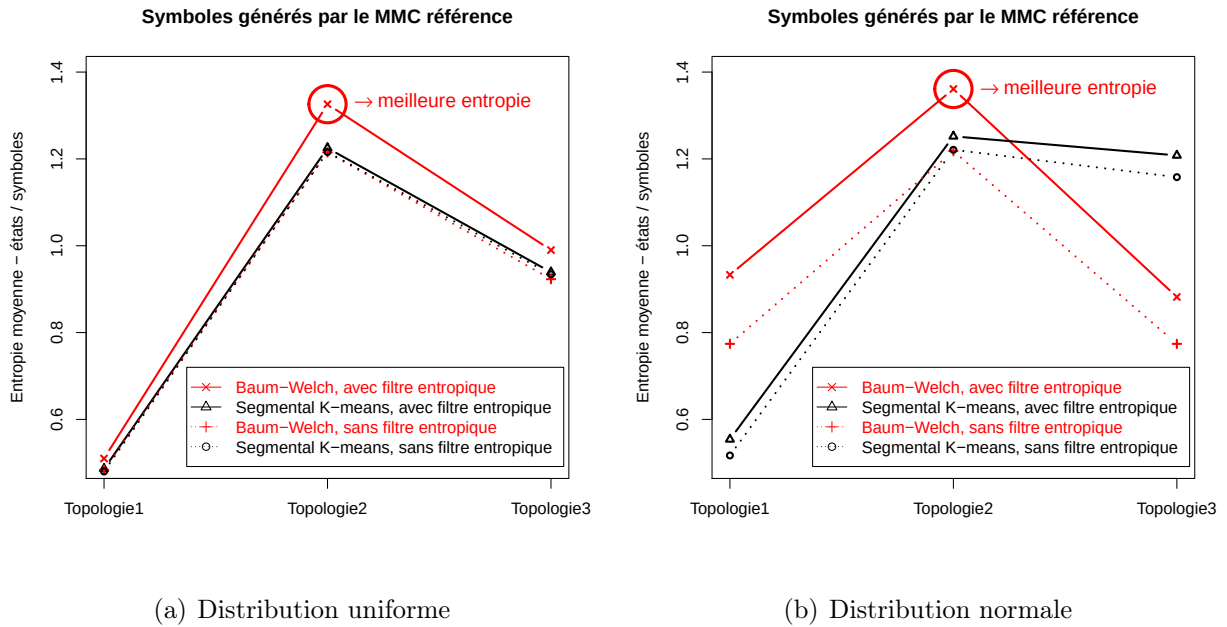


Fig 4.8 – Mesure de l'entropie de Shannon des données issues du modèle de synthèse. Figure 4.8(a) pour la distribution uniforme et Figure 4.8(b) pour la distribution normale.

Le graphe de la Figure 4.9, p 100 montre que la topologie 2 est la plus pertinente en terme de vraisemblance, pour les 2 algorithmes d'apprentissage et pour les 2 distributions étudiées. Malheureusement, concernant la distribution normale, les valeurs de la log-vraisemblance sont trop proches pour déterminer le meilleur algorithme d'apprentissage.

Tableau 4.4 – Log-Vraisemblance.

Topologie	Estimations - loi Uniforme	Log-Vraisemblance
1	Baum-Welch / Variables Forward	-1054.73
1	Segmental K-means / Viterbi	-1249.57
2	Baum-Welch / Variables Forward	-1261.89
2	Segmental K-means / Viterbi	-2019.02
3	Baum-Welch / Variables Forward	-1155.37
3	Segmental K-means / Viterbi	-1255.05
Estimations - loi Normale		
1	Baum-Welch / Variables Forward	-1893.86
1	Segmental K-means / Viterbi	-2066.17
2	Baum-Welch / Variables Forward	-2095.62
2	Segmental K-means / Viterbi	-2101.61
3	Baum-Welch / Variables Forward	-2018.01
3	Segmental K-means / Viterbi	-2029.99

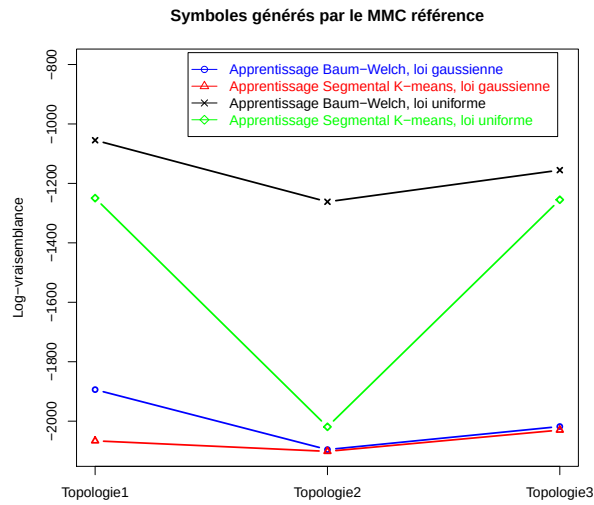


Fig 4.9 – Critère de log-vraisemblance.

Critères d'AIC (Akaike Information Criterion) et de BIC (Bayesian Information Criterion)

Les résultats du critère d'AIC sont présentés Figure 4.10(a), p 100.

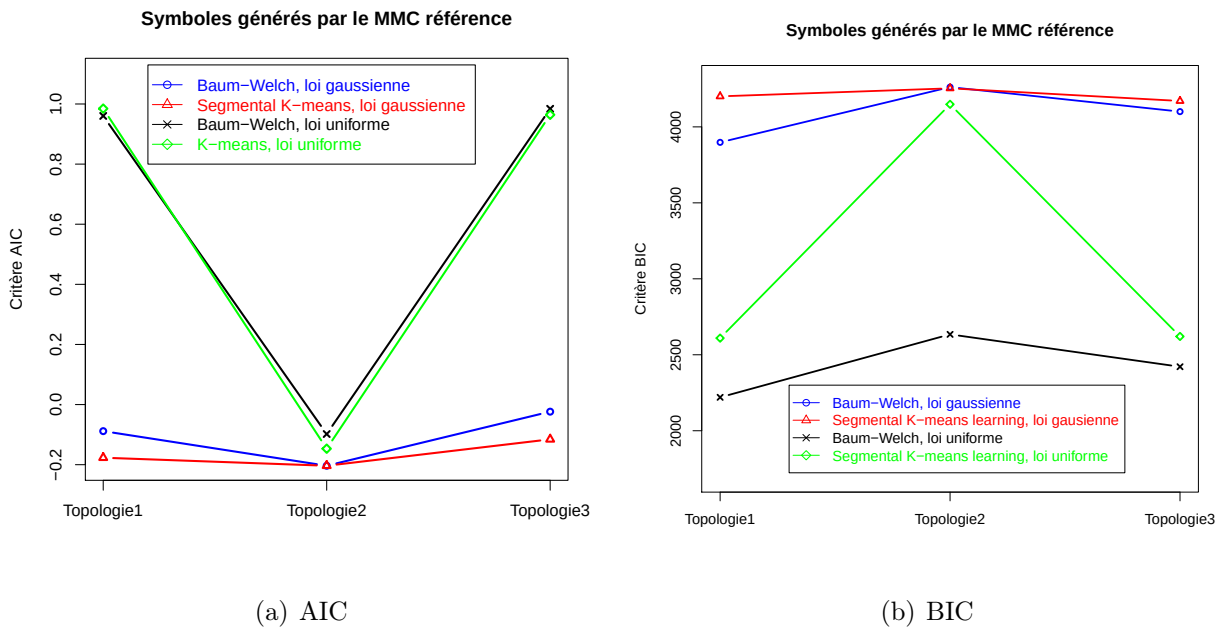


Fig 4.10 – Ces graphes nous montrent que la topologie 2 est la plus pertinente au vu du critère d'AIC et de BIC, pour les 2 algorithmes d'apprentissage et pour les 2 distributions étudiées.

Nous arrivons à la même conclusion que celle du critère de log-vraisemblance, nonobstant que pour le critère de log-vraisemblance et d'AIC, il s'agit d'un minimum et que pour le

critère de BIC, il s'agit d'un maximum (voir Figure 4.10(b), p. 100). En effet, les plus fortes valeurs du critère de BIC (malgré le terme de pénalité plus important) sont obtenues pour la topologie 2 (voir Tableau 4.5, p. 101).

Tableau 4.5 – Critère de BIC.

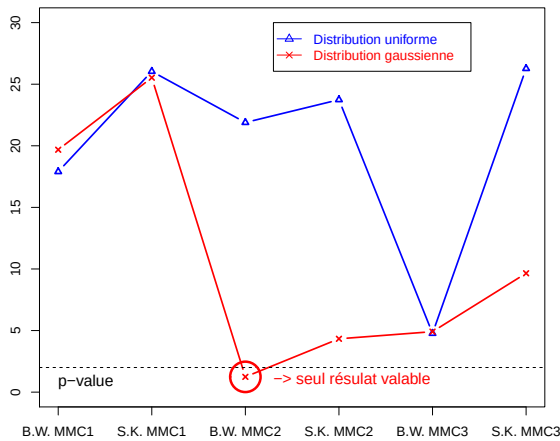
Topologie	Estimations - loi Uniforme	BIC
1	Baum-Welch / Variables Forward	2219.98
1	Segmental K-means / Viterbi	2609.65
2	Baum-Welch / Variables Forward	2634.29
2	Segmental K-means / Viterbi	4148.56
3	Baum-Welch / Variables Forward	2421.26
3	Segmental K-means / Viterbi	2620.63
Estimations - loi Normale		
1	Baum-Welch / Variables Forward	3898.24
1	Segmental K-means / Viterbi	4200.87
2	Baum-Welch / Variables Forward	4261.77
2	Segmental K-means / Viterbi	4253.73
3	Baum-Welch / Variables Forward	4100.55
3	Segmental K-means / Viterbi	4170.52

Tests statistiques

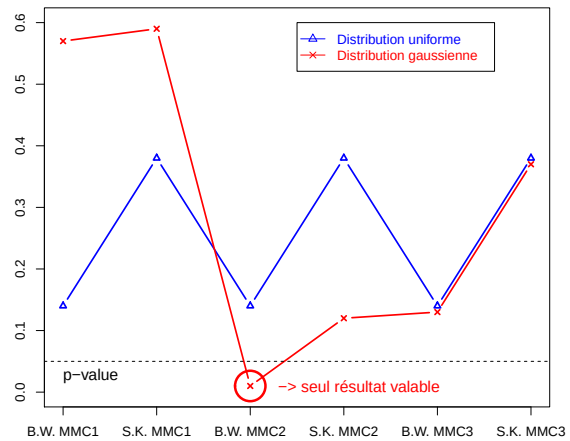
Nous avons ensuite appliqué différents tests statistiques sur les 3 topologies étudiées (présentées Figure 3.5, p. 71). Les tests de Kolmogorov-Smirnov et Aspin-Welch sont utilisés pour évaluer si deux distributions sont équivalentes. Notre but est ici de déterminer les meilleurs éléments de l'architecture de nos modèles. La Figure 4.11(b), p. 102 nous montre les résultats du test Kolmogorov-Smirnov. Ce test d'adéquation obtient la plus petite p-value pour les simulations suivantes :

- **topologie 2 ;**
- **distribution normale ;**
- **apprentissage Baum-Welch, décodé par variable Forward.**

Nous présentons Figure 4.11(a), p. 102 les résultats du test d'Aspin-Welch. Nous obtenons les mêmes conclusions que pour le test de Kolmogorov-Smirnov (Figure 4.11(b)). Les deux tests nous donnent ainsi les architectures les plus pertinentes de nos modèles. Les tests de Kolmogorov-Smirnov ainsi que celui d'Aspin-Welch déterminent si deux ensembles de données sont très différents. Un autre avantage de ce test est de ne pas faire d'hypothèses sur la distribution des données. Il est moins sensible que le test d'Aspin-Welch et il est conçu pour être utilisé sur des échantillons avec variances différentes, ce qui est le cas ici.



(a) Test statistique Aspin-Welch



(b) Test statistique Kolmogorov-Smirnov

Fig 4.11 – Nous testons les différentes architectures du modèle de synthèse à l'aide de tests statistiques d'adéquation : le test d'Aspin-Welch Figure 4.11(a) et celui de Kolmogorov-Smirnov Figure 4.11(b).

Incertitudes épistémiques

La plus petite incertitude épistémique est obtenue pour la **topologie 2** avec l'algorithme d'apprentissage **Baum-Welch décodé par variable Forward**, en utilisant la **distribution normale**, voir Figure 4.12, p. 103.

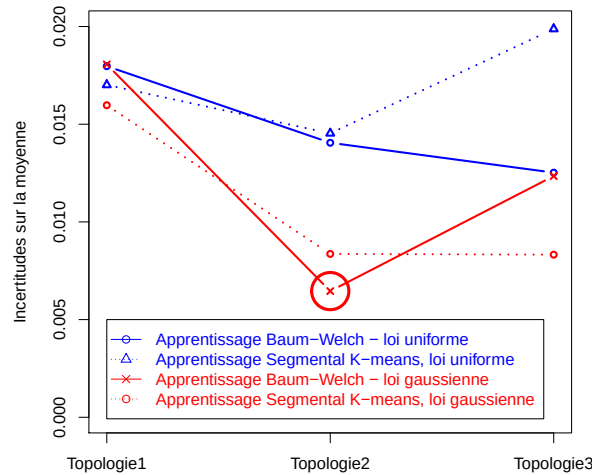


Fig 4.12 – Le calcul de l'incertitude sur la moyenne est représentatif de l'erreur épistémique dans la phase de modélisation.

Nous calculons les incertitudes moyennes (voir § 2.3.4, p. 35) sur les différentes topologies, différents algorithmes d'apprentissage et différentes distributions. Nous observons Figure 4.12, p. 103 que la plus faible incertitude est de $\pm 0.6\%$. Elle est obtenue sur la **topologie 2**. Nous pouvons conclure que cette topologie nous donne des résultats plus précis que pour les autres topologies 1 et 3 (Figures 3.5(a), p. 71 et 3.5(c), p. 71), en terme de conception de modèle. Concernant les algorithmes d'apprentissage des modèles, **Baum-Welch / décodage Variables Forward** nous donne les résultats ayant le plus faible taux d'erreur. Les résultats nous montrent aussi que la distribution **Normale** nous donne la plus faible incertitude. Le lecteur trouvera les résultats quantitatifs dans le Tableau 4.6, p. 104.

Discussion

Nous avons mesuré la pertinence des architectures des modèles étudiés, sans connaissance a priori. Nous avons utilisé une batterie de tests pour tenter d'évaluer les topologies, les algorithmes d'apprentissage et de décodage, ainsi que les distributions utilisées pour la modélisation. L'entropie de Shannon, le maximum de vraisemblance, AIC, BIC, les tests statistiques ainsi que l'incertitude épistémique nous indiquent que la **topologie 2** est la plus pertinente. Nous retrouvons bien la topologie ayant servi à simuler les données (modèle de référence basé sur la topologie 2).

En ce qui concerne les algorithmes d'apprentissage et de décodage, les résultats nous donnent **Baum-Welch décodé par Variables Forward** comme le plus pertinent. Seuls

Topologie	Estimations - loi uniforme	Incertitude (%)
1	Baum-Welch / Variables Forward	1,80%
1	Segmental K-means / Viterbi	1,70%
2	Baum-Welch / Variables Forward	1,40%
2	Segmental K-means / Viterbi	1,45%
3	Baum-Welch / Variables Forward	1,25%
3	Segmental K-means / Viterbi	1,99%
Estimations - loi normale		
1	Baum-Welch / Variables Forward	1,80%
1	Segmental K-means / Viterbi	1,60%
2	Baum-Welch / Variables Forward	0,65%
2	Segmental K-means / Viterbi	0,84%
3	Baum-Welch / Variables Forward	1,23%
3	Segmental K-means / Viterbi	0,83%

Tableau 4.6 – Résultats des incertitudes épistémiques.

les mesures de maximum de vraisemblance, BIC et AIC ne permettent pas de trancher.

La distribution **normale** apparait comme la plus pertinente avec toutes les méthodes utilisées. En accord avec le second principe (voir § 2.4.3, p. 42), nous nous attendions à trouver une meilleure entropie pour la distribution uniforme. Ce résultat est probablement dû aux valeurs extrêmes de la distribution normale comme le montre Payaro dans [127].

4.1.5 Résultats avec les autres topologies

Dans ce paragraphe, nous vérifions la concordance entre la topologie de référence utilisée pour le processus de synthèse et la topologie la plus pertinente. Nous suivons le même cheminement que pour la topologie 2 i.e. implémentation dans les trois MMC étudiés puis dans les deux algorithmes d'apprentissage.

Topologie 1

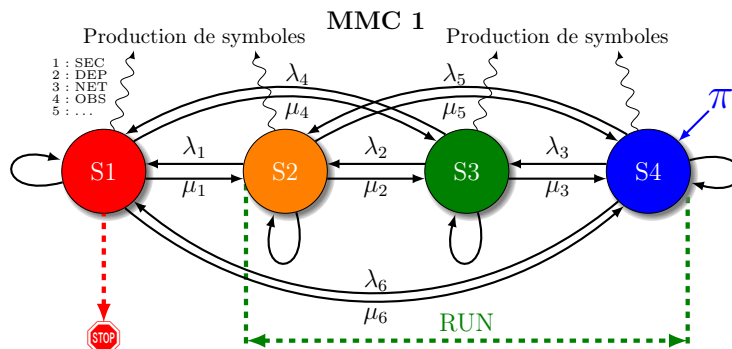


Fig 4.13 – Modèle de Markov Caché, topologie 1.

Nous utilisons dans un premier temps la topologie 1 (Figure 4.13, p. 104) pour la production des couples (*Symboles*, *Etats*) en utilisant la loi normale. Nous présentons Figure 4.14, p. 105, les résultats de l'entropie de Shannon. Ces résultats corroborent le fait que la topologie utilisée pour la modélisation se retrouve bien comme étant la plus pertinente.

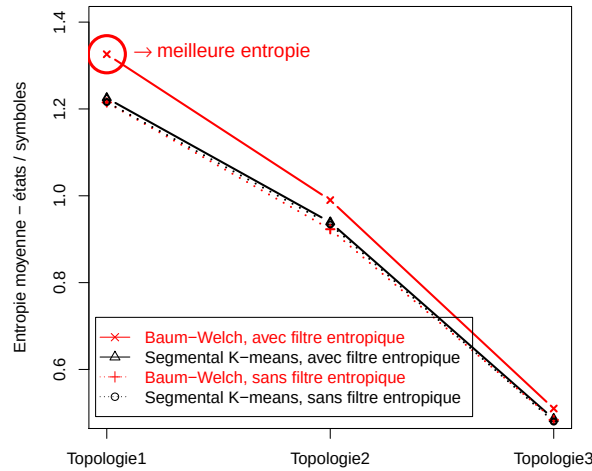


Fig 4.14 – Mesure de l'entropie de Shannon avec la topologie 1 comme référence.

Topologie 3

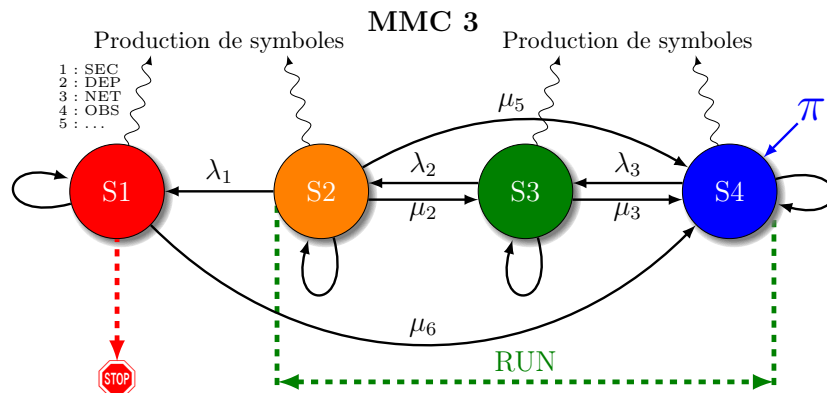


Fig 4.15 – Modèle de Markov Caché, topologie 3.

Dans un deuxième temps, nous utilisons la topologie 3 (Figure 4.15, p. 105) pour la production des couples (*Symboles*, *Etats*). Nous obtenons les mêmes conclusions que précédemment : la topologie 3 est la plus pertinente (voir Figure 4.16, p. 106).

Conclusion

Pour toutes les topologies étudiées, nous retrouvons bien la topologie de référence comme la plus pertinente lorsqu'elle est utilisée dans le modèle de synthèse. Nous pouvons donc en conclure que nos méthodes de mesures de pertinence redonnent bien la topologie ayant servi à construire le modèle.

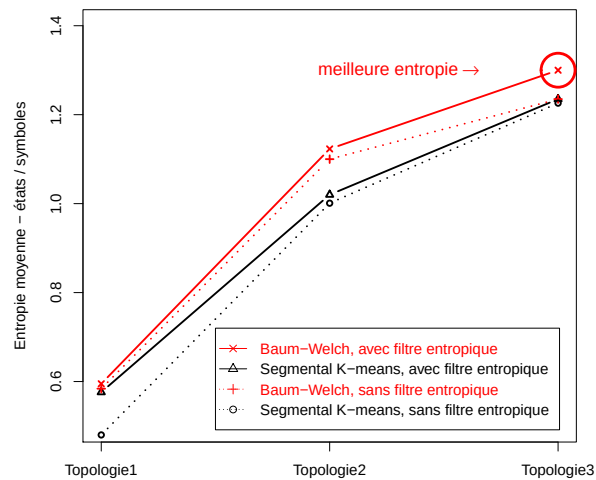


Fig 4.16 – Mesure de l'entropie de Shannon avec la topologie 3 comme référence.

