



HAL
open science

Contributions à la détection et au diagnostic de fautes dans les systèmes par réseaux Bayésiens

Mohamed Amine Atoui

► **To cite this version:**

Mohamed Amine Atoui. Contributions à la détection et au diagnostic de fautes dans les systèmes par réseaux Bayésiens. Automatique / Robotique. Université d'Angers, 2015. Français. NNT : 2015ANGE0003 . tel-01266666

HAL Id: tel-01266666

<https://theses.hal.science/tel-01266666>

Submitted on 3 Feb 2016

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

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

Thèse de Doctorat

**Mohamed Amine
ATOUI**

*Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université d'Angers
sous le label de l'Université de Nantes Angers Le Mans*

Discipline : Sciences de l'ingénieur

Laboratoire : Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS)

Soutenu le 29.09.2015

École doctorale : 503 (STIM)

Thèse n° : 77514

Contributions à la détection et au diagnostic de fautes dans les systèmes par réseaux Bayésiens

JURY

- Rapporteurs : **M. Didier THEILLIOL**, Professeur, Université de Lorraine
M. François PÉRÈS, Professeur, École Nationale d'Ingénieurs de Tarbes
- Examineurs : **M. Jean-Marc THIRIET**, Professeur, Université Joseph Fourier de Grenoble
M. Teodor TÍPLICA, Maître de conférences, HDR, Université d'Angers
- Directeur de thèse : **M. Abdessamad KOBİ**, Professeur, Université d'Angers
- Co-encadreur de thèse : **M. Sylvain VERRON**, Maître de conférences, Université d'Angers

Remerciements

Le travail présenté dans ce mémoire a été préparé au sein du Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS) de l'école d'ingénieur ISTIA, Université d'Angers.

Je tiens d'abord à exprimer ma gratitude à mon directeur de thèse, le Professeur Abdessamad Kobi, de m'avoir permis d'effectuer cette thèse. Son soutien, sa confiance et sa disponibilité m'ont permis de m'épanouir sereinement tout au long de mes travaux.

Mes remerciements vont également à mon co-encadreur Sylvain Verron pour sa gentillesse, sa disponibilité, sa patience et ses judicieux conseils durant cette thèse.

Merci à Monsieur Jean-Marc Thiriet, Professeur à l'Université Joseph Fourier de Grenoble, pour avoir accepté d'assurer la présidence du jury.

Je remercie Monsieur Didier Theilliol, Professeur à l'Université de Lorraine, ainsi que Monsieur François Pérès, Professeur à l'Université de Lorraine, d'avoir accepté de rapporter mon mémoire et pour l'intérêt qu'ils ont bien voulu porter à ce travail.

Je tiens également à remercier Monsieur Teodor Tiplica, Maître de conférence, HDR à l'Université d'Angers d'avoir accepté de prendre part au jury.

Mes remerciements s'adressent également à tous les doctorants et docteurs, membres ou anciens membres du LARIS, LASQUO et LISA (Une dédicace spéciale à Achraf, Alejandro, Fally, Julien, Khadim, Khanh, Rémy, Ulmiah) pour tous les bons moments.

J'adresse évidemment mes remerciements à mes parents, ma famille et mes amis qui m'ont soutenu et encouragé régulièrement.

Table des matières

Table des figures	7
1 Introduction générale	13
2 Réseaux Bayésiens	17
I Introduction	17
II Définition	19
III Propriétés des réseaux Bayésiens	26
IV Noeuds, relations et structures	31
IV.1 Relations entre noeuds	31
IV.2 Analyse discriminante : le cas général	34
IV.3 Classification Bayésienne : structures usuelles	37
IV.4 Classification Bayésienne : mélange de Gaussiennes	38
V Inférence	40
V.1 Notions sur la théorie des graphes	41
V.2 Algorithme de Lauritzen : première version	44
V.3 Algorithme de Lauritzen : seconde version	50
VI Conclusion	57
3 Surveillance des systèmes	59
I Introduction	59
II Méthodes à base de modèle	61
II.1 Génération des résidus	63
II.2 Détection	67
II.3 Isolation	68
III Méthodes data-driven	69
III.1 Détection	69
III.2 Diagnostic	79

III.3	Combinaison des méthodes à base de données et des méthodes à base de modèle	80
IV	Conclusion	82
4	Détection par réseau Bayésien	85
I	Contexte et objectifs	85
II	Détection par calculs des limites de surveillance	89
III	Schémas de détection	93
III.1	Schéma de détection par analyse en composantes principales sous réseau Bayésien	93
III.2	D'autres schémas de détection sous réseaux Bayésiens	103
IV	Applications	113
IV.1	Tennessee Eastman Process	114
IV.2	Hot Forming Process	118
V	Conclusion	120
5	Diagnostic par réseau Bayésien	121
I	Contexte et objectifs	121
II	Limites spécifiques de surveillance	124
II.1	Démonstration du calcul de LP	126
II.2	Détection de fautes	132
II.3	Détection et diagnostic simultanés de fautes	139
II.4	Diagnostic de fautes augmenté par un rejet de distance	144
III	Propositions de structures de réseau Bayésien	146
III.1	Modèles Gaussiens pour le diagnostic de fautes	147
III.2	Combinaison de méthodes pour le diagnostic de fautes sous réseau Bayésien	154
IV	Conclusion	164
A	Comparaison des résultats	169
	Bibliographie	191

Table des figures

2.1	La couverture de Markov	27
2.2	Noeud queue-à-queue : bloquant lorsque \mathbf{x}_3 est observée, et non bloquant lorsque on marginalise sur celle-ci	28
2.3	Noeuds tête-à-queue : le noeud \mathbf{x}_3 est bloquant dans les deux réseaux Bayésiens lorsqu'il est observé, et il ne l'est plus lorsque l'on marginalise sur lui sachant qu'il n'est pas observé.	29
2.4	Noeud tête-à-tête	30
2.5	La table de probabilités conditionnelles de \mathbf{D} étant donnés ses noeuds parents	32
2.6	La table de probabilités conditionnelles de \mathbf{x} étant donnés ses noeuds parents discrets	32
2.7	Un réseau Bayésien pour l'analyse discriminante (forme multivariée)	37
2.8	Un réseau Bayésien classifieur naïf	38
2.9	Un mélange d'analyse discriminante, où $\theta_{M_1 i_{pa(E)}}$ correspond à $p(M_1 i_{pa(E)})$	40
2.10	(a) : Réseau Bayésien et (b) :L'arbre de jonction lui correspondant	42
2.11	Graphe moral	43
2.12	Graphe triangulé dit graphe joint	44
4.1	Un réseau Bayésien classifieur pour la détection de fautes	90
4.2	Un exemple d'un réseau Bayésien pour la détection de fautes	90
4.3	Analyse en composantes principales sur réseau Bayésien, projection de \mathbf{x} dans l'espace principal, forme multivariée	95
4.4	Analyse en composantes principales sous un réseau Bayésien, projection de \mathbf{x} dans l'espace principal et résiduel, forme multivariée	95

4.5	Analyse en composantes principales sous un réseau Bayésien, forme univariée	98
4.6	Schéma de détection par analyse en composantes principales sous réseau Bayésien, forme multivariée	99
4.7	Schéma de détection par analyse en composantes principales sous réseau Bayésien : forme univariée	100
4.8	Réseau Bayésien Gaussien représentant un modèle linéaire Gaussien avec un noeud continu caché : forme multivariée	105
4.9	Schéma de détection par modèle linéaire Gaussien sous réseau Bayésien : forme multivariée	107
4.10	Un réseau Bayésien représentant une carte MEWMA	109
4.11	Un réseau Bayésien représentant un générateur de résidus : espace de parité, avec une fenêtre h donnée et où O_p^+ correspond à la ligne p dans la matrice O^+	111
4.12	Un réseau Bayésien représentant un générateur de résidus : espace de parité, où $h = 1$	112
4.13	Tennessee Eastman Process	114
4.14	La liste des variables mesurables en continu du TEP	116
4.15	La liste des variables mesurables échantillonnées du TEP	117
4.16	La liste des variables manipulées du TEP	117
4.17	La liste des différentes fautes impliquées dans le TEP	118
4.18	La liste des variables du HFP	119
4.19	Hot Forming Process	119
4.20	Un exemple illustratif lorsque une faute est apparue dans \mathbf{x}_1 (avec $ms=5$) avec $\zeta_{T_2}^F = 88.5\%$ et $\zeta_{SPE}^F = 47.9\%$. (a). Schéma de détection de fautes par analyse en composantes principales sous un réseau Bayésien lorsque toutes les variables sont disponibles. (b). Schéma de détection de fautes par analyse en composantes principales sous un réseau Bayésien lorsque la variable \mathbf{x}_2 est manquante.	120
5.1	Classification entre les classes de fautes	125
5.2	Classification entre les classes de fautes et la classe de fonctionnement normal : exemple de fausse alarme	126
5.3	Classification entre les classes de fautes et la classe de fonctionnement normal : exemple d'un rejet de distance	127

5.4	Un exemple d'un réseau Bayésien pour la détection et le diagnostic de fautes (forme multivariée)	128
5.5	Détection dans un réseau Bayésien ayant un noeud softmax	138
5.6	Un réseau Bayésien combinant une carte MEWMA et d'autres statistiques quadratiques	138
5.7	Détection et diagnostic classique	140
5.8	Détection et diagnostic de fautes intégrant la limite probabiliste de CFN	141
5.9	Description de l'ensemble de données	141
5.10	Matrice de confusion du réseau Bayésien classifieur	142
5.11	Matrice de confusion de réseau Bayésien classifieur avec une limite probabiliste	143
5.12	Erreurs de classification pour les différentes combinaisons du $CFN + 3$ fautes	143
5.13	Moyennes et écart-types des erreurs du classifieur	143
5.14	Détection et diagnostic de fautes augmentés par le rejet de distance	146
5.15	Un réseau Bayésien classifieur considérant le noeud observé \mathbf{x} comme étant fonction d'une variable Gaussienne cachée \mathbf{z} , et un bruit Gaussien ϵ pour chaque classe C_k	150
5.16	Un réseau Bayésien classifieur traitant le noeud observé \mathbf{x} et le noeud \mathbf{z}	153
5.17	Cadre proposé pour la surveillance des systèmes	155
5.18	Le réseau Bayésien à base de données pour le diagnostic de fautes	156
5.19	La table de probabilités conditionnelles de \mathbf{S}_m	157
5.20	La table de probabilités conditionnelles du noeud \mathbf{x}	157
5.21	Un exemple d'une matrice d'incidence	158
5.22	Le réseau Bayésien à base de modèle pour le diagnostic de fautes	158
5.23	La table de probabilités conditionnelles du noeud \mathbf{S}_r	159
5.24	La table de probabilités conditionnelles des noeuds \mathbf{F}_j	159
5.25	Heating water system	160
5.26	Matrice d'incidence du chauffe-eau	161
5.27	Les différentes couches d'inférence pour la combinaison de deux méthodes pour l'isolation de fautes	162
5.28	La table de probabilités conditionnelles du noeud $\mathbf{S}_{r\&m}$	162
5.29	Scénarios simulés	163

5.30	Matrices de confusion pour chaque approche proposée sur les données simulées à titre d'exemple	163
A.1	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe CFN (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TFAs	170
A.2	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_1 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	171
A.3	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_2 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	172
A.4	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_3 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	173
A.5	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_4 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	174
A.6	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_5 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	175
A.7	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_6 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	176
A.8	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_7 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	177
A.9	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_8 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	178
A.10	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_9 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	179
A.11	Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{10} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	180

A.12 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{11} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	181
A.13 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{12} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	182
A.14 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{13} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	183
A.15 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{14} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	184
A.16 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{15} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	185
A.17 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{16} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	186
A.18 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{17} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	187
A.19 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{18} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	188
A.20 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{19} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	189
A.21 Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{20} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs	190



1

Introduction générale

La nécessité permanente des entreprises et industries d'assurer des produits et/ou des services de qualité tout en garantissant la disponibilité de leurs outils et survivre face à la concurrence les ont conduits à employer de plus en plus de systèmes automatisés et/ou informatisés. De nos jours, ces systèmes sont devenus de plus en plus complexes et difficiles à gérer dû au nombre énorme de retour de données, d'interaction et d'interconnexion entre les différents composants.

Un dérèglement ou un dysfonctionnement dans le comportement nominal de ces systèmes peut conduire à des conséquences sérieuses pouvant être néfastes pour l'humain et l'environnement. De plus, essayer de remédier à ces problèmes peut s'avérer coûteux.

Afin d'éviter ces situations indésirables, on s'intéresse désormais à l'utilisation de méthodes de surveillance en imaginant l'intérêt qu'elles peuvent apporter à leurs utilisateurs en terme de sécurité, d'augmentation de productivité et de viabilité du système. Une méthode de surveillance à caractère universel n'a pas encore vu le jour. En effet, surveiller efficacement un système n'est pas une tâche aisée. Ceci dépend fortement de sa nature et de son domaine mais aussi de la qualité, de la quantité et du type d'information disponible le concernant.

De façon générale, deux cas de figures peuvent être distingués afin d'établir une stratégie de surveillance :

- la conception d’une méthode de surveillance pour un système déjà existant, une situation non évidente qu’on essaie de rendre de plus en plus rare de nos jours. Elle nécessite un effort d’adaptation afin de pouvoir surveiller correctement le système.
- la conception d’une méthode de surveillance pour un système qui est en cours de conception, une situation favorable à une surveillance optimale du système si une étude très approfondie est établie (analyse de sûreté de fonctionnement, placement des capteurs, etc.). Toutefois, étant donné les investissements temporels et financiers qu’elles représentent, ces études ne sont pas toujours approfondies suffisamment.

Les méthodes de surveillances utilisées ou proposées dans la littérature peuvent à leur tour être décomposées en deux familles : les méthodes à base de données et les méthodes à base de modèle. Les méthodes à base de modèle utilisent une représentation analytique du système alors que les méthodes à base de données ne se basent que sur les mesures prises sur le système ou leurs transformées. Ces deux familles de méthodes n’emploient donc pas des informations de même type. Ainsi, il serait judicieux de pouvoir associer ces méthodes, qu’elles soient ou non de la même famille. Ceci pourrait permettre d’améliorer la surveillance, d’augmenter le nombre de scénarios pris en compte et de bénéficier des avantages de chacune. L’objectif de cette thèse est d’apporter des contributions permettant de modéliser certaines de ces méthodes, afin de les associer et de les exploiter simultanément en se basant sur un seul outil : un Réseau Bayésien (RB).

Le premier chapitre présente les réseaux Bayésiens. Dans un premier temps, nous les définissons et présentons quelques-unes de leurs propriétés. Par la suite, nous présentons les différents noeuds pouvant composer un réseau Bayésien et qui seront utilisés pour la surveillance. Enfin, nous présentons deux algorithmes d’inférence utilisés pour faire du calcul dans les réseaux conditionnels Gaussiens, cas particulier de réseau Bayésien.

Le second chapitre est consacré à un état de l’art se voulant non-exhaustif des méthodes proposées dans le contexte de la détection et du diagnostic de fautes. Dans un premier temps, nous présentons la famille des méthodes à base de modèle et donnons quelques méthodes proposées dans la littérature pour la détection et l’isolation de fautes. Dans un second temps, nous faisons de même pour les méthodes statistiques ou à base de données.

Les troisième et quatrième chapitres sont consacrés aux contributions ap-

portées au domaine de la surveillance des systèmes par réseau Bayésien. Le chapitre 3 porte sur la détection de fautes. Nous commençons par présenter l'état de l'art de ce concept par réseau Bayésien, puis nous présentons nos propositions suivies par des cas d'applications. Le chapitre 4 présente quant à lui diverses contributions pour le diagnostic de fautes et la surveillance des systèmes. Tout d'abord, nous exposons les différents réseaux Bayésiens proposés dans la littérature permettant de faire du diagnostic de fautes et de la surveillance des systèmes. Ensuite, nous présentons une généralisation des statistiques quadratiques sous réseau Bayésien. Cette généralisation nous permet de faire de la détection et du diagnostic de façon simultanée sous réseaux Bayésiens tout en tenant compte d'un rejet de distance. Après la présentation de quelques réseaux Bayésiens pouvant être utilisés pour la surveillance des systèmes suivies par des cas d'applications, un cadre probabiliste est proposé permettant d'unifier des réseaux Bayésiens dédiés à la détection ou au diagnostic de fautes.

Enfin, dans le dernier chapitre, les conclusions et perspectives des travaux réalisés durant cette thèse sont illustrés.



2

Réseaux Bayésiens

I Introduction

L'utilisation des probabilités peut être d'un apport considérable pour la prise de décision et particulièrement pour assister l'ingénieur dans ses choix. Les probabilités peuvent être vues de deux façons différentes (Bishop, 2006). La première considère la probabilité comme étant le nombre de répétition d'un événement parmi d'autres (taux, fréquences). La seconde définit la probabilité comme une quantification révisable de l'incertitude (e.g. pour les événements se répétant rarement et ne pouvant donc pas être définis par un taux). L'utilisation de l'interprétation Bayésienne des probabilités permet l'emploi des deux définitions.

Prenons un exemple de deux variables continues non-indépendantes \mathbf{x} et \mathbf{z} , avec \mathbf{z} influant sur \mathbf{x} . Soit une observation x de \mathbf{x} . Étant donnée la non-indépendance de ces deux variables \mathbf{x} et \mathbf{z} , x agit sur \mathbf{z} et de ce fait \mathbf{z} doit être mise à jour. Cela se traduit, dans le domaine probabiliste, par le calcul de la probabilité a posteriori de \mathbf{z} étant donnée l'observation x . Ce calcul peut se faire en utilisant le théorème de Bayes :

$$p(\mathbf{z}|\mathbf{x} = x) = \frac{p(\mathbf{x} = x|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x} = x)} \quad (2.1)$$

où $p(\mathbf{x} = x|\mathbf{z})$ est la fonction de vraisemblance fonction de \mathbf{z} . Cette fonction peut être apprise, si elle n'est pas connue, à partir d'un ensemble d'apprentissage disponible (historique d'observations de \mathbf{x} et \mathbf{z}).

L'équation (2.1) permettant de calculer la probabilité a posteriori de \mathbf{z} , étant donnée une observation de \mathbf{x} , peut être écrite comme suit sans perte de généralité :

$$p(\mathbf{z}|\mathbf{x} = x) \propto p(\mathbf{x} = x|\mathbf{z})p(\mathbf{z}) \quad (2.2)$$

où \propto indique que $p(\mathbf{x} = x|\mathbf{z})p(\mathbf{z})$ est proportionnelle à $p(\mathbf{z}|\mathbf{x} = x)$. Cette écriture peut être expliquée par le fait que le terme $p(\mathbf{x} = x)$ dans l'équation (2.1) représente une constante de normalisation assurant la validité de la distribution a posteriori. En effet, celle-ci doit être une densité de probabilité et doit donc vérifier la condition suivante :

$$\int_{-\infty}^{+\infty} p(\mathbf{z}|\mathbf{x} = x)d\mathbf{z} = 1 \quad (2.3)$$

Un outil permettant de manipuler et de calculer, selon le théorème de Bayes, les probabilités de différentes variables étant données leurs nouvelles observations est le réseau Bayésien. Sous un réseau Bayésien, il est possible de gérer les deux règles primordiales pour la manipulation des probabilités, la règle des sommes et la règle du produit. Ces règles sont nécessaires pour tout calcul probabiliste, inférence ou apprentissage, impliquant des distributions de probabilités. De plus, les réseaux Bayésiens permettent une représentation graphique offrant :

- une vue simple de la structure et des propriétés du modèle probabiliste,
- une illustration visuelle des indépendances conditionnelles,
- un cadre pour les distributions discrètes et continues,
- des manipulations graphiques alternatives aux manipulations complexes nécessaires pour l'inférence et l'apprentissage dans des modèles probabilistes.

Dans ce qui suit, nous allons définir ce qu'est un réseau Bayésien, ses propriétés et les différents noeuds pouvant le composer. Ensuite, nous détaillerons deux algorithmes d'inférence exacte basés sur l'arbre de jonction et proposés dans le cadre des réseaux Bayésiens hybrides (réseaux Bayésiens composés de noeuds discrets et/ou Gaussiens).

II Définition

Un réseau Bayésien est une représentation graphique permettant de modéliser des relations incertaines entre variables décrivant l'influence qu'une variable peut avoir sur une autre. Un réseau Bayésien (Jensen and Nielsen, 2007) est composé de ce qui suit :

- un graphe acyclique dirigé \mathbf{G} , $\mathbf{G}=(V, E)$, où V est l'ensemble des sommets de \mathbf{G} (noeuds) pouvant être décomposé en deux sous-ensembles : Υ l'ensemble des noeuds discrets et Γ l'ensemble des noeuds continus. Ainsi, un noeud représente une variable aléatoire pouvant être discrète ou continue, univariée ou multivariée. E correspond à l'ensemble des arêtes (arcs) de \mathbf{G} . Un arc lie un noeud "père" à un noeud "enfant" expliquant ainsi la relation causale (relation de cause à effet) ou la régression pouvant exister entre eux deux.
- un espace probabiliste fini (Ω, \mathbb{Z}, p) , avec Ω un espace non-vide, \mathbb{Z} un sous espace de Ω et, p une mesure de probabilité dans \mathbb{Z} avec $p(\Omega) = 1$. Nous utilisons la même notation pour la distribution de probabilités (cas discret) et la fonction de densité de probabilité (cas continu). Il n'y aura pas d'ambiguïté selon le contexte.
- un ensemble de variables aléatoires $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_m$ associées aux sommets du graphe \mathbf{G} et définies dans (Ω, \mathbb{Z}, p) , de façon à capturer la décomposition de leur distribution de probabilité jointe sous forme d'un produit de distributions conditionnelles, comme suit :

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \prod_{i=1}^m p(\mathbf{x}_i | pa(\mathbf{x}_i)) \quad (2.4)$$

où $pa(\mathbf{x}_i)$ est l'ensemble des noeuds parents de \mathbf{x}_i dans le graphe \mathbf{G} . Notons qu'un réseau Bayésien caractérise une et une seule distribution jointe, alors que cette dernière peut être représentée par plusieurs réseaux Bayésiens.