4.2 Résultats des études réelles

4.2.1 Présentation des environnements d'étude

Les données récoltées pour cette étude sont issues d'une peseuse volumétrique sur une ligne de production dans l'agroalimentaire. Les autres données industrielles sont issues d'une presse basse pression d'un process industriel pour l'automobile. Tous les processus des usines respectivement étudiées, sont liés entre eux de manière séquentielle. L'arrêt d'un processus engendre donc l'arrêt des éléments situés en aval. Ce processus de production est maintenu par la mise en place d'une politique de maintenance préventive. Pour ce faire, les agents de maintenance doivent consigner leurs actions ou observations dans une base de données centralisée (voir un exemple dans le Tableau 4.8). Un échantillonnage toutes les 6 heures a été choisi conformément à la politique de maintenance interne. Si aucune donnée n'est saisie, nous insérons dans cette base de données le symbole RAS (= tout va bien). Cet échantillonnage temporel régulier permet de positionner notre étude dans le champ d'application des processus markoviens. Nous utilisons ensuite ces séquences de symboles (voir Tableau 4.7) pour modéliser le niveau de dégradation du processus. Nous modélisons cette « signature » à l'aide de MMC. Les données récoltées sur 2 ans comprennent environ 2000 évènements (voir Tableau 4.8).

VEP	VEP	TEP	TEP	SEC	TEP	TEP	DEP	AU	jour/heure/...
-----	-----	-----	-----	-----	-----	-----	-----	----	----------------

Tableau 4.7 – Séquence d'un message issue des données de maintenance.

Noms	Equipe	Date	Machine	Opération	Cd	Ti	N°	Code
Dupond	AM	11/01/2007	Peseuse	Huilage	VEP	20	1	9
Dupond	AM	11/01/2007	Peseuse	Huilage	VEP	20	2	9
Dupond	N	12/01/2007	Peseuse	Huilage	TEP	30	3	5
Dupond	N	12/01/2007	Peseuse	Huilage	TEP	30	4	5
Dupond	M	13/01/2007	Peseuse	Cadenas	SEC	10	5	6
Dupond	VSD	13/01/2007	Peseuse	Cadenas	TEP	30	6	5
Dupond	VSD	13/01/2007	Peseuse	Cadenas	TEP	30	7	5
Dupond	M	16/01/2007	Peseuse	Huilage	DEP	90	8	1
Dupond	AM	19/01/2007	Peseuse	Cadenas	AU	10	9	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tableau 4.8 – Exemple de consignation des évènements.

4.2.2 Modélisation des processus industriels

Dans le cadre d'activités de maintenance, Vrignat et al. [178] modélisent ces dysfonctionnements à l'aide de MMC. Nous rappelons dans le Tableau 4.9, p. 108, la signification des symboles choisis compte tenu des observations.

Ces symboles définissent les actions de maintenance menées sur le processus. Par exemple, le symbole DEP correspond à un arrêt de la production. C'est un état critique qu'il faut

minimiser. L'étude de Vrignat [178] considère deux modèles différents avec deux corpus d'apprentissage, l'un sur l'année 2005 et l'autre sur les deux années 2005–2006. Les symboles RAS sont insérés pour avoir un échantillonnage à la journée ou toutes les 6 heures. Pour la suite nous adopterons les dénominations suivantes :

- 05M1 | 1j : Corpus d'apprentissage 2005 modèle 1 (topologie 1) / 1 donnée par jour ;
- 0506M1 | 1j : Corpus 2005–2006 modèle 1 (topologie 1) / 1 donnée par jour ;
- 05M2 | 1j : Corpus 2005 modèle 2 (topologie 2) / 1 donnée par jour ;
- 0506M2 | 1j : Corpus 2005–2006 modèle 2 (topologie 2) / 1 donnée par jour ;
- 05M1 | 6h : Corpus 2005 modèle 1 (topologie 1) / 1 donnée toutes les 6 heures ;
- 0506M1 | 6h : Corpus 2005–2006 modèle 1 (topologie 1) / 1 donnée toutes les 6 heures ;
- 05M2 | 6h : Corpus 2005 modèle 2 (topologie 2) / 1 donnée toutes les 6 heures ;
- 0506M2 | 6h : Corpus 2005–2006 modèle 2 (topologie 2) / 1 donnée toutes les 6 heures.

Etat du processus		
MARCHE		
ARRET		
N°Obs.	Symboles	Nature des interventions
1	DEP	(Dépannage / arrêt de la production)
2	RM	(Réglage Machine)
3	AU	(Autre)
4	OBS	(Observation)
5	TEP	(Travaux Entretien Préventif)
6	SEC	(Sécurité)
7	RAN	(Remise A Niveau / planifié)
8	NET	(Nettoyage Machine)
9	VEP	(Visite Entretien Préventif)
10	RAS	(pas d'intervention)

Tableau 4.9 – Codification symbolique des interventions de maintenance.

La topologie 3 n'a pas été testée car elle n'apportait rien par rapport à la topologie 2. En effet, la seule différence avec la topologie 2 concerne la transition entre S1 et S2. L'état S1 devient alors un état absorbant. Nous considérons empiriquement, que le modèle « redémarre » dans tous les cas en S4 et non en S2. Cette transition n'a de sens qu'en termes d'apprentissage.

4.2.3 Description des MMC utilisés

Les modèles se présentent sous la forme d'un automate stochastique. Les états représentent les niveaux de dégradation du processus. Les symboles de l'automate représentent les observations du processus (nous redonnons pour mémoire, les différentes topologies étudiées Figure 4.17, p. 109).

Le modèle donne la probabilité d'être dans un des quatre états S1, S2, S3 ou S4, en fonction des symboles (noté « Cd » dans le Tableau 4.10, p. 109), selon les hypothèses de Markov pour les modèles d'ordre 1 vues au paragraphe 3.4.2, p. 68. Les probabilités des

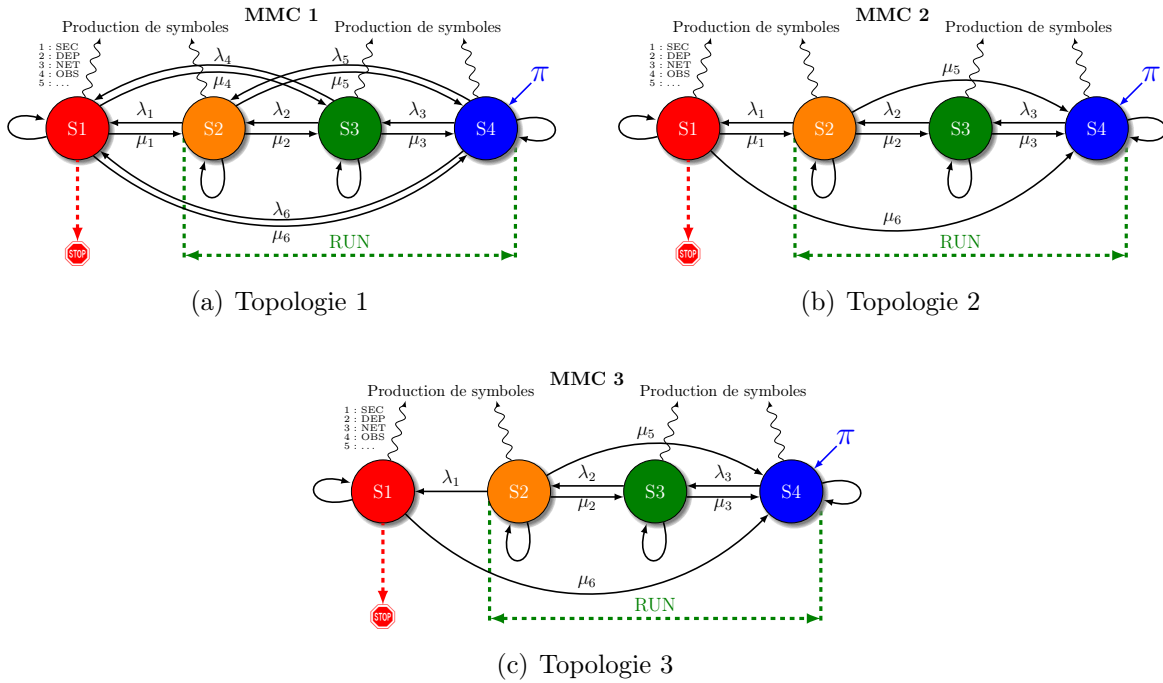


Fig 4.17 – Modèles de Markov Cachés, topologies à 4 états

quatre états sont données par la variable Forward et les différents niveaux par l'algorithme de Viterbi. Nous utilisons les données issues de l'étude de Vrignat [178]. Nous obtenons ainsi les niveaux de dégradation probables du processus étudié, sur une échelle de 1 à 4.

N°	DATE	Cd	Symb.	S1	S2	S3	S4	Niveau
1	09/01/2007	RAS	10	0,0%	0,0%	0,0%	100,0%	4
2	10/01/2007	RAS	10	0,0%	0,0%	0,1%	99,9%	4
3	11/01/2007	VEP	9	0,0%	0,0%	99,9%	0,1%	3
4	11/01/2007	VEP	9	0,0%	66,3%	33,7%	0,0%	2
5	12/01/2007	TEP	5	0,0%	11,7%	86,8%	1,5%	3
6	12/01/2007	TEP	5	0,0%	8,3%	90,7%	1,0%	3
7	13/01/2007	SEC	6	0,0%	49,5%	48,8%	1,7%	3
8	13/01/2007	TEP	5	0,0%	14,9%	81,5%	3,6%	3
9	13/01/2007	TEP	5	0,0%	8,4%	89,7%	1,9%	3
10	14/01/2007	RAS	10	0,0%	99,5%	0,0%	0,5%	2
11	15/01/2007	RAS	10	0,0%	98,8%	0,0%	1,2%	2
12	16/01/2007	DEP	1	100,0%	0,0%	0,0%	0,0%	1

Tableau 4.10 – Séquence de symboles / niveaux de dégradation.

Les séquences de symboles ainsi produites, nous donnent selon le modèle, une estimation du niveau de dégradation du processus (voir Figure 4.18, p. 110). Nous pouvons alors quantifier ces informations au moyen de différents critères et ainsi établir une évaluation des modèles, selon ces critères.

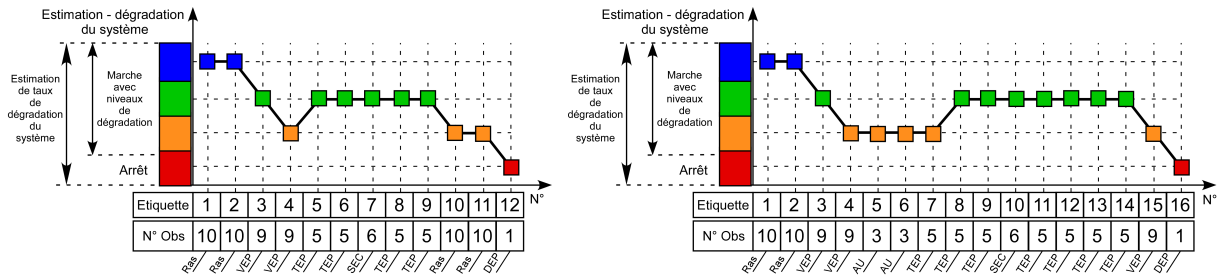


Fig 4.18 – Exemple de dégradation d'un processus [178].

4.2.4 Pertinence des observations empiriques

Dans cette section, nous voulons déterminer les symboles ou les séquences de symboles les plus pertinents, à partir de séquences d'observations empiriques. Nous utilisons pour cela les différents principes de l'entropie de Shannon (voir § 2.4.3, p. 41). Nous essayons ensuite d'étudier d'éventuels liens entre symboles. Nous utilisons l'« information mutuelle » vue au § 2.4.3, p. 45. Une « information mutuelle » nulle prouverait leur indépendance. Nous tentons également de déterminer si une suite de symboles peut engendrer un état significativement discernable i.e. un phénomène particulier (par exemple, un arrêt identifié par le symbole DEP) est précédé d'une série identifiable de symboles. Nous effectuons pour cela des calculs d'entropie sur des chaînes de Markov de différents ordres. Nous utilisons aussi des techniques couramment utilisées dans la littérature comme les arbres de décision ou les graphes de généralisation / spécialisation.

Maximum d'entropie

Les entropies de chaque symbole (chaîne de Markov d'ordre 1) sont calculées par rapport aux états S1, S2, S3 et S4 des modèles de Markov présentés au § 4.2.3, p. 108. Nous établissons ainsi les entropies pour chaque symbole, en fonctions de tous les états du modèle. Nous présentons dans le Tableau 4.11, p. 111, un exemple de résultat pour 9 séquences de dégradations issues du modèle de topologie 2 (MMC2 Figure 4.17(b), p. 109). Ce tableau nous montre que le symbole « RAS » obtient une entropie maximale de **1,418**. Ceci prouve que ce symbole est le plus utilisé dans tous les états du système. Nous remarquons aussi que le symbole « DEP » a une entropie nulle. Ce symbole est donc totalement discriminé i.e. l'état S1 correspond toujours à l'émission du symbole « DEP » et uniquement ce symbole. Comme pour le modèle de synthèse, nous utilisons la notion de filtre entropique vu au § 2.4.3, p. 42 pour éliminer les symboles ayant une entropie nulle et ceux ayant une entropie maximale. Nous éliminons ainsi les deux symboles « DEP » et « RAS » pour les calculs avec ce filtre. Le symbole « RAS » est utilisé notamment pour combler les champs non renseignés de la base de données. Il permet également d'avoir un échantillonnage régulier nécessaire à la théorie markovienne. Le symbole « DEP » est totalement discriminé, ces deux symboles n'apportent aucune autre information sur la pertinence des observations.

Une fois les symboles précédents éliminés du processus d'évaluation, nous pouvons établir un classement des symboles les plus pertinents par ordre d'entropie décroissante. Ainsi, si l'on classe les symboles du plus pertinent au moins pertinent, nous obtenons la séquence suivante :

– AU, RAN, SEC, VEP, TEP, RM, OBS.

Cette séquence est obtenue en utilisant l'algorithme d'élimination successive des symboles (voir §A.5, p. 146).

Symboles	S1	S2	S3	S4	Total	Entropie
AU		4	3	1	8	1,097
DEP	9				9	0,000
OBS		3			3	0,184
RAN			2	1	3	0,817
RAS		31	5	6	42	1,418
RM		4			4	0,224
SEC		6		1	7	0,641
TEP		12			12	0,420
VEP		16			16	0,473

Tableau 4.11 – Mesure de l'entropie, données empiriques, évaluation modèle MMC 2 décodage Viterbi.

Codage symbolique		États				Total	Entropie
		S1	S2	S3	S4		
AU	AU				2	2	0,174
AU	DEP	2				2	0,482
OBS	VEP		1			1	0,176
RAN	RAN				2	2	0,174
RAN	RAS				1	1	0,105
RAS	AU				3	3	0,229
RAS	DEP	5				5	0,471
RAS	OBS		1			1	0,176
RAS	RAS		2	3	21	26	1,309
RAS	SEC		1		2	3	0,350
RAS	TEP				1	1	0,105
⋮	⋮						⋮
VEP	VEP		8	1		9	0,819

Tableau 4.12 – Entropie pour les chaînes de Markov d'ordre 2, données empiriques, évaluation modèle MMC 2 décodage Viterbi.

Codage symbolique			États				Total	Entropie
			S1	S2	S3	S4		
AU	AU	OBS				1	1	0,107
AU	AU	RAS				1	1	0,107
AU	DEP	RAS				2	2	0,176
RAS	DEP	SEC				1	1	0,107
RAS	OBS	VEP		1			1	0,176
RAS	RAS	AU				1	1	0,107
RAS	RAS	DEP	5				5	0,471
RAS	RAS	OBS		1			1	0,176
RAS	RAS	RAS		1	1	13	15	0,969
RAS	RAS	SEC				2	2	0,176
RAS	RAS	VEP				2	2	0,176
RAS	SEC	AU				1	1	0,107
⋮	⋮	⋮						⋮
VEP	VEP	DEP	1				1	0,352
VEP	VEP	RAS		1	1		2	0,475
VEP	VEP	RM			2		2	0,431
VEP	VEP	TEP			1		1	0,299
VEP	VEP	VEP		1	1		2	0,475

Tableau 4.13 – Entropie pour les chaînes de Markov d’ordre 3, données empiriques, évaluation modèle MMC 2 décodage Viterbi.

Codage symbolique				États				Total	Entropie
				S1	S2	S3	S4		
AU	AU	OBS	OBS				1	1	0,108
AU	OBS	OBS	RAS				1	1	0,108
AU	RAS	RAS	RAS				1	1	0,108
AU	RAS	VEP	VEP		1			1	0,176
AU	TEP	TEP	DEP	1				1	0,352
DEP	AU	AU	OBS				1	1	0,108
DEP	RAN	RAN	RAN				1	1	0,108
DEP	RAS	AU	AU				1	1	0,108
DEP	RAS	AU	RAS				1	1	0,108
DEP	RAS	RAS	AU				1	1	0,108
DEP	RAS	RAS	RAS				2	2	0,178
DEP	SEC	RAS	RAS				1	1	0,108
OBS	OBS	RAS	TEP				1	1	0,108
OBS	RAS	TEP	TEP		1			1	0,176
OBS	VEP	VEP	DEP	1				1	0,352
RAS	RAS	DEP	RAS				2	2	0,178
RAS	RAS	DEP	SEC				1	1	0,108
RAS	RAS	OBS	VEP		1			1	0,176
RAS	RAS	RAS	DEP	2				2	0,482
RAS	RAS	RAS	OBS		1			1	0,176
RAS	RAS	RAS	RAS			1	8	9	0,711
RAS	RAS	RAS	SEC				2	2	0,178
RAS	RAS	RAS	VEP				1	1	0,108
⋮	⋮	⋮	⋮						⋮
VEP	VEP	TEP	TEP		1			1	0,176
VEP	VEP	VEP	RAS			1		1	0,299
VEP	VEP	VEP	VEP		1			1	0,176

Tableau 4.14 – Entropie pour les chaînes de Markov d’ordre 4, données empiriques, évaluation modèle MMC 2 décodage Viterbi.

Entropie d'une chaîne de Markov d'ordre 2, 3 et 4

Nous tentons ensuite de calculer de nouveau les entropies de toutes les chaînes de Markov possibles d'ordre > 1 . Nous utilisons les mêmes données que pour le calcul précédent pour l'ordre 1. Les extraits des résultats sont donnés dans les Tableaux 4.12, p. 111 ; 4.13, p. 112 et 4.14, p. 112.

Les résultats nous donnent les séquences de symboles donnant le plus d'information selon le principe du maximum d'entropie. Soit une entropie de **1,309** pour la séquence « **RAS - RAS** » d'ordre 2, **0,969** pour la séquence « **RAS - RAS - RAS** » à l'ordre 3 et **0,711** pour la séquence « **RAS - RAS - RAS - RAS** », d'ordre 4.

Nous avons ensuite recommencé les calculs d'entropie en éliminant les séquences précédentes pour chaque ordre (avec le filtre entropique). Le calcul est effectué indépendamment du passage d'un état à un autre. Malheureusement, ces différents ordres n'ont pas donné de résultats significatifs. Il y a trop de valeurs identiques pour en tirer des conclusions intéressantes.

Néanmoins, si nous nous concentrons sur les observations précédentes, une situation de dépannage (i.e. des séquences de 4 symboles se terminant par le symbole **DEP**), nous obtenons des résultats très intéressants pour le cas d'une chaîne de Markov d'ordre 4. En effet, si nous calculons de nouveau l'entropie après passage dans le filtre entropique, nous obtenons des entropies maximales pour les séquences particulières de symboles. Ces séquences sont illustrées dans le Tableau 4.15, p. 113. Nous observons des entropies maximales pour certaines séquences récurrentes de 4 symboles se terminant par le symbole **DEP**. Un extrait des résultats du 4^{ème} ordre est donné dans le Tableau 4.14, p 112.

Séquences de symboles				Entropie
AU	TEP	TEP	DEP	0,352
OBS	VEP	VEP	DEP	0,352
VEP	VEP	AU	DEP	0,352
Tous les autres 4-uplets				$\leq 0,178$

Tableau 4.15 – Séquences de symboles (ordre 4) ayant une entropie maximale, après passage dans le filtre entropique.

Nous observons que le symbole **DEP** apparaît uniquement dans toutes les séquences où l'entropie est maximale. Nous pouvons en conclure que ces signatures particulières de 3 symboles, induisent un état d'arrêt de production (**DEP** = dépannage), soit un état **S1** de l'automate considéré. Certaines séquences pourraient nous donner une indication de l'ordre des opérations à ne pas faire pour ne pas être dans une situation critique. Une séquence « **AU - TEP - TEP** », « **OBS - VEP - VEP** » ou « **VEP - VEP - AU** », peut nous amener à cette situation critique. D'après l'hypothèse de Markov présentée au paragraphe §3.4.2 p. 68, la connaissance des 3 états précédents peut en effet nous renseigner sur l'état suivant. En effet, plusieurs actions préventives (« **TEP, VEP** ») indiquent que le service maintenance avait pressenti la panne imminente.