- une Table de Probabilités Conditionnelles (TPC) associée à chaque noeud. Une table de probabilités conditionnelles définit et décrit pour chaque variable les relations causales/dépendances probabilistes que cette dernière peut avoir avec ses noeuds parents. Les probabilités ou distributions de probabilité associées à une table de probabilités conditionnelles peuvent être connues (à l'aide d'experts) ou estimées selon les

données disponibles en utilisant des méthodes statistiques d'échantillonnage (cas non-supervisés) ou des méthodes statistiques supervisées.

- des calculs nommés inférence, utilisés étant donnée la disponibilité d'une ou plusieurs nouvelle(s) observation(s) (évidences) concernant une ou plusieurs variable(s) de \mathbf{G} , afin de mettre à jour le réseau (e.g. calculer les probabilités a posteriori des noeuds non observés étant donnée(s) la ou les nouvelle(s) information(s) disponible(s)) ainsi que la distribution de probabilité jointe (et celle de chaque variable). Ces calculs sont également nécessaires dans le cas où l'on recherche uniquement (sans avoir d'évidence) la distribution d'une seule ou plusieurs variables de \mathbf{G} . Ceci correspond à ce que l'on appelle une opération de marginalisation. Elle correspond à sommer (cas des variables discrètes) ou intégrer (cas des variables continues) sur les variables de \mathbf{G} différentes des variables recherchées. Par exemple, soit \mathbf{x}_1 , \mathbf{x}_2 et \mathbf{x}_3 des variables continues. La distribution de probabilité marginale de \mathbf{x}_3 revient à calculer ceci :

$$p(\mathbf{x}_3) = \int p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) d(\mathbf{x}_1, \mathbf{x}_2) \quad (2.5)$$

Parmi les différentes distributions de probabilités continues uni/multivariées, nous nous intéressons et concentrons sur l'une des plus importantes et également la plus utilisée (Bishop, 2006) : la distribution de probabilité normale uni/multivariée. En effet, dans le cadre de la détection et du diagnostic de fautes dans les systèmes, nous considérerons tout au long du manuscrit que les observations ou mesures prises sur le système suivent une distribution Gaussienne (si ce n'est pas le cas, nous le mentionnerons). Cela est dû et se justifie par plusieurs de ses propriétés analytiques (Bishop, 2006; Ahrendt, 2005) ainsi qu'à nos perspectives envisagées. Les propriétés intéressantes sont les suivantes :

- une distribution Gaussienne d'une variable \mathbf{x} de dimension m est une distribution exponentielle symétrique définie comme ci-dessous :

$$\mathcal{N}(\mathbf{x}|\mu; \Sigma) = \frac{1}{2\pi^{\frac{m}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)}{2}} > 0 \quad (2.6)$$

avec, pour tout \mathbf{x} :

$$\int_{-\infty}^{+\infty} \mathcal{N}(\mathbf{x}|\mu; \Sigma) d\mathbf{x} = 1 \quad (2.7)$$

où la probabilité que \mathbf{x} appartienne à un intervalle $(-\infty, v)$ est donnée par la fonction de distribution cumulative définie comme suit :

$$p(v) = \int_{-\infty}^v \mathcal{N}(\mathbf{x}|\mu; \Sigma) d\mathbf{x} \quad (2.8)$$

- une distribution Gaussienne est gouvernée, que ce soit une distribution pour une variable \mathbf{x} multivariée ou univariée, par uniquement deux paramètres indépendants (μ : moyenne, Σ : matrice de variance-covariance (variance σ^2 dans le cas univariée, où σ est son écart-type)), correspondant à ses moments d'ordre 0 et 1. Ces paramètres peuvent être facilement estimés à partir d'un nombre suffisamment élevé de N observations x_i indépendantes et identiquement distribuées, en utilisant par exemple l'estimation par maximum de vraisemblance (voir (Duda et al., 2001)). Ceci correspond à calculer la fonction du logarithme de vraisemblance, donnée ci-dessous :

$$\ln(p(x_1, \dots, x_N|\mu; \Sigma)) = \sum_{i=1}^N \ln(\mathcal{N}(x_i|\mu; \Sigma)) \quad (2.9)$$

et d'en déduire les solutions la maximisant, permettant ainsi d'estimer la moyenne et la matrice de variance-covariance. Deux solutions indépendantes en découlent (dont l'une intervient dans le calcul de l'autre) présentées ci-dessous :

$$\mu_{MV} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i, \quad (2.10)$$

$$\Sigma_{MV} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_{MV})(x_i - \mu_{MV})^T \quad (2.11)$$

L'espérance de la première donne μ , alors que celle de la deuxième donne lieu à la matrice de variance-covariance biaisée, pouvant être corrigée en ajustant Σ_{MV} .

- une distribution Gaussienne peut être écrite sous la représentation stan-

dard fonction de ses moments (moyenne et matrice de variance-covariance) mais également sous la représentation canonique. Ces deux représentations sont équivalentes, à partir de l'une on peut en déduire l'autre. Cependant, comme on le verra par la suite, la transformation d'écriture peut engendrer une approximation voire même être impossible dans certains cas.

La représentation canonique d'une distribution de probabilités Gaussienne $p(\mathbf{x})$ d'une variable \mathbf{x} , de paramètres μ et Σ , peut être définie comme suit :

$$p(\mathbf{x}) = e^{g + \mathbf{x}^T h - \frac{1}{2} \mathbf{x}^T K \mathbf{x}} \quad (2.12)$$

où

$$g = \ln(\text{cst}) - \frac{1}{2} \mu^T \Sigma^{-1} \mu, \quad (2.13)$$

$$h = \Sigma^{-1} \mu, \quad (2.14)$$

$$K = \Sigma^{-1} \quad (2.15)$$

et cst est la constante de normalisation.

- une distribution Gaussienne peut s'écrire, étant donnée la matrice de variance-covariance Σ et sa décomposition en vecteurs propres, comme le produit de m distributions Gaussiennes univariées en projetant \mathbf{x} (variable multivariée de m dimensions) dans un nouvel espace \mathbf{z} , $\mathbf{z} = P^T(\mathbf{x} - \mu)$ comme ci-dessous :

$$p(\mathbf{z}) = \prod_{j=1}^m \frac{1}{2\pi(\lambda_j)^{1/2}} e^{-\frac{z_j^2}{2\lambda_j}}, \quad (2.16)$$

$$\Sigma = P^T \Lambda P, \quad P^T P = I \quad (2.17)$$

où λ_j est la $j^{\text{ème}}$ valeur propre de Σ , P est la matrice des vecteurs propres de Σ et I est la matrice identité.

- la somme de deux variables Gaussiennes indépendantes est une variable Gaussienne. Autrement dit, la convolution de deux Gaussiennes est une Gaussienne.
- une distribution Gaussienne est associée à une variable liée linéairement à une autre variable qui suit une loi de distribution Gaussienne.

Soit $\mathbf{z} = A\mathbf{x} + e$ avec \mathbf{x} une variable Gaussienne et e un bruit Gaussien, alors $p(\mathbf{z}) = \int_{\mathbf{x}} p(\mathbf{z}|\mathbf{x})p(\mathbf{x})$ suit une distribution Gaussienne. Cette propriété peut être primordiale pour les systèmes dynamiques linéaires Gaussiens respectant la propriété de Markov. Elle leur permet, étant donné l'état initial, de prédire et d'estimer les différents états et sorties passé(e)s ou futur(e)s.

- une distribution Gaussienne possède une forme géométrique dépendant d'une forme quadratique représentant la distance de Mahalanobis :

$$T^2 = (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \quad (2.18)$$

- une distribution Gaussienne jointe de deux ensembles de variables implique que la distribution conditionnelle de l'une conditionnée par l'autre, ainsi que leurs distributions marginales, sont également Gaussiennes.
- la multiplication de deux distributions Gaussiennes donne lieu à une distribution Gaussienne non normalisée (une Gaussienne multipliée par un facteur, voir [Ahrendt \(2005\)](#)).

Soit une variable multivariée \mathbf{x} , où pour chaque composante k ($k \in 1, \dots, K$), elle suit une distribution Gaussienne :

$$x_k \sim \mathcal{N}(\mu_k, \Sigma_k) \quad (2.19)$$

Le produit de ces Gaussiennes donne lieu à ceci :

$$\prod_{k=1}^K \mathcal{N}(\mu_k, \Sigma_k) = v \mathcal{N}(\mu, \Sigma) \quad (2.20)$$

où :

$$\mu = \Sigma \left(\sum_{k=1}^K \Sigma_k^{-1} \mu_k \right), \quad (2.21)$$

$$\Sigma = \sum_{k=1}^K (\Sigma_k^{-1})^{-1}, \quad (2.22)$$

$$v = \frac{1}{\prod_{k=1}^K 2\pi |\Sigma_k|^{1/2}} \prod_{i < j} e^{\frac{1}{2}(\mu_i - \mu_j)^T \Sigma_{ij} (\mu_i - \mu_j)}, \quad (2.23)$$

$$\Sigma_{ij} = \Sigma_i^{-1} \Sigma \Sigma_j^{-1} \quad (2.24)$$

— à partir de la forme quadratique faisant partie du terme de l'exponentielle de la distribution jointe d'une variable Gaussienne $\mathbf{x} = [\mathbf{x}_1 \ \mathbf{x}_2]^T$, avec une moyenne :

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad (2.25)$$

et une matrice de variance-covariance, partitionnée comme-çeci :

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Sigma_{12} = \Sigma_{21}^T \quad (2.26)$$

dont la matrice inverse $\Lambda = \Sigma^{-1}$, nommée matrice de précision, correspond à :

$$\Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \quad (2.27)$$

avec :

$$\Lambda_{11} = (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})^{-1}, \quad (2.28)$$

$$\Lambda_{12} = \Lambda_{11} \Sigma_{12} \Sigma_{22}^{-1}, \quad (2.29)$$

$$\Lambda_{21} = -\Sigma_{22}^{-1} \Sigma_{21} \Lambda_{11}, \quad (2.30)$$

$$\Lambda_{22} = \Sigma_{22}^{-1} + \Sigma_{22}^{-1} \Lambda_{12} \quad (2.31)$$

il est facile de déterminer les paramètres de la distribution conditionnelle Gaussienne $p(\mathbf{x}_1 | \mathbf{x}_2)$ de la variable \mathbf{x}_1 ou bien la distribution marginale $p(\mathbf{x}_2)$ de \mathbf{x}_2 en utilisant une technique dite "completing the square" utile

pour la manipulation des distributions Gaussiennes. Elle se base sur la décomposition suivante du terme quadratique :

$$\begin{aligned} -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= -\frac{1}{2}(\mathbf{x}_1-\boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_{11}(\mathbf{x}_1-\boldsymbol{\mu}_1) \\ &\quad -\frac{1}{2}(\mathbf{x}_1-\boldsymbol{\mu}_1)^T \boldsymbol{\Lambda}_{12}(\mathbf{x}_2-\boldsymbol{\mu}_2) -\frac{1}{2}(\mathbf{x}_2-\boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_{21}(\mathbf{x}_1-\boldsymbol{\mu}_1) \\ &\quad -\frac{1}{2}(\mathbf{x}_2-\boldsymbol{\mu}_2)^T \boldsymbol{\Lambda}_{22}(\mathbf{x}_2-\boldsymbol{\mu}_2) \end{aligned} \quad (2.32)$$

et regroupe les termes selon leur ordre comme ceci :

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = -\frac{1}{2}\mathbf{x}^T \boldsymbol{\Sigma}^{-1}\mathbf{x} + \mathbf{x}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \text{const} \quad (2.33)$$

où *const* est une constante indépendante de \mathbf{x} . Ensuite, elle vient considérer les termes impliquant une variable \mathbf{z} , $\mathbf{z} \in (\mathbf{x}_1, \mathbf{x}_2)$, et les réécrire étant donnée la formule suivante :

$$-\frac{1}{2}(\mathbf{z}^T A\mathbf{z} + b^T \mathbf{z}) = -\frac{1}{2}(\mathbf{z} - A^{-1}b)^T A(\mathbf{z} - A^{-1}b) - \frac{1}{2}b^T A^{-1}b \quad (2.34)$$

où A et b sont respectivement une matrice et un vecteur n'incluant pas la variable \mathbf{z} . Ceci afin de pouvoir marginaliser facilement sur \mathbf{z} (sachant les propriétés d'une densité de probabilité) et déduire ses paramètres.

- une variable représentant une somme de m autres variables aléatoires, étant donné le théorème central limite, suit une distribution de probabilité qui tend vers une distribution Gaussienne. En d'autres termes, plus le nombre m est grand, plus la distribution de leur somme suit une distribution Gaussienne. De plus, du fait qu'une seule Gaussienne permet d'approximer des distributions de probabilités, la somme de plusieurs d'entre elles le permet d'autant plus. En effet, dans le cas de distributions multimodales, l'utilisation d'une Gaussienne pose problème car elle ne contient qu'un seul maximum. Ainsi, l'utilisation de cette somme de Gaussiennes, dite mélange ou mixture de Gaussienne, peut être une solution à considérer.
- étant donnée sa forme exponentielle et sa souplesse analytique, la distribution Gaussienne permet, contrairement à d'autres distributions continues, de bénéficier d'algorithmes d'inférence exacte au sein de réseaux Bayésiens comportant des noeuds Gaussiens (réseaux conditionnels Gaussiens).

III Propriétés des réseaux Bayésiens

Les réseaux Bayésiens sont issus de l'hybridation de deux domaines différents : la théorie des graphes et la théorie des probabilités. Un réseau Bayésien est une représentation graphique d'un modèle probabiliste exposant les différentes relations que peuvent avoir les variables du modèle. Sa structure permet des calculs locaux de probabilités en utilisant toute l'information de la distribution jointe. Néanmoins, sur un réseau Bayésien on ne peut pas exprimer toutes les relations possibles. Cependant, il reste un cadre idéal et naturel pour représenter les relations causales et les hypothèses d'indépendance entre variables. En d'autres termes, un réseau Bayésien exprime et factorise la probabilité jointe de m variables en m indépendances conditionnelles.

Ces indépendances conditionnelles permettent de réduire les calculs nécessaires pour l'inférence et l'apprentissage d'un modèle probabiliste en réduisant tout simplement sa structure. Par exemple, une probabilité jointe de m variables s'écrivant en utilisant la règle de chaîne (general product rule) comme ceci :

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = p(\mathbf{x}_m | \mathbf{x}_{m-1}, \dots, \mathbf{x}_1) \dots p(\mathbf{x}_2 | \mathbf{x}_1) p(\mathbf{x}_1) \quad (2.35)$$

Cette équation peut être réduite en introduisant ou en définissant les indépendances conditionnelles entre ses variables. De plus, en illustrant ces indépendances sous un réseau Bayésien, il suffit de multiplier les distributions de probabilité conditionnelle de chaque variable en fonction de ses parents, pour réécrire la distribution jointe comme ceci :

$$p(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m) = \prod_{i=1}^m p(\mathbf{x}_i | pa(\mathbf{x}_i)) \quad (2.36)$$

Toutefois, le fait que la distribution conditionnelle de chaque variable dans cette écriture soit définie en fonction de ses parents ne signifie pas que d'autres variables n'influent pas sur elle. En d'autres termes, un noeud peut être influencé par d'autres noeuds que ses parents dans le réseau Bayésien. Ces noeuds respectent la condition de Markov qui stipule que chaque variable est isolée par un sous-ensemble de variables de l'ensemble V appelé couverture de Markov. Une variable est donc conditionnellement indépendante des autres variables en dehors de sa couverture. Celle-ci inclut ses noeuds parents, enfants et les co-

parents de ses enfants (e.g. figure 2.1). Ses noeuds, une fois observés, bloquent le noeud en question des autres noeuds hors du périmètre.

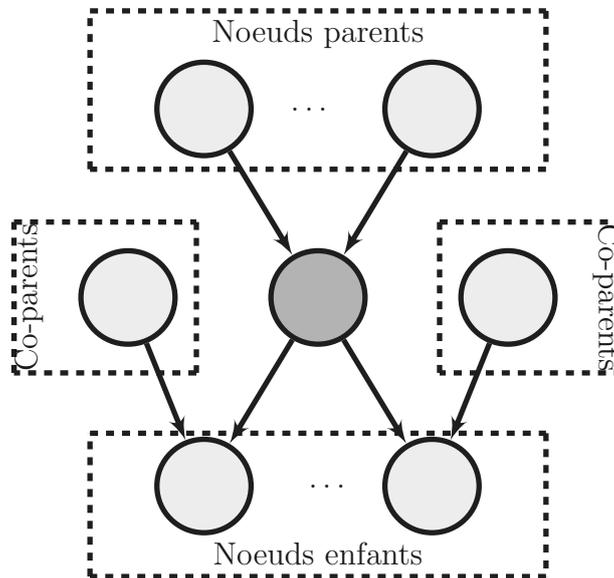


FIGURE 2.1 – La couverture de Markov

Cette notion de noeuds bloquants est une façon de décrire ou de reconnaître dans un réseau Bayésien qu’un ensemble de variables est conditionnellement indépendant étant donnée(s) une ou plusieurs autres variables. Cette propriété qu’offrent les réseaux Bayésiens est utile et primordiale pour tout calcul d’inférence, mais de plus elle permet de déterminer visuellement et immédiatement si un ensemble de variables est conditionnellement indépendant d’un autre ensemble de variables (Barber, 2012). Elle permet ainsi d’éviter le recours à des tests répétitifs impliquant la sommation et la multiplication de probabilités, et une perte de temps pour déterminer la validité ou non de la décomposition de la distribution jointe. Des algorithmes (Nielsen and Jensen, 2009; Jensen and Nielsen, 2007) se basant sur cette notion de noeuds bloquants et chemins bloqués ont été mis en place afin de permettre d’interpréter directement à partir du graphe l’indépendance conditionnelle d’une variable sur une ou plusieurs autres variable(s).

Soit une distribution jointe $p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ de trois variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$. Cette distribution jointe peut être factorisée de différentes façons, donnant ainsi lieu à différents graphes acycliques dirigés. Ces graphes peuvent être classés selon les directions des flèches présentes sur le noeud correspondant à la variable \mathbf{x}_3 . Ainsi, on peut distinguer trois réseaux Bayésiens ayant des chemins différents.

Le premier est un réseau Bayésien avec un chemin contenant un noeud avec une connexion divergente (noeud queue-à-queue), comme montré dans la figure 2.2.

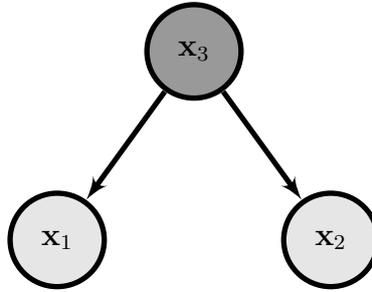


FIGURE 2.2 – Noeud queue-à-queue : bloquant lorsque \mathbf{x}_3 est observée, et non bloquant lorsque on marginalise sur celle-ci

La distribution lui correspondant est donnée ci-dessous :

$$p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = p(\mathbf{x}_1|\mathbf{x}_3)p(\mathbf{x}_2|\mathbf{x}_3)p(\mathbf{x}_3) \quad (2.37)$$

Lorsque \mathbf{x}_3 est observée, il est facile de constater que les variables \mathbf{x}_1 et \mathbf{x}_2 sont conditionnellement indépendantes. Cela peut s'écrire comme ceci :

$$p(\mathbf{x}_1, \mathbf{x}_2|\mathbf{x}_3) = p(\mathbf{x}_1|\mathbf{x}_3)p(\mathbf{x}_2|\mathbf{x}_3) \quad (2.38)$$

Ainsi, on peut dire que \mathbf{x}_3 bloque le chemin entre \mathbf{x}_1 et \mathbf{x}_2 . Cependant, lorsqu'aucune des variables n'est observée, la marginalisation de $p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ sur \mathbf{x}_3 , ne permet généralement pas d'obtenir deux termes indépendants.

Deux autres réseaux Bayésiens donnant lieu à deux distributions équivalentes à celle du premier graphe sont ceux avec un noeud \mathbf{x}_3 présentant une connexion en série (tête-à-queue), comme montré sur la figure 2.3. Leurs distributions jointes sont équivalentes et peuvent s'écrire comme ci-dessous :

$$p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = p(\mathbf{x}_2)p(\mathbf{x}_3|\mathbf{x}_2)p(\mathbf{x}_1|\mathbf{x}_3) \quad (2.39)$$

$$= p(\mathbf{x}_1)p(\mathbf{x}_3|\mathbf{x}_1)p(\mathbf{x}_2|\mathbf{x}_3) \quad (2.40)$$

Lorsqu'aucune des variables les composant sont observées alors les variables \mathbf{x}_1 et \mathbf{x}_2 ne sont pas conditionnellement indépendantes étant donnée la variable

\mathbf{x}_3 , comme on peut le voir ci-dessous :

$$p(\mathbf{x}_1, \mathbf{x}_2) = \sum_{\mathbf{x}_3} p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = p(\mathbf{x}_2)p(\mathbf{x}_1|\mathbf{x}_2) \quad (2.41)$$

$$= p(\mathbf{x}_1)p(\mathbf{x}_2|\mathbf{x}_1) \quad (2.42)$$

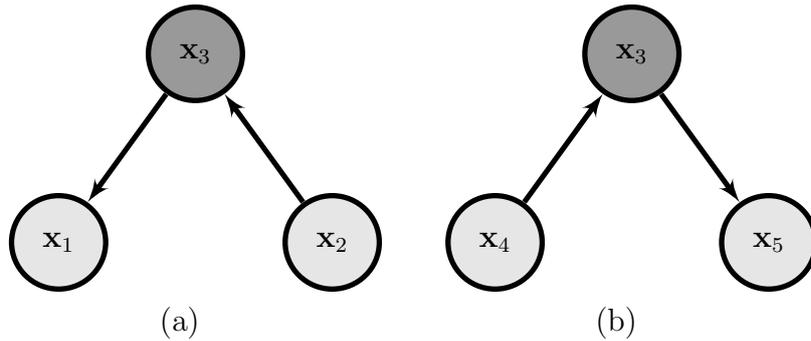


FIGURE 2.3 – Noeuds tête-à-queue : le noeud \mathbf{x}_3 est bloquant dans les deux réseaux Bayésiens lorsqu’il est observé, et il ne l’est plus lorsque l’on marginalise sur lui sachant qu’il n’est pas observé.