Information mutuelle sur les données industrielles

Nous avons réalisé des calculs d'indépendance entre les différentes variables étudiées (symboles et états). Nous utilisons l'information mutuelle vue au § 2.4.3, p. 45. Le Tableau 4.16, p. 114 présente ce test d'indépendance entre chaque symbole. Les symboles indépendants ont une information mutuelle nulle. On trouve ainsi que les symboles **RM**, **OBS**, **TEP** et **VEP** sont indépendants les uns aux autres. Pour les autres valeurs différentes de 0, nous pouvons conclure d'une certaine dépendance entre les symboles. Par exemple, le symbole **AU** est en partie dépendant des symboles **RAS**, **TEP** et **VEP**, avec une valeur $> 0,5$. Les valeurs « NC » sont non calculables (division par zéro).

Symboles	AU	DEP	OBS	RAN	RAS	RM	SEC	TEP	VEP
AU		NC	0,221	NC	0,646	0,277	0,425	0,542	0,614
DEP	NC		NC	NC	NC	NC	NC	NC	NC
OBS	0,221	NC		NC	0,025	0,000	0,063	0,000	0,000
RAN	NC	NC	NC		NC	NC	NC	NC	NC
RAS	0,646	NC	0,025	NC		0,033	0,193	0,087	0,109
RM	0,277	NC	0,000	NC	0,033		0,077	0,000	0,000
SEC	0,425	NC	0,063	NC	0,193	0,077		0,137	0,151
TEP	0,542	NC	0,000	NC	0,087	0,000	0,137		0,000
VEP	0,614	NC	0,000	NC	0,109	0,000	0,151	0,000	

Tableau 4.16 – Information mutuelle pour les chaînes de Markov d'ordre 1, données empiriques, évaluation modèle MMC 2 décodage Viterbi.

Nous avons ensuite effectué des tests de dépendance entre états et symboles. Des extraits des résultats sont présentés Tableau 4.17, p. 115. Ce tableau ne donne aucun résultat exploitable.

Pour terminer, nous avons tenté de trouver d'autres inter-dépendances entre des séquences de symboles. Nous avons réalisé les mêmes calculs que précédemment en utilisant les chaînes de Markov d'ordre 2 à 4. Nous n'avons pas présenté ces résultats car le nombre de données des tableaux de contingence est assez conséquent : 6400 pour l'ordre 2, 18000 pour l'ordre 3, 33200 pour l'ordre 4.

Malheureusement, nous n'avons trouvé aucune conclusion pertinente pour ces informations mutuelles calculées à partir des chaînes de Markov d'ordre > 1 . Aucune dépendance ou indépendance entre les 2-uplets, 3-uplets et 4-uplets n'ont pu être mises en évidence.

Arbre de décision et espace de versions

Dans ce paragraphe, nous mettons en œuvre les deux approches d'apprentissage vues au § 2.4.4, p. 48 : l'**arbre de décision** associé à la méthode ID3 (voir l'algorithme en annexe A.1, p. 141) et la **généralisation** avec l'« espace de versions » :

- l'arbre de décision sur des chaînes de Markov de différents ordres est construit en utilisant les propriétés de l'entropie. Il permet d'identifier le symbole ou la suite de symboles comme étant déclencheur d'un état particulier du système (où l'entropie est minimale) ;

	États					États			
05M1-1J	S1	S2	S3	S4	05M1-6H	S1	S2	S3	S4
AU		0,019	0,065	0,012	AU				0,144
DEP	0,424				DEP	0,233			
OBS		0,018			OBS		0,001	0,010	
RAN			0,083	0,036	RAS		0,339	0,747	0,018
RAS		0,174	0,038	0,062	RM			0,033	0,031
RM		0,024			SEC		0,006		0,014
SEC		0,035		0,009	TEP		0,001	0,117	
TEP		0,072			VEP		0,014		0,037
VEP		0,097							
0506M1-1J	S1	S2	S3	S4	0506M1-6H	S1	S2	S3	S4
AU		0,023	0,063	0,010	AU		0,011	0,004	0,002
DEP	0,424				DEP	0,233			
OBS		0,004	0,022	0,022	OBS		0,002	0,010	
RAN			0,032	0,076	RAS		0,344		0,040
RAS		0,203	0,035	0,048	RM		0,003	0,021	
RM		0,011	0,017	0,016	SEC		0,004		0,027
SEC		0,041		0,007	TEP		0,001	0,117	
TEP		0,085			VEP		0,006	0,111	
VEP		0,113							
05M2-1J	S1	S2	S3	S4	05M2-6H	S1	S2	S3	S4
AU				0,095	AU		0,005	0,006	0,006
DEP	0,424				DEP	0,233			
OBS		0,010		0,020	OBS		0,002		
RAN				0,036	RAS		0,159	0,009	0,009
RAS		0,011	0,023	0,347	RM		0,003		
RM		0,018	0,108		SEC		0,005		
SEC		0,020	0,011	0,038	TEP		0,009		
TEP		0,109	0,058	0,013	VEP		0,012		
VEP		0,223	0,005	0,014					
0506M2-1J	S1	S2	S3	S4	0506M2-6H	S1	S2	S3	S4
AU				0,014	AU			0,004	0,006
DEP	0,424				DEP	0,233			
OBS				0,005	OBS			0,001	
RAN				0,005	RAS			0,109	0,007
RAS				0,073	RM			0,002	
RM				0,007	SEC			0,003	
SEC				0,012	TEP			0,006	
TEP				0,021	VEP			0,008	
VEP				0,028					

Tableau 4.17 – Information mutuelle pour les chaînes de Markov d'ordre 1 - Modèle MMC 2

- l'espace de versions ou graphe de généralisation/spécialisation pour des chaînes de Markov, permet de classer, sous forme de tableau, les symboles selon un ordre de pertinence. Le graphe est construit en éliminant successivement les éléments les plus pertinents (où l'entropie est minimale). Cela nous donne ainsi un aperçu global des symboles en fonction des états ou des états en fonction des symboles.

Arbre de décision pour des chaînes de Markov de différents ordres

La construction d'un arbre de décision permet de corriger le sur-apprentissage en effectuant un « élagage », a posteriori, des éléments inutiles. Il faut construire l'arbre entier puis supprimer les nœuds inutiles. C'est aussi la méthode d'apprentissage la plus utilisée. Des centaines d'applications couvrant des domaines comme le diagnostic ou encore le contrôle de processus utilisent les arbres de décision. Quelques propriétés de ce type d'apprentissage :

- les instances doivent être représentées par des couples (*attributs, valeurs*) ;
- les valeurs sont discrètes ;

– les données pour l'apprentissage (instances) peuvent contenir des erreurs.
En tout état de cause, cette méthode d'apprentissage est surtout utilisée en classification.

Nous avons réalisé des arbres de décision à partir des différents corpus d'apprentissage. Nous utilisons les 3 modèles étudiés, pour des chaînes de Markov d'ordre 1 à 4. Au delà de l'ordre 4, les données sont diluées et les résultats ne présentent que peu d'intérêt. En effet, la probabilité de retrouver plusieurs fois une même série de 5 symboles sur l'ensemble des données est très faible. Nous détaillons les calculs jusqu'à l'ordre 3 (3-uplets) dans les tableaux 4.18, p. 116, pour le corpus d'apprentissage 2005–2006 / modèle MMC 1 et dans le Tableau 4.19, p. 117, pour le corpus d'apprentissage 2005–2006 / modèle MMC 2. Nous voyons que plus l'ordre augmente, plus le nombre de n-uplets dont l'entropie est minimale augmente. De plus, sur le modèle MMC 2, les entropies minimales montrent que l'état S2 est le plus discriminant pour l'ensemble de ces n-uplets. Contrairement au modèle 1, où l'état S3 est le plus discriminant.

Corpus 2005-2006 Modèle 1	Entropie min	États Entropie min	Entropie max	États Entropie max
Symboles 1-uplet (ordre 1)	DEP	S1	VEP	S2,S3
Symboles 2-uplet (ordre 2)	OBS RAS OBS VEP RAS OBS	S2 S2 S2	RAS RAS	S2,S4
Symboles 3-uplet (ordre 3)	AU AU RAS AU OBS RAS OBS RAS RAS RAS AU AU RAS OBS VEP RAS RAS OBS RAS RM RAS RAS SEC RAS RM RM RAS SEC RAS AU SEC RAS RAS SEC SEC RAS	S2 S2 S2 S2 S2 S2 S2 S2 S2 S2 S2 S2	RAS RAS RAS	S2,S4

Tableau 4.18 – Exemple d'arbre de décision pour une chaîne de Markov d'ordre 1, 2 et 3 / Corpus 2005–2006 Modèle MMC 1.

La construction de tels arbres avec le critère d'entropie maximale, ne donne pas de résultats pertinents (voir dernière colonne des Tableaux 4.18, p. 116 et 4.19, p. 117).

Espace de versions ou graphe de généralisation/spécialisation pour des chaînes de Markov d'ordre 1

Le graphe est construit de manière incrémentale à partir d'un espace d'hypothèses i.e. les niveaux de dégradations du processus et d'un ensemble d'instances i.e. l'ensemble des symboles. Nous donnons Tableau 4.20, p. 118, un exemple de construction d'un graphe de généralisation/spécialisation. Malheureusement, nous ne pouvons en tirer aucune conclusion pertinente sur une « classification » des symboles par ordre d'importance. Nous pouvons juste conclure sur les symboles **TEP** et **VEP**. En effet, ils obtiennent une entropie la plus élevée pour les états S2 et S3. Ils sont donc les symboles les plus représentatifs de ces états.

Le Tableau 4.21, p. 118 décrit la construction d'un graphe de généralisation/spécialisation par rapport aux états du système. Il permet de déterminer les états S1 à S4 les plus pertinents. Nous utilisons ici une chaîne de Markov d'ordre 1. Pour l'état S4, nous voyons que l'entropie

Corpus 2005-2006 Modèle 2	Entropie min	États	Entropie min	Entropie max	États	Entropie max
Symboles 1-uplet (ordre 1)	DEP		S1	AU		S3,S4
Symboles 2-uplet (ordre 2)	AU OBS		S3	DEP AU		S4
	OBS RAS		S3			
	OBS VEP		S3			
	RAS OBS		S3			
	RM RM		S3			
	RM SEC		S3			
	SEC TEP		S3			
	SEC VEP		S3			
	TEP VEP		S3			
	VEP RM		S3			
	VEP TEP		S3			
Symboles 3-uplet (ordre 3)	AU AU OBS		S3	RAS RAS RAS		S3
	AU AU RAS		S3			
	AU OBS RAS		S3			
	DEP AU AU		S3			
	OBS RAS RAS		S3			
	OBS VEP VEP		S3			
	RAS AU AU		S3			
	RAS OBS VEP		S3			
	RAS RAS OBS		S3			
	RAS RM RAS		S3			
	RAS RM SEC		S3			
	RAS SEC RAS		S3			
	RAS SEC SEC		S3			
	RAS SEC TEP		S3			
	RM RM RAS		S3			
	RM SEC SEC		S3			
	SEC RAS AU		S3			
	SEC RAS RAS		S3			
	SEC SEC RAS		S3			
	SEC SEC VEP		S3			
	SEC TEP TEP		S3			
	SEC VEP VEP		S3			
	TEP TEP VEP		S3			
	TEP VEP VEP		S3			
	VEP RM RM		S3			
	VEP TEP TEP		S3			
	VEP VEP RM		S3			
	VEP VEP TEP		S3			

Tableau 4.19 – Exemple d’arbre de décision pour une chaîne de Markov d’ordre 1, 2 et 3 / Corpus 2005–2006 Modèle MMC 2 .

n’est pas nulle (1,236) nous ne pouvons donc pas donner de conclusion sur cet état. Par ailleurs, nous avons des entropies élevées sur les états S2 et S3. Ces états sont donc représentés par un maximum de symbole (voir définition de l’entropie § 2.4.3, p. 41).

	États						
Codage symbolique	S1	S2	S3	S4	Total	Majoritaire	Entropie
AU		6	1	1	8	6	0,532
OBS		1	1		2	1	0,247
RAS		182		24	206	182	0,320
RM		2	2		4	2	0,389
SEC		3		3	6	3	0,436
TEP		2	8		10	8	0,597
VEP		5	9		14	9	0,656
Codage symbolique	S1	S2	S3	S4	Total	Majoritaire	Entropie
AU		6	1	1	8	6	0,540
RAS		182		24	206	182	0,314
RM		2	2		4	2	0,399
SEC		3		3	6	3	0,436
TEP		2	8		10	8	0,595
VEP		5	9		14	9	0,651
Codage symbolique	S1	S2	S3	S4	Total	Majoritaire	Entropie
AU		6	1	1	8	6	1,244
RM		2	2		4	2	0,684
SEC		3		3	6	3	0,742
TEP		2	8		10	8	0,881
VEP		5	9		14	9	1,032
Codage symbolique	S1	S2	S3	S4	Total	Majoritaire	Entropie
AU		6	1	1	8	6	1,262
SEC		3		3	6	3	0,764
TEP		2	8		10	8	0,895
VEP		5	9		14	9	1,024
Codage symbolique	S1	S2	S3	S4	Total	Majoritaire	Entropie
AU		6	1	1	8	6	0,746
TEP		2	8		10	8	0,935
VEP		5	9		14	9	1,030
Codage symbolique	S1	S2	S3	S4	Total	Majoritaire	Entropie
TEP		2	8		10	8	1,028
VEP		5	9		14	9	0,832

Tableau 4.20 – Exemple de construction d’un graphe de généralisation/spécialisation pour une chaîne de Markov d’ordre 1.

États	AU	DEP	OBS	RAS	RM	SEC	TEP	VEP	Total	Entropie
S1		9							9	0,000
S2	6		1	182	2	3	2	5	201	2,964
S3	1		1		2		8	9	21	2,042
S4	1			24		3			28	1,236
États	AU		OBS	RAS	RM	SEC	TEP	VEP	Total	Entropie
S2	6		1	182	2	3	2	5	201	2,964
S3	1		1		2		8	9	21	2,042
S4	1			24		3			28	1,236
États	AU		OBS	RAS	RM	SEC	TEP	VEP	Total	Entropie
S2	6		1	182	2	3	2	5	201	2,186
S3	1		1		2		8	9	21	2,068

Tableau 4.21 – Exemple de construction d’un graphe de généralisation/spécialisation sur les états (chaîne de Markov d’ordre 1).

Discussion

Nous avons appliqué l'entropie de Shannon sur les symboles isolés, sur les chaînes de Markov d'ordre 2, 3 et 4 (bigrammes, trigrammes, etc.). Nous avons trouvé que le symbole DEP, ayant une entropie nulle, était totalement discriminé. Dans le même ordre d'idée, le symbole RAS est représentatif de tous les états S2, S3 et S4 car il possède une entropie maximale. Nous avons ainsi décidé d'éliminer ces deux symboles pour la mesure de pertinence des autres symboles. Nous avons pu ainsi trouver une liste de symboles du plus pertinent au moins pertinent. Nous avons mis en évidence des séquences 3-uplet de symboles donnant suite au symbole DEP. Ces séquences sont donc à surveiller car elles ont une probabilité plus élevée de produire une situation de panne.

Les calculs de l'information mutuelle entre les symboles nous ont donné des résultats de dépendance entre symboles. Le symbole AU est en partie dépendant des symboles TEP et VEP. Les symboles RM, OBS, TEP et VEP sont indépendants entre eux.

Les calculs de graphes de généralisation/spécialisation ou les arbres de décision ne nous ont pas permis d'obtenir des conclusions exploitables sur la pertinence des observations.

4.2.5 Évolution de la modélisation

Découpage temporel

Nous étudions dans cette section, la pertinence de la périodicité des observations. Nous répondons à la question : quel est l'échantillonnage le plus pertinent pour les observations ? Nous étudions ici les deux découpages temporels proposés dans Vrignat [177] : un échantillonnage toutes les 6 heures ou un échantillonnage à la journée.

Mesure de l'entropie de Shannon

Sans connaissance a priori, nous calculons les entropies moyennes des symboles seuls (colonne « Ordre 1 » du Tableau 4.22, p. 120), pour chaque mode d'échantillonnage (1 jour et 6 heures). Nous mesurons ensuite l'entropie des chaînes de Markov d'ordre 2, 3 et 4. Rappelons que pour une chaîne de Markov d'ordre k : l'état suivant dépend des k états précédents. Nous voyons alors Tableau 4.22, p. 120, que les modèles avec un échantillonnage de 1 jour sont les plus pertinents (où l'entropie est maximale). Ce tableau montre qu'il y a moins de dispersion de l'information dans les modèles avec un échantillonnage à la journée, où l'entropie moyenne est maximale.

Discussion : un échantillonnage 6 heures devrait donner plus d'informations au système. L'entropie devrait alors être supérieure pour cette période de scrutation des observations. L'explication se trouve dans le remplissage de la base de données. Pour fixer un échantillonnage régulier toutes les 6 heures, nous avons rempli les champs sans observation par des champs « RAS » (Rien A Signaler). Ces informations n'apportent donc, selon Shannon, aucune information au système.

Dans un second temps, nous calculons l'entropie moyenne de chaque modèle. Selon le 2^{ème} principe de l'entropie maximale énoncé au § 2.4.3, p. 42, nous choisirons alors le modèle dont l'entropie moyenne est maximale. Nous calculons l'entropie moyenne du modèle afin d'en évaluer la pertinence des séquences d'observations. Le meilleur échantillonnage sera celui dont l'entropie moyenne est maximale après filtrage entropique.

Échantillonnage	Entropie moyenne			
	Ordre 1	Ordre 2	Ordre 3	Ordre 4
1 jour	0,292	0,112	0,084	0,072
6 heures	0,158	0,074	0,060	0,051

Tableau 4.22 – Entropie moyenne des modèles de Markov sur les 2 types d'échantillonnage.

Le Tableau 4.23, p. 120, présente les résultats des mesures d'entropies moyennes des modèles avant et après le filtre entropique. Nous effectuons un filtrage de niveau 1 i.e. nous éliminons un seul symbole dont l'entropie est maximale ainsi que toutes les entropies nulles au travers de ID3 [135]. Le modèle le plus pertinent apparaît alors : « 0506M1 | 1 jour » où l'entropie moyenne est maximale.

Échantillonnage	Modèles	Entropie moy sans filtre	Entropie moy avec filtre
1 jour	05M1	0,586	0,695
1 jour	0506M1	0,699	0,881
1 jour	05M2	0,698	0,857
1 jour	0506M2	0,268	0,365
6 heures	05M1	0,524	0,741
6 heures	0506M1	0,397	0,828
6 heures	05M2	0,264	0,393
6 heures	0506M2	0,200	0,394

Tableau 4.23 – Résultats de l'entropie moyenne avec et sans filtre entropique.

Sur la Figure 4.19 (a), p. 121, nous voyons que le filtre entropique permet de mettre en valeur un maximum pour le modèle 0506M1 sur la base de temps de 1 jour. La Figure 4.19 (b), p. 121, nous montre de manière beaucoup plus flagrante que le modèle 0506M1 sur la base de temps de 6 heures est le plus pertinent après application du filtre entropique. D'après [178], c'est bien le modèle 0506M1 qui fournit les meilleurs résultats de prédiction de pannes.

Résultats du maximum de vraisemblance

Nous utilisons ici une partie de la méthode de Bourguignon [26], qui permet de sélectionner des « modèles de Markov parcimonieux » en utilisant le principe du maximum de vraisemblance. Nous allons maximiser le logarithme de la fonction de vraisemblance i.e. pour notre cas discret, nous calculons le maximum de vraisemblance sur les probabilités de transition de chaque modèle : voir Tableau 4.24, p. 121. Nous observons un maximum pour le modèle 0506M1 | 6 heures (voir Figure 4.20, p. 122). Les résultats pour les modèles de base de temps 1 jour ne nous donnent pas des résultats exploitables. La variance trop faible indique que la dispersion des résultats n'est pas satisfaisante pour tirer des conclusions pertinentes.

Évolution de l'entropie pour les données empiriques

Les données industrielles, ont été collectées de 2005 à 2007. Les résultats de l'entropie de Shannon sont donnés Figure 4.21, p. 122. La valeur maximale est atteinte pour 180 symboles.

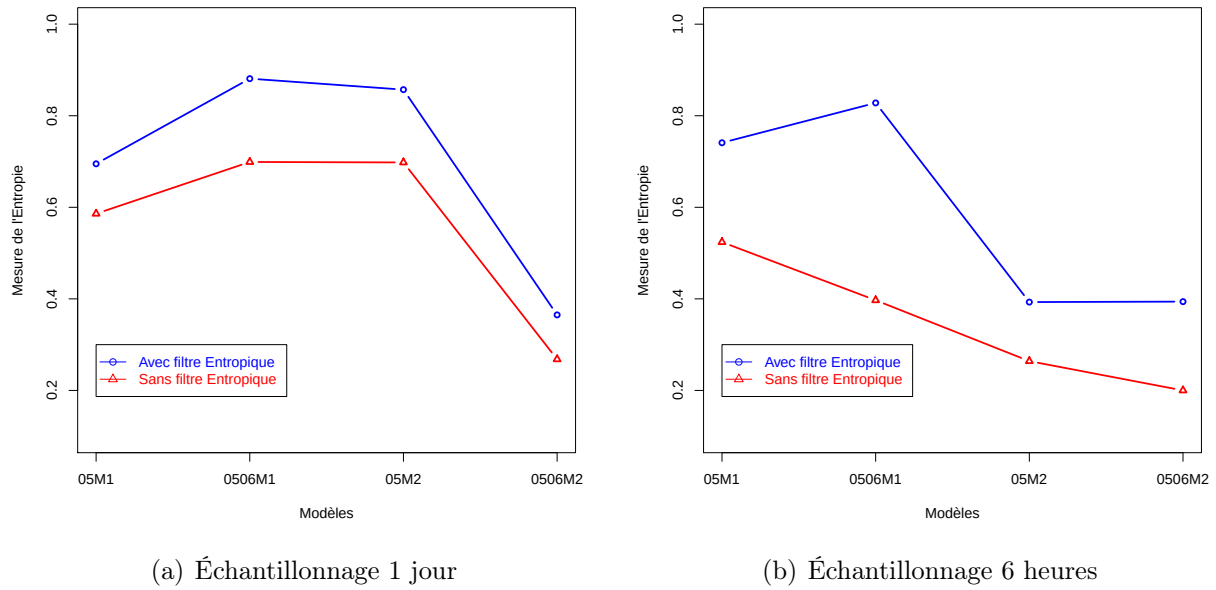


Fig 4.19 – Entropies moyennes des modèles

Échantillonnage	Modèles	Log-Vraisemblance
1 jour	05M1	930
1 jour	0506M1	820
1 jour	05M2	570
1 jour	0506M2	484
6 heures	05M1	2260
6 heures	0506M1	3626
6 heures	05M2	1446
6 heures	0506M2	1350

Tableau 4.24 – Résultats de vraisemblance.