Finalement, le dernier réseau Bayésien donnant lieu à une autre factorisation de la distribution jointe inclue un chemin avec un noeud \mathbf{x}_3 convergeant (tête-à-tête), tel qu’illustré sur la figure 2.4.

La distribution jointe lui correspondant est donnée ci-dessous :

$$p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = p(\mathbf{x}_1)p(\mathbf{x}_2)p(\mathbf{x}_3|\mathbf{x}_1, \mathbf{x}_2) \quad (2.43)$$

Dans ce cas de figure, lorsqu’aucune des variables n’est observée, la marginalisation de la probabilité jointe en fonction de \mathbf{x}_3 permet aux deux variables \mathbf{x}_1 et \mathbf{x}_2 d’être indépendantes, $p(\mathbf{x}_1, \mathbf{x}_2) = p(\mathbf{x}_1)p(\mathbf{x}_2)$. Cependant, si les variables \mathbf{x}_1 et \mathbf{x}_2 sont conditionnées par \mathbf{x}_3 (\mathbf{x}_3 est observée) alors elles sont dépendantes graphiquement. En d’autres termes, \mathbf{x}_3 ne bloque pas le chemin entre \mathbf{x}_1 et \mathbf{x}_2 et ne permet donc pas à \mathbf{x}_1 et \mathbf{x}_2 d’être indépendants. Il faut noter que si le noeud \mathbf{x}_3 possède des noeuds enfants, alors observer ceux-ci n’induit pas une indépendance entre \mathbf{x}_1 et \mathbf{x}_2 . Toutefois, dans le cas où \mathbf{x}_3 possède des noeuds enfants observés \mathbf{x}_1 et \mathbf{x}_2 se retrouvent conditionnellement indépendants.

Les différents exemples vus précédemment nous permettent de voir les effets engendrés par une marginalisation ou un conditionnement sur une variable étant donnée la structure d’un modèle probabiliste et son équivalent en réseau

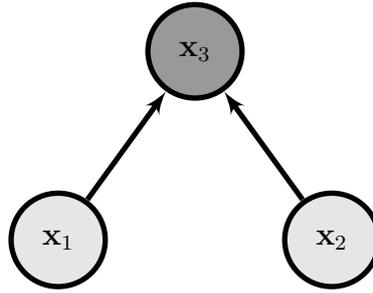


FIGURE 2.4 – Noeud tête-à-tête

Bayésien. Pour pouvoir généraliser ces conclusions sur d'autres structures de réseaux Bayésiens plus complexes et en considérant cette fois des ensembles de variables, un concept essentiel est introduit, nommé la séparation directe (D-separation), permettant de déterminer les indépendances conditionnelles entre ces ensembles. Ce concept peut être défini comme ceci :

Soit A, B, C trois ensembles de variables faisant partie d'un réseau Bayésien, $A, B, C \subseteq V$, définis de la sorte

$$A \cap B = \emptyset, \quad (2.44)$$

$$B \cap C = \emptyset, \quad (2.45)$$

$$A \cap C = \emptyset \quad (2.46)$$

On peut dire que les noeuds de A et B sont conditionnellement indépendants de C (A est directement séparé de B étant donné C) uniquement si tous les chemins possibles d'un noeud de A à un noeud de B passant par un noeud de C sont bloquant. Un chemin est dit bloquant si celui-ci comporte un noeud respectant au moins l'une des deux conditions suivantes :

- un noeud appartenant à l'ensemble C avec une connexion en série ou une connexion divergente,
- un noeud ayant une connexion convergente et ne faisant pas parti (ni lui, ni ses noeuds enfants) de l'ensemble C .

Une autre propriété intéressante des réseaux Bayésiens est leur capacité à intégrer la notion de temps. Cette notion est exprimée par le biais d'arcs dits temporels. Ces arcs viennent lier deux noeuds appartenant à deux couches successives selon la propriété de Markov (le futur est conditionnellement indépendant du passé étant donné le présent). Ces couches représentent chacune un réseau Bayésien. Chaque réseau représente un instant du temps. L'ensemble

de ces réseaux liés par des arcs temporels représente un intervalle borné dans le temps. Cet ensemble est dit réseau Bayésien dynamique. Plusieurs algorithmes de calculs dans ce type de réseau ont été proposés dans la littérature (voir (Murphy, 2002)).

IV Noeuds, relations et structures

+ Dans le cadre de ces travaux de recherche, nous nous intéressons à des systèmes impliquant des variables continues. Afin de pouvoir les manipuler dans le contexte des réseaux Bayésiens, nous allons utiliser le réseau Conditionnel Gaussien (RCG). Le réseau conditionnel Gaussien est une forme particulière de réseau Bayésien. Dans ce type de réseau, chaque noeud représente une variable aléatoire pouvant être discrète (suivant une loi de probabilité discrète) ou continue (univariée/ multivariée, sous l'hypothèse Gaussienne). Les différents noeuds peuvent être reliés entre eux, cependant pour assurer des calculs exacts (voir (Cowell, 2005; Lauritzen and Jensen, 2001; Lauritzen, 1992)), les noeuds discrets ne sont pas autorisés à avoir des noeuds continus comme parents. Ainsi, chaque noeud Gaussien, étant donné ses parents Gaussiens, suit un modèle Gaussien de régression linéaire (une combinaison linéaire des observations de ses parents continus) dont les paramètres dépendent également de ses parents discrets. Contrairement aux noeuds discrets, les noeuds Gaussiens peuvent avoir des noeuds Gaussiens et/ou discrets comme parents. Nous allons détailler plus précisément toutes ces relations. Par la suite, les noeuds continus seront représentés graphiquement par des cercles, alors que les noeuds discrets seront représentés par des carrés.

IV.1 Relations entre noeuds

Noeud discret avec parents discrets

Considérons un noeud discret \mathbf{D} avec K valeurs $k, k = 1, \dots, K$, et d parents $pa(\mathbf{D})$ où, par exemple, chacun d'entre eux prend aussi K valeurs (soit K^d combinaisons différentes). Chaque noeud discret, étant donné ses parents, suit une loi de distribution généralement représentée sous une table de probabilités conditionnelles. La table de probabilités conditionnelles de \mathbf{D} est présentée dans la figure 2.5, où $p(k|k_{pa(\mathbf{D})})$ est la probabilité de la valeur k correspondant à la $k_{pa(\mathbf{D})}^{ème}$ valeur de ses parents. Cette table augmente exponentiellement étant

donnés le nombre de variables aléatoires prises en compte et le nombre de leurs valeurs.

$k_{pa(\mathbf{D})}$	\mathbf{D}		
	1	\dots	K
1	$p(k = 1 k_{pa(\mathbf{D})} = 1)$	\dots	$p(k = K k_{pa(\mathbf{D})} = 1)$
\vdots	\vdots	\vdots	\vdots
K	$p(k = 1 k_{pa(\mathbf{D})} = K)$	\dots	$p(k = K k_{pa(\mathbf{D})} = K)$
\vdots	\vdots	\vdots	\vdots
K^d	$p(k = 1 k_{pa(\mathbf{D})} = K^d)$	\dots	$p(k = K k_{pa(\mathbf{D})} = K^d)$

FIGURE 2.5 – La table de probabilités conditionnelles de \mathbf{D} étant donnés ses noeuds parents

Noeud Gaussien avec parents discrets

Soit un noeud Gaussien \mathbf{x} ayant seulement des noeuds discrets comme parents, $pa(\mathbf{x})$, où nous assumons à titre d'exemple que chaque parent possède K valeurs. Ce noeud est linéaire et Gaussien pour chaque valeur $k_{pa(\mathbf{x})}$ de ses d parents $pa(\mathbf{x})$. Sa distribution conditionnelle peut s'écrire :

$$p(\mathbf{x} | k_{pa(\mathbf{x})}) = \mathcal{N}(\mu_{k_{pa(\mathbf{x})}}; \Sigma_{k_{pa(\mathbf{x})}}), \quad k_{pa(\mathbf{x})} \in I_{pa(\mathbf{x})} \quad (2.47)$$

où $\mu_{k_{pa(\mathbf{x})}}$ et $\Sigma_{k_{pa(\mathbf{x})}}$ sont respectivement la moyenne de \mathbf{x} et sa matrice de variance-covariance pour chaque valeur $k_{pa(\mathbf{x})}$ de l'ensemble des configurations possibles $I_{pa(\mathbf{x})}$ de ses d parents. Cette distribution pour chaque valeur des parents de \mathbf{x} peut être représentée par une table de probabilités conditionnelles comme le montre la figure 2.6.

$i_{pa(\mathbf{x})}$	\mathbf{x}
1	$\mathbf{x} \sim \mathcal{N}(\mu_{k_{pa(\mathbf{x})}=1}; \Sigma_{k_{pa(\mathbf{x})}=1})$
\vdots	\vdots
K	$\mathbf{x} \sim \mathcal{N}(\mu_{k_{pa(\mathbf{x})}=K}; \Sigma_{k_{pa(\mathbf{x})}=K})$
\vdots	\vdots
K^d	$\mathbf{x} \sim \mathcal{N}(\mu_{k_{pa(\mathbf{x})}=K^d}; \Sigma_{k_{pa(\mathbf{x})}=K^d})$

FIGURE 2.6 – La table de probabilités conditionnelles de \mathbf{x} étant donnés ses noeuds parents discrets

Noeud Gaussien avec parents Gaussiens

Soit un noeud Gaussien \mathbf{x} avec c noeuds Gaussiens Φ_1, \dots, Φ_c comme parents. Ce type de noeud est appelé noeud linéaire Gaussien. Sa distribution conditionnelle est donnée par :

$$p(\mathbf{x}|\Phi_1, \dots, \Phi_c) = \mathcal{N}(\mu^{\mathbf{x}} + W_1\Phi_1 + \dots + W_c\Phi_c; \Sigma^{\mathbf{x}}) \quad (2.48)$$

où $\mu^{\mathbf{x}}$ est un paramètre qui gouverne la moyenne de \mathbf{x} , $\Sigma^{\mathbf{x}}$ est la matrice de variance-covariance de \mathbf{x} , et W_1, \dots, W_c sont les coefficients de régression.

La distribution de probabilité jointe $p(\mathbf{x}, pa(\mathbf{x}))$ est aussi Gaussienne. Dans le cas où $\Sigma^{\mathbf{x}}$ est nulle alors (2.48) représente une relation linéaire déterministe entre \mathbf{x} et ses parents.

Noeud Gaussien avec parents Gaussiens et discrets

Soit un noeud Gaussien \mathbf{x} avec c noeuds parents Gaussiens $\Gamma_{\mathbf{x}} = pa(\mathbf{x}) \cap \Gamma$, $\Gamma_{\mathbf{x}} = \Phi_1, \dots, \Phi_c$, et d autres discrets $\Upsilon_{\mathbf{x}} = pa(\mathbf{x}) \cap \Upsilon$. Ce type de noeud est appelé noeud linéaire conditionnel Gaussien. Sa distribution est donnée par :

$$p(\mathbf{x}|pa(\mathbf{x})) = \mathcal{N}(\mu_{k_{\Upsilon_{\mathbf{x}}}}^{\mathbf{x}} + W_{k_{\Upsilon_{\mathbf{x}}}}^1 \Phi_1 + \dots + W_{k_{\Upsilon_{\mathbf{x}}}}^c \Phi_c; \Sigma_{k_{\Upsilon_{\mathbf{x}}}}^{\mathbf{x}}), \quad (2.49)$$

$$k_{\Upsilon_{\mathbf{x}}} \in [1, \dots, K^d]$$

où $k_{\Upsilon_{\mathbf{x}}}$ est une configuration parmi les K^d configurations possibles que peuvent avoir les parents discrets de \mathbf{x} . $\mu_{k_{\Upsilon_{\mathbf{x}}}}^{\mathbf{x}}$ est un paramètre qui gouverne la moyenne de \mathbf{x} , $\Sigma_{k_{\Upsilon_{\mathbf{x}}}}^{\mathbf{x}}$ est la matrice de variance-covariance de \mathbf{x} , et $W_{k_{\Upsilon_{\mathbf{x}}}}^1, \dots, W_{k_{\Upsilon_{\mathbf{x}}}}^c$ sont les coefficients de régression.

La distribution de probabilité jointe $p(\mathbf{x}, pa(\mathbf{x}))$ correspond à une somme (mixture) de Gaussiennes et peut s'écrire comme ceci :

$$p(\mathbf{x}) = \sum_{k_{\Upsilon_{\mathbf{x}}=1}^{K^d} p(k_{pa(\mathbf{x})})p(\mathbf{x}, \Gamma_{\mathbf{x}}|k_{\Upsilon_{\mathbf{x}}}) \quad (2.50)$$

Noeud discret avec parents Gaussiens

Le noeud discret avec parents Gaussiens est rarement utilisé dans la littérature. D'ailleurs, il n'est pas pris en compte par la plupart des algorithmes d'inférence et non implémenté dans la plupart des logiciels et boîtes à outils dédiés aux réseaux Bayésiens (e.g. BNT, Hugin, etc.). Par exemple dans (Lauritzen and Jensen, 2001; Lauritzen, 1992), deux algorithmes fréquemment utilisés pour les calculs dans les réseaux Bayésiens hybrides, un noeud Gaussien est assumé n'avoir que des enfants Gaussiens. Toutefois, ces deux algorithmes ont été étendus pour tenir compte du fait qu'un noeud Gaussien peut être parent d'un noeud discret. Dans (Murphy, 1998), un algorithme non exact basé sur un rapprochement variationnel a été proposé. Quelques années plus tard, (Lerner et al., 2001) suggère un algorithme exact gérant ces noeuds discrets (également appelés noeuds softmax) dépendant d'un ou plusieurs noeuds continus.

L'utilisation de ces différents types de noeuds dans un réseau Bayésien pourrait permettre, par exemple, de modéliser des techniques statistiques permettant la détection et le diagnostic de fautes. Une méthode classique parmi ces techniques est l'analyse discriminante, ayant pour but de réaliser de la classification supervisée.

IV.2 Analyse discriminante : le cas général

Afin de modéliser ou de prédire la nature des données, des outils statistiques sont utilisés. Parmi ces outils, on peut distinguer les méthodes de classification. Ces méthodes peuvent être discriminées de façon générale en deux classes : les méthodes supervisées et les méthodes non-supervisées. Les méthodes non-supervisées manipulent un ensemble de données brut sans l'aide d'aucun a priori. Ces méthodes, dites méthodes de clustering, essayent de discriminer cet ensemble en K classes (non connues a priori). Les méthodes de classification supervisées, contrairement au cas non-supervisé, disposent d'un ensemble d'apprentissage complet, c'est à dire qu'une classe (parmi les différentes K classes connues) est attribuée à chaque observation appartenant à l'ensemble d'apprentissage. Parmi les outils pouvant être utilisés pour la classification nous pouvons citer : KPPV : K-Plus Proche Voisin (Cover and Hart, 1967), RNA : Réseaux de Neurones Artificiels (Zhang, 2000), SVM : Séparateurs à Vaste Marge (Byun and Lee, 2002), etc.

Les réseaux Bayésiens peuvent être également utilisés pour la classification supervisée de données multidimensionnelles. Ces réseaux sont dits réseaux Bayésiens classifieurs (Friedman et al., 1997). L'ensemble de leurs noeuds V inclut toujours un noeud discret indexant les différentes classes. Certaines structures prises par ces réseaux donnent lieu à une Analyse Discriminante (AD) ou à une mixture d'analyse discriminante.

L'analyse discriminante (McLachlan, 2004; Fukunaga, 1990) est une technique statistique supervisée (un apprentissage supervisé pour la classification) utilisée pour prédire la classe d'une nouvelle observation en discriminant entre $m, m > 1$, classes ou ensembles (voir (Duda et al., 2001)), généralement distribué(e)s sous l'hypothèse de normalité.

Soit x un nouveau vecteur d'observation indépendant et identiquement distribué (i.i.d.) d'une variable continue $\mathbf{x} \in \mathbb{R}^m$, et un ensemble de données d'apprentissage où chaque observation de \mathbf{x} est associée à sa classe d'appartenance (une des valeurs ou états de la variable discrète \mathbf{D} ayant comme valeurs $C_k, k \in \{1, \dots, K\}$). Contrairement à d'autres méthodes (multi-class logistic régression, SVM, RN, etc.) modélisant directement la relation entre les deux variables d'entrée (\mathbf{x}) et de sortie (la classe d'appartenance), l'analyse discriminante (Jebara, 2001) calcule la probabilité a posteriori (voir équation (2.52)) donnée par $p(C_k|x)$ après avoir défini la fonction de densité conditionnelle et ses paramètres ainsi que la probabilité a priori de \mathbf{D} , $p(\mathbf{D} = C_k)$, de chaque classe.

$$p(\mathbf{D} = C_k|x) \propto p(\mathbf{D} = C_k)p(x|\mathbf{D} = C_k) \quad (2.51)$$

$$= \frac{p(\mathbf{D} = C_k)p(x|\mathbf{D} = C_k)}{p(x)}, \quad p(x) > 0 \quad (2.52)$$

où $p(x) = \sum_{k=1}^K p(\mathbf{D} = C_k)p(x|\mathbf{D} = C_k)$ est un facteur de normalisation pouvant être omis comme mentionné précédemment (voir équation (2.2)).

Selon l'hypothèse émise (distribution considérée) et la manière dont les données se présentent, la probabilité a priori de la classe C_k peut être définie de plusieurs façons. Cependant, la manière la plus répandue et naturelle est de prendre en compte la proportion de chaque classe dans l'ensemble d'apprentissage. Ainsi, on peut écrire :

$$p(\mathbf{D} = C_k) = \frac{N_{C_k}}{\sum_{j=1}^K N_{C_j}} \quad (2.53)$$

où N_{C_k} est le nombre d'observations appartenant à la classe C_k .

La densité de probabilité Gaussienne multivariée de x étant donnée la classe C_k , sous l'hypothèse de normalité, peut être calculée comme ceci :

$$p(x|\mathbf{D} = C_k) = \frac{1}{2\pi^{\frac{m}{2}} |\Sigma_{C_k}|^{\frac{1}{2}}} \mathbf{e}^{-\frac{(x-\mu_{C_k})^T \Sigma_{C_k}^{-1} (x-\mu_{C_k})}{2}} \quad (2.54)$$

avec μ_{C_k} le vecteur de moyennes de la classe C_k , et Σ_{C_k} la matrice de variance-covariance de la classe C_k . Ces deux paramètres représentent les paramètres de la fonction de vraisemblance de la classe C_k et doivent être estimés s'ils ne sont pas définis. Pour cela, on pourra utiliser l'estimation par maximum de vraisemblance (voir équations (2.10) et (2.11)).

Une fois les paramètres estimés et la probabilité a priori de chaque classe définie, l'analyse discriminante affecte x à la classe C_k ayant la plus grande probabilité a posteriori $p(C_k|x)$ comme ci-dessous :

$$\delta : x \in C_{k^*}, \text{ si } k^* = \underset{k=1, \dots, K}{\operatorname{argmax}} p(\mathbf{D} = C_k) p(x|\mathbf{D} = C_k) \quad (2.55)$$

La règle du maximum a posteriori donnée dans (2.55) représente une Analyse Discriminante Quadratique (ADQ). Ceci est dû au fait que la fonction de décision entre les deux classes est quadratique en x , et de ce fait les limites séparant les classes le sont aussi. A partir de ce cas général, plusieurs autres règles de discrimination peuvent être élaborées en émettant d'autres hypothèses sur la matrice de variance-covariance associée à chaque classe. Par exemple, en considérant une matrice de variance-covariance commune pour les différentes classes C_k , la fonction de décision entre deux classes devient linéaire en x décrivant alors des frontières linéaires entre les classes. Cette règle de décision est nommée analyse discriminante linéaire.

Une analyse discriminante quadratique peut être réalisée sous un réseau conditionnel Gaussien comme celui présenté dans la figure (2.7). Ce réseau conditionnel Gaussien est composé d'un noeud racine discret \mathbf{D} (représentant les K classes), et d'un noeud Gaussien multivarié $\mathbf{x} \in \mathbb{R}^m$ (prenant en compte les corrélations pouvant exister entre les m variables).

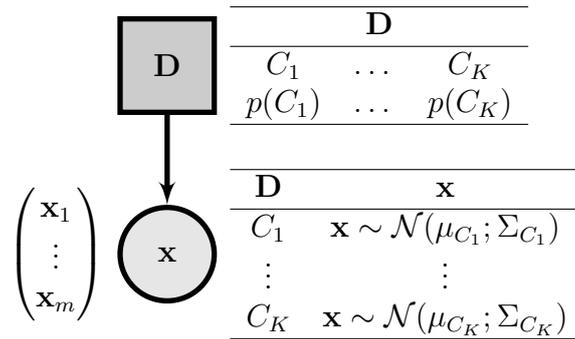


FIGURE 2.7 – Un réseau Bayésien pour l’analyse discriminante (forme multi-variée)

IV.3 Classification Bayésienne : structures usuelles

Plusieurs structures de réseaux Bayésiens peuvent être déclinées pour résoudre un problème de classification (Nielsen and Jensen, 2009; Friedman et al., 1997). Nous allons présenter un bref aperçu de ces différentes structures.

Réseau Bayésien Naïf

Une première structure est le Réseau Bayésien Naïf (RBN, voir figure 2.8). Il fait l’hypothèse, très forte, que les variables observées (noeuds enfants du noeud racine discret de décision \mathbf{D}) sont conditionnellement indépendantes étant donnée \mathbf{D} . En d’autres termes, les noeuds observés (descripteurs) ne sont pas interconnectés. Malgré cette hypothèse simpliste, cet algorithme est très utilisé : il est performant malgré le peu de paramètres à estimer. D’ailleurs, dans le cas où toutes les variables observées impliquées sont discrètes, il peut surpasser certaines méthodes plus complexes. L’hypothèse émise sur ce réseau peut être partiellement ou complètement relaxée via le biais des réseaux Bayésiens présentés par la suite.

Réseaux Bayésiens Naïfs Augmentés

D’autres types de structures peuvent être envisagées et ont été proposées afin d’alléger l’hypothèse d’indépendance des variables observées permettant alors une amélioration de la performance du classifieur. Une de ces structures est le TAN (Tree Augmented Naïve Bayes). Cela correspond à un réseau Bayésien naïf dont les noeuds observés sont liés sous forme d’un arbre. Ces relations explicites entre noeuds peuvent être établies par un expert ou apprises à partir

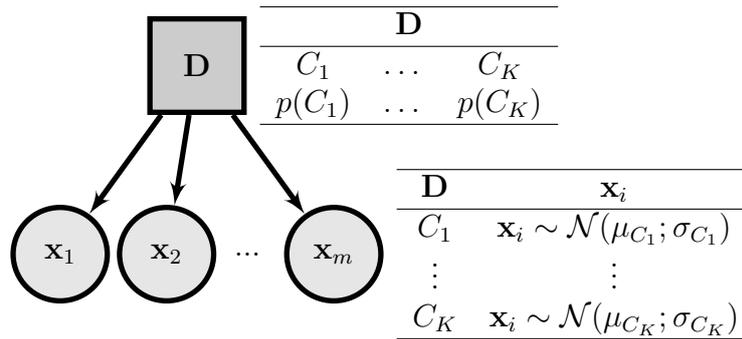


FIGURE 2.8 – Un réseau Bayésien classifieur naïf

des données disponibles pour un problème donné.

Réseau Bayésien Naïf Semi-Condensé

Les structures de réseau Bayésien vues précédemment peuvent être généralisées sous un Réseau Bayésien Naïf Semi Condensé (RBNSC). La structure de ce type de réseau permet de généraliser les relations de dépendances entre les différentes variables (noeuds reliés entre eux) et ainsi relaxer les hypothèses effectuées précédemment. Un Réseau Bayésien Naïf Semi Condensé est caractérisé par un noeud joint regroupant les différentes variables. La matrice de variance-covariance de ce noeud permet de définir les corrélations entre variables.

Ce type de réseau offre une structure simple permettant de prendre en compte les corrélations pouvant exister dans un groupe de variables (e.g. une variable jointe $\mathbf{x} = [x_1, \dots, x_m]^T$ au lieu de m variables indépendantes $\mathbf{x}_1, \dots, \mathbf{x}_m$). On pourra constater que le réseau Bayésien correspondant à une analyse discriminante quadratique présentée dans la figure 2.7 correspond à un réseau Bayésien naïf semi condensé.

IV.4 Classification Bayésienne : mélange de Gaussiennes

Les réseaux Bayésiens classifieurs font l'hypothèse que leurs variables observées suivent des distributions Gaussiennes. Ces méthodes restent applicables en pratique même lorsque cette hypothèse est non respectée, mais engendrent alors une perte d'efficacité. Pour remédier à cela, quelques solutions existent. Parmi elles, une solution consiste à considérer une combinaison linéaire d'un ensemble fini de densités locales Gaussiennes (un mélange de Gaussiennes),

permettant de capturer des propriétés de données (Chamroukhi, 2010) telles que la multimodalité, l'hétérogénéité non-identifiée, etc. Ce mélange, permettant d'approximer une variable non-Gaussienne, peut s'écrire comme ceci :

$$p(x|\mathbf{D} = C_k) = \sum_{r=1}^R p(x|\mathbf{D} = C_k, \mathbf{E} = M_r)p(\mathbf{E} = M_r|\mathbf{D} = C_k) \quad (2.56)$$

où $\sum_{r=1}^R p(\mathbf{E} = M_r|\mathbf{D} = C_k) = 1$ et \mathbf{E} est une variable discrète, et r est le nombre maximal de composantes (modes ou sous-classes) qu'une classe peut avoir. $p(\mathbf{x}|\mathbf{D} = C_k, \mathbf{E} = M_r)$ est la distribution Gaussienne d'une composante de la classe C_k et est caractérisée par deux paramètres $\mu_{C_k.M_r}$ et $\Sigma_{C_k.M_r}$. Ces paramètres caractérisant chaque densité (la fonction de densité d'un mode donné d'une classe donnée) peuvent être fixés indépendamment les uns des autres. Ils sont estimés, étant donnée la disponibilité des données nécessaires, en utilisant des méthodes comme le maximum de vraisemblance, le maximum a posteriori ou alors EM (Expectation and Maximisation).

Concernant la détermination du nombre de Gaussiennes impliquées, plusieurs solutions peuvent être utilisées. On peut recourir, par exemple, à une validation croisée pour choisir le modèle le plus adéquat. Cependant, bien qu'efficace cette solution est coûteuse en temps. Une autre alternative est de comparer les vraisemblances biaisées de chaque modèle en utilisant par exemple un critère d'information pour la sélection du modèle (e.g. AIC (Akaike Information Criterion, (Akaike, 1974)), BIC (Bayesian Information Criterion, (Schwarz et al., 1978)), etc.).

Un mélange de Gaussiennes se modélise facilement sur réseau Bayésien. Il suffit d'utiliser un réseau conditionnel Gaussien représentant une analyse discriminante auquel on ajoute un noeud discret représentant les différentes composantes du mélange. Par exemple, le réseau représentant une analyse discriminante quadratique (voir figure 2.7) peut être augmenté par un noeud discret permettant de gérer les observations non-Gaussiennes. Le réseau correspondant à cela est donné dans la figure 2.9. De la même manière, on pourra venir augmenter les autres structures vues précédemment (réseau Bayésien naïf, réseau Bayésien naïf augmenté, etc.) pour prendre en compte un mélange de Gaussiennes.

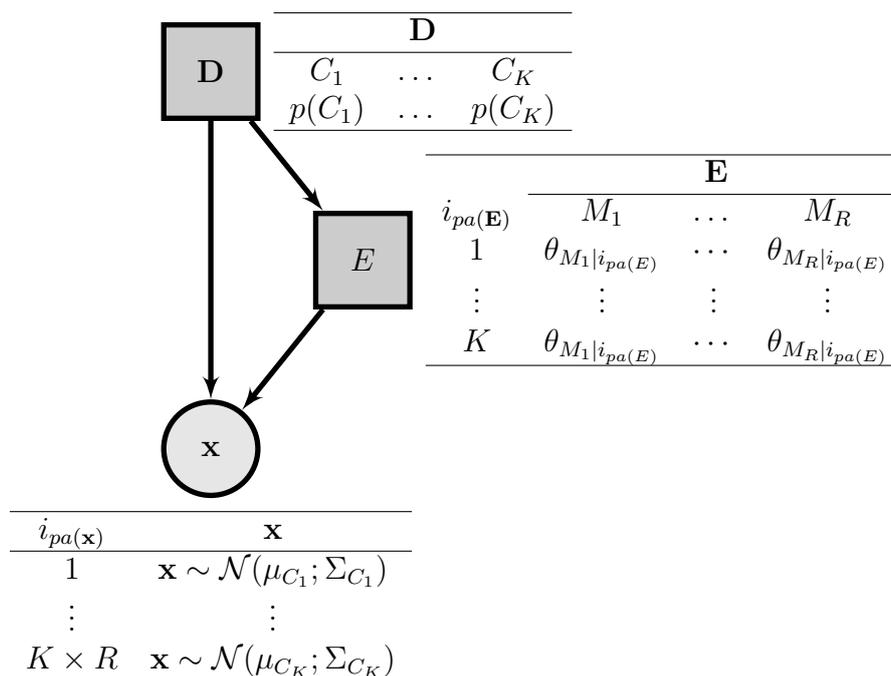


FIGURE 2.9 – Un mélange d’analyse discriminante, où $\theta_{M_1|i_{pa(\mathbf{E})}}$ correspond à $p(M_1|i_{pa(\mathbf{E})})$

V Inférence

Comme nous l’avons mentionné précédemment, nous allons principalement utiliser des réseaux Bayésiens et plus précisément des réseaux conditionnels Gaussiens, et les employer pour la détection et le diagnostic de fautes dans les systèmes. Nous allons donc avoir besoin de variables représentant l’état du système (des variables nous permettant la prise de décision), mais aussi d’autres variables permettant de manier et gérer les paramètres, mesures et observations prises sur le système.

Nous allons utiliser des noeuds discrets pour représenter et modéliser les états d’un système donné. Concernant les autres variables de nature continue, nous choisissons de ne pas les discrétiser afin d’éviter toute perte d’information. En effet, nous allons les traiter comme telles sous l’hypothèse qu’elles suivent une loi de distribution Gaussienne et qu’elles ne sont liées entre elles qu’avec des relations de dépendances linéaires. Notons que ces hypothèses peuvent évoluer par l’intermédiaire d’un noeud discret ou d’inférence spécifique (cas non-linéaire, non Gaussien).

Une fois la structure du réseau et ses paramètres définis (à l’aide d’al-

algorithmes d'apprentissage et/ou d'experts), afin de calculer les distributions marginales ainsi que les distributions de probabilités de chaque variable après l'introduction d'une évidence, il est impératif d'utiliser des algorithmes de propagation adéquats à la structure choisie. Plusieurs algorithmes d'inférence ont été proposés dans la littérature dans le cadre des réseaux conditionnels Gaussiens (Cobb et al., 2007). Nous allons détailler deux algorithmes basés sur l'arbre de jonction (Lauritzen and Jensen, 2001; Lauritzen, 1992). Ces derniers sont ceux implémentés dans la boîte à outils BNT/Matlab proposée par Murphy, et dans le logiciel Hugin, puisque ce sont les deux logiciels de réseaux Bayésiens utilisés pour ces travaux de thèse.

Ces algorithmes permettent une inférence exacte dans les réseaux conditionnels Gaussiens. L'un utilise un schéma basé sur deux représentations distinctes des distributions de probabilités et l'autre sur une représentation unifiée. Dans certains cas, dû à l'utilisation de ces deux représentations, le premier est numériquement instable (inversion de matrices lors du passage d'une représentation à une autre).

Dans ce qui suit, en s'inspirant de (Cobb et al., 2007; Jensen and Nielsen, 2007; Kjærulff and Madsen, 2005; Deventer, 2004; Barber, 2003; Lauritzen and Jensen, 2001; Murphy, 1998), nous présentons de façon concise les outils, calculs et mécanismes nécessaires au fonctionnement des deux algorithmes. Dans un premier temps, nous commençons par l'algorithme proposé par (Lauritzen, 1992) (implémenté sur la BNT) et puis son alternative proposée par (Lauritzen and Jensen, 2001) (implémenté sur Hugin) par le même premier auteur, au vu de l'instabilité du premier. Notons tout de même que les deux algorithmes peuvent être utilisés pour l'inférence dans les réseaux conditionnels Gaussiens.

V.1 Notions sur la théorie des graphes

Les deux algorithmes mentionnés précédemment s'appuient sur la théorie des graphes. En d'autres termes, pour faire de la marginalisation et/ou de la mise à jour de la distribution jointe des variables (la propagation des évidences), ils se basent sur un formalisme appelé arbre de jonction. Cet arbre est le résultat de la transformation d'un réseau Bayésien en un hypergraphe composé de noeuds qui sont aussi des hypergraphes. Dans ce qui suit, nous allons voir en quoi consiste cet arbre et comment l'obtenir.

Soit un réseau Bayésien composé de m variables $\mathbf{x}_1, \dots, \mathbf{x}_m$, l'hypergraphe

ou l'arbre de jonction lui correspondant devrait conserver et reformuler la distribution jointe de ses variables tout en respectant la propriété suivante :

$$p(\mathbf{x}_1, \dots, \mathbf{x}_m) = \frac{\prod_i \phi_{C_i}}{\prod_j \phi_{S_j}} \quad (2.57)$$

où ϕ représente un potentiel (une notion décrite dans ce qui suit). C_i et S_j , sont les noeuds composant l'hypergraphe. Les noeuds C_i sont appelés des noeuds cliques et les S_j sont appelés des noeuds séparateurs. Chacun de ces noeuds est composé des noeuds du réseau Bayésien d'origine.

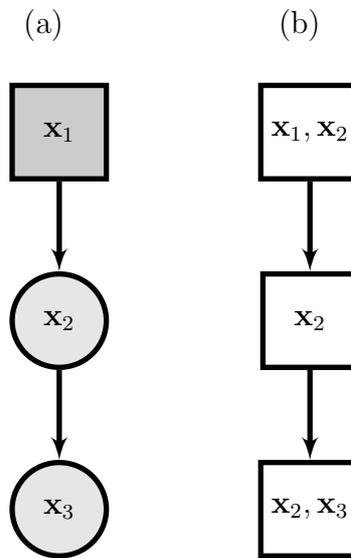


FIGURE 2.10 – (a) : Réseau Bayésien et (b) :L'arbre de jonction lui correspondant

Prenons l'exemple d'un réseau Bayésien composé de trois noeuds complètement connectés. Deux d'entre eux représentent des variables Gaussiennes, alors que le troisième représente une variable discrète. Dans la figure 2.10 nous illustrons ce réseau Bayésien et l'arbre de jonction lui correspondant.

Plusieurs étapes sont nécessaires pour aboutir à un tel arbre. Dans un premier temps, le réseau Bayésien concerné est rendu non dirigé (suppression du sens des arcs). Ensuite, les noeuds parents partageant un noeud fils sont liés. Cette opération est appelée moralisation. Le graphe obtenu après cette étape est appelé graphe moral. Le graphe moral correspondant au réseau Bayésien précédent est donné dans la figure 2.11. Une fois le graphe moral obtenu, il est aussitôt triangulé (étape de triangulation (Kjærulff, 1990)). Contrairement à la moralisation, à l'issue de la triangulation, le résultat obtenu n'est pas unique.

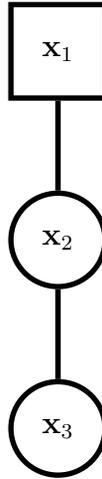


FIGURE 2.11 – Graphe moral

En effet, obtenir un graphe triangulé revient à résoudre un problème NP complet nécessitant le recours à des heuristiques dont la solution n'est pas toujours optimale (pour plus de détails concernant la triangulation, le lecteur peut consulter ([Jensen and Nielsen, 2007](#); [Cobb et al., 2007](#))). L'objectif de la triangulation est d'augmenter les cycles du graphe moral par un lien associant deux noeuds non consécutifs sous certaines contraintes tout en rajoutant le moins de liens possibles afin d'éviter d'augmenter la complexité et donc le temps de calculs. Dans le cas hybride, pour le bon déroulement de la propagation dans le graphe triangulé, ce dernier se doit de ne pas avoir un chemin contenant des noeuds continus entre deux noeuds discrets non-voisins (strong triangulation).

Le résultat de ces opérations (moralisation et triangulation) aboutit à la formation des cliques (des hypergraphes composés des noeuds du réseau Bayésien d'origine) comme celles présentées dans l'exemple précédent, figure 2.12.

Les cliques sont un ensemble maximal de noeuds liés deux à deux. Ces cliques sont générées de sorte à former un hypergraphe, nommé arbre joint, et respectant la contrainte qu'une variable présente dans deux cliques doit également être présente dans les cliques les séparant.

Selon ([Leimer, 1988](#)) les cliques d'un graphe décomposable (un graphe de cliques) peuvent former un hypergraphe dont au moins un de ses noeuds est une racine forte. Une racine forte est une clique respectant les règles suivantes :

$$A \cap B = \Upsilon, B \setminus A = \Gamma \quad (2.58)$$

où A, B et C sont des cliques. Avec A voisin de B et plus proche du noeud racine que l'est B .

Par ailleurs, il est toujours possible d'obtenir un arbre joint en organisant les cliques d'un graphe triangulé (Jensen, 2001). Cependant, l'arbre joint seul n'est pas capable de réaliser parfaitement la propagation. L'idée est alors d'augmenter l'arbre joint par des noeuds séparateurs (ensemble de variables résultat de l'intersection des cliques voisines). Le résultat est alors appelé arbre de jonction.

Une variable du réseau Bayésien d'origine définit une clique si elle et ses parents appartiennent à la même clique. A une variable, discrète ou continue, est associé un potentiel. Ainsi, chaque clique et chaque séparateur sont définis par un potentiel. C'est principalement la définition et la manipulation des différents potentiels qui différencie les deux algorithmes d'inférence choisis dans le cadre de cette thèse.



FIGURE 2.12 – Graphe triangulé dit graphe joint

V.2 Algorithme de Lauritzen : première version

Définition des potentiels

Dans (Lauritzen, 1992), le potentiel est défini comme une fonction non-négative et une généralisation de la distribution de probabilité satisfaisant le critère de normalisation. Un potentiel est caractérisé par son domaine, correspondant aux variables assignées au potentiel. Un potentiel, comme une distribution de probabilité, peut subir des opérations telles que la marginalisation, la multiplication et la division.

Le potentiel, dans le cas discret, est représenté par une table de probabilités (conditionnelles ou marginales). Ainsi, la multiplication de deux potentiels ϕ_1 et ϕ_2 , ayant chacun un domaine respectif : $Dom(\phi_1) = \mathbf{x}_1$ et $Dom(\phi_2) = \mathbf{x}_2$,

se fait en multipliant leurs entrées comme ceci :

$$\phi_3(x_1, x_2) = \phi_1(x_1)\phi_2(x_2) \quad (2.59)$$

où x_1, x_2 font partie respectivement du domaine de $\mathbf{x}_1, \mathbf{x}_2$.

Quant à la marginalisation, celle-ci est réalisée de la même manière que les distributions de probabilités, c'est à dire en sommant le potentiel en fonction de la variable marginalisée. Par exemple, la marginalisation du potentiel ϕ_3 en fonction de la variable \mathbf{x}_1 , $\phi_3^\downarrow^{\mathbf{x}_2}$, s'écrit comme ceci :

$$\phi_3^\downarrow^{\mathbf{x}_2} = \sum_{\mathbf{x}_1} \phi_3(\mathbf{x}_1, \mathbf{x}_2) \quad (2.60)$$

D'un autre coté, le potentiel Gaussien univarié/multivarié est représenté par une distribution conditionnelle Gaussienne univariée/multivariée pouvant être déclinée sous différentes représentations : une représentation en fonction des moments ou une représentation canonique.

Comme nous nous intéressons au réseau Bayésien hybride, il est possible que des noeuds continus et des noeuds discrets soient assignés à une même clique faisant partie de son domaine. Ce type de clique est dit clique hybride. Son potentiel correspond à une distribution conditionnelle Gaussienne univariée/multivariée dont les paramètres dépendent des différentes configurations possibles de ses noeuds parents discrets. Comme toute clique, la clique hybride, pour des fins de propagation, est sujette à des opérations de marginalisation. Le résultat de la marginalisation sur ce type de potentiel (ou clique) diffère selon la(es) variable(s) impliquée(s) (marginalisées). Les auteurs dans ([Lauritzen, 1992](#)), distinguent deux types de marginalisation : la marginalisation forte (strong) et la marginalisation faible (weak).

La marginalisation forte survient lorsque l'on marginalise un potentiel conditionnel Gaussien, étant donnée la configuration des variables discrètes de la clique, en fonction des variables continues. Elle est dite forte car ce qui en résulte est un potentiel de même type que le potentiel marginalisé, c'est à dire un potentiel conditionnel Gaussien.

Par exemple, soit un potentiel $\phi_{\{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2\}}$, ayant comme domaine $Dom(\phi)$ deux variables continues $\mathbf{y}_1, \mathbf{y}_2$ et une variable discrète \mathbf{x} . La forme canonique

correspondant à ce potentiel peut être représentée comme ceci :

$$\phi_{\{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2\}} = (g, \mathbf{h}, \mathbf{K}) \quad (2.61)$$

où \mathbf{h} et \mathbf{K} peuvent être décomposés comme ceci :

$$\mathbf{h} = [\mathbf{h}_1, \mathbf{h}_2]^T, \quad (2.62)$$

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{pmatrix} \quad (2.63)$$

La marginalisation de $\phi_{\{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2\}}$ par la variable \mathbf{y}_2 peut ainsi être définie et écrite, sous une forme canonique, comme ci-dessous :

$$\phi_{\{\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2\}}^{\downarrow\{\mathbf{x}, \mathbf{y}_1\}} = (g', \mathbf{h}', \mathbf{K}') \quad (2.64)$$

avec

$$g'(\mathbf{x}) = g(\mathbf{x}) + \frac{|\mathbf{y}_1| \ln(2\pi) - \ln(\det(\mathbf{K}_{11}(\mathbf{x}))) + \mathbf{h}_1^T(\mathbf{x}) \mathbf{K}_{11}(\mathbf{x})^{-1} \mathbf{h}_1(\mathbf{x})}{2}, \quad (2.65)$$

$$\mathbf{h}'(\mathbf{x}) = \mathbf{h}_2(\mathbf{x}) - \mathbf{K}_{21}(\mathbf{x}) \mathbf{K}_{11}^{-1}(\mathbf{x}) \mathbf{h}_1(\mathbf{x}), \quad (2.66)$$

$$\mathbf{K}'(\mathbf{x}) = \mathbf{K}_{22}(\mathbf{x}) - \mathbf{K}_{21}(\mathbf{x}) \mathbf{K}_{11}^{-1}(\mathbf{x}) \mathbf{K}_{12}(\mathbf{x}) \quad (2.67)$$

Un autre type de marginalisation est celui fonction de variable(s) discrète(s). Cela conduit de façon générale à une mixture de Gaussienne (marginalisation faible) sauf dans certains cas (e.g. absence de noeuds continus).

Par exemple, soit un potentiel $\phi_{\{\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2\}} = (g, \mathbf{h}, \mathbf{K})$, ayant comme domaine $Dom(\phi)$ deux variables discrètes $\mathbf{x}_1, \mathbf{x}_2$ et une variable continue \mathbf{y} . La marginalisation de ce potentiel par la variable \mathbf{x}_2 donne lieu au potentiel (représentation avec moments) suivant :

$$\phi(\mathbf{y}, \mathbf{x}_1, \mathbf{x}_2)^{\downarrow\{\mathbf{x}_1, \mathbf{y}\}} = (g', \mathbf{h}', \mathbf{K}') \quad (2.68)$$

$$(2.69)$$

avec

$$g'(\mathbf{x}_1) = \sum_{\mathbf{x}_2} g(\mathbf{x}_1, \mathbf{x}_2), \quad (2.70)$$

$$h'(\mathbf{x}) = \sum_{\mathbf{x}_2} \frac{h(\mathbf{x}_1, \mathbf{x}_2)g(\mathbf{x}_1, \mathbf{x}_2)}{g'(\mathbf{x}_1)}, \quad (2.71)$$

$$K'(\mathbf{x}) = \sum_{\mathbf{x}_2} \frac{K(\mathbf{x}_1, \mathbf{x}_2)g(\mathbf{x}_1, \mathbf{x}_2)}{g'(\mathbf{x}_1)} + \quad (2.72)$$

$$\sum_{\mathbf{x}_2} \frac{(h(\mathbf{x}_1, \mathbf{x}_2) - h'(\mathbf{x}_1))^T (h(\mathbf{x}_1, \mathbf{x}_2) - h'(\mathbf{x}_1))g(\mathbf{x}_1, \mathbf{x}_2)}{g'(\mathbf{x}_1)} \quad (2.73)$$

Les opérations précédemment employées et définies sont utilisées pour le calcul exact de la distribution des variables discrètes ainsi que les moyennes et matrices de variance-covariance des variables continues. Une définition complète de l'algèbre utilisée pour les potentiels est donnée dans (Jensen, 2001).