Cette valeur maximale de l'entropie correspond à la valeur asymptotique pour un nombre élevé de symboles. L'écart entre les deux courbes est dû à une cardinalité différente des alphabets de symboles, entre les deux processus étudiés.

Essais avec une fenêtre glissante normée

Pour tenter de valider la fenêtre normée précédemment trouvée, nous avons utilisé des séquences de 200 symboles d'une GMAO industrielle. Nous avons ensuite réévalué ces symboles en utilisant l'algorithme **Baum-Welch** avec un décodage **variable Forward**. Nous étudions ainsi de nouveau l'entropie, en utilisant cette quantité minimale de symboles pour des données empiriques :

- les résultats sont donnés Figure 4.22, p. 123. La courbe en rouge représente l'entropie des données empiriques et celle en noire, l'entropie issue de l'apprentissage (présentée

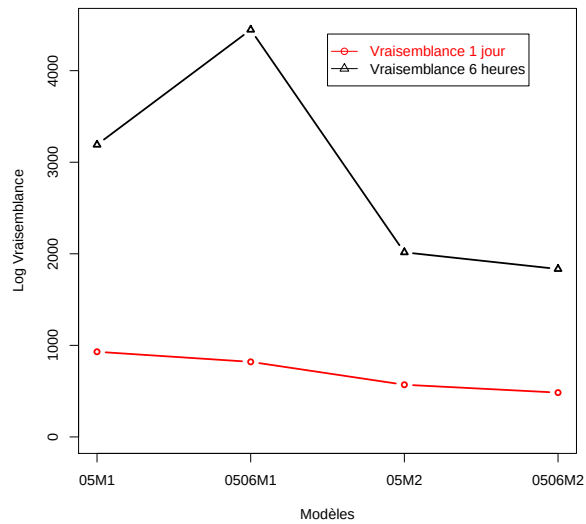


Fig 4.20 – Mesures de la vraisemblance.

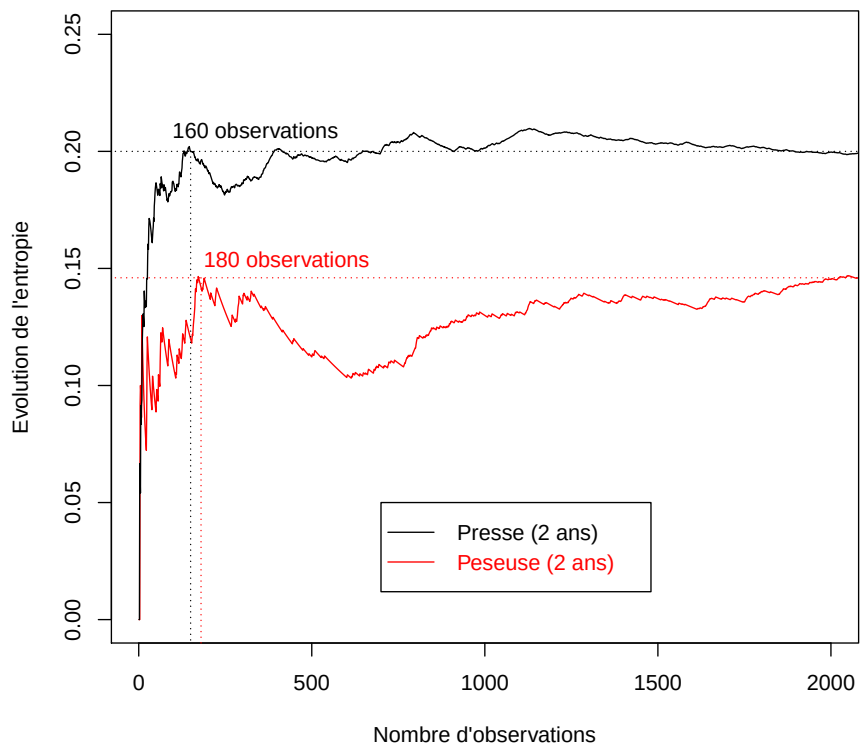


Fig 4.21 – Nombre minimal de données par entropie de Shannon, données issues d'une GMAO industrielle.

ci-dessus) en utilisant une fenêtre glissante normée de 200 symboles. Nous observons une phase de transition qui correspond à l'apprentissage de la première séquence des symboles. En effet, au fur et à mesure que la base de données s'enrichit de nouveaux symboles, l'entropie de la séquence augmente. Une fois cette phase terminée, l'entropie reste supérieure à l'entropie issue des données empiriques. De plus, elle ne se stabilise pas et oscille dans l'intervalle $[0, 12; 0, 27]$. Nous remarquons également une importante diminution de l'entropie entre 500 et 700 symboles. En effet, les séquences $[400; 600]$ comportent beaucoup de symboles **RAS** (186 sur 200) ce qui engendre un appauvrissement en symboles, de la séquence d'apprentissage.

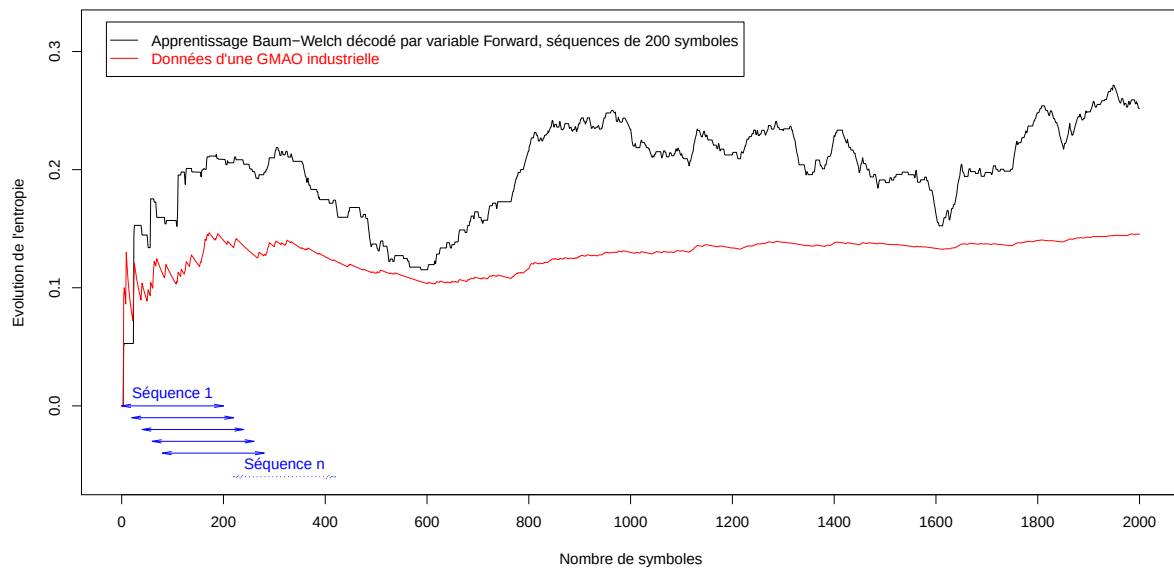


Fig 4.22 – Évolution de l'entropie évaluée avec une fenêtre glissante normée (200 symboles). Les données d'apprentissage sont issues d'une GMAO industrielle.

Conclusion

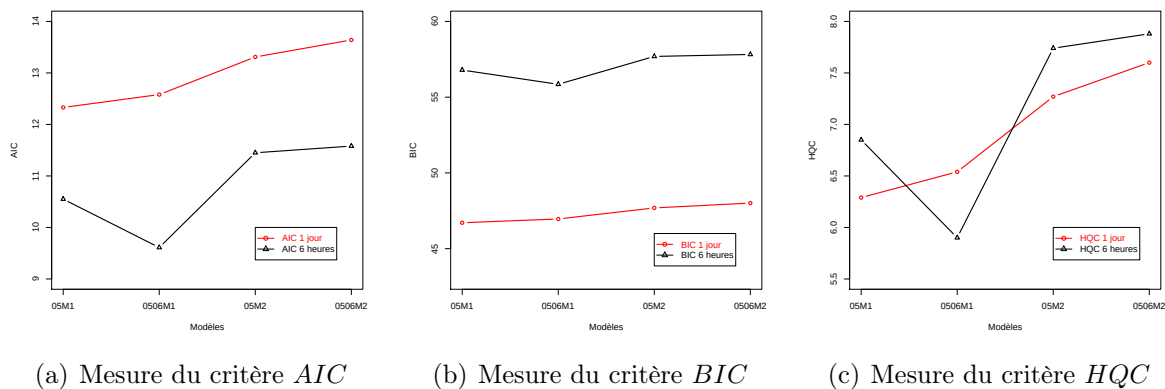
Le principe du maximum de vraisemblance nous donne comme modèle le plus pertinent : 0506M1 | 6 heures. Les résultats concernant 0506M1 | 1 jour ne sont pas éloquentes. Comme pour l'entropie, le principe du maximum de vraisemblance est intéressant pour la sélection de modèle dans la mesure où le nombre de données est suffisamment important.

Mesures des critères d'Akaike, de Bayes et d'Hannan-Quinn

Nous avons vu au § 2.4.3, p. 43, que chaque critère possède sa spécificité. Les critères *BIC* et *HQC* pénalisent plus les modèles ayant un grand nombre de données. Le modèle le plus pertinent étant celui qui obtient la valeur minimale. Nous obtenons les résultats du Tableau 4.25, p. 124, pour les modèles étudiés. Ce qui nous donne graphiquement la Figure 4.23, p. 124.

Échantillonnage	Modèles	AIC	BIC	HQC
1 jour	05M1	12,33	46,71	6,29
1 jour	0506M1	12,58	46,96	6,54
1 jour	05M2	13,31	47,69	7,27
1 jour	0506M2	13,64	48,01	7,60
6 heures	05M1	10,55	56,79	6,85
6 heures	0506M1	9,61	55,85	5,90
6 heures	05M2	11,45	57,69	7,74
6 heures	0506M2	11,58	57,82	7,88

Tableau 4.25 – Résultats des différents critères.


Fig 4.23 – Critères AIC , BIC et HQC .

Nous voyons que les 3 critères nous amènent aux mêmes conclusions que précédemment pour le modèle 0506M1 | 6 heures. C'est donc le plus pertinent au sens des critères AIC , BIC et HQC . Les résultats de [178] montrent que ce modèle fonctionne. Par contre, il n'a pas été retenu car il est trop sensible dans la détection des pannes. En effet, le passage d'un niveau de dégradation à un autre se fait avec plusieurs rebonds.

Algorithme de calcul de l'entropie moyenne d'un modèle de Markov avec filtre entropique

Description de l'algorithme (voir Annexe A.5, p. 146) : nous calculons les valeurs d'entropie de chaque symbole du modèle de Markov en éliminant les valeurs nulles à chaque itération. Le symbole ayant une entropie nulle est totalement discriminé par rapport à l'état qui lui correspond. Nous éliminons ensuite les symboles dont l'entropie est maximale. En effet, ces symboles correspondent en général à des données trop présentes, comme par exemple dans notre cas d'étude, le symbole « RAS ». Trop nombreux, ces symboles peuvent empêcher une bonne estimation de la pertinence du modèle. Une fois le modèle passé au travers de ce filtre entropique, nous calculons son entropie moyenne. La valeur maximale nous donne alors le modèle le plus pertinent.

4.2.6 Confrontation du modèle de synthèse avec l'application réelle

Nous comparons les données issues du modèle synthétique avec celles provenant d'une GMAO industrielle. La codification symbolique choisie étant identique dans les deux cas (MMC de synthèse et MMC situation réelle), nous avons une cohérence au niveau des observations du processus (voir Tableau 4.9, p. 108). Les données du modèle de synthèse sont du type *Symb_U* pour la loi uniforme et *Symb_N* pour la loi normale.

Dans l'étude du cas réel (voir contexte § 4.2.1, p. 107), nous avons trouvé une fenêtre de 180 symboles afin d'obtenir un maximum d'entropie. Comme nous pouvons le voir sur les figures 4.2 et 4.21, la mesure de l'entropie (informations issues du contexte réel) suit le même comportement que l'entropie « synthétique ». Par ailleurs, toutes les valeurs du maximum d'entropie trouvées précédemment correspondent aux valeurs asymptotiques des évolutions des entropies respectives. Ainsi, ces valeurs d'entropie maximales nous donnent une idée sur le nombre minimal d'observations pour estimer au mieux, les modèles. Ce nombre minimal de symboles nous donne des éléments pour choisir la taille minimale de la fenêtre glissante permettant d'optimiser la conceptualisation des modèles de maintenance préventive.

4.2.7 Conclusion

Nous avons étudié la pertinence des différents découpages temporels. De la même façon, l'entropie a été évaluée selon les états des modèles afin de vérifier que le plus pertinent obtient également un bon score « entropique ». Nous illustrons ainsi que sans connaissance a priori, le modèle 0506M1 avec un découpage temporel de 6 heures est le plus pertinent. Les autres méthodes d'évaluation utilisant le principe de maximum de vraisemblance présentées ici, identifient aussi le modèle 0506M1 comme le plus pertinent. Par contre, l'échantillonnage de 6 heures est préféré à l'échantillonnage à la journée. Vignat et al. [180] montrent que le modèle 0506M1 | 6 heures fonctionne mais il est trop réactif (par rapport aux contraintes fixées dans [180]) et a donc une forte propension dans la prédiction de fausses pannes. Un sous-échantillonnage apporte une certaine stabilité dans les changements d'états et diminue donc le risque de converger trop vite dans un état S1. Ceci engendre alors des « rebonds » lors du passage d'un état à un autre. La méthode par la mesure de l'entropie réagit comme un filtre anti-rebond et donne comme le plus pertinent le modèle préconisé par [180]. Par contre, les méthodes utilisant le principe de maximum de vraisemblance nous désignent le

modèle le plus réactif.

Dans un deuxième temps, nous avons montré que l'on peut trouver une quantité minimale de données pour estimer correctement un modèle. Il faudra valider ce nombre minimal d'observations avec de nouveaux tests sur le modèle de synthèse. Ainsi, dans une future phase d'exploitation, nous pourrions donc réévaluer le modèle régulièrement. Dans ces conditions, l'expert pourra bénéficier d'informations lui permettant d'effectuer une programmation dynamique des interventions de maintenance puisque les actions programmées par la méthode vont modifier le comportement du système.

4.3 Comparaison du modèle de synthèse avec le cas industriel

Nous avons comparé le modèle de synthèse avec les données de maintenance fournies par une entreprise d'agro-alimentaire. Les deux distributions sont testées : distribution normale et uniforme. Nous avons au préalable utilisé l'algorithme du filtre entropique vu au § 2.4.3, p. 42 afin d'utiliser uniquement les symboles les plus pertinents. Nous donnons Tableau 4.26, p. 126 les densités respectives du modèle de synthèse (distributions Gaussienne et Uniforme), simulant des données de maintenance, et celles d'un cas industriel. Les données sont issues de la GMAO d'un sous-système pour l'agro-alimentaire (de 2005 à 2006).

Symboles	Densités des symboles			
	Peseuse	Modèle Gauss	Modèle Ajusté	Modèle Uniforme
AU	0,0769	0,0590	0,0590	0,0883
OBS	0,0288	0,0150	0,0150	0,0883
RAN	0,0288	0,0210	0,0210	0,0910
RM	0,0385	0,0530	0,0530	0,0831
SEC	0,0673	0,0760	0,1110	0,0857
TEP	0,0922	0,0990	0,0760	0,0934
VEP	0,1395	0,1110	0,0990	0,0902
Coef. Corrélation : Peseuse VS modèles		0,7554	0,9611	0,3750
Kolmogorov-S : Peseuse VS modèles		93.75%	93.75%	5.62%

Tableau 4.26 – Comparaison du modèle de synthèse, avec les données de maintenance provenant d'une GMAO d'un sous-système pour l'agro-alimentaire (année 2005-2006).

Les résultats d'adéquation avec un test de Kolmogorov-Smirnov (voir § 2.4.5, p. 50) donnent 93,75% pour le modèle de synthèse utilisant une distribution Gaussienne des symboles. Nous trouvons 5,62% d'adéquation pour une distribution Uniforme des symboles. Nous observons Figure 4.24, p. 127 que le modèle de synthèse avec une distribution Gaussienne des symboles se rapproche le plus des données récoltées en milieu industriel.

4.3.1 Ajustement du modèle de synthèse

Nous avons ajusté le modèle de synthèse en proposant la modification suivante :

- lors de l'émission du symbole TEP, il faut réémettre le symbole SEC ;
- lors de l'émission du symbole SEC, il faut réémettre le symbole VEP ;
- lors de l'émission du symbole VEP, il faut réémettre le symbole TEP.

En effet, les noms des symboles sont affectés aléatoirement au départ. Il s'agit alors de réaffecter les noms pour améliorer les correspondances avec les données empiriques.

Nous donnons Figure 4.25, p. 128 (courbe en pointillés rouges), le nouveau modèle de synthèse après l'ajustement proposé précédemment. Le coefficient de corrélation s'avère être meilleur après ajustement (0,9611). Le modèle est donc plus proche des données empiriques.

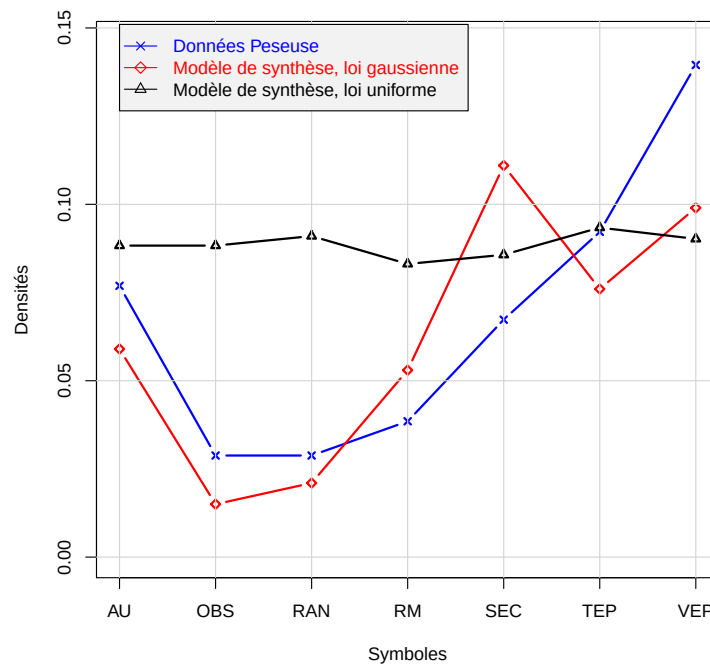


Fig 4.24 – Modèles de synthèse non ajustés.

Nous réaffectons ainsi chaque action à chacun des symboles afin que le modèle de synthèse ressemble à la situation industrielle. Ce réajustement nous permettra d'avoir un modèle de synthèse plus proche de la réalité de terrain. Si le modèle de synthèse ressemble à la situation industrielle alors on peut supposer que le système réel se dégrade comme nous l'avons défini dans le modèle de synthèse.

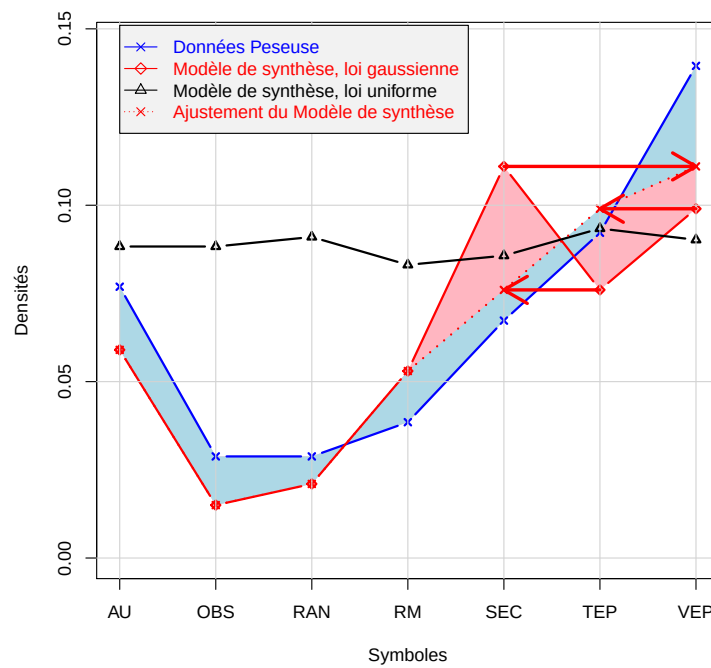


Fig 4.25 – Comparaison du modèle de synthèse ajusté, avec les données de maintenance provenant d’une entreprise du secteur de l’agro-alimentaire.