Une fois les potentiels de chaque variable déterminés, chaque potentiel des noeuds de l'arbre joint est initialisé à 1, puis multiplié par les potentiels des variables lui étant assignées (sachant ses noeuds parents). Ainsi, le potentiel d'une clique hybride est associé aux potentiels conditionnels Gaussiens de ses variables Gaussiennes pour chaque valeur de ses noeuds parents discrets.

De même, le potentiel de chaque noeud séparateur est initialisé à 1. Ceci permet, dans un premier temps, de respecter le fait que le produit de l'ensemble des potentiels des cliques divisé par celui des séparateurs donne la distribution jointe des variables du réseau Bayésien. Cela se traduit par :

$$p(\mathbf{x}_1, \dots, \mathbf{x}_m) = \prod_i p(\mathbf{x}_i | pa(\mathbf{x}_i)) \quad (2.74)$$

$$= \frac{\prod_k \phi_{C_k}}{\prod_j \phi_{S_j}} \quad (2.75)$$

$$= \prod_k \phi_{C_k} \quad (2.76)$$

Cette importante propriété est assurée lors de l'initialisation et est maintenue par la suite.

Une autre propriété que l'algorithme cherche à respecter est la consistance globale. Celle-ci stipule que le potentiel marginal d'un séparateur entre deux cliques doit correspondre aux potentiels marginaux de ces dernières.

Après l'initialisation, cette propriété n'est pas vérifiée. Ainsi, pour l'assurer,

les cliques de l'arbre de jonction s'envoient des messages (propagation). L'algorithme proposé par les auteurs permet, à la fin des propagations, de représenter les vraies distributions marginales des variables discrètes (consistance globale) et dans le cas hybride de s'assurer que la moyenne et la variance d'une clique C soient égales (consistance) à celles générées par le marginal du potentiel joint comme suit :

$$\phi_C \approx \phi_{\mathbf{x}_1, \dots, \mathbf{x}_m}^{\downarrow C} \quad (2.77)$$

Propagations

Une clique reçoit ou absorbe l'information de ses cliques voisines par le biais des noeuds de l'arbre de jonction les séparant. Par exemple, soit une clique émettrice C_i d'un message vers une clique réceptrice C_j , séparés par le noeud séparateur S_{ij} . Le message transmis de C_i vers C_j est tout d'abord reçu par le noeud séparateur. Ceci se traduit formellement par la marginalisation du potentiel de la clique C_i sachant le domaine du séparateur comme ceci :

$$\phi_{S_{ij}}^{nouveau} = \phi_{C_i}^{\downarrow S_{ij}} \quad (2.78)$$

où $\phi_{S_{ij}}^{nouveau}$ est le nouveau potentiel du séparateur S_{ij} . A son tour S_{ij} transmet le message vers C_j qui l'absorbe. Cela se traduit par une multiplication du potentiel de la clique réceptrice par un coefficient α :

$$\alpha = \frac{\phi_{S_{ij}}^{nouveau}}{\phi_{S_{ij}}^{ancien}} \quad (2.79)$$

où $\phi_{S_{ij}}^{ancien}$ est l'ancien potentiel du séparateur S_{ij} avant sa mise à jour, de telle sorte que :

$$\phi_{C_j}^{nouveau} = \alpha \phi_{C_j}^{\downarrow S_{ij}} \quad (2.80)$$

où $\phi_{C_j}^{nouveau}$ correspond au potentiel mis-à-jour de la clique C_j .

Afin de permettre au minimum la propriété de consistance, les messages sont envoyés selon un ordre bien établi. En effet, les messages entre cliques sont propagés en deux temps, vers et à partir du noeud racine. Le choix du noeud racine dépend du réseau Bayésien utilisé et de la direction de l'arbre déterminé. Dans le cas hybride, une racine forte doit être choisie comme point

de départ.

Dans un premier temps, les messages sont envoyés des cliques feuilles vers la clique racine. Une clique est par la suite habilitée à envoyer un message à ses voisins, une fois qu'elle a fini d'absorber l'information envoyée par ses cliques voisines éloignées de la racine. Cette opération s'appelle la collection des évidences. Lors de cette phase, les potentiels Gaussiens des variables sont représentés sous forme canonique. Une fois que les messages arrivent à destination, le noeud racine est mis à jour. A la fin de cette étape, l'arbre de jonction n'est toujours pas globalement consistant, car une clique peut ne pas avoir observée tous les messages transmis et peut ainsi ne pas être cohérente vis à vis des autres cliques. Pour remédier à cela, une autre propagation est nécessaire, celle-ci est nommée distribution des évidences. Elle consiste à mettre à jour les cliques feuilles à partir de la clique racine.

En présence d'une évidence, le réseau doit se mettre à jour, en utilisant les deux propagations vues précédemment, afin de maintenir ses propriétés. L'introduction de l'évidence se traduit par la modification des potentiels des différents noeuds de l'arbre. Dans le cas d'une évidence sur une variable discrète, il suffit juste de mettre à zéro les potentiels correspondant aux configurations impossibles. Concernant le cas Gaussien, il faut mettre à jour le potentiel des cliques dont les variables observées appartiennent à leurs domaines. Cela conduit à de nouveaux potentiels avec de nouveaux domaines plus restreints.

Par exemple, soit une clique composée de l variables, dont \mathbf{y} , au potentiel $\phi = (g, \mathbf{h}, \mathbf{K})$, où :

$$\mathbf{h} = \begin{pmatrix} h_{\mathbf{r}}(x) \\ h_{\mathbf{y}}(x) \end{pmatrix}, \quad (2.81)$$

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{\mathbf{r}\mathbf{r}} & \mathbf{K}_{\mathbf{r}\mathbf{y}} \\ \mathbf{K}_{\mathbf{y}\mathbf{r}} & \mathbf{K}_{\mathbf{y}\mathbf{y}} \end{pmatrix} \quad (2.82)$$

avec x une configuration des noeuds discrets faisant partie des l variables, et $\mathbf{K}_{\mathbf{r}\mathbf{r}}$ une matrice de dimension $(l-1) \times (l-1)$ où \mathbf{r} représente les l variables sauf \mathbf{y} .

En présence d'une observation $\mathbf{y} = y$, la modification de la clique se fait

ainsi :

$$g'(x) = g(x) + h_{\mathbf{y}}(x)y - \frac{K_{\mathbf{y}\mathbf{y}}y^2}{2}, \quad (2.83)$$

$$h'(x) = h_{\mathbf{r}}(x) - yK_{\mathbf{y}\mathbf{r}}(x), \quad (2.84)$$

$$K'(x) = K_{\mathbf{r}\mathbf{r}}(x) \quad (2.85)$$

Ainsi, le nouveau potentiel obtenu $\phi' = (g', h', K')$ est calculé en éliminant les composantes de \mathbf{y} .

En résumé, l'algorithme proposé par (Lauritzen, 1992) manipule deux formes de représentation pour traiter les potentiels conditionnels Gaussiens, notamment d'une clique hybride. La représentation avec les moments est plus naturelle et usuelle que la représentation canonique. Cependant, lors de l'initialisation, la forme canonique est préférée à celle des moments car l'initialisation d'une variable peut être assignée à une clique alors que ses parents ne le sont pas.

Ainsi, il n'est pas concevable de représenter la distribution de cette variable par ses moments, notamment sa moyenne, puisqu'elle dépend des valeurs de ses parents Gaussiens. Pour remédier à cela, la représentation sous forme canonique est employée, possédant également l'avantage d'être plus pratique pour les opérations de multiplication et division précédant l'initialisation. Elle est également associée à la propagation, collection et distribution des évidences. Cependant, l'écriture sous forme de moments est plus représentative d'où la transformation de la forme canonique vers celle-ci après la collection des évidences. Toutefois, cette transformation est numériquement instable. Cette instabilité est principalement due aux opérations d'inversion de matrices.

Pour remédier à cela, quelques années plus tard, (Lauritzen and Jensen, 2001) présentent un nouvel algorithme également basé sur l'arbre de jonction. Cet algorithme propose une solution à l'instabilité numérique du précédent et permet, de plus, de gérer le cas d'une matrice de variance-covariance nulle (intégration du cas déterministe).

V.3 Algorithme de Lauritzen : seconde version

Les auteurs dans (Lauritzen and Jensen, 2001) ont pensé à une nouvelle représentation des potentiels, alternative à celles précédemment utilisées. Celle-ci permet une compréhension directe des quantités utilisées comme les probabi-

lités, moyennes, coefficients de régression et matrices de variance-covariance. Cette nouvelle présentation définit deux types de variables, les variables dites de tête (head) et les variables dites de queue (tail). En se basant sur cette représentation, les auteurs proposent un nouveau schéma de propagation permettant, contrairement au précédent, de fournir sans effort des distributions marginales complètes. De plus, il permet de gérer les matrices de variance-covariance lorsqu'elles sont nulles ou singulières. Ceci représentent des améliorations considérables par rapport au schéma original.

Les deux algorithmes proposés dans (Lauritzen, 1992) et (Lauritzen and Jensen, 2001), ont en commun la construction de l'arbre de jonction mais aussi les étapes d'initialisation, de propagation et d'incorporation des évidences. Cependant, ils diffèrent dans la manière dont leurs potentiels sont définis et manipulés.

Définition des potentiels

Dans (Lauritzen and Jensen, 2001), que ce soit dans le cas Gaussien ou discret, un potentiel est représenté sous un même formalisme. Ce formalisme définit un potentiel d'un ensemble de variables comme étant proportionnel à la distribution de ses variables tête conditionnées à ses variables queue. Les variables tête et queue associées à un potentiel correspondent aux variables Gaussiennes appartenant et définissant son domaine.

Soit \mathbf{x} une variable discrète avec k valeurs et $\mathbf{y}_1, \mathbf{y}_2$ des variables continues, respectivement de dimension m et n , liées pour chaque valeur de \mathbf{x} par la régression conditionnelle Gaussienne suivante :

$$p(\mathbf{y}_1 | \mathbf{x} = x_i, \mathbf{y}_2 = y_2) = \mathcal{N}(A(x_i) + B(x_i)y_2, C(x_i)) \quad (2.86)$$

où $i \in 1, \dots, k$ et A est la table regroupant les différents vecteurs gouvernant la moyenne des variables tête étant donnée les variables queue pour chaque configuration x_i du noeud discret \mathbf{x} . De même, B est une table dont chaque case correspond à la matrice de régression de dimension $m \times n$ associée à une valeur du noeud discret. C est une autre table regroupant les k matrices de variance-covariance de \mathbf{x}_1 .

Le potentiel correspondant à ces variables est défini et représenté comme

ci-dessous :

$$\phi = [p; A; B; C](T|Q)$$

où p regroupe les probabilités correspondant aux différentes configurations de \mathbf{x} . T et Q dénotent respectivement les variables tête et queue représentant l'ensemble des variables continues appartenant au domaine du potentiel ϕ . Ainsi, T représente la variable \mathbf{y}_1 et Q représente la variable \mathbf{y}_2 . Dans un réseau Bayésien, une variable tête est un parent d'une variable queue. Dans l'arbre de jonction, une variable Gaussienne se présente comme variable tête uniquement dans un seul potentiel d'une clique parmi les cliques auxquelles elle appartient. Concernant la variable \mathbf{x} , $\mathbf{x} \in pa(T)$, elle définit et appartient également au domaine de ϕ .

Un potentiel peut être augmenté par des variables discrètes pour étendre son domaine. Ce dernier peut être également étendu par le rajout de variables continues à l'ensemble Q . Ceci se fait bien entendu en redéfinissant la dimension de B . D'un autre côté, des variables continues peuvent être retirées si les paramètres leurs correspondant sont égaux à zéros. D'ailleurs, les auteurs définissent un potentiel comme étant minimal si les colonnes de B sont non nulles. Dans (Lauritzen and Jensen, 2001) les auteurs ont introduit des opérations permettant de manipuler la nouvelle représentation des potentiels et exploiter ses propriétés.

De même que l'algorithme précédent, une des opérations nécessaires pour la transmission des messages est la marginalisation des potentiels. Dans le nouveau schéma proposé par les auteurs, on peut distinguer deux cas de figures : la marginalisation d'un potentiel en fonction de variables continues et celle en fonction de variables discrètes. Dans le schéma proposé, les auteurs marginalisent un potentiel en fonction des variables continues avant celles discrètes. Ceci est dû au fait que marginaliser un potentiel en fonction de ses variables tête fait émerger les variables discrètes (potentiel minimal).

La marginalisation d'un potentiel en fonction de variables continues donne lieu généralement à un potentiel de même type que celui du marginalisé. Ce type de marginalisation est dite forte. Par exemple, soit le potentiel suivant $\phi = [p, A, B, C](T_1, T_2|Q)$, son marginal en fonction de T_2 est donné par :

$$\phi^{\downarrow T_2} = [p, A_1, B_1, C_{11}](T_1|Q) \quad (2.87)$$

où A_1, B_1, C_{11} font parties de A, B et C .

Un autre type de marginalisation est la marginalisation faible. Elle émerge lors de la marginalisation en fonction de variables discrètes.

Soit le potentiel suivant $\phi = [p, A, -, C](T|-)$ dont le domaine discret est composé de variables discrètes \mathbf{d}_1 et \mathbf{d}_2 avec respectivement comme valeurs $i_{\mathbf{d}_1}, i_{\mathbf{d}_2}$ et où les tirets correspondent à l'absence de paramètres. Le résultat de sa marginalisation en fonction de \mathbf{d}_2 est donné par :

$$\phi^{\downarrow(\mathbf{d}_1 \cup T)} = [\tilde{p}, \tilde{A}, -, \tilde{C}(T|-)] \quad (2.88)$$

avec :

$$\tilde{p}(i_{\mathbf{d}_1}) = \sum_{i_{\mathbf{d}_2}} p(\{i_{\mathbf{d}_1}, i_{\mathbf{d}_2}\}), \quad (2.89)$$

$$\tilde{A}(i_{\mathbf{d}_1}) = \frac{1}{p_{i_{\mathbf{d}_1}}} \sum_{i_{\mathbf{d}_2}} A(\{i_{\mathbf{d}_1}, i_{\mathbf{d}_2}\}) p(\{i_{\mathbf{d}_1}, i_{\mathbf{d}_2}\}), \quad (2.90)$$

$$\begin{aligned} \tilde{C}(i_{\mathbf{d}_1}) = \frac{1}{\tilde{p}_{i_{\mathbf{d}_1}}} \sum_{i_{\mathbf{d}_2}} \{ & C(\{i_{\mathbf{d}_1}, i_{\mathbf{d}_2}\}) + [A(\{i_{\mathbf{d}_1}, i_{\mathbf{d}_2}\}) - A(i_{\mathbf{d}_1})] \times \\ & [A(\{i_{\mathbf{d}_1}, i_{\mathbf{d}_2}\}) - A(i_{\mathbf{d}_2})]^T \} p(\{i_{\mathbf{d}_1}, i_{\mathbf{d}_2}\}) \end{aligned} \quad (2.91)$$

Toutefois, la marginalisation en fonction de variables discrètes peut donner également, dans certains cas, lieu à une marginalisation forte. Par exemple, lorsque le potentiel est composé uniquement de variables discrètes.

En plus de la marginalisation, les auteurs proposent une nouvelle opération dite combinaison des potentiels. Elle est équivalente à la multiplication dans l'algorithme précédent. Deux types de combinaison peuvent être distingués : la combinaison directe et la combinaison récursive.

La combinaison directe \otimes de deux paires arbitraires de potentiels n'est possible que lorsque la tête de l'une d'entre elles est disjointe du domaine de l'autre et que les deux potentiels ont une queue minimale ou susceptible de l'être après réduction.

Par exemple, soit le potentiel ϕ résultat de la combinaison direct de deux potentiels $\phi_1 = [p, A, B, C](T_1|Q)$ et $\phi_2 = [q, E, F, G](T_2|Q_1 \cup Q)$. Ce potentiel,

en fonction de ces deux derniers, peut s'écrire comme ci-dessous :

$$\phi = [p, A, B, C](T_1|Q) \dot{\otimes} [q, E, F, G](T_2|T_1 \cup Q) \quad (2.92)$$

$$= [pq, U, V, W](H|T \setminus H) \quad (2.93)$$

avec :

$$T = T_1 \cup T_2, \quad (2.94)$$

$$U = \begin{pmatrix} A \\ E + F_1 A \end{pmatrix}, \quad (2.95)$$

$$V = \begin{pmatrix} B \\ F_2 + F_1 B \end{pmatrix}, \quad (2.96)$$

$$W = \begin{pmatrix} C & CF_1^T \\ F_1 C & G + F_1 CF_1^T \end{pmatrix} \quad (2.97)$$

où T_2 est disjointe du domaine du potentiel ϕ_2 , $T_2 \cap (T_1, Q_1) = \emptyset$ et F est une matrice partitionnée en $[F_1; F_2]$ correspondant aux deux ensembles composant la variable queue en fonction de la dimension de la tête et la queue ϕ_2 . $T \setminus H$ signifie toutes les variables représentant T sauf celles appartenant H .

La combinaison directe de deux potentiels est une opération associative et commutative. Cette combinaison correspond à un montage ordinaire de distributions conditionnelles. Elle peut être vue comme le résultat de la décomposition de ϕ en deux potentiels ϕ_1 et ϕ_2 . En effet, en considérant l'exemple précédent, dans le cas où $q = 1$ ou $T_1 = \emptyset$, le potentiel ϕ_2 est égale à $\phi^{\downarrow(T_1 \cup Q)}$. Ceci correspond à la marginalisation forte de ϕ . Par conséquent, le potentiel ϕ_2 est dit potentiel complément de ϕ_1 , noté $\phi^{\downarrow(T_1 \cup Q)}$, et peut se déduire du potentiel :

$$\phi = \phi^{\downarrow(T_1 \cup Q)} \dot{\otimes} \phi^{\downarrow(T_1 \cup Q)} \quad (2.98)$$

comme ceci :

$$q = pq/p, \quad (2.99)$$

$$E = U_2 - W_{21}W_{11}U_1, \quad (2.100)$$

$$F = [W_{21}W_{11}^+ : V_2 - W_{21}W_{11}V_1], \quad (2.101)$$

$$G = W_{22} - W_{21}W_{11}^+W_{12} \quad (2.102)$$

avec $T = (T_1, T_2)$ et $(M)^+$ est l'inverse généralisé (pseudo-inverse de Moore-Penrose, où $((M)^+)^+ = M$) d'une matrice (M) . Cette décomposition correspond à celle d'une distribution de probabilité en distribution marginale et une autre conditionnelle.

Lors de la propagation, on est parfois amené à combiner deux potentiels dont les ensembles tête sont disjoints. Dans ce cas là, on ne peut appliquer une combinaison directe. Pour remédier à cela, les auteurs proposent une généralisation de la combinaison directe définie uniquement lorsque les ensembles tête des potentiels combinés sont disjoints. Cette généralisation est dite combinaison récursive. Elle correspond à une combinaison plus générale des potentiels (utile lors de l'étape d'initialisation)

Prenons les potentiels ϕ_1 et ϕ_2 précédents. Si leurs ensembles tête sont disjoints, leur combinaison récursive \otimes peut s'écrire :

$$\phi_1 \otimes \phi_2 = \phi_1 \dot{\otimes} \phi_2 \quad \text{ou} \quad \phi_1 \otimes \phi_2 = \phi_2 \dot{\otimes} \phi_1 \quad (2.103)$$

Dans le cas où aucune des combinaisons directes n'est définie, les deux potentiels sont décomposés, jusqu'à ce que les potentiels obtenus vérifient les définitions précédentes, comme ceci :

$$\phi_1 \otimes \phi_2 = \phi_1 \dot{\otimes} (\phi_2' \otimes \phi_2'') \quad (2.104)$$

avec :

$$\phi_2 = \phi_2' \dot{\otimes} \phi_2'' \quad (2.105)$$

où ϕ_2'' correspond à la marginalisation forte de ϕ_2 donnée ci-dessous :

$$\phi_2'' = \phi_2^{\downarrow com} \quad (2.106)$$

où com représente les variables que l'ensemble tête de ϕ_2 a en commun avec le domaine de ϕ_1 . ϕ_2'' correspond au complément de ϕ_2' .

Propagations

Une fois l'arbre de jonction défini et sa racine forte déterminée, il faut alors associer à chaque clique un potentiel. De même que pour l'algorithme précédent, le potentiel de chaque noeud du réseau Bayésien d'origine est associé à une clique arbitraire contenant sa famille. Par définition, le potentiel de chaque clique correspond à la combinaison directe des potentiels des variables lui étant assignées. Ceci fait partie de l'initialisation, une étape primordiale à la propagation. Elle implique l'envoi de messages vers la racine et de la racine vers les noeuds feuilles. Dans les deux cas, une clique envoie un message à une de ses cliques voisines étant donné le noeud séparateur qu'ils partagent. Dans le nouvel algorithme, le contenu des messages et la mise à jour des différents noeuds de l'arbre diffèrent du premier algorithme proposé.

Par exemple, soit une clique C_2 qui envoie un message à une autre clique C_1 dans le cadre d'une collection de messages. Le potentiel de la clique réceptrice est alors combiné récursivement au résultat de la marginalisation forte du potentiel de la clique émettrice, comme ceci :

$$\phi_{C_1}^{nouveau} = \phi_{C_1} \otimes \phi_{C_2}^{\downarrow S_{C_2 \rightarrow C_1}} \quad (2.107)$$

où $S_{C_2 \rightarrow C_1}$ est le noeud séparateur entre les deux cliques.

Dans le cas où la clique C_1 correspond à la racine, son potentiel est alors la distribution marginale correcte de son domaine. Quant à la clique émettrice, elle conserve uniquement le complément de la marginalisation de son potentiel :

$$\phi_{C_2}^{nouveau} = \phi_{C_2}^{|S_{C_2 \rightarrow C_1}} \quad (2.108)$$

Une fois la collection terminée, la combinaison des potentiels des cliques d'un arbre donné est égale à la distribution jointe ϕ de toutes les variables du réseau Bayésien, comme ceci :

$$\phi = \bigotimes_i \phi_{C_i} \quad (2.109)$$

avec i le numéro de la clique.

La distribution des messages résulte en des noeuds séparateurs avec des potentiels correspondants à la marginalisation du potentiel joint :

$$\phi_{S_{C_j \rightarrow C_k}} = (\phi_{S_{C_i \rightarrow C_j}} \otimes \phi_{C_j})^{\downarrow S_{C_j \rightarrow C_k}} \quad (2.110)$$

$$= \phi^{\downarrow S_{C_j \rightarrow C_k}} \quad (2.111)$$

avec i le numéro de clique la plus proche de la clique racine et $S_{C_j \rightarrow C_k}$ et $S_{C_i \rightarrow C_j}$ sont les séparateurs partagés par les cliques C_i, C_j, C_k .

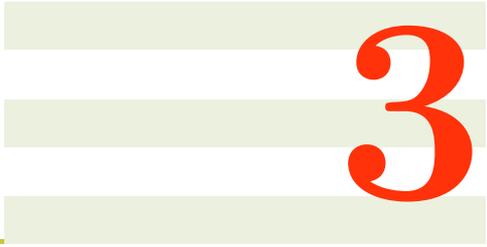
A la fin des deux propagations, le potentiel de chaque clique correspond à celui obtenu lors de la collection des messages. Concernant les noeuds séparateurs, leurs potentiels calculés lors de la distribution des évidences sont maintenus.

Une fois l'initialisation effectuée, il est possible d'établir des requêtes (marginalisation/calcul de probabilités après ou avant l'introduction des évidences) sur le réseau Bayésien considéré. En effet, le schéma proposé par les auteurs permet de traiter plusieurs types de calculs sur réseau Bayésiens et notamment réseaux conditionnels Gaussiens. Celui-ci s'appuie particulièrement sur un mécanisme appelé "push". Ce dernier réarrange l'arbre (rapproche les variables encore plus de la racine tout en modifiant les potentiels sans influencer le potentiel joint) afin de répondre à une requête donnée. Ainsi, il permet, par exemple, de calculer le marginal des variables se trouvant dans des cliques différentes. Il est aussi utilisé lors de l'incorporation des évidences. En effet, contrairement aux évidences des noeuds discrets, les nouvelles observations de noeuds continus sont introduites, une à la fois, dans les cliques concernées menant parfois à l'utilisation de ce mécanisme.

VI Conclusion

Dans ce premier chapitre nous avons introduit les réseaux Bayésiens. Tout d'abord nous les avons défini puis expliqué quelques unes de leurs propriétés. Suite à cela, nous avons présenté les différents noeuds pouvant composer un réseau Bayésien. Nous avons vu qu'ils permettent de prendre en compte des variables discrètes et continues. Dans le contexte de la classification supervisée, plus précisément de l'analyse discriminante, nous avons présenté dans un premier temps quelques réseaux Bayésiens classifieurs. Dans la suite de la thèse nous allons nous appuyer sur cet outil pour faire de la détection et du

diagnostic. Nous considérons la détection et le diagnostic comme un problème de discrimination entre respectivement deux classes et K classes. Finalement, nous avons présenté deux algorithmes d'inférence permettant de faire du calcul dans les réseaux Bayésiens. Ces algorithmes permettent particulièrement de faire de l'inférence exacte dans les réseaux conditionnels Gaussiens, un cas particulier des réseaux Bayésiens.



3

Surveillance des systèmes

I Introduction

La course des entreprises et industries à optimiser le rapport gain/perte ainsi que la demande croissante pour des produits et/ou services de qualité et conformes aux spécifications ont énormément contribué à l'évolution et l'utilisation impérative de méthodes de surveillance automatique des systèmes rarement déterministes. En effet, ces derniers sont sujets à plusieurs facteurs les rendant stochastiques. On distingue deux types de facteurs : ceux non contrôlables et ceux contrôlables. Les facteurs non contrôlables contrairement aux facteurs contrôlables nuisent au fonctionnement idéal du système : ils correspondent généralement aux bruits et perturbations pouvant survenir sur le système. Les facteurs contrôlables correspondent aux différentes fautes pouvant causer des décentrages et des dispersions de la production. Ces dernières sont la raison pour laquelle la surveillance est employée. En effet, le rôle de la surveillance est essentiellement de les maîtriser. En d'autres termes, elle essaie d'identifier et de déceler le plus rapidement possible la présence d'une faute sur le système, en vérifiant la cohérence des diverses informations disponibles sur celui-ci. Ainsi, le résultat de la surveillance contribue à la diminution des fautes ou dans le cas idéal à leur élimination. Cette dernière s'effectuera à l'aide d'une entité (opérateur, ingénieur, automate) qui étudiera l'état de fonction-

nement anormal causé par la faute et prendra une décision concernant l'avenir du système (maintenance, fermeture, adaptation de réglages sur la commande du système).

La surveillance se décompose de façon générale en deux phases extrêmement importantes pour le fonctionnement efficace, économique et sûr d'un système donné : la détection et le diagnostic.

La détection consiste à déterminer le plus rapidement possible si le processus est toujours en fonctionnement normal ou bien si il est en fonctionnement anormal, dû à l'apparition d'une ou plusieurs fautes (capteurs, actionneurs et processus) qu'elle soit soudaine, naissante ou transitoire. La phase de détection est loin d'être simple ([Dubuisson, 2001](#)), et cela pour plusieurs raisons : une vision et/ou une modélisation complète de tout le système est rarement possible (une faute peut ne pas être observable), il existe des interactions et interconnexions entre composantes (plusieurs fautes peuvent se compenser et ne plus être observables).

Concernant la phase de diagnostic, on peut la définir de différentes façons selon la précision à décrire la faute survenue sur les actionneurs et/ou capteurs et/ou le système, ainsi que le type de la méthode de surveillance utilisée. Selon ([Chiang et al., 2001](#)), on peut distinguer trois définitions : Identification, Diagnostic et Isolation du système. L'identification permet de déterminer parmi les mesures prises, celles qui sont les plus pertinentes pour expliquer la faute survenue. Ainsi, elle diminue l'espace de recherche. Il est nécessaire de noter que l'identification peut parfois être une passerelle à d'autres méthodes pour la phase du diagnostic. Le diagnostic consiste à expliquer les fautes survenues sur le système en exprimant au minimum leur type, leur emplacement, leur amplitude ainsi que leur durée. L'isolation, quant à elle, se situe entre les deux définitions précédentes. En effet, l'isolation fournit plus d'informations que l'identification et moins que le diagnostic. En d'autres termes, elle permet de distinguer la mesure responsable de l'état de fonctionnement anormal du système et ainsi localiser l'emplacement de composantes fautives.

La surveillance, de façon générale, consiste à expliquer le comportement du système en se basant sur les observations prises sur celui-ci (entrées du système, mesures entre composantes, sorties du système). Il semble donc évident que la qualité de la surveillance dépend fortement de la qualité de ces observations. Ainsi, les méthodes de surveillance dites à base de redondance physique ou matérielle sont considérées comme des méthodes très fiables. D'ailleurs, elles

représentent une valeur sûre pour les systèmes qui emploient des composants critiques et pour lesquelles une panne ou un dysfonctionnement représente un risque économique et/ou un risque important pour la sécurité des humains et de l'environnement (usine chimique, base nucléaire). Elle consiste de façon générale à mesurer une seule grandeur ou composante à l'aide de plusieurs capteurs. Cependant, son principe relativement simple la rend coûteuse économiquement (coût des capteurs et de leurs maintenances) et très encombrante (difficulté de duplication des capteurs). Pour remédier à cela, d'autres méthodes ont été développées et améliorées, chacune adaptée à des scénarios différents (informations disponibles, fautes à détecter). Ces méthodes sont, de façon générale, répertoriées dans deux classes principales : les méthodes à base de modèle et les méthodes à base de données.

II Méthodes à base de modèle

Les contraintes matérielles ainsi que les coûts d'installation instaurés par les méthodes utilisant la redondance physique sont diminués de façon considérable grâce aux méthodes à base de modèle. Ces méthodes reposent sur une connaissance a priori du système. Cette connaissance correspond à un ensemble d'équations mathématiques précises issues des principes premiers (modèle du système). Ces équations représentent les dépendances physiques ou temporelles existantes entre les différentes variables du système expliquant son comportement dynamique (Isermann and Ballé, 1997). De plus, elle permet d'extraire l'information sur les possibles changements du fonctionnement nominal du système.

Le modèle analytique correspondant à un système donné se doit d'être réaliste. En effet, il doit intégrer tous les facteurs influençant de façon significative ou non l'évolution du système. En d'autres termes, il doit pouvoir prendre en compte les entrées inconnues (les différentes perturbations et bruits pouvant influencer le système) ainsi que les fautes. Toutefois, obtenir un modèle efficace et complet pour un système complexe est loin d'être chose aisée (incertitudes sur les paramètres du modèle, non linéarité, etc.) : l'effort et le coût demandé pour le réaliser sera proportionnel à la complexité du système.

La connaissance du modèle du système permet assez souvent d'identifier les différentes fautes systèmes, actionneurs ou/et capteurs pouvant le faire dévier de son fonctionnement normal. Ainsi, la surveillance avec des méthodes à base

de modèle consiste dans un premier temps à déceler un changement survenu étant données les entrées, sorties et le modèle du système. Dans un deuxième temps, si un changement est survenu, elles essaient de l'isoler étant donnée la connaissance a priori des différentes fautes.

Ces fautes peuvent être classées selon leur nature en deux catégories : additives et multiplicatives. Les fautes additives sont les fautes venant s'ajouter aux mesures prises sur le système (sorties du modèle) et concerne la plupart du temps les fautes capteurs et actionneurs. Les fautes multiplicatives sont décrites comme des changements survenues sur les paramètres du système.

De façon générale, dans les méthodes à base de modèle, le fonctionnement du système est considéré comme étant caractérisé par des grandeurs qui dépendent du temps et non de l'espace (Isermann, 2004; Zwingelstein, 1995). Ceci dit, deux classes de représentations décrivant le fonctionnement du système peuvent être distinguées : la représentation classique (les équations différentielles, les fonctions de transfert) et la représentation d'état (Jaulin, 2005).

La représentation classique, contrairement à la représentation d'état, ne permet pas de décrire le comportement d'un système dans le temps. En effet, la représentation d'état introduit une notion d'état qui représente une mémoire minimum du passé à partir d'un instant donné, permettant de déterminer l'évolution future du système (Fossard and Biannic, 2006). Ainsi, pour représenter l'évolution dynamique du système, il semble évident d'utiliser cette représentation (elle est composée de deux équations : l'équation d'état et l'équation d'observation, un exemple d'une représentation d'état sans bruits d'un système linéaire discret à temps invariant est donné dans (3.1)). D'ailleurs, elle est la plus utilisée dans les méthodes de détection et diagnostic de fautes à base de modèle (Ding, 2008; Isermann, 2006; Chiang et al., 2001).

$$\begin{cases} \mathbf{z}(t+1) = \mathbf{A}\mathbf{z}(t) + \mathbf{B}\mathbf{u}(t) \\ \mathbf{y}(t) = \mathbf{C}\mathbf{z}(t) + \mathbf{D}\mathbf{u}(t) \end{cases} \quad (3.1)$$

où, pour un instant donné t , $\mathbf{z}(\cdot) \in \mathbb{R}^n$ est le vecteur d'état, $\mathbf{y}(\cdot) \in \mathbb{R}^s$ est le vecteur des sorties, $\mathbf{u}(\cdot) \in \mathbb{R}^l$ est le vecteur des entrées. $\mathbf{A} \in \mathbb{R}^{n \times n}$ est la matrice d'évolution, $\mathbf{B} \in \mathbb{R}^{n \times l}$ est la matrice de commande, $\mathbf{C} \in \mathbb{R}^{s \times n}$ est la matrice d'observabilité et $\mathbf{D} \in \mathbb{R}^{s \times l}$ est la matrice de transmission directe.

La surveillance à base d'un modèle représenté sous une représentation d'états peut être décomposée en deux phases : la génération des résidus et

la prise de décision. Dans ce qui suit, nous allons voir que, pour chaque phase, plusieurs méthodes ont été proposées dans la littérature.

II.1 Génération des résidus

La génération des résidus est une étape majeure dans la surveillance à base de modèle. Elle consiste à générer des résidus permettant la description de l'état de fonctionnement du système. En d'autres termes, elle calcule l'écart entre le comportement observé et le comportement estimé du système lorsqu'il fonctionne correctement dans le but d'une exploitation ultérieure.

La génération des résidus dépend fortement du modèle choisi. En effet, indépendamment du fait que l'obtention du modèle peut parfois être compliquée, le modèle choisi peut ne pas prendre en compte certaines relations existantes entre mesures. Il semble ainsi évident qu'un bon générateur est synonyme d'une bonne détection et/ou diagnostic de fautes. Notons aussi qu'un changement du modèle (e.g. système ayant plusieurs configurations) et/ou du système de mesure (e.g. ajout ou suppression d'un ou plusieurs capteurs) peut impliquer un changement du modèle et par conséquent une reformulation du générateur des résidus ([Dubuisson, 2001](#)).

Plusieurs méthodes existent pour la génération des résidus, on distingue principalement trois méthodes : générateur de résidus à base d'estimation paramétrique, générateur de résidus à base de l'espace de parité et générateur de résidus à base d'observateurs/filtres. Toutefois, selon ([Ding et al., 2009](#); [Theilliol, 2003](#)) ces méthodes peuvent être équivalentes dans certains cas.

Générateur de résidus à base d'estimation paramétrique

Lorsque la structure du modèle est connue, le générateur de résidus, dans un premier temps, essaye d'estimer les paramètres du modèle du système en se basant sur les mesures prises sur ce dernier. En d'autres termes, il cherche à minimiser l'écart entre les entrées et les sorties mesurées. Dans la littérature, deux méthodes principales peuvent être distinguées pour l'estimation des paramètres : minimisation de l'erreur d'équation et minimisation de l'erreur de sortie.

Après avoir déterminé pour un instant donné les paramètres dictant le comportement dynamique du système $\hat{\theta}(t)$, une comparaison est effectuée entre ces paramètres et les paramètres physiques réels $\theta(t)$ du modèle lors de son

fonctionnement normal. Le résultat de cette comparaison $\theta(t) - \hat{\theta}(t)$ est dit résidus $\mathbf{r}(t)$. Selon (Isermann, 2006), ces résidus permettent une compréhension profonde de la faute.

Toutefois, il est rare que les paramètres du modèle du système correspondent exactement aux coefficients physiques du système. Ceci est vrai lorsque le modèle est fidèlement construit à partir des principes premiers. Cependant, ceci n'est pas toujours le cas. En effet, un recours à la simplification des hypothèses et au regroupement des paramètres physiques du modèle est généralement de rigueur. Pour plus d'informations et de détails sur ce type de générateur, le lecteur pourra se référer aux travaux d'Isermann des deux dernières décennies. Il est à noter que ce type de générateur est beaucoup plus adapté aux fautes multiplicatives. En d'autres termes, il est approprié lorsque les conséquences d'une faute sur le système se répercutent sur ses paramètres physiques.

Générateur de résidus à base de l'espace de parité

Le générateur de résidus par l'espace de parité (Ding, 2008; Ding et al., 1999; Gertler, 1997) est très utilisé dans la littérature, il est simple à concevoir, à implémenter et à utiliser pour des systèmes statiques ou dynamiques. Cependant, contrairement au générateur précédent, il est dédié aux fautes additives. Il permet de concevoir des résidus dépendant des relations linéaires existantes entre les mesures. Ces relations sont appelées équations de parité.

Soit la représentation d'état d'un système discret à temps invariant présentée ci-dessous :

$$\begin{cases} \mathbf{z}(t+1) = A\mathbf{z}(t) + B\mathbf{u}(t) + B_n\mathbf{n}(t) + B_f\mathbf{f}(t) \\ \mathbf{y}(t) = C\mathbf{z}(t) + D\mathbf{u}(t) + D_n\mathbf{n}(t) + D_f\mathbf{f}(t) \end{cases} \quad (3.2)$$

où les vecteurs $\mathbf{f}(\cdot)$ et $\mathbf{n}(\cdot)$ correspondent respectivement aux fautes et aux perturbations (entrées inconnues). B_n , B_f , D_n et D_f sont des matrices associées aux entrées inconnues et aux fautes.

Les deux équations qui composent cette représentation, pour un horizon d'observation de taille $(h+1)$, $h \geq 0$, peuvent être regroupées et réécrites sous la forme suivante :

$$\mathbf{y}_h(t) = \Gamma\mathbf{z}(t-h) + H_u\mathbf{u}_h(t) + H_d\mathbf{d}_h(t) + H_f\mathbf{f}_h(t) \quad (3.3)$$

où

$$\Gamma = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^h \end{pmatrix}, \quad (3.4)$$

$$H_u = \begin{pmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{h-1}B & \cdots & CB & D \end{pmatrix} \quad (3.5)$$

et H_d , H_f sont des matrices ayant la même structure que H_u .

En se basant sur ce type d'équations, le principe de la génération des résidus à partir de l'espace d'état consiste à décrire les résidus de façon indépendante du vecteur d'état $\mathbf{z}(t - h)$. Ceci consiste à résoudre l'équation suivante :

$$W\Gamma = 0 \quad (3.6)$$

Cela revient à trouver le complément orthogonal W à Γ . Ainsi, en multipliant (3.6) par la matrice de parité W on obtient :

$$W\mathbf{y}_h(t) - WH_u\mathbf{u}_h(t) = WH_d\mathbf{d}_h(t) + WH_f\mathbf{f}_h(t) \quad (3.7)$$

A partir de cette équation, les résidus peuvent alors être définis en fonction des mesures prises sur le système (entrées et sorties) ou alors en fonction des fautes et des entrées inconnues (et forcément du bruit non pris en compte ici par la représentation d'état pour des raisons de clarté). Ceci est illustré ci-dessous :

$$\mathbf{r}_h(t) = W\mathbf{y}_h(t) - WH_u\mathbf{u}_h(t) = WH_d\mathbf{d}_h(t) + WH_f\mathbf{f}_h(t) \quad (3.8)$$

Par ailleurs, il semble évident que la qualité des résidus générés et les fautes prises en compte dépend fortement du choix de la matrice de parité.

Générateur de résidus à base d'observateurs/filtres

Un générateur de résidus à base d'observateurs ou de filtres consiste respectivement à reconstruire (la reconstruction se fait par un observateur comme celui de Luenberger lorsque le modèle est déterministe) ou à estimer (l'estimation se fait par un filtre comme celui de Kalman, lorsque le modèle est stochastique) le vecteur d'état ou celui de la sortie faisant partie du modèle du système (e.g. représentation d'état à temps discret) modélisant respectivement l'effet de la faute sur les états et les sorties. Dans ce cas, les résidus sont obtenus en comparant ces estimations à leurs mesures réelles prises sur le système. En pratique selon (Dubuisson, 2001), ces générateurs de résidus sont considérés comme une alternative aux générateurs utilisant l'espace de parité lorsqu'il est difficile de se défaire des vecteurs d'état (une matrice de parité conforme est introuvable).

Soit la représentation d'état donnée dans (3.2). Celle-ci ne présente pas de bruits. Dans ce cas, l'utilisation d'un observateur de Luenberger (voir équation (3.9)) pour la génération du vecteur des résidus d'états et de sorties semble plus judicieuse que celle utilisant un filtre de Kalman. De plus, les observateurs contrairement aux filtres de Kalman conduisent, grâce à leur flexibilité dans le choix du gain L , à une riche variété de structure pour la détection et le diagnostic de fautes (Baïkeche, 2007).

$$\begin{cases} \hat{\mathbf{z}}(t+1) = A\hat{\mathbf{z}}(t) + B\mathbf{u}(t) - L(\mathbf{y}(t) - \hat{\mathbf{y}}(t)) \\ \hat{\mathbf{y}}(t) = C\hat{\mathbf{z}}(t) + D\mathbf{u}(t) + D_n\mathbf{n}(t) + D_f\mathbf{f}(t) \end{cases} \quad (3.9)$$

où $\hat{\mathbf{z}}(\cdot)$ et $\hat{\mathbf{y}}(\cdot)$ sont respectivement le vecteur d'état et le vecteur de sortie, estimés à un instant t . L est la matrice de gain (se doit de satisfaire les conditions de stabilité $(A - (L \times C))$ de l'observateur. Cette dernière détermine la dynamique de l'observateur et la sensibilité/robustesse des résidus aux fautes (Chiang et al., 2001). Ci-dessous nous donnons un exemple de résidus pouvant être générés en utilisant un simple observateur :

$$\begin{aligned} \mathbf{r}^e(t) = \mathbf{z}(t) - \hat{\mathbf{z}}(t) &= (A - (L \times C))\mathbf{r}^e(t-1) + B_n\mathbf{n}(t) + B_f\mathbf{f}(t) \\ &\quad - L(D_n\mathbf{n}(t) + D_f\mathbf{f}(t)), \end{aligned} \quad (3.10)$$

$$\mathbf{r}^s(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t) = C\mathbf{r}^s(t-1) + D_n\mathbf{n}(t) + D_f\mathbf{f}(t) \quad (3.11)$$

où $\mathbf{r}^e(\cdot)$, $\mathbf{r}^s(\cdot)$ sont respectivement les résidus d'états et de sorties à l'instant t .

Ces résidus générés peuvent être sujets à des perturbations, et cela peut être handicapant pour les étapes succédant à la génération des résidus (détection et isolation de fautes). Pour obtenir des résidus robustes aux bruits et aux perturbations, et/ou sensibles aux fautes, plusieurs techniques peuvent être utilisées (Isermann, 2006). Parmi ces techniques, on peut citer : les observateurs à entrées inconnues (Unknown Input Observer) (Simani et al., 2013) et les fonctions de transformations matricielles (Fellouah, 2007).