4.3.2 Résultats après ajustement du modèle de synthèse

Nous ajustons le modèle de simulation avec la permutation des symboles proposée ci-dessus. Nous comparons alors de nouveau les données empiriques avec celles issues des trois topologies du nouveau modèle de simulation. Nous donnons les résultats dans le Tableau 4.27, p. 129, et sur la Figure 4.26, p. 129. En comparant les densités des 3 topologies avec celle du cas industriel, nous trouvons la **topologie 2** comme étant la plus proche du modèle de synthèse (coefficient de corrélation le plus élevé). Notre réajustement des symboles s'avère donc être plus efficace pour le modèle de synthèse utilisant la topologie 2.

Symboles	Densités des symboles			
	Peseuse	Topologie 1	Topologie 2	Topologie 3
AU	0,0769	0,1137	0,0862	0,1027
OBS	0,0288	0,0339	0,0301	0,0341
RAN	0,0288	0,01756	0,0156	0,0244
RM	0,0385	0,0436	0,0489	0,0317
SEC	0,0673	0,0735	0,0604	0,0747
TEP	0,0922	0,0719	0,0843	0,0813
VEP	0,1395	0,0652	0,0907	0,0779
Coef. Corrélation : Peseuse VS topologies		0,5631	0,8793	0,7370

Tableau 4.27 – Comparaison des topologies après réajustement des symboles.

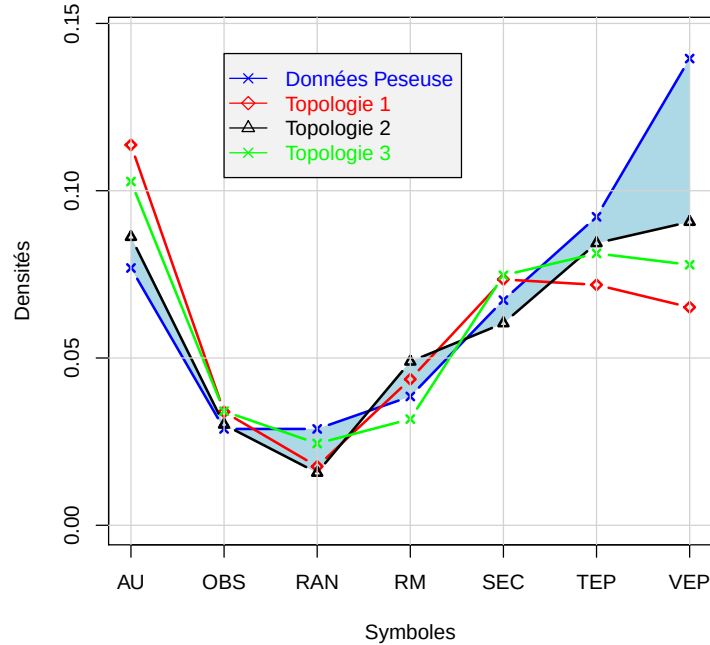


Fig 4.26 – Comparaison des topologies après réajustement des symboles.

4.4 Conclusion

Les études sur les séquences d'observations les plus pertinentes nous ont donné des séquences très différentes entre le modèle de simulation et l'étude de cas réels. Ce qui paraît normal pour un modèle de simulation de type stochastique.

Les résultats sur la fenêtre glissante nous amènent aux mêmes conclusions pour le modèle de simulation ainsi que pour le cas concret. En effet, nous trouvons une entropie maximale pour un nombre fini de symboles. Cette fenêtre permettra d'estimer en temps réel les modèles sans utiliser l'historique complet de la base de données de l'entreprise. Nous pourrions ainsi faire évoluer la modélisation en utilisant désormais **200** observations pour l'estimation des états du système.

Nous avons également déterminé la pertinence de l'architecture des modèles à l'aide d'outils issus de la littérature. Des critères comme l'entropie de Shannon, le maximum de vraisemblance, AIC ou des tests statistiques, ont permis de mesurer la pertinence sur les topologies, les algorithmes d'apprentissage et de décodage, ainsi que sur les distributions.

Le lecteur trouvera le résumé des résultats dans le Tableau 4.28, p. 130. La croix (×) indique la meilleure topologie, le meilleur algorithme d'apprentissage ou la meilleure distribution, en fonction des critères présentés.

Critères d'évaluation	Topologie			Apprentissage		Distribution	
	1	2	3	B.W.	S.K.	Normale	Uniforme
Entropie de Shannon 1 ^{er} ordre	×			×		×	
Entropie de Shannon 2 nd ordre	aucune			aucun		aucune	
Entropie de Shannon 3 ^{ième} ordre	aucune			aucun		aucune	
Entropie de Shannon 4 ^{ième} ordre	aucune			aucun		aucune	
Maximum de Vraisemblance	×			aucun		×	
Akaike Information Criterion	×			aucun		×	
Bayesian Information Criterion	×			aucun		×	
Test d'Aspin-Welch	×			×		×	
Test de Kolmogorov-Smirnov	×			×		×	
Meilleure incertitude	×			×		×	

Tableau 4.28 – Résumé des résultats sur les différents critères de selection de modèles.

Ainsi, la topologie 2 s'avère être celle qui donne le meilleur score avec de nombreux critères ou tests statistiques. L'algorithme d'apprentissage Baum-Welch décodé par Variables Forward donne les meilleurs résultats. Enfin, la distribution normale, par rapport à la distribution uniforme est la plus pertinente.

Nous avons ensuite tenté d'ajuster les données du modèle de synthèse afin que celui-ci puisse s'identifier le mieux possible à la situation industrielle. Les données issues de la loi uniforme ne reflétant pas la situation réelle, nous avons proposé des réaffectations de symboles pour le modèle utilisant la loi Gaussienne. **Cela nous permet de supposer que le processus industriel se dégrade de la même manière que le processus de synthèse fondé sur la topologie 2.**

Conclusion générale et perspectives

Conclusion

Nous avons présenté dans un premier temps, les principes généraux de la maintenance dans le domaine industriel. En se basant sur des normes, nous avons défini les objectifs, les politiques de maintenance ainsi que les différentes typologies adoptées dans ce domaine. Nous avons ensuite introduit les outils de GMAO ainsi que les avancées technologiques récentes utilisées dans ce secteur de l'industrie. Enfin, nous avons présenté les travaux initiés par Pascal Vignat dans sa thèse [177], sur une modélisation de la maintenance en utilisant des Modèles de Markov Cachés, point de départ de ce manuscrit. Ces travaux ont montré qu'il était possible de modéliser la dégradation d'un processus industriel quelconque. Au travers de cas concrets, il a pu donner à l'expert une nouvelle approche de sa gestion de la maintenance préventive à l'aide d'indicateurs de niveaux de dégradation du processus à maintenir.

Dans un second temps, nous avons réalisé un état de l'art sur les approches classiques de mesures de pertinence de modèles. Au travers des méthodes d'analyse de sensibilité ou des méthodes de mesures de pertinence de modèles dont dispose la littérature, nous avons défini une stratégie d'utilisation de certaines de ces méthodes dans notre problématique.

Nous avons ensuite défini les attentes théoriques de nos travaux, ainsi que plusieurs approches classiques visant à évaluer des modèles selon une approche Markovienne. Après avoir défini les différents types de modélisation que nous avons choisis, nous avons présenté un modèle de synthèse ainsi que les architectures utilisées (topologies, algorithmes d'apprentissage, algorithmes de décodage). Ce dernier nous a permis de comparer les données empiriques avec les données simulées.

Enfin, les résultats obtenus nous ont renseignés sur les architectures les plus pertinentes des MMC étudiés. Celles-ci nous ont donné dans un premier temps, les symboles les plus pertinents ainsi que ceux qui n'apportent aucune information supplémentaire à la construction d'un modèle. Nous avons pu trouver dans un second temps, des séquences pertinentes (notamment pour des chaînes de Markov d'ordre 4) annonçant une panne probable. Nous avons ainsi proposé un ajustement du modèle de synthèse (loi Gaussienne), afin qu'il se comporte comme celui de la situation industrielle. Cela nous permet alors de supposer que le processus réel se dégrade comme nous l'avons défini avec ce modèle de synthèse, améliorant ainsi les probabilités de prédiction des pannes.

Des études sur les caractéristiques de la modélisation nous ont montré qu'une période d'échantillonnage des observations toutes les 6 heures était plus pertinente qu'un échantillonnage à la journée. Elles nous ont aussi montré que nous pouvions utiliser un nombre minimal de symboles pour que le modèle puisse donner une information la plus pertinente possible. Ainsi, dans une future phase d'exploitation de ces indicateurs de niveaux de dégradation, cette quantité minimale de données nous permettra de réajuster les modèles utilisés.

Les études sur la pertinence de l'architecture d'un modèle nous ont permis de choisir la topologie, les algorithmes d'apprentissage et de décodage ainsi que la distribution des observations la plus en adéquation avec notre problématique industrielle.

Situation scientifique dans le contexte National et International actuel

De nombreuses approches récentes dans la maintenance industrielle tentent d'appréhender la problématique d'anticipation des pannes par diverses techniques telles que :

- optimiser les performances en fonction du risque donné et les stratégies de fiabilité i.e. établir une politique de surveillance drastique sur les équipements les plus sensibles aux risques de défaillances ;
- organiser le choix des tâches de maintenance en fonction des conséquences des défaillances (Maintenance Steering Groupe MSG-3) ;
- utiliser des techniques comme l'« Asset Management » pour obtenir l'efficacité maximale d'un équipement en optimisant toute la chaîne humaine et matérielle liée à cet équipement.

Notre démarche tente d'appréhender ce pronostic de panne en présentant des outils de choix et d'optimisation de modélisations markoviennes. Ainsi ces outils pourront aider l'expert en fiabilité (qui est devenu de nos jours expert en statistiques), d'enrichir la panoplie d'outils disponibles.

Bénéfices et originalité de l'approche Markovienne :

L'utilisation des MMC dans la détection de défaillances est souvent basée sur l'estimation des lois de survie d'un matériel afin d'en déterminer sa durée de vie utile restante. Du point de vue de l'expert en maintenance, l'intérêt fondamental de notre approche est son aspect visuel qui permettra de contrôler facilement l'état de son système. En outre, bénéficiant de toute la théorie Markovienne, cette approche reste très efficace pour l'estimation des défaillances et assez simple à mettre en œuvre dans un système de gestion de la maintenance industrielle informatisée. Son originalité réside en outre sur les points suivants :

- utilisation d'indicateurs de disponibilité par MMC : les quatre niveaux utilisés (S1 à S4), nous permettent de visualiser très facilement la situation de dégradation d'un processus industriel. L'approche par MMC nous donne accès aux propriétés d'un processus de Markov ainsi qu'à celles des états cachés que nous avons développées dans ce manuscrit ;
- pertinence des observations : en utilisant des méthodes de la littérature courante, nous avons développé des techniques pour tester et comparer la pertinence des observations. Ces techniques pourront servir dans d'autres domaines comme par exemple alerter un routeur¹ informatique ou l'administrateur réseau d'un engorgement futur de son réseau informatique ;

1. Son rôle est de faire circuler des informations d'un réseau vers un autre réseau le plus efficacement possible

- nous avons ensuite déterminé un échantillonnage optimal pour la gestion des données en maintenance industrielle. Nous avons aussi trouvé qu'un nombre fini de données était nécessaire et suffisant afin d'obtenir des réponses pertinentes de la part de nos modèles. Nous avons ainsi pu déterminer une fenêtre glissante optimale de symboles qui permet de réévaluer le modèle dynamiquement ;
- enfin, par des études de pertinence sur les différentes architectures des MMC, nous avons développé des techniques de mise en évidence du meilleur MMC. La validation de ces architectures de modèle, a permis de déterminer le modèle de synthèse le plus réaliste par rapport à un processus industriel réel.

Développement durable :

Dans nos travaux sur la prédiction des pannes industrielles, la détermination du meilleur modèle devrait permettre d'abaisser significativement le taux de pannes du matériel de production. Ce qui veut dire dans un premier temps, moins d'intervention humaine. En effet, dans le cas d'une fonderie industrielle, ces interventions nécessitent la manipulation de produits à haute température par les techniciens de maintenance et peuvent s'avérer très dangereuses. Dans un deuxième temps, dans le cadre du développement durable, nos études pourront engendrer une baisse de CO_2 rejeté dans l'atmosphère. Toujours dans notre exemple d'industries possédant des fours, l'arrêt d'un four pour cause de panne, est une perte d'énergie considérable. La minimisation des pannes pour la presse basse pression permettra d'éviter la refonte de l'aluminium. La consommation électrique ainsi que les rejets de CO_2 seraient donc diminués.

Limites de cette étude

La fiabilité des observations fluctue en fonction de la personne qui saisit les informations de maintenance dans la base de données de la GMAO de l'entreprise. Ces données sont utilisées pour l'apprentissage de nos modèles. Nous devons donc vérifier au préalable la fiabilité de ces informations.

Enfin, à chaque changement d'équipement ou d'évolution du processus de fabrication, le modèle doit évoluer pour s'adapter à ces modifications. Il faudrait alors pouvoir faire des mises à jour dynamiques du modèle.

Les perspectives d'approfondissement

- les **symboles manquants** : nous avons trouvé les observations (ou symboles) les plus pertinents ainsi que ceux apportant peu d'information au modèle. Comme nous l'avons évoqué au § 3.4.6, p. 72, nous pourrions par la suite essayer de trouver d'autres symboles (donc d'autres actions de maintenance), qui nous permettraient d'affiner la qualité de prédiction des dégradations d'un processus ;

- la **robustesse** du modèle : nous avons testé différentes « qualités » du modèle par l'intermédiaire de divers critères de sélection de modèle. Nous pourrions traiter l'aspect robustesse i.e. étudier la stabilité du modèle de synthèse face à des perturbations externes. Nous pourrions par exemple introduire un bruit de type Gaussien dans les observations et étudier le comportement des différentes architectures des modèles ;
- l'utilisation de chaînes **semi-markoviennes** : cette technique pourrait nous permettre d'étudier des processus dont le temps de passage d'un état à un autre peut suivre une loi discrète quelconque. Ce processus doit garder néanmoins une évolution de type markovienne i.e. un processus sans mémoire.
- les réseaux bayésiens dynamiques, avec les travaux de Philippe Weber [183]. Il utilise ce procédé pour des applications en sûreté de fonctionnement et d'analyse de fiabilité de systèmes. Nous pourrions ainsi modéliser l'évolution des variables aléatoires en fonction des évènements de maintenance.

Annexe A

Algorithmes

A.1 Apprentissage Interactive Dichotomizer 3 (ID3)

Entrée : **une partie du corpus d'études (entre 40 et 70 %)** ;
 Sortie : **un arbre de décision** ;
 Validation : **la deuxième partie du corpus qui valide les règles** ;

Appel : **ID3(Ce, F_{ob} , Attributs)** ;
 Corpus d'études Ce : **ensemble des individus** ;
 Un individu a des Attributs : **Valeur(Ind, Att)** ;
 Une fonction-objectif F_{ob} par attribut : $F_{ob}(Ind) = Obj$;
 L'ensemble d'individus qui ont pour l'Attribut la valeur Val : **set(Attribut, Val)** ;

[Créer un noeud]
Si (F_{ob} est égale à une seule valeur) **Alors**
 | Etiquette = valeur ;
Sinon
 Soit A $\in$ Attributs, l'Attribut *le plus discriminant* ;
 Etiquette = A ;
 Pour chaque valeur possible de A **faire**
 | **Si** (set(A, Valeur) $\neq \emptyset$) **Alors**
 | Etiquette-arc = Valeur ;
 | Fils = ID3(set(A, Valeur), F_{ob} , Attributs / A)
 | **Fin Si**
 Fin Pour
Fin Si
 Renvoyer le noeud ;

Algorithme 1: ID3

A.2 Compression de Lempel-Ziv-Welch

```
w : nombre de symboles ayant une entropie max à éliminer ;  
wc : nombre de symboles du modèle ;  
[initialisation]  
w ← Nul ;  
Tant que (lecture d'un caractère c) faire  
    Si (wc existe dans le dictionnaire) Alors  
        | w = wc ;  
    Sinon  
        | ajouter wc au dictionnaire ;  
        | écrire le code de w ;  
        | w = c ;  
    Fin Si  
Fait  
    écrire le code de w ;
```

Algorithme 2: Lempel-Ziv-Welch

A.3 Itératif d'Espace de Version (IVSA)

Entrée : I , un ensemble d'instances ;
 Sortie : H , une liste ordonnée d'hypothèses régionales ;
 Initialisation : $I' := I$, $H := \langle \text{liste vide} \rangle$, $A_{cc_0} := 0$;

Répéter

Pour $i = 1, 2, \dots$ **faire**

$CH := \text{Générateur}(I')$;

$(H, A_{cc_i}) := \text{Assembler}(I, H, A_{cc_{i-1}}, CH)$;

$I' := \text{Incorrect}(I, H)$;

Fin Pour

jusqu'à ce que ($I' = \phi$ ou ($i > \text{max}$ et $A_{cc_i} = A_{cc_{i-\text{max}}}$))

Algorithme 3: IVSA

Entrée : I' , un ensemble d'instances d'entrée ;
 Sortie : CH , un ensemble d'hypothèses régionales candidates ;
 Initialisation : $CH = \phi$;

Pour chaque valeur possible de décision **faire**

$X := \text{instance de } I' \text{ classée négative ou positive}$;

Répéter

$(S, G, x_m) := \text{VSA}(X)$;

$X := X - \{x_1, \dots, x_{m-1}\}$;

$CH := CH \cup S \cup G$;

jusqu'à ce que ($X = \phi$)

Fin Pour

Algorithme 4: algorithme « Générateur »

Entrée : I , un ensemble d'instances d'entrée ;
 Entrée : H , la liste initiale des hypothèses régionales acceptées ;
 Entrée : A_{cc} , la précision initiale de classification ;
 Entrée : CH , un ensemble d'hypothèses régionales candidates ;
 Sortie : mise à jour de H et A_{cc} ;

Pour chaque valeur d'hypothèse candidate $h_i \in CH$ **faire**

$R_i := (|P_c| + |N_c|)/|I|$

Fin Pour

Pour $h \in CH$, ordonnées par valeurs de R décroissantes **faire**

$BestA'_{cc} := 0$;

Pour $j = 1, 2, \dots, |H| + 1$ **faire**

$H' := H$ avec h inséré avant la position j ;

$A'_{cc} := |Correct(I, H')|/|I|$;

Si $(A_{cc} > BestA'_{cc})$ **Alors**

$BestH' := H'$; $BestA'_{cc} := A'_{cc}$;

Fin Si

Fin Pour

Si $(BestA'_{cc} > A_{cc})$ **Alors**

$H := BestH'$; $A_{cc} := BestA'_{cc}$;

Fin Si

Fin Pour

[Supprime les hypothèses inutiles de la liste H]

Pour $h \in H$ **faire**

$H' := H - \{h\}$;

$A'_{cc} := |Correct(I, H')|/|I|$;

Si $(A'_{cc} \geq A_{cc})$ **Alors**

$H := H'$; $A_{cc} := A'_{cc}$;

Fin Si

Fin Pour

Algorithme 5: algorithme « Assembler »

A.4 Viterbi

```

Entrée :  $MMC, e_1, \dots, e_n$ ;
Sortie :  $Viterbi\_path, p_{max}$ ;

[Initialisation]
( $Paths, Probas$ )  $\leftarrow$  initialize( $e_1, MMC$ );
 $E \leftarrow \{e_1, e_2, \dots, e_1, \dots, e_n\}$ ;
Pour  $e_1, \dots, e_i \in E$  faire
    Pour  $s_l \in S$  faire
         $(p(e_i)_{s_l}, s_m) \leftarrow$ 
             $\max_{s_K \in S} (p(e_1, \dots, e_{i-1})_{s_K} \cdot p(s_K; s_l) \cdot p(s_l; e_i))$ ;
        update_path ( $s_l, s_m, Paths, New\_Paths$ );
        update_proba ( $s_l, Probas, New\_Probas$ );
    Fin Pour
     $Patch \leftarrow New\_Paths$ ;
     $Probas \leftarrow New\_Probas$ ;
Fin Pour
( $Viterbi\_path, p_{max}$ )  $\leftarrow$  most_probable( $Paths, Probas$ );
return  $Viterbi\_path, p_{max}$ 

```

Algorithme 6: algorithme « Viterbi »

A.5 Algorithme de calcul de l'entropie moyenne

```

n-max : nombre de symboles ayant une entropie max à élimi-
ner ;
nb-symb : nombre de symboles du modèle ;
nb-etat : entier ; nombre d'états du modèle de Markov
S[ ] : liste des différents symboles du modèle ;
entropie, entropie-max : réel ; calcul de l'entropie
i,j : entier ; variables d'itération
i-max : entier ; indice du symbole d'entropie max
[Calcul des entropies nulles]
i ← 1 ;
Tant que (i ≤ nb-symb) faire
    Si (entropie(S[i]) = 0) Alors
        S[i] = NULL ; on élimine le symbole du tableau
        nb-symb = nb-symb - 1 ;
        i ← i++ ;
    Fin Si
Fait
[Calcul des entropies maxi du modèle]
i ← 1 ;
entropie-max = 0 ;
Pour (j = 0 ; j ≤ n-max ; j++) faire
    Tant que (i ≤ nb-symb) faire
        Si (entropie(S[i]) ≥ entropie-max) Alors
            entropie-max = entropie(S[i]) ;
            i-max = i ;
            i ← i++ ;
        Fin Si
    Fait
    S[i-max] = NULL ; on élimine le symbole du tableau
    nb-symb = nb-symb - 1 ;
Fin Pour
[Calcul de l'entropie moyenne du modèle]
Pour (j = 0 ; j ≤ n-max ; j++) faire
    entropie = entropie(S[i]) + entropie ;
Fin Pour
entropie = entropie / nb-etat ;

```

Algorithme 7: sélection du modèle le plus pertinent.

Annexe B

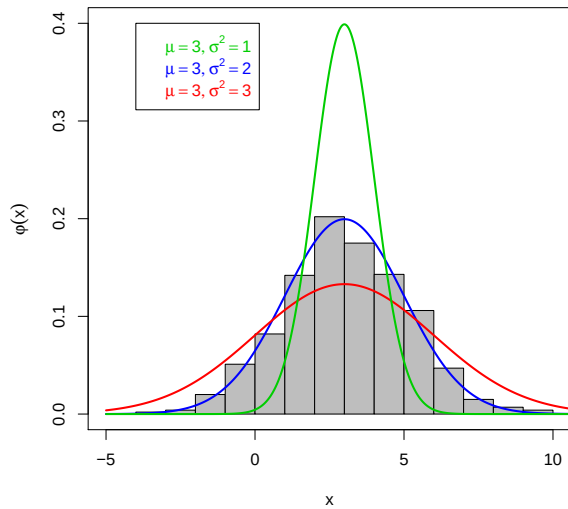
Principales lois de probabilités

B.1 Loi Normale (Laplace – Gauss)

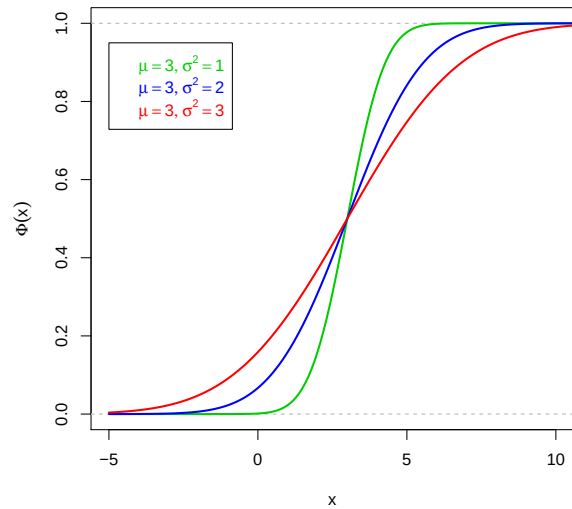
Cette loi permet de modéliser de nombreuses répartitions de variables aléatoires. La loi normale est définie par la densité de probabilité $\varphi : \mathbb{R} \rightarrow \mathbb{R}^+$, d'espérance μ et d'écart type σ :

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad (\text{B.1})$$

Nous donnons Figure B.1(a), p. 149 un exemple d'ajustement de modèle à l'aide d'une loi normale.



(a) Loi Normale, densité de probabilité.



(b) Loi Normale, fonction de répartition.

Fig B.1 – Loi Normale.

Nous donnons Figure B.1(b), p. 149 la fonction de répartition de la loi de Gauss. Elle est définie comme suit :

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dt. \quad (\text{B.2})$$

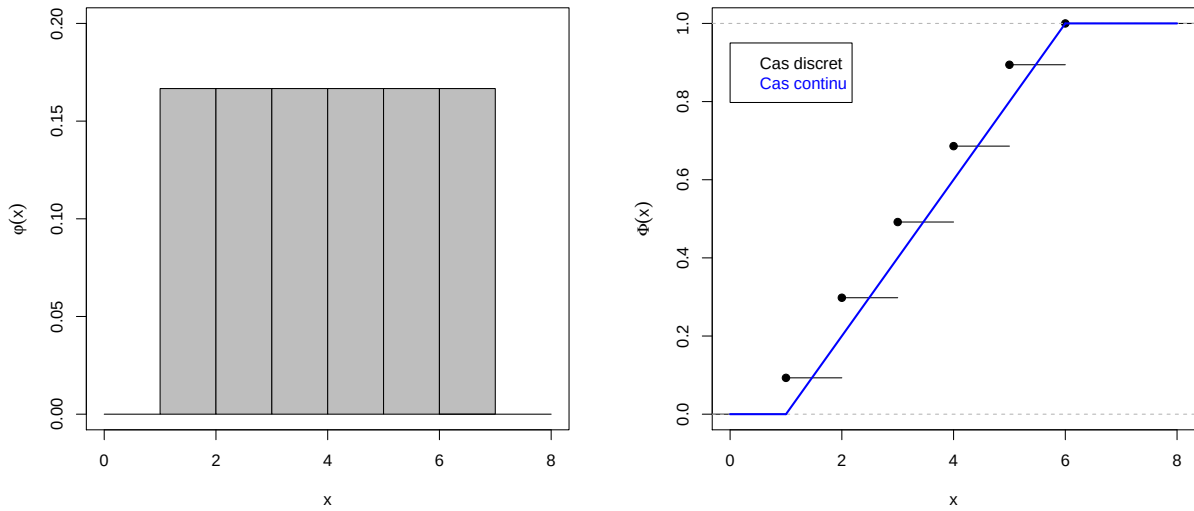
B.2 Loi Uniforme

La loi uniforme permet de modéliser des variables aléatoires uniformément réparties sur un intervalle.

Cette loi est définie par la densité de probabilité sur l'intervalle $[\alpha, \beta]$:

$$\varphi(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{pour } \alpha \leq x \leq \beta, \\ 0 & \text{sinon.} \end{cases} \quad (\text{B.3})$$

Nous donnons Figure B.2(a), p. 150 un exemple dans le cas discret d'un dé non biaisé ($\alpha = 1, \beta = 6$) :



(a) Loi Uniforme, densité de probabilité.

(b) Loi Uniforme, fonction de répartition.

Fig B.2 – Loi Uniforme.

La fonction de répartition est donnée par :

$$\Phi(x) = \begin{cases} 0 & \text{pour } x < \alpha, \\ \frac{x - \alpha}{\beta - \alpha} & \text{pour } \alpha \leq x < \beta, \\ 1 & \text{pour } x \geq \beta. \end{cases} \quad (\text{B.4})$$

Nous donnons Figure B.2(b), p. 150 la fonction de répartition de la loi uniforme.

B.3 Loi Exponentielle

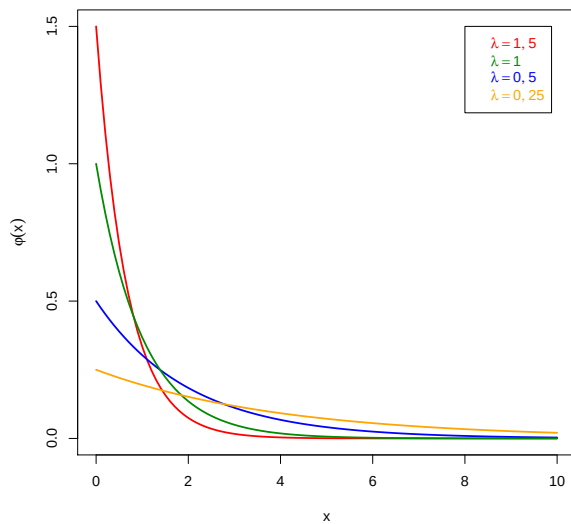
La loi exponentielle correspond souvent à des évènements dont la probabilité de survenue diminue avec le temps. Elle est également utilisée pour modéliser des durées de vie.

La densité de probabilité est de la forme :

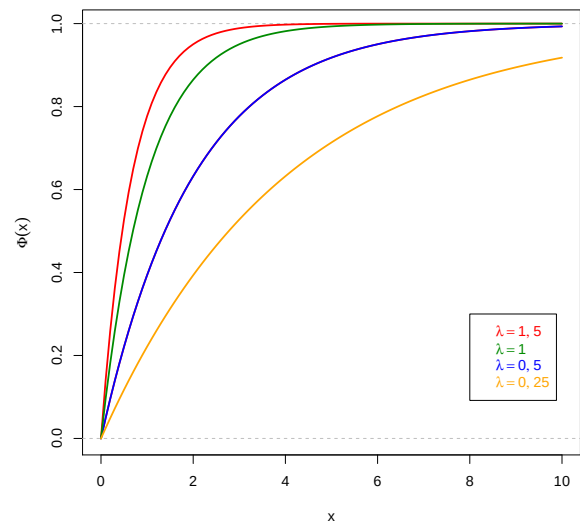
$$\varphi(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0, \\ 0 & , x < 0. \end{cases} \quad (\text{B.5})$$

La fonction de répartition est donnée par :

$$\Phi(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & , x \geq 0, \\ 0 & , x < 0. \end{cases} \quad (\text{B.6})$$



(a) Loi exponentielle, densité de probabilité.



(b) Loi exponentielle, fonction de répartition.

Fig B.3 – Loi Exponentielle.

Annexe C

Notions de mathématiques utilisées

Matrice Jacobienne

Définition

Soit F une fonction vectorielle de $\mathbb{R}^n$ dans $\mathbb{R}^m$ (i.e. F est définie par m fonctions à valeurs dans $\mathbb{R}$) :

$$F : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{pmatrix}, \quad (\text{C.1})$$

la matrice Jacobienne (du mathématicien Charles Jacobi) $J_F(M)$ est la matrice aux dérivées partielles suivante :

$$J_F(M) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}. \quad (\text{C.2})$$

Cette matrice intervient dans la résolution de problèmes non-linéaires, notamment pour résoudre des systèmes d'équations de Principe de Maximum d'Entropie.

Exemple

La matrice Jacobienne de la fonction $F : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ définie par :

$$F(x, y, z) = (x, 2y + 5z, 3x^2 - 4y + 10z, z \sin(x)), \quad (\text{C.3})$$

est :

$$J_F(x, y, z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 5 \\ 6x & -4y & 10 \\ z \cos(x) & 0 & \sin(x) \end{pmatrix}. \quad (\text{C.4})$$

Bibliographie

- [1] ABRIAL, J. R. *The B Book*. Cambridge University Press, 2005.
- [2] AGOUZAL, A. ET LAFOUGE, T. On the relation between the maximum entropy principle and the principle of least effort : The continuous case. *J. Informetrics* 2, 1 (2008), pages 75–88.
- [3] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. 2nd inter. symp. on information theory. *2nd Inter. Symp. on Information Theory* (1973), pages 267–281.
- [4] ALAM, F., MCNAUGHT, K. ET RINGROSE, T. Using morris’ randomized oat design as a factor screening method for developing simulation metamodels. In *Simulation Conference, 2004. Proceedings of the 2004 Winter* (dec. 2004), volume1.
- [5] ARMINJON, M. ET IMBAULT, D. Maximum entropy principle and texture formation. *Zeitschrift für angewandte Mathematik und Mechanik*, 80, Suppl. N°1 (2000), pages 13–16.
- [6] ARTUSI, R., VERDERIO, P. ET MARUBINI, E. Bravais-pearson and spearman correlation coefficients : meaning, test of hypothesis and confidence interval. *Int J Biol Markers* 17, 2 (2002), pages 148–151.
- [7] ASH, R. Information theory. *Dover Publications* (1990).
- [8] AUPETIT, S. *Contributions aux Modèles de Markov Cachés : métaheuristiques d’apprentissage, nouveaux modèles et visualisation de dissimilarité*. These, Université François Rabelais - Tours, Nov. 2005.
- [9] AUPETIT, S., MONMARCHÉ, N. ET SLIMANE, M. Hidden Markov models training using population based metaheuristics. In *Advances in Metaheuristics for Hard Optimization*, P. Siarry and Z. Michalewicz, éditeurs, Natural Computing Series. Springer, 2007, pages 415–438.
- [10] AVILA, M. *Optimisation de modèles Markoviens pour la reconnaissance de l’écrit*. Thèse de doctorat, Université de Rouen, 1996.
- [11] BACHIR, S. *Contribution au diagnostic de la machine asynchrone par estimation paramétrique*. Thèse de doctorat, Université de Poitiers, 03 2002.
- [12] BAIER, C., HAVERKORT, B. R., HERMANN, H. ET KATOEN, J.-P. Model-checking algorithms for continuous-time Markov chains. *IEEE Transaction on Software Engineering* 29, 6 (2003), pages 524–541.
- [13] BARDOU, L. Soutien logistique intégré. *Techniques de l’ingénieur Gestion des flux logistiques base documentaire : TIB119DUO*. (2013).

- [14] BAUM, L. E., PETRIE, T., SOULES, G. ET WEISS, N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics* 41, 1 (1970), pages 164–171.
- [15] BEIRLANT, J., DUDEWICZ, E. J., GYÖRFI, L. ET MEULEN, E. C. Nonparametric entropy estimation : An overview. *International Journal of the Mathematical Statistics Sciences* 6 (1997), pages 17–39.
- [16] BENTLER, P. M. *EQS structural equations program manual*. BMDP Statistical Software, 1995.
- [17] BERCU, B. ET CHAFAÏ, D. *Modélisation stochastique et simulation - Cours et applications*. Collection Sciences Sup - Mathématiques appliquées pour le Master, Société de Mathématiques Appliquées et Industrielles (SMAI), éditions Dunod, 2007.
- [18] BEVEN, K. ET BINLEY, A. The future of distributed models : Model calibration and uncertainty prediction, 1992.
- [19] BIERNACKI, C. ET GOVAERT, G. Choosing models in model-based clustering and discriminant analysis. 0 RR-3509, INRIA, 10 1998.
- [20] BLASONE, R.-S., VRUGT, J. A., MADSEN, H., ROSBJERG, D., ROBINSON, B. A. ET ZYVOLOSKI, G. A. Generalized likelihood uncertainty estimation (glue) using adaptive Markov chain monte carlo sampling. *Advances in Water Resources* 31, 4 (2008), pages 630–648.
- [21] BLOCH, I. Information combination operators for data fusion : a comparative review with classification. *IEEE Transactions on Systems, Man and Cybernetics*, A-26(1) (1996), pages 52–67.
- [22] BLOCH, I. Fusion d’informations numériques : panorama méthodologique. *Proceedings of the Journées Nationales de la Recherche en Robotique* (2005), pages 79–88.
- [23] BOCHAROVA, I. E., HUG, F., JOHANNESSEN, R. ET KUDRYASHOV, B. D. A closed-form expression for the exact bit error probability for viterbi decoding of convolutional codes. *IEEE Transactions on Information Theory* 58, 7 (2012), pages 4635–4644.
- [24] BOUCHARD, G. ET CELEUX, G. Model selection in supervised classification. Rapport technique, INRIA, 2004.
- [25] BOUQUET, F. Méthodes et outils pour l’intelligence artificielle. *Cours de master informatique MOIA* (2003).
- [26] BOURGUIGNON, P. Y. ET ROBELIN, D. Modèles de Markov parcimonieux : sélection de modèle et estimation. *Statistique et Génome* (2004).
- [27] BOX, G. E. P. *Science and statistics*. J. Am. Stat. Assoc. 71, 1976.
- [28] BOZDOGAN, H. Determining the number of clusters in the standart multivariate normal mixture model using model selection criteria. *Rapport technique, Department of Mathematics, University, of Illinois, Chicago* (1983).
- [29] BOZDOGAN, H. On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics, Theory and Methods*, 19(1) (1983), pages 221–278.
- [30] BRISBOIS, J. études probabilistes de sûreté. *B 3 831* (1995).

-
- [31] BURNHAM, K. P. ET ANDERSON, D. R. *Model selection and multimodel inference : a practical information-theoretic approach*, 2nd édition. Springer, July 2002.
 - [32] BURNHAM, K. P. ET ANDERSON, D. R. Multimodel inference, understanding AIC and BIC in model selection. *Sociological Methods and Research* (2004), pages 261–304.
 - [33] CADINI, F., AVRAM, D. ET ZIO, E. System State Estimation by Particle Filtering for Fault Diagnosis and Prognosis. *Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability* 224, 3 (2010), pages 149–158.
 - [34] CAMPOLONGO, F., CARIBONI, J. ET SALTELLI, A. An effective screening design for sensitivity analysis of large models. *Environ. Model. Softw.* 22 (Octobre 2007), pages 1509–1518.
 - [35] CAO, D., KANG, J., ZHAO, J. ET ZHANG, X. Fault diagnosis of gearbox based on eemd and hmm. In *Prognostics and System Health Management (PHM), 2012 IEEE Conference on* (may 2012), pages 1–9.
 - [36] CAPPE, O. ET ROUEFF, F. Evaluation numérique de l’information de fisher pour des observations irrégulières de l’état d’une file d’attente. *GRETSI, Groupe d’Etudes du Traitement du Signal et des Images* (2003).
 - [37] CASTANIER, B. *Modélisation stochastique et optimisation de la maintenance conditionnelle des systèmes à dégradation graduelle*. Thèse de doctorat, Université de Technologie de Troyes, 2002.
 - [38] CAVANAUGH, J. E. Unifying the derivations for the akaike and corrected akaike information criteria. *Statistics & Probability Letters* 33, 2 (1997), pages 201–208.
 - [39] CAVANAUGH, J. E. A large-sample model selection criterion based on kullback’s symmetric divergence. *Statistics & Probability Letters* 42, 4 (1999), pages 333–343.
 - [40] CELEUX, G. ET DIEBOLT, J. The sem algorithm : A probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Comput. Statist. Quarterly*, 2 (1986), pages 73–82.
 - [41] CELEUX, G. ET DIEBOLT, J. The em and the sem algorithms for mixtures : statistical and numerical aspects. *Rapport de recherche, INRIA* (1987).
 - [42] CELEUX, G. ET GOVAERT, G. A classification em algorithm for clustering and two stochastic versions. Research Report RR-1364, INRIA, 1991. Projet CLOREC.
 - [43] CHANDRASEKARAN, V., JOHNSON, J. K. ET WILLSKY, A. S. Maximum entropy relaxation for graphical model selection given inconsistent statistics. *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology Cambridge, MA 02139* (2007).
 - [44] CHAUVEAU, D. *Méthodes de Monte-Carlo par chaînes de Markov et algorithmes de restauration de données manquantes*, 2001. Habilitation à diriger des recherches, Université de Marne la vallée, France.
 - [45] CHEN, S. S. ET GOPALAKRISHNAN, P. S. Speaker, environment and channel change detection and clustering via the bayesian information criterion. In *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop* (Lansdowne, Virginia, USA, February 1998).

- [46] CLAESKENS, G., CROUX, C. ET VAN KERCKHOVEN, J. An information criterion for variable selection in support vector machines. *J. Mach. Learn. Res.* 9 (June 2008), pages 541–558.
- [47] CLAESKENS, G. ET HJORT, N. L. *Model selection and model averaging*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge, New York, 2008.
- [48] COVER, T. M. ET THOMAS, J. A. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.
- [49] CRAMÉR, H. *Mathematical Methods of Statistics*. Princeton University Press, 1946.
- [50] DAVIES, R. H., TWINING, C. J., COOTES, T. F., WATERTON, J. C. ET TAYLOR, C. J. A minimum description length approach to statistical shape modelling. *IEEE Transactions on Medical Imaging* 21 (2001), pages 525–537.
- [51] DAVIS, R. I. A., LOVELL, B. C. ET CAELLI, T. Improved estimation of hidden Markov model parameters from multiple observation sequences. In *ICPR (2)* (2002), pages 168–171.
- [52] DE ROCQUIGNY, E. La maîtrise des incertitudes dans un contexte industriel 1^{ère} partie : Une approche méthodologique globale basée sur des exemples. *Journal de la Société française de statistique* 147(3) (2006), pages 33–71.
- [53] DE ROCQUIGNY, E. La maîtrise des incertitudes dans un contexte industriel 2^{ème} partie : Une approche méthodologique globale basée sur des exemples. *Journal de la Société française de statistique* 147(3) (2006), pages 73–106.
- [54] DE WAELE, S. ET BROERSEN, P. Order selection for vector autoregressive models. *Trans. Sig. Proc.* 51, 2 (Feb. 2003), pages 427–433.
- [55] DEMPSTER, A. P. *Upper and lower probabilities induced by multivalued mapping*. Annals of Mathematical Statistics, AMS-38, 1967.
- [56] DEMPSTER, A. P., LAIRD, N. M. ET RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39, 1 (1977), pages 1–38.
- [57] DESPUJOLS, A. *Approche Fonctionnelle de la Maintenance*. Ed. Techniques Ingénieur, 2004.
- [58] DESPUJOLS, A. *Optimisation de la maintenance par la fiabilité (OMF)*. Ed. Techniques Ingénieur, 2009.
- [59] DEZERT, J. ET SMARANDACHE, F. An introduction to dsmt. *CoRR abs/0903.0279* (2009).
- [60] DREZNER, Z., TUREL, O. ET ZEROM, D. A modified kolmogorov-smirnov test for normality. *Communications in Statistics - Simulation and Computation* 39 (2010), pages 693–704.
- [61] DYTAM, C. *Choosing and Using Statistics : A Biologist's Guide*. John Wiley & Sons, 2011.
- [62] EASTERLING, R. G. ET BERGER, J. O. Statistical foundations for the validation of computer models. *Sandia National Laboratories* (2003).