D'autres types d'observateurs existent, appelés reconSTRUCTEURS d'entrées. Au lieu de modéliser les effets des fautes sur les états ou les sorties, ils les modélisent sur les entrées. En d'autres termes, les résidus générés par ces observateurs sont issus de la comparaison des entrées du système avec les entrées estimées.

Après avoir choisi un générateur de résidus adéquat, les résidus générés peuvent être utilisés pour prendre une décision sur l'état de fonctionnement du système. La décision intervient après la génération d'un vecteur de résidus de qualité ayant une sensibilité requise aux fautes considérées. Elle consiste à déterminer de façon générale quelle est ou quelles sont les fautes responsables de l'état de fonctionnement anormal du système. En d'autres termes, il s'agit tout simplement de surveiller les résidus. Et pour cela, deux phases sont nécessaires : la détection et l'isolation.

II.2 Détection

Après la génération des résidus, ceux-ci sont évalués par un test de cohérence. Il s'agit de déterminer si un ou plusieurs résidus sont nuls ou non. En effet, lors du fonctionnement normal, les résidus sont censés être nuls. Cependant, ce n'est pas toujours le cas. En effet, les résidus ne sont pas sensibles qu'aux fautes mais également aux bruits de mesures effectuées sur le système ainsi qu'aux erreurs de modélisation du générateur de résidus. A cet effet, les résidus sont considérés comme étant statistiquement nuls avec une variation donnée lors du fonctionnement normal. Ainsi, lorsque l'on dispose d'informations a priori sur les propriétés statistiques du bruit, des tests statistiques s'imposent (Dubuisson, 2001). La prise de décision peut également se faire de façon déterministe grâce à des comparaisons à des seuils prédéfinis à l'aide d'études de sensibilité (Zwingelstein, 1995).

II.3 Isolation

L'isolation des fautes consiste à localiser les fautes responsables d'un état de fonctionnement anormal du système. L'isolation peut être étendue dans le but de déterminer l'amplitude de la faute dans le cas où les fautes ne varient pas au cours du temps (Dubuisson, 2001). L'isolation, contrairement à la détection, nécessite la présence de plusieurs résidus (un seul résidu permet la détection d'une faute survenue sur le système) sensibles aux fautes. Plusieurs méthodes d'isolation ont été proposées dans la littérature. La méthode la plus connue est celle basée sur les résidus renforcés. Cette méthode consiste à donner aux résidus des propriétés spéciales leur permettant d'acquérir une capacité d'isolation entre les fautes. On peut distinguer deux types de résidus générés : les résidus directionnels et les résidus structurés (Isermann, 2006).

Les résidus structurés

Les résidus structurés sont construits de façon à ce qu'ils soient sensibles à certaines fautes et non à d'autres. Dans ce cas, l'isolation consiste à comparer les résidus évalués et les caractéristiques de chaque faute (chaque faute est expliquée par un ensemble de résidus). Ces caractéristiques sont généralement assemblées dans un tableau binaire (matrice d'incidence), expliquant les dépendances entre les fautes et les résidus (traduisant la sensibilité ou la robustesse des résidus aux fautes), obtenus pendant l'étape d'apprentissage en utilisant, selon (Theilliol, 2003), des connaissances heuristiques et/ou des connaissances analytiques du comportement du système.

Pour une meilleure localisation, il semble évident que chaque faute doit être unique. En d'autres termes, chaque faute est caractérisée par un ensemble de résidus différents des autres fautes, d'ailleurs plus la différence est grande, meilleure est l'isolation. Notons aussi que lorsqu'un seul résidu permet d'expliquer à lui seul une faute, une localisation des fautes multiples survenues sur le système, est possible.

Les résidus directionnels

Les résidus directionnels sont conçus de telle façon que dès qu'un défaut apparaît ils prennent une direction particulière dans l'espace des résidus, généralement associés aux filtres comme générateur de résidus. Ainsi, la signature

directionnelle de la faute la plus proche de la signature directionnelle des résidus sera admise comme responsable de l'état de fonctionnement anormal du système.

III Méthodes data-driven

Contrairement aux méthodes nécessitant un modèle mathématique représentant le système, les méthodes à base de données (data-driven methods) s'appuient uniquement sur les données collectées sur le système (Venkatasubramanian et al., 2003). Elles s'en servent efficacement pour expliquer une nouvelle observation provenant du système.

Ces méthodes sont considérées comme une alternative aux méthodes à base de modèle. Elles ont fait l'objet de plusieurs simulations et comparaison sur plusieurs cas d'applications ou benchmarks : TEP (Downs and Vogel, 1993), Damadics (Blimes, 2004), Wind turbine (Odgaard et al., 2013), etc. Toutefois, dans le cas des systèmes hyper-dynamiques ainsi que pour des systèmes à boucle fermée, cette classe de méthodes est moins performante que les méthodes à base de modèle (Ding, 2012). Cependant, les méthodes qui se basent sur un modèle analytique précis expliquant la dynamique du système ont parfois de la difficulté à établir le modèle. Or, la précision du modèle est un élément crucial pour prévenir les fausses alarmes et les mauvaises décisions. En effet, pour des systèmes complexes ou hybrides cela peut être une mission difficile, très coûteuse, longue et même parfois impossible. En d'autres termes, le développement de méthodes de surveillance à base de modèle nécessite un effort considérable (Isermann, 2006). Ceci explique l'intérêt porté sur les méthodes data-driven ces dernières années (Ruz-Hernandez et al., 2007; Isermann, 2006; Kempowski, 2004; Venkatasubramanian et al., 2003; Chiang et al., 2004, 2001). La surveillance à base de données peut être décomposée en deux étapes : détection et diagnostic. Dans ce qui suit, nous allons présenter quelques méthodes à base de données proposées pour la détection de fautes suivies par d'autres pour le diagnostic.

III.1 Détection

Plusieurs méthodes ont été proposées pour la détection de fautes à base de données. Ces méthodes sont considérées comme des méthodes statistiques

simples mais très puissantes. La majorité de ces méthodes mettent en évidence les relations existantes entre les variables du système sans formuler de façon explicite son modèle (relations entre les variables d'entrées et sorties du système).

Analyse en Composantes Principales

Parmi les méthodes dites data-driven, l'une des plus utilisées est l'Analyse en Composantes Principales (ACP). L'analyse en composantes principale est une technique statistique linéaire multivariée assez répandue (Ding, 2011; Jackson, 2005; Tipping and Bishop, 1999; Jolliffe, 2005; Hotelling, 1933) employée dans divers domaines pour des usages variés. Elle est utilisée pour la réduction de dimension des données (pré-traitement) ou alors comme un outil de modélisation en exprimant les dépendances existantes entre les variables d'un ensemble de données multivariées.

Soit $X \in \mathbb{R}^{N \times m}$, un ensemble normalisé (pour la définition des directions et l'unification des amplitudes) de N échantillons collectés sur un ensemble de variables d'entrées et sorties $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_m]^T$ suivant une distribution normale, d'un système donné. La matrice de variance-covariance Σ de \mathbf{x} est donnée par :

$$\Sigma \approx \frac{1}{N-1} X^T X \quad (3.12)$$

Ceci correspond à une estimation sans biais de la matrice de variance-covariance (ajustement du résultat biaisée obtenu par le maximum de vraisemblance, vu dans le chapitre précédent).

L'analyse en composantes principales peut être définie formellement de différentes manières. Une première définition, la plus répandue, consiste à définir l'analyse en composantes principales comme une technique recherchant les axes principaux permettant d'expliquer la quasi-totalité de la variabilité perçue dans l'ensemble d'observations multivariées X . En d'autres termes, l'analyse en composantes principales cherche à déterminer un nombre a d'axes orthogonaux, avec $a < m$, de sorte à projeter orthogonalement X dans un espace linéaire de dimension réduite maximisant sa variance tout en minimisant l'effet du bruit. Les axes définissant cet espace sont appelés axes principaux ou composantes principales. Pour les déterminer, il suffit de trouver les points stationnaires résolvant le problème d'optimisation (3.13) et appartenant à une

matrice de passage projetant une observation $x \in \mathbb{R}^m$ dans le nouvel espace orthogonal.

$$\max_{p \neq 0} \frac{p^T X^T X p}{p^T p} \quad (3.13)$$

où $p \in \mathbb{R}^m$.

Une seconde définition de l'analyse en composantes principales la décrit comme une projection linéaire dans un sous espace (principal) minimisant la somme des moindres-carrés des erreurs de projection. Résoudre ce problème d'optimisation consiste à chercher une reconstruction optimale \hat{x}_n de x_n , un échantillon de l'ensemble X ,

$$\min \sum_{n=1}^N \|x_n - \hat{x}_n\|^2 \quad (3.14)$$

en maximisant $\hat{x}_n, \hat{x}_n = \hat{A}t_n + \tilde{A}e$ étant donné \hat{A}, \tilde{A}, e et t_n qui sont des paramètres à définir, avec e est supposé constant pour les N observations.

Formellement, que ce soit la première ou la seconde définition, les deux problèmes d'optimisation convergent vers des solutions communes. L'une des solutions possibles est de décomposer la matrice de variance-covariance de X en valeurs propres à l'aide d'une décomposition en valeurs singulières (singular values decomposition (SVD)) comme suit :

$$\Sigma = P \Lambda P^T, \quad (3.15)$$

$$P = [P_1 \dots P_m], P_j \in \mathbb{R}^m, P^T = [B_1^T \dots B_m^T], B_i^T \in \mathbb{R}^m, P P^T = I, \quad (3.16)$$

$$\Lambda = \text{diag}(\sigma_1^2, \dots, \sigma_m^2), \sigma_1^2 \geq \dots \geq \sigma_m^2 > 0 \quad (3.17)$$

où P est la matrice des vecteurs propres, axe géométrique orthogonal $P_j, j = 1, \dots, m$, et $B_i, i = 1, \dots, m$ sont ses lignes. $\sigma_j^2 = \lambda_j$ sont les valeurs réelles propres non-négatives correspondant à la matrice de variance-covariance de X . A chaque valeur propre correspond un vecteur propre, dont elle explique la variabilité. La matrice Λ comporte l'ensemble des valeurs propres. Cette matrice est diagonale ce qui explique l'orthogonalité des axes et la non-corrélation des variables lui étant associées.

Une fois les m axes déterminés, un autre problème réside en la détermination du nombre a , le nombre dominant a des vecteurs propres de Σ dans

lequel la variance retenue lors de la projection est maximale (les axes principaux $\hat{P} \subset P$ correspondant aux plus grandes valeurs propres $\hat{\Lambda}$ classées par ordre décroissant de l'axe ayant la plus grande variabilité (la plus grande valeur propre) à l'axe en capturant le moins (la plus petite valeur propre)), comme ci-dessous :

$$\lambda_1 > \lambda_a > \dots > \lambda_m \quad (3.18)$$

Fixer une valeur pour a peut se faire de multiples façons, le lecteur, pour plus de détails, pourra consulter (Mnassri, 2012; Bishop, 2006; Valle et al., 1999). Sa détermination permet de diviser Λ en deux parties, et notamment P en deux matrices de passages associées à leur variabilité, l'une représentant l'espace réduit et l'autre l'espace principale, comme ci-dessous :

$$\Lambda = \begin{bmatrix} \hat{\Lambda} & 0 \\ 0 & \tilde{\Lambda} \end{bmatrix}, \tilde{\Lambda} = \text{diag}(\sigma_{a+1}^2, \dots, \sigma_m^2), \quad (3.19)$$

$$P = [\hat{P} \tilde{P}] \in \mathbb{R}^{m \times m}, \hat{P} \in \mathbb{R}^{m \times a}, \quad (3.20)$$

$$P^T = [B_1^T \dots B_m^T], B_i^T = [\hat{B}_i^T \tilde{B}_i^T], \hat{B}_i^T \in \mathbb{R}^a \quad (3.21)$$

Ainsi, l'analyse en composantes principales permet de projeter linéairement et orthogonalement X , l'espace original dans un autre espace orthogonal Z , comme ceci :

$$Z = XP \quad (3.22)$$

où les colonnes orthonormales de $P \in \mathbb{R}^{m \times m}$ sont ses axes.

Cette projection (rotation orthonormal) permet la séparation de l'espace original des observations en deux parties (sous-espaces) $X = \hat{X} + \tilde{X}$: une partie principale $\hat{X} = \hat{Z}\hat{P}^T$, où $\hat{P} \in \mathbb{R}^{m \times a}$, et une partie résiduelle $\tilde{X} = \tilde{Z}\tilde{P}^T$, où $\tilde{P} \in \mathbb{R}^{m \times (m-a)}$. \hat{Z} et \tilde{Z} pouvant être déduites de la sorte :

$$X = \hat{X} + \tilde{X} = \hat{Z}\hat{P}^T + \tilde{Z}\tilde{P}^T, Z = [\hat{Z} \tilde{Z}], \quad (3.23)$$

$$\hat{Z} = X\hat{P} = [Z_1 \dots Z_a] = [XP_1 \dots XP_a], \hat{Z} \in \mathbb{R}^{N \times a}, \quad (3.24)$$

$$\tilde{Z} = X\tilde{P} = [Z_{a+1} \dots Z_m] = [XP_{a+1} \dots XP_m], \tilde{Z} \in \mathbb{R}^{N \times (m-a)} \quad (3.25)$$

où \hat{X} est le résultat de la transformation arrière de X , pouvant être également

écrite de la sorte $X\hat{P}\hat{P}^T$, et correspond aux échantillons de l'ensemble des variables d'origine sans bruit.

Une fois les sous-espaces principal et résiduel identifiés, l'analyse en composantes principales peut être utilisée pour la détection de fautes ([Daszykowski, 2007](#)). Elle opère de manière relativement semblable aux méthodes à base de modèle, en modélisant le comportement du système en fonctionnement normal où les fautes sont détectées en comparant le comportement observé et celui donné par le modèle. En d'autres termes, ceci consiste à surveiller les deux espaces générés par l'analyse en composantes principales. En effet, une faute peut affecter un des deux. D'ailleurs, un espace peut être sensible à des fautes alors que l'autre ne l'est pas. Pour surveiller ces espaces, plusieurs façons de faire existent. Selon [Isermann \(2006\)](#), la détection de fautes avec analyse en composantes principales peut se faire généralement de trois façons

- détection d'un changement dans l'espace principal,
- détection d'un changement dans l'espace réduit,
- détection par la comparaison des reconstructions de la variable une fois projeté dans l'espace principal et réduit,

Ceci peut se faire en utilisant des tests statistiques. Selon ([Ding, 2011](#)), ces tests prennent une forme quadratique. Les plus utilisées sont ceux employant les statistiques T^2 de Hotelling ([Tracy et al., 1992](#)) et SPE (Squared Prediction Errors, [Jackson and Mudholkar \(1979\)](#)). Cependant d'autres statistiques ont été proposées correspondant à leurs améliorations et/ou associations ([Mnassri, 2012](#); [Ding, 2011](#); [Qin and Li, 2001](#)).

Ces statistiques, avec les limites de contrôles qui leurs sont associées, détectent différents types de fautes, et leurs avantages peuvent être utilisés en les employant ensemble tirant ainsi avantage de chacune ([Chiang et al., 2001](#)). La statistique T^2 permet de détecter les observations éloignées des données alors que la statistique SPE détecte les résidus très grands par rapport au modèle de l'analyse en composantes principales ([Daszykowski, 2007](#)).

La détection utilisant ces deux statistiques peut se faire en définissant en premier les limites de contrôle (LC) pour SPE et la statistique T^2 pour un niveau de signification donné α , comme ceci :

$$\text{SPE} : LC_{\text{SPE}} = \theta_1 \left(\frac{g_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{\frac{1}{h_0}}, \quad (3.26)$$

$$\theta_i = \sum_{j=a+1}^m \sigma_j^{2i}, i = 1, 2, 3, h_0 = 1 - \frac{2\theta_1 \theta_3}{2\theta_2^2}$$

où g_α est la déviation de la loi normale correspondant au quantile $1 - \alpha$,

$$\text{T}^2 : LC_{\text{T}^2} = \frac{a(N^2 - 1)}{N(N - a)} F_\alpha(a, N - a) \quad (3.27)$$

où F_α est la distribution de Fisher. Si N est élevé la limite de T^2 peut être approximée par la loi du χ^2 .

Ensuite pour chaque nouvel échantillon normalisé, étant donnée la moyenne et les variances de X , $x = [x_1, \dots, x_m]^T$, on calcule les deux statistiques SPE et T^2 comme ci-dessous :

$$\text{T}^2 = x^T \hat{P} \hat{\Lambda}^{-1} \hat{P}^T x = \hat{t}^T \Lambda^{-1} \hat{t}, \hat{t} = \hat{P}^T x, \quad (3.28)$$

$$\text{SPE} = x^T \tilde{P} \tilde{P}^T x = \tilde{t}^T \tilde{t}, \tilde{t} = \tilde{P}^T x \quad (3.29)$$

où T^2 est associée aux composantes principales $\hat{t} \in \mathbb{R}^a, \hat{t} = \hat{P}^T \mathbf{x}$ et SPE aux dernières $\tilde{t} \in \mathbb{R}^{(m-a)}, \tilde{t} = \tilde{P}^T \mathbf{x}$.

Finalement, afin de détecter les fautes système, les images de chaque observation dans les deux espaces sont calculées et comparées à leurs limites de contrôle. Ainsi, une décision peut être prise en utilisant la règle logique suivante : si $\text{SPE} \leq LC_{\text{SPE}}$ et $\text{T}^2 \leq LC_{\text{T}^2}$ alors le système est déclaré sous contrôle, sinon le système est en faute.

Toutefois, l'analyse en composantes principales est connue pour être sensible aux observations aberrantes. En effet, la présence de quelques observations aberrantes dans un échantillon peut influencer considérablement le calcul de sa moyenne et sa matrice de variance-covariance, paramètres dont elle dépend. L'analyse en composantes principales est également limitée par sa nature linéaire (ne gère pas les relations non-linéaires), déterministe (l'absence d'une représentation probabiliste) et statique (considère que les données sont indépendantes et identiquement distribuées), et l'hypothèse de normalité émise sur les variables faisant parties de son modèle.

Ces insuffisances prêtées à l'analyse en composantes principales ont inté-

ressé un bon nombre de chercheurs. En effet, dans la littérature on peut trouver un nombre considérable de méthodes proposant des améliorations à l'analyse en composantes principales. Par exemple, plusieurs méthodes ont été proposés afin de rendre l'analyse en composantes principales robuste aux observations aberrantes. Dans ([Chen and Sun, 2009](#)), les auteurs discriminent et distinguent ces méthodes comme ceci :

- une estimation robuste de la matrice de variance-covariance (e.g. des poids sont associés aux observations).
- une catégorie basée sur la poursuite de projection (une stratégie de recherche séquentielle d'axes de projection maximisant une mesure robuste de variance).
- émettre d'autres hypothèses sur la distribution des variables. Par exemple, remplacer la distribution Gaussienne par la distribution de Student (t-distribution) qui est, étant donnée sa densité, plus robuste aux observations aberrantes.

D'autres méthodes ont également été proposées afin de traiter l'absence d'un modèle génératif, les données manquantes (e.g. erreurs humaines lors de la saisie des données, capteurs hors service, impossibilité de prendre des mesures à un moment donné, etc.), le cas dynamique, le cas non Gaussien et multi-modes (e.g. l'état de fonctionnement correspond à plusieurs modes de fonctionnement), et le cas non linéaire.

D'autres méthodes dans le même principe que l'analyse en composantes principales, peuvent également être utilisées (voir e.g. ([Tiplica, 2012](#); [Yin et al., 2012](#); [Ding, 2011](#); [Chiang et al., 2001](#))).

Méthodes des sous-espaces

Uniquement à partir des données, certaines méthodes peuvent être utilisées pour estimer le modèle du système. Ces méthodes sont dites de sous-espaces ([Qin, 2006](#); [Favoreel et al., 2000](#); [Ljung, 1998](#)) également connues sous la notation SMI (Subspace Model Identification) ou encore SIM (Subspace Identification Methods). Elles correspondent à des algorithmes linéaires d'identification, mis au point pour faire face aux problèmes de construction d'un modèle analytique précis pour des systèmes complexes. Dans ([Van Overschee and De Moor, 1996](#)), les auteurs affirment que ces méthodes de sous-espaces peuvent effectivement approximer un système donné par des modèles suffisamment précis

(la précision du modèle est un facteur important pour la performance de tout système de surveillance afin d'éviter les fausses alarmes et un diagnostic erroné)

Les SMI sont également considérés comme une alternative aux méthodes d'erreur de prédiction (PEM) (McKelvey, 1995). Ces méthodes sont plus rapides (n'effectuent aucune recherche non linéaire) mais en pratique elles sont plutôt moins précises que les PEM. Toutefois, les deux méthodes ne sont pas considérées comme des opposants mais plutôt comme deux méthodes complémentaires. En effet, pour une estimation d'un modèle, les SMI peuvent se charger de fournir l'estimation initiale des paramètres du modèle, qui sera ensuite optimisée grâce aux PEM. Une comparaison approfondie entre les deux méthodes est effectuée dans (Favoreel et al., 2000).

Les méthodes des sous-espaces connues pour leur simplicité et stabilité permettent de modéliser le système sous une représentation d'état (Wang and Qin, 2002). Cette structure est largement utilisée pour la surveillance des méthodes à base de modèle. Elle est également une structure très pratique, soit pour l'estimation, le filtrage, la prévision ou pour le contrôle.

Les méthodes SMI génèrent un modèle utilisé uniquement pour les régions d'opérations correspondant aux données disponibles. De plus, pour garantir un modèle suffisamment précis pour des systèmes dynamiques multivariables, il est souhaitable que les données choisies pour identifier un modèle SMI découlent, dans le cas idéal, des expériences conçues, garantissant un niveau adéquat d'excitation (Schubert et al., 2011; Van Overschee and De Moor, 1996). Ces méthodes sont en général dédiées aux systèmes linéaires discrets dynamiques invariants au temps. Soit la représentation d'état suivante :

$$\left\{ \begin{array}{l} \mathbf{z}(t+1) = A\mathbf{z}(t) + B\mathbf{u}(t) + w(t) \\ \mathbf{y}(t) = C\mathbf{z}(t) + D\mathbf{u}(t) + v(t) \\ E \left[[w(p)v(p)]^T [w(q)^T v(q)^T] \right] = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix}, \delta_{pq} \geq 0 \end{array} \right. \quad (3.30)$$

où $v(\cdot)$ et $w(\cdot)$ sont deux vecteurs non observables correspondant à des bruits blancs stationnaires de moyenne nulle et non corrélés avec les entrées $\mathbf{u}(\cdot)$. Les matrices Q et S appartiennent à la matrice de covariance des deux vecteurs $v(\cdot)$ et $w(\cdot)$. δ_{pq} correspond au symbole de Kronecker qui est égale à 1 si $p = q$.