-
- [63] EFRON, B. ET TIBSHIRANI, R. J. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
 - [64] EL KORSO, M., BOYER, R., RENAUX, A. ET S., M. Expressions non-matriciellles des bornes de Cramér-Rao pour la localisation de source en champ proche. In *Colloque GRETSI 2009* (Dijon France, 2009). Ces travaux sont financés par la région Île de France et le groupe de recherche Digiteo.
 - [65] FLETCHER, R. ET XU, C. Hybrid methods for nonlinear least squares. *IMA Journal of Numerical Analysis* 7, 3 (1987), pages 371–389.
 - [66] FOATA, D. ET FUCHS, A. *Processus stochastiques : processus de Poisson, chaînes de Markov et martingales : cours et exercices corrigés*. Sciences sup. Dunod, 2002.
 - [67] FOX, M., GHALLAB, M., INFANTES, G. ET LONG, D. Robot introspection through learned hidden Markov models. *Artif. Intell.* 170, 2 (2006), pages 59–113.
 - [68] FRECHET, M. Sur l’extension de certaines evaluations statistiques au cas de petits echantillons. *Review of the International Statistical Institute, Vol. 11, No. 3/4* (1943), pages 182–205.
 - [69] FUMAGALLI, L., MACCHI, M. ET RAPACCINI, M. Computerized maintenance management systems in SMEs : A survey in Italy and some remarks for the implementation of Condition Based Maintenance. *13th IFAC Symposium on Information Control Problems in Manufacturing, INCOM2009, Moscow 3–5 June 2009 ; 13* (2009), pages 1615–1619.
 - [70] GALINDO, P. The competitive forward-backward algorithm (cfb). In *Artificial Neural Networks, 1995., Fourth International Conference on* (jun 1995), pages 82–85.
 - [71] GEIST, M., PIETQUIN, O. ET FRICOUT, G. Différences temporelles de kalman : le cas stochastique. In *Actes des Journées Francophones de Planification, Décision et Apprentissage pour la conduite de systèmes JFPDA 2009* (Paris France, 06 2009).
 - [72] GOSAVI, A. A risk-sensitive approach to total productive maintenance. *Automatica* 42, 8 (2006), pages 1321–1330.
 - [73] GOUPY, J. ET CREIGHTON, L. *Introduction aux plans d’expériences*, 3e édition édition. Dunod, Nov. 2006.
 - [74] GUEDJ, B. ET ALQUIER, P. PAC-Bayesian Estimation and Prediction in Sparse Additive Models. page 1, Aug. 2012.
 - [75] HAMILTON, H. J. The iterated version space algorithm. In *In Proc. of Ninth Florida Artificial Intelligence Research Symposium (FLAIRS-96)* (1996), pages 209–213.
 - [76] HANEBECK, U. D. ET HORN, J. New estimators for mixed stochastic and set theoretic uncertainty models : The general case. In *Proceedings of the European Control Conference (ECC 2001)* (Porto, Portugal, Sept. 2001).
 - [77] HANNA, M. S. A close look at the ifo data model. *SIGMOD Rec.* 24, 1 (Mar. 1995), pages 21–26.
 - [78] HANNAN, E. J. ET QUINN, B. G. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)* 41, 2 (1979), pages 190–195.

- [79] HELD, L., RUFIBACH, K. ET BALABDAOUI, F. A score regression approach to assess calibration of continuous probabilistic predictions. *Biometrics* 66, 4 (2010), pages 1295–1305.
- [80] HENG, S., ZHANG, S., TAN, A. ET MATHEW, J. Rotating machinery prognostics. state of the art, challenges and opportunities. *Mechanical Systems and Signal Processing* 23, 3 (2009), pages 724–739.
- [81] HERNANDO, D., CRESPI, V. ET CYBENKO, G. Efficient computation of the hidden Markov model entropy for a given observation sequence. *IEEE transactions on information theory* (2005), pages 2681–2685.
- [82] HIJAZI, H. ET ROS, L. Borne de Cramér-Rao Bayésienne associée à l’estimation des gains complexes d’un canal de Rayleigh avec spectre de Jakes pour les récepteurs OFDM à grande mobilité. In *Actes du 22ème colloque GRETSI sur le Traitement du Signal et des Images, GRETSI 2009 22ème Colloque GRETSI Traitement du Signal & des Images, GRETSI 2009* (Dijon France, 09 2009). Département Images et Signal.
- [83] HOCKER, D., XIAOHU, L. ET IYENGAR, S. S. Shannon entropy based time-dependent deterministic sampling for efficient on-the-fly quantum dynamics and electronic structure. *J. Chem. Theory Comput.* 1 (2011), pages 256–268.
- [84] HONG, T.-P. ET TSANG, S.-S. A generalized version space learning algorithm for noisy and uncertain data. *Knowledge and Data Engineering, IEEE Transactions on* 9, 2 (mar/apr 1997), pages 336–340.
- [85] HUDA, S., YEARWOOD, J. ET TOGNERI, R. A stochastic version of expectation maximization algorithm for better estimation of Hidden Markov Model. *Pattern Recogn. Lett.* 30 (October 2009), pages 1301–1309.
- [86] HURVICH, C. M. ET TSAI, C. L. Regression and time series model selection in small samples. *Biometrika* 76, 2 (1989), pages 297–307.
- [87] HURVICH, C. M. ET TSAI, C. L. Model selection for extended quasi-likelihood models in small samples. *Biometrics* 51 (1995), pages 1077–1084.
- [88] INOUE, M. ET UEDA, N. Exploitation of unlabeled sequences in Hidden Markov Models. *IEEE Trans. On Pattern Analysis and Machine Intelligence* 25 (2003), pages 1570–1581.
- [89] IOOSS, B. Revue sur l’analyse de sensibilité globale de modèles numériques. *e-articles server (based on gBUS)* (2010).
- [90] IOOSS, B. ET RIBATET, M. Global sensitivity analysis of computer models with functional inputs. *Reliability Engineering & System Safety* 94, 7 (07 2009), pages 1194–1204.
- [91] J., M. F. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association* 46, 253 (1951), pages 68–78.
- [92] JARDINE, A., LIN, D. ET BANJEVIC, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20, 7 (Oct. 2006), pages 1483–1510.
- [93] JARDINE, A., LIN, D. ET BANJEVIC, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20, 7 (Oct. 2006), pages 1483–1510.

-
- [94] JAYNES, E. Information theory and statistical mechanics. *Physical Review*, vol. 16, no. 4 (1957), pages 620–630.
- [95] JIN, F. ET DING, S. An improved pca algorithm based on wif. In *IJCNN'08* (2008), pages 1576–1578.
- [96] JOSHI, A., JAMES, S., MECKL, P., KING, G. ET JENNINGS, K. Information-theoretic feature selection for classification. In *American Control Conference, 2007. ACC '07* (2007), pages 2000–2005.
- [97] JUANG, B. H. ET RABINER, L. R. The segmental k-means algorithm for estimating parameters of hidden Markov models. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 38, 9 (Sept. 1990), pages 1639–1641.
- [98] KADANE, J. B. ET LAZAR, N. A. Methods and criteria for model selection. *Journal of the American Statistical Association* 99 (2004), pages 279–290.
- [99] KADOTA, K. ET TAKAHASHI, K. Detection of genes with tissue-specific patterns using akaike's information criterion. *Genome Informatics* 14 (2003), pages 308–309.
- [100] KALMAN, R. E. A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering, Série D, VOL. 82* (1960), pages 34–45.
- [101] KANS, M. An approach for determining the requirements of computerised maintenance management systems. *Comput. Ind.* 59, 1 (Jan. 2008), pages 32–40.
- [102] KAPETANIOS, G. Information criteria, model selection uncertainty and the determination of cointegration rank. NIESR Discussion Papers 166, National Institute of Economic and Social Research, 2000.
- [103] KLEIJNEN, J. P. C. Verification and validation of simulation models. *European Journal of Operational Research* 82 (1995), pages 145–162.
- [104] KULLBACK, S. ET LEIBLER, R. A. On information and sufficiency. *The Annals of Mathematical Statistics* 22, 1 (1951), pages 79–86.
- [105] LABADI, K., CHEN, H. ET AMODEO, L. Modeling and performance evaluation of inventory systems using batch deterministic and stochastic petri nets. *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 37, 6 (2007), pages 1287–1302.
- [106] LE DUY, T. D. *Traitement des incertitudes dans les applications des études probabilistes de sûreté nucléaire*. Thèse de doctorat, Université de technologie (Troyes), 2011.
- [107] LEBARBIER, E. ET MARY-HUARD, T. Le critère BIC : fondements théoriques et interprétation. Research Report RR-5315, INRIA, 2004.
- [108] LEBOLD, M. ET THURSTON, M. Open standards for Condition-Based Maintenance and Prognostic Systems. In *5th Annual Maintenance and Reliability Conference* (2001).
- [109] LI, F., QI, F., SHI, G. ET ZHANG, L. Optimization-based particle filter for state and parameter estimation. *Journal of Systems Engineering and Electronics* (2009), pages 479–484.
- [110] LINARÈS, G., LÉVY, C. ET PLAGNIOL, J.-C. Estimation rapide de modèles semi-continus discriminants. In *Journées d'études de la Parole, JEP* (Dinard, France, 2006).

- [111] LOW-KAM, C., LAURENT, A. ET TEISSEIRE, M. Détection de séquences atypiques basée sur un modèle de Markov d'ordre variable. In *EGC'09 : Extraction et Gestion de Connaissances* (2009), page 12.
- [112] MALLOWS, J. Some comments on cp. *Echnometrics*, 15 (1973), pages 661–675.
- [113] MARCO, D. Markov random processes are neither bandlimited nor recoverable from samples or after quantization. *Information Theory, IEEE Transactions on* 55, 2 (feb. 2009), pages 900–905.
- [114] MCKAY, M. D., MORRISON, J. D. ET C., U. S. Evaluating prediction uncertainty in simulation models. *Los Alamos National Laboratory Los Alamos, New Mexico 87545-0600 USA* (1998).
- [115] MCQUARRIE, A. The model selection criterion AICu. *Statistics & Probability Letters* 34, 3 (June 1997), pages 285–292.
- [116] MITCHELL, T. M. Version spaces : a candidate elimination approach to rule learning. In *IJCAI'77 : Proceedings of the 5th international joint conference on Artificial intelligence* (San Francisco, CA, USA, 1977), Morgan Kaufmann Publishers Inc., pages 305–310.
- [117] MORRIS, M. D. Factorial sampling plans for preliminary computational experiments. *Technometrics* 33 (April 1991), pages 161–174.
- [118] MULLER, A., SUHNER, M.-C. ET IUNG, B. Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system. *Reliability Engineering & System Safety* 93, 2 (Feb. 2008), pages 234–253.
- [119] NEUMAN, S. P. Maximum likelihood bayesian averaging of uncertain model predictions. *Stochastic Environmental Research and Risk Assessment* 17 (2003), pages 291–305. 10.1007/s00477-003-0151-7.
- [120] NILSSON, J. Reliability and cost centered maintenance methods - nuclear power and reliability centered maintenance (RCM), July 2007.
- [121] NORME. Maintenance - terminologie de la maintenance. *NF EN 13306 (indice de classement : X60319)* (2010).
- [122] OLIVIER, C., JOUZEL, F., EL MATOUAT, A. ET COURTELLEMONT, P. Prediction with vague prior knowledge. *Communications in Statistics 25- Theory and Methods* 1 (1996), pages 601–608.
- [123] OLIVIER, C., JOUZEL, F., EL MATOUAT, A. ET COURTELLEMONT, P. Un nouveau critère pour la sélection de l'ordre d'un modèle. *Seizième colloque Grets* (1997).
- [124] OULD EL MEDHI, A., LECLERCQ, E. ET LEFEBVRE GREAH, D. Identification of stochastic and deterministic stochastic petri net models for reliability analysis. *Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability* 223, 1 (2009), pages 13–26.
- [125] PAKYARI, R. ET BALAKRISHNAN, N. A general purpose approximate goodness-of-fit test for progressively type-ii censored data. *IEEE Transactions on Reliability* 61, 1 (2012), pages 238–244.
- [126] PARDOUX, E. *Processus de Markov et applications*. Dunod, 2006.

-
- [127] PAYARÓ, M. ET PALOMAR, D. P. Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector gaussian channels. *IEEE Trans. Inf. Theor.* 55, 8 (2009), pages 3613–3628.
- [128] PHUA, C. W. ET FITCH, R. Tracking value function dynamics to improve reinforcement learning with piecewise linear function approximation. In *Proceedings of the 24th international conference on Machine learning* (New York, NY, USA, 2007), ICML '07, ACM, pages 751–758.
- [129] PIBOULEAU, L. *Assimiler et utiliser les statistiques*. Ellipses Marketing, technosup, 2010.
- [130] PIGEAU, A. *Structuration géo-temporelle de données multimédia personnelles en vue de la navigation sur un appareil mobile*. Thèse de doctorat, Université de Nantes, 12 2005.
- [131] POSADA, D. ET BUCKLEY, T. R. Model selection and model averaging in phylogenetics : Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology* 53, 5 (2004), pages 793–808.
- [132] POWELL, M. J. D. A hybrid method for non-linear equations. *Numerical methods for non-linear algebraic equations* (1970).
- [133] PRZYTULA, K. ET THOMPSON, D. Construction of bayesian networks for diagnostics. In *Aerospace Conference Proceedings, 2000 IEEE* (2010), volume5, pages 193–200 vol.5.
- [134] PUJOL, G., LE RICHE, R., BAY, X. ET ROUSTANT, O. Minimisation de quantiles - application en mécanique. In *9e colloque national en calcul des structures* (May 2009).
- [135] QUINLAN, J. R. Discovering rules by induction from large collections of examples. In *D. Michie (Eds.), Expert Systems in the Micro-Electronic Age*. Edinburgh : Edinburgh University Press. (1979), pages 168–201.
- [136] QUINLAN, J. R. *C4.5 : Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*, 1 édition. Morgan Kaufmann, January 1993.
- [137] RABINER, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceeding of the IEEE, 77(2) SIAM interdisciplinary journal* 1 (1989), pages 257–286.
- [138] RAMASSO, E. Contribution of belief functions to hidden Markov models with an application to fault diagnosis. In *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on* (sept. 2009), pages 1–6.
- [139] RAMASSO, E., ROMBAUT, M. ET PELLERIN, D. Forward-Backward-Viterbi procedures in the Transferable Belief Model for state sequence analysis using belief functions. In *Forward-Backward-Viterbi procedures in the Transferable Belief Model for state sequence analysis using belief functions ECSQARU* (Hammamet Tunisie, 2007). Département Images et Signal.
- [140] RISSANEN, J. Modelling by the shortest data description. *Automatica* 14 (1978), pages 465–471.
- [141] RONCHETTI, E. Robust model selection in regression. *Statis. Probab. Lett.* 3 (1985), pages 21–23.

- [142] ROSEN, R. *Life Itself : A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. Columbia University Press, 1991.
- [143] ROUCOULES, L., YANNOU, B. ET EYNARD, B. *Ingénierie de la conception et cycle de vie des produits*. Traité IC2. Hermes science publ. Lavoisier, Paris, 2006.
- [144] RUKHIN, A., SOTO, J., NECHVATAL, J., BARKER, E., LEIGH, S., LEVENSON, M., BANKS, D., HECKERT, A., DRAY, J., VO, S., RUKHIN, A., SOTO, J., SMID, M., LEIGH, S., VANGEL, M., HECKERT, A., DRAY, J. ET III, L. E. B. *A statistical test suite for random and pseudorandom number generators for cryptographic applications*, 1 édition. Wiley, 2010.
- [145] SALTELLI, A., CHAN, K. ET SCOTT, E. M. *Sensitivity Analysis*, 1 édition. Wiley, 2000.
- [146] SARGENT, R. G. Verification and validation of simulation models. In *Winter Simulation Conference* (2010), pages 166–183.
- [147] SCALAS, E., MARTIN, E. ET GERMANO, G. The ehrenfest urn revisited : Playing the game on a realistic fluid model. *Physical Review E - Statistical, Nonlinear and Soft Matter Physics* 76, 1 Pt 1 (2007).
- [148] SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics* 6 (1978), pages 461–464.
- [149] SCILAB ENTERPRISES ET CONSORTIUM SCILAB. *Scilab : Le logiciel open source gratuit de calcul numérique*. Scilab Enterprises et Consortium Scilab, Digiteo, Version 5.3.0, Paris, France, 2010.
- [150] SERGENT, M., DUPUY, D., CORRE, B. ET CLAEYS-BRUNO, M. Comparaison de méthode criblage pour la simulation numérique. In *41èmes Journées de Statistique, SFdS, Bordeaux* (Bordeaux, France, 2009).
- [151] SERIR, L., RAMASSO, E. ET ZERHOUNI, N. Time-Sliced temporal evidential networks : the case of evidential HMM with application to dynamical system analysis. In *Prognostics and Health Management*. (Denver - Colorado, États-Unis, 2011), I. C. N. . CPF11PHM-CDR, éditeur, volumesur CD ROM, page 10.
- [152] SHAFER, G. *A mathematical theory of evidence*. Princeton university press, 1976.
- [153] SHANG, J. ET CAVANAUGH, J. E. Bootstrap variants of the akaike information criterion for mixed model selection. *Comput. Stat. Data Anal.* 52 (2008), pages 2004–2021.
- [154] SHANNON, C. E. A mathematical theory of communication. *Bell system technical journal* 27 (1948).
- [155] SHI, P. ET TSAI, C. L. A note on the unification of the akaike information criterion. *Statist. Soc. B* 60 (1998), pages 551–558.
- [156] SHORACK, G. R. ET WELLNER, J. A. *Empirical Processes With Applications to Statistics*. Society for Industrial & Applied Mathematics, 2009.
- [157] SMARANDACHE, F. ET DEZERT, J. *Advances and applications of DS_mT for information fusion (Collected works)*. Vol. 1-3, American Research Press, 2009.
- [158] SMETS, P. ET KENNES, R. The transferable belief model. *Artificial Intelligence* 66(2) (1994), pages 191–234.

-
- [159] SOBOL, I. M. Sensitivity analysis for non linear mathematical models. *Mathematical Modeling and Computational Experiment*, 1 (1993), pages 407–414.
 - [160] STEINEBACH, J., LEHMANN, E. L. ET ROMANO, J. P. Testing statistical hypotheses. *Metrika* 64 (2006), pages 255–256.
 - [161] STOICA, P. ET ARYE, N. Music, maximum likelihood, and cramér-rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 37, 5 (Mai 1989), pages 720–741.
 - [162] STOICA, P. ET BABU, P. On the proper forms of BIC for model order selection. *IEEE Transactions on Signal Processing* 60, 9 (2012), pages 4956–4961.
 - [163] SUDRET, B. *Uncertainty propagation and sensitivity analysis in mechanical models – Contributions to structural reliability and stochastic spectral methods*, 2007. Habilitation à diriger des recherches, Université Blaise Pascal, Clermont-Ferrand, France.
 - [164] SUGIURA, N. Further analysts of the data by akaike’ s information criterion and the finite corrections – further analysts of the data by akaike’ s. *Communications in Statistics - Theory and Methods* 7, 1 (1978), pages 13–26.
 - [165] TACNET, J.-M., BATTON-HUBERT, M. ET DEZERT, J. A two-step fusion process for multi-criteria decision applied to natural hazards in mountains. *CoRR abs/1005.0896* (2010).
 - [166] TARANTOLA, S., SALTELLI, A. ET ANNONI, P. Sensitivity analysis. A coming of age. In *41èmes Journées de Statistique, SFdS, Bordeaux* (Bordeaux, France France, 2009).
 - [167] TOBON-MEJIA, D., MEDJAHAR, K., ZERHOUNI, N. ET TRIPOT, G. Hidden Markov models for failure diagnostic and prognostic. In *Prognostics and System Health Management*. (Shenzhen, Chine, 2011), IEEE Catalog Number : CFP116IH-PRT, pages 1–8.
 - [168] TOBON-MEJIA, D., MEDJAHAR, K., ZERHOUNI, N. ET TRIPOT, G. A data-driven failure prognostics method based on mixture of gaussians Hidden Markov Models. *Reliability, IEEE Transactions on* 61, 2 (june 2012), pages 491–503.
 - [169] TURÁNYI, T. Sensitivity analysis of complex kinetic systems. tools and applications. *Journal of Mathematical Chemistry* 5 (1990), pages 203–248. 10.1007/BF01166355.
 - [170] UEDA, T. Simple method for the detection of outliers. *Japanese J. Appl. Stat.* (2009), pages 67–76.
 - [171] VALLISNERI, M. *A User Manual for the Fisher Information Matrix*. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, 2007.
 - [172] VANDEWALLE, V. Sélection prédictive d’un modèle génératif par le critère AICp. In *41èmes Journées de Statistique, SFdS, Bordeaux* (Bordeaux, France, 2009).
 - [173] VENDITTI, V. *Aspects du principe de maximum d’entropie en modélisation statistique*. Thèse de doctorat, Université Joseph Fourier - Grenoble 1, 1998.
 - [174] VIDYASAGAR, M. Bounds on the kullback-leibler divergence rate between Hidden Markov Models. *Decision and Control, 2007 46th IEEE Conference on* (2007), pages 6160–6165.

- [175] VITERBI, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* 13, 2 (Apr. 1967), pages 260–269.
- [176] VRIGNAT, P. Modélisation des dysfonctionnements d'un système dans le cadre d'activités de maintenance. Mémoire de D.E.A., Université d'Orléans - LVR, 2007.
- [177] VRIGNAT, P. *Génération d'indicateurs de maintenance par une approche semiparamétrique et par une approche Markovienne*. Thèse de doctorat, Université d'Orléans, 2010.
- [178] VRIGNAT, P., AVILA, M., DUCULTY, F. ET KRATZ, F. Modélisation des dysfonctionnements d'un système dans le cadre d'activités de maintenance. *16ème Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement, Avignon, Communication 4A-1* (2008).
- [179] VRIGNAT, P., AVILA, M., DUCULTY, F. ET KRATZ, F. Use of HMM for evaluation of maintenance activities. *IJAIS, International Journal of Adaptive and Innovative Systems, Vol. 1, Nos. 3/4 1* (2010), pages 216–232.
- [180] VRIGNAT, P., AVILA, M., DUCULTY, F., ROBLÈS, B. ET KRATZ, F. Utilisation des chaînes de Markov cachées pour une évaluation des activités de maintenance dans le cadre d'un processus industriel pour l'agroalimentaire. *CNRIUT* (2009).
- [181] WANG, J. ET SCHAALE, G. B. Model selection for linear mixed models using predictive criteria. *Communications in Statistics - Simulation and Computation* 38 (2009), pages 788–801.
- [182] WEBER, M. D., LEEMIS, L. M. ET KINCAID, R. K. Minimum kolmogorov-smirnov test statistic parameter estimates. *Journal of Statistical Computation and Simulation* 76 (2006), pages 196–206.
- [183] WEBER, P. ET JOUFFE, L. Complex system reliability modelling with dynamic object oriented bayesian networks (doobn). *Rel. Eng. Sys. Safety* 91, 2 (2006), pages 149–162.
- [184] WELCH, B. L. The generalization of student's problem when several different population variances are involved. *Biometrika* 34, 1-2 (1947), pages 28–35.
- [185] WELCH, B. L. Welch's k-sample test. *Biometrika* 38 (1951), pages 330–336.
- [186] WELCH, T. A. A technique for high-performance data compression. *Computer* 17 (June 1984), pages 8–19.
- [187] WOLSZTYNSKI, E., THIERRY, E. ET PRONZATO, L. Minimum-entropy estimation in semi-parametric models. *Signal Process.* 85 (May 2005), pages 937–949.
- [188] WOODROOFE, M. On model selection and the arc sine laws. *The Annals of Statistics* 10 (1982), pages 1182–1194.
- [189] YANG, Y. Can the strengths of AIC and BIC be shared? *BIOMETRICA* 92 (2003), pages 937–950.
- [190] YING, J., KIRUBARAJAN, T., PATTIPATI, K. R. ET PATTERSON-HINE, A. A hidden Markov model-based algorithm for fault diagnosis with partial and imperfect tests. *Trans. Sys. Man Cyber Part C* 30, 4 (Nov. 2000), pages 463–473.
- [191] YU, D. ET YAU, K. K. W. Conditional akaike information criterion for generalized linear mixed models. *Comput. Stat. Data Anal.* 56, 3 (Mar. 2012), pages 629–644.

- [192] ZHOU, J. ET ZHANG, X.-P. Hidden Markov model framework using independent component analysis mixture model. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on* (may 2006), volume 5, page V.
- [193] ZHOU, Z.-J., HU, C.-H., XU, D.-L., CHEN, M.-Y. ET ZHOU, D.-H. A model for real-time failure prognosis based on hidden Markov model and belief rule base. *European Journal of Operational Research* 207, 1 (2010), pages 269–283.
- [194] ZWINGELSTEIN, G. *La maintenance basée sur la fiabilité : guide pratique d'application de la RCM*. Collection Diagnostic et maintenance. Hermès, Paris, 1996. RCM : Reliability centered maintenance.

Communications

- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. HMM Framework, for Industrial Maintenance Activities, *Proceedings of the Institution of Mechanical Engineers, Part O : Journal of Risk and Reliability* 2014.
- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. Quantité minimale d'observations pour l'estimation de Modèles de Markov Cachés, dans le cadre d'une politique de maintenance préventive, *5^{èmes} Journées Doctorales / Journées Nationales MACS, 11–12 Juillet 2013 Strasbourg, (présentation orale)*.
- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. Evaluation of Minimal Data Size by Using Entropy, in a HMM Maintenance Manufacturing Use, *MIM'2013 IFAC (International Federation of Automatic Control) Manufacturing Modelling, Management and Control, 19–21 Juin 2013 Saint Petersburg (Russie)*.
- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. Mise à jour dynamique de Modèles de Markov Cachés : Application dans l'Aide à la Décision pour une Maintenance Préventive Industrielle. *CNRIUT 12–14 Juin 2013 Corté, (présentation orale)*.
- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. HMM Framework, for Industrial Maintenance Activities, *QUALITA'2013 10ème Congrès International Pluridisciplinaire Qualité et Sûreté de Fonctionnement (Qualité, Sûreté de Fonctionnement, Développement Durable), 20–22 Mars 2013 Compiègne, page 43, (présentation orale en Anglais)*.
- VRIGNAT, P., DUCULTY, F., BÉGOT, S., MILLET, J.F., ROBLÈS, B., AVILA, M. Solution Bluetooth : utilisation d'une architecture logicielle dans le cadre d'une mise en place d'une solution multi-clients à partir d'un serveur OPC, *CETISIS, 20–22 Mars 2013 Caen*.
- VRIGNAT, P., AVILA, M., DUCULTY, F., ROBLÈS, B., BÉGOT, S., KRATZ, F. Génération d'indicateurs dans le cadre d'une politique de maintenance préventive conditionnelle, $\lambda\mu$, *18ème Congrès de Maîtrise des Risques et de Sûreté de Fonctionnement, 16–18 Octobre 2012 Tours*.
- ROBLÈS, B., AVILA, M., DUCULTY. ET KRATZ, F. Study of the relevance of stochas-

- tic parameters on Hidden Markov Model parametric and semi-parametric measures of uncertainty, search for missing symbols in model development. *Journée des Jeunes Chercheurs du laboratoire Prisme 2-3 Juillet 2012 Roiffé*, (**présentation orale**).
- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. Incertitudes Stochastiques sur des Modèles de Markov Cachés : Application dans l'Aide à la Décision pour une Maintenance Préventive Industrielle. *CNRIUT 6-8 Juin 2012 Tours*, (**présentation orale**).
 - ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. Methods to choose the best Hidden Markov Model topology for improving maintenance policy. *MOSIM 2012, the 9th International Conference on Modeling, Optimization and SIMulation (Bordeaux, France), 6-8 June 2012*, (**présentation orale en Anglais**).
 - ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P. ET KRATZ, F. Statistical evaluation of Hidden Markov Models topologies, based on industrial synthetic model. *14th IFAC (International Federation of Automatic Control) Symposium on Information Control Problems in Manufacturing (INCOM), volume 14, 1029-1034. Elsevier Ltd on IFAC-PapersOnLine.net doi :10.3182/20120523-3-RO-2023.00052 ISBN : 978-3-902661-98-2 Bucharest, Romania, May 23-25 2012*, (**présentation orale en Anglais**).
- A obtenu le prix du meilleur article de la session « Intelligent Integrated Maintenance and Quality Strategies ».**
- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P., BÉGOT, S. ET KRATZ, F. Mesures de pertinence par les critères du « maximum de vraisemblance » de « BIC » et d'« AIC » appliqués à l'évaluation des paramètres stochastiques de Modèles de Markov Cachés. **Journal National de la Recherche en IUT N°3 (2012)**.
 - ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P. ET KRATZ, F. Evaluation of relevance of stochastic parameters on hidden markov models. *In Advances in Safety, Reliability and Risk Management, European Safety and Reliability Conference (ESREL), ISBN : 978-0-415-68379-1 (Troyes, France, Sept. 2011), Taylor & Francis Group, page 71*, (**présentation orale en Anglais**).
 - VRIGNAT, P., AVILA, M., DUCULTY, F., ROBLÈS, B. ET KRATZ, F. Decison support with a markovian approach for maintenance context activities. *In Advances in Safety, Reliability and Risk Management, European Safety and Reliability Conference (ESREL), ISBN : 978-0-415-68379-1 (Troyes, France, Sept. 2011), Taylor & Francis Group, page 66*.
 - ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P. ET KRATZ, F. Mesures de pertinence par les critères du « maximum de vraisemblance » et de « BIC » appliqués à l'évaluation des paramètres stochastiques de Modèles de Markov Cachés. *CNRIUT 8-10 Juin 2011 Cherbourg-Octeville*, (**présentation orale**).

- ROBLÈS, B., AVILA, M., DUCULTY, F., VRIGNAT, P. ET KRATZ, F. Evaluation de la pertinence des paramètres stochastiques sur des Modèles de Markov Cachés. *CNRIUT 9–11 Juin 2010 Angers, (présentation orale)*.
- VRIGNAT, P., AVILA, M., DUCULTY, F., ROBLÈS, B. ET KRATZ, F. Utilisation des chaînes de Markov cachées pour une évaluation des activités de maintenance dans le cadre d'un processus industriel pour l'agroalimentaire. *CNRIUT 8–10 Juin 2009 Villeneuve d'Ascq*.

Liste des figures

1.1	Nombres d'accidents mortels et de victimes liés à un manque de suivi des équipements, pour les principales activités concernées.	8
1.2	Typologies des actions de maintenance (NF EN 13306 (indice de classement : X60319)) [121].	15
1.3	Exemple d'optimisation de la Maintenance Basée sur la Fiabilité.	17
1.4	Exemple d'Optimisation de la Maintenance par la Fiabilité.	19
1.5	Modèle de Markov Caché à deux états.	21
1.6	Diagnostic vs Pronostic de panne.	22
2.1	Études de pertinences de Modèles de Markov Cachés.	29
2.2	Principe d'analyse de sensibilité d'un modèle [145].	32
2.3	schéma général d'analyse de l'incertitude probabiliste inspiré de [163].	36
2.4	Méthode du <i>Bootstrap</i>	52
3.1	Exemple d'un réseau de Petri à 4 places, 2 transitions et 2 jetons.	64
3.2	Chaîne de Markov à deux états.	66
3.3	Chaîne de Markov : modèle de diffusion d'Ehrenfest.	67
3.4	Relations de dépendance entre les variables aléatoires d'un MMC. Pour chaque état q_t à un instant t , il y a émission d'un symbole V_t pris dans l'ensemble $\mathbb{V}$	68
3.5	Modèles de Markov Cachés, topologies à 4 états. Les λ_k , μ_k sont des a_{ij} illustrant les transitions S_i vers S_j . Les λ_k détériorent l'état et les μ_k améliorent l'état. La matrice d'initialisation π pointe obligatoirement sur l'état S4 puisque nous supposons démarrer toujours dans l'état optimal (S4 est l'état optimal, S1 est l'état du processus arrêté).	71
3.6	Modèle de Markov Caché à 4 états, topologie 2, référence du modèle de synthèse.	73
3.7	Séquences V des T observations du modèle de synthèse. La Figure 3.7(a) représente la distribution normale et la Figure 3.7(b) représente la distribution uniforme. Sur l'abscisse, nous pouvons voir la liste des 11 séquences d'observations émisent par le modèle de synthèse. Pour chaque séquence, on trouve les 10 symboles (DEP, RM, AU, OBS, TEP, SEC, RAN, NET, VEP et RAS). Ces figures illustrent la distribution de chaque séquence pour chaque symbole. Nous remarquons que chaque séquence se termine par le symbole d'arrêt DEP.	75

3.8	Distribution des symboles par état du modèle de synthèse. La Figure 3.8(a) représente la distribution normale et la Figure 3.8(b) représente la distribution uniforme. Sur l'abscisse, nous pouvons voir la liste des 10 symboles. Pour chaque symbole, on trouve les 4 états (S1, S2, S3 et S4). Ces figures illustrent la distribution de chaque symbole pour chaque état. Nous remarquons que le premier symbole n'est émis que par l'état S1 (symbole d'arrêt pour l'état de non fonctionnement). Nous remarquons aussi que l'état S1 n'émet aucun des 9 autres symboles.	76
3.9	Dégradation d'un processus.	77
3.10	Étapes d'évaluation de l'architecture des modèles, à l'aide d'un modèle de synthèse. Le modèle de synthèse utilise la topologie 2 pour générer des 2-uplet (<i>Symboles</i> , <i>Etats</i>) en utilisant les distributions (uniforme et normale). Nous injectons alors ces signatures dans les 3 topologies étudiées. Nous utilisons les 2 algorithmes d'apprentissage et de décodage pour obtenir de nouvelles séquences que nous analysons pour en évaluer la pertinence. Nous essayons ainsi de trouver la meilleure architecture des modèles.	79
3.11	Distribution des symboles par état. La Figure 3.11(a) représente la distribution normale des observations émises par le modèle de synthèse. La Figure 3.11(b) représente la distribution des observations après apprentissage Baum-Welch et décodage par Variables Forward. La Figure 3.11(c) représente la distribution des observations après apprentissage Segmental K-means et décodage par Viterbi.	80
3.12	Production des symboles. Cela nous permet de tester différentes distributions de symboles, différentes topologies et différents algorithmes d'apprentissage et de décodage.	81
3.13	Modélisation Newton de la densité des états.	85
4.1	Modèle de Markov caché à 4 états, topologie 2, référence du modèle de synthèse.	89
4.2	Nombre minimal de données par entropie de Shannon. Les données sont issues du modèle de synthèse.	93
4.3	Évaluation du nombre minimal de données.	94
4.4	Nombre minimal de données par entropie de Shannon en utilisant les 2 algorithmes d'apprentissage.	94
4.5	Évaluation du nombre minimal de données pour une utilisation optimale des algorithmes d'apprentissage.	95
4.6	La fenêtre glissante en rouge contient un nombre minimal de symboles. L'entropie est maximale pour ce nombre de symboles.	96
4.7	Évolution de l'entropie évaluée avec une fenêtre glissante normée (200 symboles). Les données d'apprentissage sont issues du modèle de synthèse en utilisant la distribution normale.	97
4.8	Mesure de l'entropie de Shannon des données issues du modèle de synthèse. Figure 4.8(a) pour la distribution uniforme et Figure 4.8(b) pour la distribution normale.	99
4.9	Critère de log-vraisemblance.	100

4.10	Ces graphes nous montrent que la topologie 2 est la plus pertinente au vu du critère d'AIC et de BIC, pour les 2 algorithmes d'apprentissage et pour les 2 distributions étudiées.	100
4.11	Nous testons les différentes architectures du modèle de synthèse à l'aide de tests statistiques d'adéquation : le test d'Aspin-Welch Figure 4.11(a) et celui de Kolmogorov-Smirnov Figure 4.11(b).	102
4.12	Le calcul de l'incertitude sur la moyenne est représentatif de l'erreur épistémique dans la phase de modélisation.	103
4.13	Modèle de Markov Caché, topologie 1.	104
4.14	Mesure de l'entropie de Shannon avec la topologie 1 comme référence.	105
4.15	Modèle de Markov Caché, topologie 3.	105
4.16	Mesure de l'entropie de Shannon avec la topologie 3 comme référence.	106
4.17	Modèles de Markov Cachés, topologies à 4 états	109
4.18	Exemple de dégradation d'un processus [178].	110
4.19	Entropies moyennes des modèles	121
4.20	Mesures de la vraisemblance.	122
4.21	Nombre minimal de données par entropie de Shannon, données issues d'une GMAO industrielle.	122
4.22	Évolution de l'entropie évaluée avec une fenêtre glissante normée (200 symboles). Les données d'apprentissage sont issues d'une GMAO industrielle.	123
4.23	Critères <i>AIC</i> , <i>BIC</i> et <i>HQC</i>	124
4.24	Modèles de synthèse non ajustés.	127
4.25	Comparaison du modèle de synthèse ajusté, avec les données de maintenance provenant d'une entreprise du secteur de l'agro-alimentaire.	128
4.26	Comparaison des topologies après réajustement des symboles.	129
B.1	Loi Normale.	149
B.2	Loi Uniforme.	150
B.3	Loi Exponentielle.	151

Liste des tableaux

1.1	Évolution des valeurs et des grands ratios de la maintenance. Source : l'Observatoire Réseau Maintenance [®] 2012.	10
1.2	Proportion d'utilisation des logiciels de « Gestion de la Maintenance Assistée par Ordinateur », dans les différents secteurs économiques.	16
1.3	Exemple d'une base de données utilisée dans une Gestion de la Maintenance Assistée par Ordinateur.	16
2.1	Caractère de sur-apprentissage de l'AIC (distribution Arc Sinus [188]). . . .	56
3.1	Codification symbolique des interventions de maintenance.	70
3.2	Séquence d'un message issue des données de maintenance.	77
3.3	p-value des états générée par le modèle de synthèse.	78
4.1	Mesure de l'entropie, données issues du modèle de synthèse, loi normale. . .	91
4.2	Mesure de l'entropie, données issues du modèle de synthèse, loi uniforme. . .	91
4.3	Entropie moyenne pour les algorithmes d'apprentissage et décodage.	98
4.4	Log-Vraisemblance.	99
4.5	Critère de BIC.	101
4.6	Résultats des incertitudes épistémiques.	104
4.7	Séquence d'un message issue des données de maintenance.	107
4.8	Exemple de consignation des événements.	107
4.9	Codification symbolique des interventions de maintenance.	108
4.10	Séquence de symboles / niveaux de dégradation.	109
4.11	Mesure de l'entropie, données empiriques, évaluation modèle MMC 2 décodage Viterbi.	111
4.12	Entropie pour les chaînes de Markov d'ordre 2, données empiriques, évaluation modèle MMC 2 décodage Viterbi.	111
4.13	Entropie pour les chaînes de Markov d'ordre 3, données empiriques, évaluation modèle MMC 2 décodage Viterbi.	112
4.14	Entropie pour les chaînes de Markov d'ordre 4, données empiriques, évaluation modèle MMC 2 décodage Viterbi.	112
4.15	Séquences de symboles (ordre 4) ayant une entropie maximale, après passage dans le filtre entropique.	113
4.16	Information mutuelle pour les chaînes de Markov d'ordre 1, données empiriques, évaluation modèle MMC 2 décodage Viterbi.	114
4.17	Information mutuelle pour les chaînes de Markov d'ordre 1 - Modèle MMC 2	115

4.18	Exemple d'arbre de décision pour une chaîne de Markov d'ordre 1, 2 et 3 / Corpus 2005–2006 Modèle MMC 1.	116
4.19	Exemple d'arbre de décision pour une chaîne de Markov d'ordre 1, 2 et 3 / Corpus 2005–2006 Modèle MMC 2	117
4.20	Exemple de construction d'un graphe de généralisation/spécialisation pour une chaîne de Markov d'ordre 1.	118
4.21	Exemple de construction d'un graphe de généralisation/spécialisation sur les états (chaîne de Markov d'ordre 1).	118
4.22	Entropie moyenne des modèles de Markov sur les 2 types d'échantillonnage. .	120
4.23	Résultats de l'entropie moyenne avec et sans filtre entropique.	120
4.24	Résultats de vraisemblance.	121
4.25	Résultats des différents critères.	124
4.26	Comparaison du modèle de synthèse, avec les données de maintenance prove- nant d'une GMAO d'un sous-système pour l'agro-alimentaire (année 2005-2006).126	
4.27	Comparaison des topologies après réajustement des symboles.	129
4.28	Résumé des résultats sur les différents critères de selection de modèles. . . .	130

Bernard ROBLES

Étude de la pertinence des paramètres stochastiques sur des Modèles de Markov Cachés.

Résumé :

Le point de départ de ce travail est la thèse réalisée par Pascal Vrignat sur la modélisation de niveaux de dégradation d'un système dynamique à l'aide de Modèles de Markov Cachés (MMC), pour une application en maintenance industrielle. Quatre niveaux ont été définis : S1 pour un arrêt de production et S2 à S4 pour des dégradations graduelles. Recueillant un certain nombre d'observations sur le terrain dans divers entreprises de la région, nous avons réalisé un modèle de synthèse à base de MMC afin de simuler les différents niveaux de dégradation d'un système réel. Dans un premier temps, nous identifions la pertinence des différentes observations ou symboles utilisés dans la modélisation d'un processus industriel. Nous introduisons ainsi le filtre entropique.

Ensuite, dans un but d'amélioration du modèle, nous essayons de répondre aux questions : Quel est l'échantillonnage le plus pertinent et combien de symboles sont ils nécessaires pour évaluer au mieux le modèle ?

Nous étudions ensuite les caractéristiques de plusieurs modélisations possibles d'un processus industriel afin d'en déduire la meilleure architecture. Nous utilisons des critères de test comme les critères de l'entropie de Shannon, d'Akaike ainsi que des tests statistiques. Enfin, nous confrontons les résultats issus du modèle de synthèse avec ceux issus d'applications industrielles. Nous proposons un réajustement du modèle pour être plus proche de la réalité de terrain.

Mots clés : Modèles de Markov Cachés, sélection de modèles, test statistique, algorithmes d'apprentissage et de décodage, entropie de Shannon, incertitudes de modélisation, maintenance prédictive.

Study of the relevance of stochastic parameters on Hidden Markov Models.

Summary :

As part of preventive maintenance, many companies are trying to improve the decision support of their experts. This thesis aims to assist our industrial partners in improving their maintenance operations (production of pastries, aluminum smelter and glass manufacturing plant). To model industrial processes, different topologies of Hidden Markov Models have been used, with a view to finding the best topology by studying the relevance of the model outputs (also called signatures). This thesis should make it possible to select a model framework (a framework includes : a topology, a learning & decoding algorithm and a distribution) by assessing the signature given by different synthetic models. To evaluate this « signature », the following widely-used criteria have been applied : Shannon Entropy, Maximum likelihood, Akaike Information Criterion, Bayesian Information Criterion and Statistical tests.

Keywords : Hidden Markov Models, model selection, statistical test, learning and decoding algorithms, Shannon entropy, uncertainties, predictive maintenance.



Laboratoire PRISME, ENSI 88 boulevard Lahitolle 18020
Bourges Cedex