Une équation d'entrées et sorties correspondant à cette représentation d'états

et décrivant la relation linéaire existante entre la matrice représentant les sorties futures \mathbf{Y}_f et les entrées futures \mathbf{U}_f , étant donnée une fenêtre de temps $h + 1$, est donnée ci-dessous :

$$\mathbf{Y}_f(t) = \Gamma \mathbf{Z}_h + H_u \mathbf{U}_f + H_m M_f + N_f \quad (3.31)$$

où $\mathbf{Y}_f, \mathbf{U}_f$ sont, respectivement, les matrices de Hankel des sorties et entrées futures, Γ est la matrice d'observabilité étendue, \mathbf{Z}_h est la séquence d'état à estimer, H_u est la matrice déterministe triangulaire et inférieure de Toeplitz et H_m la matrice stochastique, triangulaire et inférieure de Toeplitz, M_f et N_f décrivent deux matrices de Hankel représentant respectivement l'impact du bruit du système et celui des mesures.

Le but des méthodes SMI est, par exemple, d'utiliser les données collectées sur le système pour trouver une estimation de la matrice étendue Γ et/ou une estimation de la séquence d'état $\hat{\mathbf{Z}}_h$ du modèle inconnu représentant le comportement dynamique du système. Cette étape est commune à la plupart des algorithmes d'identification des sous espaces (Le Rohellec, 1998). Par la suite, selon l'estimation de l'une ou de l'autre, on peut distinguer deux classes d'algorithmes : les uns se basent sur la séquence d'état du modèle pour estimer les matrices du modèle A, B, C et D (on peut citer N4SID (Numerical Algorithms For Subspace state space System Identification, Van Overschee and De Moor (1996)), CVA (Canonical Variate Aanalysis, Larimore (1990))), tandis que d'autres se basent sur la matrice d'observabilité étendue, on peut citer MOESP (MIMO Output-Error State sPace model identification, (Verhaegen and Dewilde, 1992)). Ces deux classes d'algorithmes permettent également d'estimer l'ordre du modèle, et les paramètres statistiques décrivant les caractéristiques du bruit du système $w(\cdot)$ et du bruit de mesure $v(\cdot)$.

Les méthodes d'identification des sous espaces peuvent être utilisées pour la surveillance des systèmes de différentes manières (Ding, 2014, 2012; Wang et al., 2011; Lieftucht et al., 2009; Dong and Verhaegen, 2009; Kulcsar et al., 2009; Huang et al., 2008; Qin and Li, 2001). La façon la plus naturelle est de les utiliser comme une alternative aux méthodes à base de modèle, pour faire face à la difficulté de concevoir des modèles mathématiques pour des systèmes multi-physiques. Les modèles sont alors estimés à partir des données d'entrées et de sorties et seront ainsi utilisés par des méthodes de surveillance à base de modèle. Ces méthodes à base de modèle, comme vues précédemment,

produisent des résidus issus de la comparaison des mesures estimées lors du fonctionnement normal du système et celles observées. Ensuite, ces résidus (considérés comme indépendants) sont évalués pour décider si le système est en état de fonctionnement normal ou en état de faute. Pour évaluer ces résidus, une analyse monovariée est utilisée (des seuils de tolérance ou des méthodes statistiques monovariées pour optimiser le compromis entre les fausses alarmes et les mauvaises détections). Cependant, lors de l'utilisation des méthodes à base de modèle se basant sur le SMI, des dépendances peuvent avoir lieu entre les résidus. Cela est dû aux incertitudes sur les dynamiques non-modélisées dans le modèle estimé par le SMI. Ainsi, avant d'effectuer une localisation, une analyse multivariée de ces résidus est conseillée pour garantir une bonne détection des fautes survenues sur le système. Pour cela, des méthodes utilisant la statistique T^2 multivariée peuvent être utilisées ([Schubert et al., 2011](#)).

D'autres méthodes utilisant l'analyse en composantes principales et le SMI ont été introduites, contribuant à l'amélioration de la surveillance à base de données. On peut notamment citer les travaux de ([Treasure et al., 2004](#)) qui utilise tout d'abord le SMI pour estimer la séquence d'état décrivant le comportement dynamique du système. Ensuite, l'analyse en composantes principales est appliquée à un ensemble de données contenant la séquence d'état, les entrées et les sorties. Ainsi, une réduction de dimension est effectuée et par conséquent des composantes principales sont déterminées dans lesquelles la variation à la fois du système et des variables d'état sont capturées permettant la surveillance en utilisant les statistiques T^2 et SPE.

Contrairement aux schémas précédent, dont les méthodes de sous espace sont une étape à part entière dans la conception, une nouvelle stratégie de surveillance a été récemment proposée par ([Ding, 2014, 2012; Ding et al., 2009](#)). Les auteurs suggèrent l'incorporation des méthodes sous espace dans la conception des méthodes à base de modèle. En d'autres termes, ceci consiste, par exemple, à définir un générateur de résidus basé sur les observateurs ou l'espace de parité directement à partir des données, en s'inspirant de l'étroite relation entre les méthodes sous-espace et l'espace de parité, et ce dernier et l'analyse en composantes principales. En s'appuyant sur les équations de parité générées à partir des données on peut déduire plusieurs types d'observateurs (à boucle ouverte ou fermée).

III.2 Diagnostic

Lorsqu'un défaut est détecté, il est nécessaire d'identifier la ou les variables qui sont en cause ou le type de faute. Pour cela, plusieurs méthodes à base de données ont été proposées. Généralement, on les associe aux méthodes statistiques de détection vu précédemment. Ces méthodes de diagnostics peuvent être distinguées en trois approches : l'approche par structuration des résidus, l'approche utilisant des bancs de modèles (e.g. principe de reconstruction) et l'approche des contributions aux indices de détection (exemples d'indices de détection : T^2 , SPE, etc.).

D'autres méthodes de diagnostics peuvent également être utilisées, comme celles se basant sur les méthodes de classification. En effet, le problème de diagnostic peut être formulé, pour une observation donnée comme un problème de discrimination entre les différentes fautes du système. Ce dernier consiste à définir les frontières entre différentes classes (fautes) de données et ensuite, en utilisant un classifieur, décider à quelle classe correspond le mieux une nouvelle observation. La première étape de la classification correspond à l'apprentissage. La seconde étape est la prise de décision.

Trois types de problèmes de décisions peuvent être distingués ([Chiang et al., 2004](#)). Les deux premiers : lorsque les classes sont séparées de façon linéaire et non-linéaire. Ces deux problèmes sont considérés comme les plus simples car il est facile de classer correctement la majorité des données. Le troisième problème de décision concerne le cas où les classes sensées représenter les données se chevauchent. Ce dernier cas est loin d'être simple à traiter, il complique la décision sur le classifieur à choisir. En effet, malgré le fait que plusieurs méthodes de classification (classifieurs) ont été proposées dans la littérature, un classifieur optimal permettant de classer une observation correctement dans tous les cas de figures possibles n'a pas à notre connaissance vu le jour. Ainsi, pour une situation donnée le choix d'un classifieur dépend de plusieurs facteurs (e.g. connaissance de la nature du problème à traiter et/ou des indices de performance représentatifs des résultats de classification lors des essais, etc.)

Aussi, pour assurer une bonne généralisation (des classifieurs souples qui généralisent leur ensemble d'apprentissage et évitent le sur-apprentissage) et tenir compte du bruit, un choix judicieux de l'espace de représentation doit être effectué. Il est conseillé d'effectuer un pré-traitement de données ([Chiang et al., 2001](#)) ainsi que de recourir à des méthodes de réduction de dimensions

de données. Ces dernières permettent l'obtention de régions de décisions plus représentatives mais ne garantissant pas toujours une meilleure classification (Bouveyron, 2013). Une meilleure solution est de sélectionner des variables étant donné le problème et le classifieur utilisé. Ceci peut être utile lorsque la dimension du problème considéré est très grande. D'ailleurs, ceci assure de meilleurs résultats qu'une réduction de données. Notons, par exemple, que la méthode statistique dite d'analyse discriminante de Fisher (et ses différentes déclinaisons) est parfaitement adaptée à la classification supervisée. Elle permet de réduire la dimension du système sous certaines contraintes (Fukunaga, 1990) tout en préservant les propriétés discriminantes des données.

Il est important de mentionner que contrairement à la classification, le nombre de classes dans le diagnostic (les classes représentant les différentes fautes) n'est pas fixé. En effet, une connaissance exhaustive du système est parfois impossible (e.g. des modes de fonctionnement inconnues, des modes de fonctionnement dangereux). Ainsi des méthodes dites évolutives (Dubuisson, 2001) intégrant des notions de rejets doivent remplacer les méthodes usuelles de classification pour être appliquées pour le diagnostic. On distingue deux notions de rejet :

- le rejet d'ambiguïté : il intervient lorsqu'une observation est affectée à une zone commune à plusieurs modes de fonctionnement. Ainsi soit un nouveau mode de fonctionnement est créé temporairement soit on fait appel à un expert pour prendre une décision.
- le rejet de distance : il intervient lorsqu'une observation ne correspond pas aux modes de fonctionnements existants. Ainsi, une décision devrait être prise concernant la création d'un nouveau mode de fonctionnement.

III.3 Combinaison des méthodes à base de données et des méthodes à base de modèle

Comme vu précédemment, plusieurs méthodes de détection et de diagnostic ont été proposées dans la littérature. Chacune possède ses avantages et ses inconvénients. Généralement, une méthode fonctionnant bien dans une circonstance pourrait ne pas bien fonctionner dans une autre lorsque les caractéristiques du système différent. Manifestement, il est difficile de concevoir une méthode parfaite surveillant efficacement un système dans tous les scénarios probables. Ceci dit, il semble évident que les méthodes de surveillance peuvent

être plus ou moins complémentaires. D'ailleurs, certains chercheurs disent que la création d'un cadre utilisant les deux classes de méthodes de surveillance permettrait une amélioration de la surveillance des systèmes (Ding et al., 2009; Venkatasubramanian et al., 2003). En d'autres termes, un cadre de travail se basant simultanément sur les données et le modèle du système, en utilisant une combinaison des deux méthodes permettrait de bénéficier des avantages des deux méthodes et de lutter contre les lacunes individuelles de chacune lorsqu'elles sont utilisées séparément. En effet, la capacité des méthodes à base de données à gérer un nombre important de données dans les systèmes complexes, associée à la capacité des méthodes à base de modèles (décrivant bien le comportement dynamique du système et offrant une compréhension physique du système) permettrait d'assurer une meilleure surveillance, et de lutter contre les imperfections de chacune lorsqu'un manque de précision sur les modes de fonctionnement existe ou un modèle précis du système est difficile à obtenir.

Cependant, une combinaison des deux méthodes, malgré le potentiel qu'elle peut offrir, reste largement inexplorée. En effet, peu d'articles dans la littérature s'intéressent à ce sujet. La majorité des contributions des chercheurs dans la littérature est focalisée sur le développement ou l'amélioration d'une des deux méthodes de surveillance.

Dans (Schubert et al., 2011), un schéma de détection et d'isolation unifiant les méthodes à base de modèles et les méthodes à base de données et combinant leurs avantages a été proposé. En effet, ce schéma consiste principalement à intégrer des méthodes à base de données dans des méthodes à base de modèle et cela pour la détection et la localisation des fautes capteurs, actionneurs, ou du système. Dans un premier temps, le modèle dynamique du système est identifié grâce à une méthode à base de données (méthodes des sous espace). Le modèle, une fois identifié, est utilisé différemment selon les fautes prises en compte. Concernant les fautes survenues sur les capteurs, un observateur de type UIO (Unknown Input Observer) est utilisé pour chaque capteur. Chaque UIO est sensible à toutes les fautes exceptée la faute du capteur qu'elle surveille (une méthode à base de modèle). Les résidus générés sont alors examinés à l'aide d'une statistique SPE. Lors de la présence d'une faute capteur, tous les résidus signalent un fonctionnement anormal du système, sauf un seul. Celui-ci étant conçu pour être robuste à la faute détectée, cela permet de localiser le capteur en état de fonctionnement anormal. Par ailleurs, une méthode de reconstruction des entrées est utilisée pour prendre en compte les fautes issues

des actionneurs et du système. Les résidus obtenus correspondent à la différence entre les entrées reconstruites et les entrées mesurées. Après avoir généré les résidus, une carte de contrôle multivarié T^2 est appliquée sur le vecteur des résidus dans le but de détecter une éventuelle faute survenue sur le système. Lors de l'apparition d'une faute, les résidus s'éloignent considérablement de leur moyenne. Les résidus conçus par cette méthode offrent une capacité d'explication, qui peut être utile à l'opérateur ou l'ingénieur pour localiser la cause de la faute. Le schéma proposé par les auteurs permet ainsi de contourner les faiblesses des méthodes à base de modèle (difficultés de modélisation du système) et celle des méthodes à base de données.

Dans (Ghosh et al., 2011), plusieurs méthodes de surveillance ont été utilisées simultanément. Une fusion des décisions de ces méthodes est effectuée. Dans cet article, une philosophie similaire à celles des classifieurs multiples en reconnaissance des formes est utilisée. L'idée est qu'en fusionnant des décisions de différentes méthodes de surveillance, s'employant dans des scénarios différents et où la performance de chacune est insuffisante, cela va aboutir à une meilleure surveillance. Par exemple, pour la surveillance d'une colonne de distillation de laboratoire, les auteurs utilisent quatre méthodes de surveillance : une méthode à base de modèle et trois méthodes à base de données. Un filtre de Kalman étendu comme méthode à base de modèle, SOM pour (Self Organized Map), un réseau de neurones artificiels et enfin l'analyse en composantes principales. Chaque méthode permet de détecter et de diagnostiquer les fautes survenues sur le système. Ces quatre méthodes employées ont accès à la même information (les entrées sont toutes les mêmes pour chaque méthode). La sortie de chaque méthode correspond à une assignation à une classe de faute. Une stratégie de fusion est alors appliquée pour prendre la meilleure décision issue des quatre méthodes. Une autre approche hybride a été proposée dans (Luo et al., 2010). Les auteurs proposent d'utiliser les équations de parité et un observateur non-linéaire pour générer les résidus. Une fois les résidus générés, des tests statistiques sont utilisés pour détecter et isoler, avec l'aide de machines à vecteurs de support, les différentes fautes d'un système de freinage antiblocage.

IV Conclusion

Dans ce chapitre, nous avons abordé la détection et le diagnostic de fautes dans les systèmes multivariés. Cela nous a permis de présenter de façon non-

exhaustive les différentes méthodes utilisées. Nous avons vu que ces méthodes pouvaient être décomposées en deux familles : les méthodes à base de données et les méthodes à base de modèle. Les méthodes à base de données se basent, comme l'indique leurs noms uniquement sur les données collectées sur le système tandis que les méthodes à base de modèle s'appuient sur le modèle analytique, représentation mathématique, du système. Chacune d'entre elles possède ses avantages et ses inconvénients. Généralement, une méthode qui fonctionne bien dans une circonstance pourrait ne pas bien fonctionner dans une autre lorsque les caractéristiques du système différent. Ainsi, comme de nombreux chercheurs le suggèrent, combiner ces méthodes permettrait une amélioration dans la prise de décision. En d'autres termes, une combinaison des deux méthodes permettrait de bénéficier de leurs avantages et de lutter contre les lacunes individuelles de chacune lorsqu'elles sont utilisées séparément. Nous avons vu que cet axe de recherche reste inexploré malgré les travaux prometteurs proposés. En effet, la majorité des contributions dans le contexte de la détection et diagnostic de fautes est focalisée sur le développement ou l'amélioration des méthodes issues ou faisant partie de ces deux familles de méthodes.

Détection par réseau Bayésien

I Contexte et objectifs

Plusieurs méthodes ont été proposées pour la détection de fautes dans les systèmes. Ces méthodes, se basant pour la plupart sur les observations et/ou les résidus (générés à partir du modèle du système ou des données collectées et leurs transformations), sont généralement associées à des statistiques quadratiques leur permettant la prise de décision. Nous appellerons cette association un schéma de détection. Dans ce qui suit, nous allons proposer un cadre commun à différents schémas de détection. Concrètement, nous proposons de représenter ces schémas de détection sur un seul outil, un réseau Bayésien. Le réseau Bayésien est un outil puissant permettant de modéliser l'incertitude ainsi que les relations de régression et de causes à effets pouvant exister entre variables. Il offre un cadre probabiliste et statistique permettant la modélisation et la prise de décision. Tout cela a amené certains chercheurs à s'intéresser de plus en plus aux réseaux Bayésiens et à tester leurs potentiels que ce soit pour la détection ou le diagnostic de fautes. Dans ce chapitre nous nous intéressons particulièrement à la détection de fautes.

En effet, aujourd'hui, énormément de travaux proposent d'utiliser les réseaux Bayésiens pour la détection. ([Rojas-Guzman and Kramer, 1993](#)) propose un réseau Bayésien pour la détection de fautes dans les capteurs. Le réseau

Bayésien comporte plusieurs noeuds discrets. Ces noeuds modélisent la différence entre la valeur vraie d'un paramètre surveillé et sa valeur fournie par le capteur. Une grande différence entre ses valeurs permet d'accroître la probabilité de présence d'une anomalie dans le capteur. Une extension de ce modèle est donnée dans (Heger and Aradhya, 2002; Aradhya, 1997). Dans (Hood and Ji, 1997), les auteurs emploient les réseaux Bayésiens pour la détection de fautes (e.g. crash dans un serveur de fichiers) dans les réseaux informatiques. Ils utilisent un réseau Bayésien, un arbre causal composé de noeuds discrets, comme un mécanisme pour combiner des informations issues de différentes variables sélectionnées, les plus discriminatives, dans le but de détecter des anomalies. La méthode proposée n'utilise pas de modèles de fautes spécifiques. La structure et les tables de probabilités conditionnelles du réseau sont apprises à partir des connaissances a priori (les différentes observations et les connaissances d'experts) résumant le comportement nominal des variables mesurables. Le réseau a été testé sur des données réelles contenant des crashes dans un serveur de fichier.

Dans (Lerner et al., 2000), les auteurs proposent un réseau Bayésien dynamique hybride pour la détection de fautes. La structure du réseau proposé est tout d'abord généré à partir d'un graphe causal temporel (TCG, Temporal Causal Graph). A ce stade-là, le réseau généré permet de modéliser l'état de fonctionnement normal du système considéré. Les auteurs dans ce travail se concentrent sur les systèmes composés de plusieurs sous-systèmes faiblement interconnectés. Ensuite, ils viennent augmenter le réseau obtenu par d'autres noeuds permettant de détecter les ruptures par éclatement, les fautes de mesures et les variations de paramètres. Dans (Neil et al., 2001), un réseau Bayésien a été proposé pour détecter les fautes capteurs. Dans (Schwall and Gerdes, 2002), les auteurs proposent un réseau Bayésien dynamique hybride pour modéliser le comportement temporel des fautes considérées, analyser et évaluer les résidus générés en présence d'un historique de fautes et d'une matrice d'incidence. Leur réseau a été testé sur un système de gestion de voitures. Dans (Matsuura and Yoneyama, 2004), un réseau Bayésien pour la détection des fautes capteurs est proposé. Leur réseau est composé de noeuds discrets représentant les entrées, les sorties et les variables d'états d'un observateur Luenberger d'ordre complet du système considéré. Les observations et les estimations de ces variables sont sauvegardées lors du fonctionnement normal du système et utilisées pour apprendre le réseau. La détection de fautes se fait en

examinant la probabilité jointe du réseau.

Dans (Mehranbod et al., 2005), les auteurs proposent des réseaux Bayésiens pour la détection des fautes capteurs. Ces réseaux correspondent à une amélioration des travaux effectués dans (Heger and Aradhye, 2002; Aradhye, 1997; Rojas-Guzman and Kramer, 1993). La méthodologie employée consiste en l'utilisation de réseaux Bayésiens, se basant sur les mêmes noeuds que ceux des réseaux Bayésiens présentés précédemment, augmentés par des indices permettant de détecter une défaillance de capteurs une fois dépassant des seuils empiriques. Dans (Kawahara et al., 2005), les auteurs proposent une méthode pour la détection de fautes dans un engin spatial en utilisant un réseau Bayésien dynamique. La méthode proposée s'appuie dans un premier temps sur les connaissances physiques (équations différentielles) et les connaissances d'experts (règles empiriques) pour construire le réseau et dans un deuxième temps sur les données récoltées pour venir le compléter ou le modifier par le biais de méthodes d'apprentissages statistiques.

Dans (Roychoudhury et al., 2006), les auteurs proposent une amélioration des travaux de (Lerner et al., 2000) à partir d'un modèle d'espace d'états déduit d'un graphe causal temporel, résultat de la transformation d'un Bond Graph (BG, voir Samantaray and Bouamama (2008)). Les auteurs, dans un premier temps, construisent un observateur sous forme de réseau Bayésien dynamique en faisant l'hypothèse que les données de fonctionnement normal sont disponibles (notamment pour l'estimation des paramètres). Puis, pour effectuer la détection, ils s'appuient sur le Z-test appliqué sur la moyenne des résidus calculés dans une fenêtre donnée. Les résidus sont la différence entre les observations et les estimations fournies par l'observateur. Dans (Nunnari et al., 2006), les auteurs proposent un réseau Bayésien original pour la détection de fautes. Le réseau Bayésien proposé permet de lier les noeuds correspondants aux entrées et sorties aux noeuds correspondants aux fautes (des noeuds softmax) par le biais d'une boîte noire composée de noeuds cachés. Il est employé sur un actionneur hydraulique.

Dans (Huang et al., 2008), les auteurs proposent un cadre probabiliste et une architecture générale pour la surveillance des systèmes. Parmi les modules faisant partie de cette architecture, un module de détection de fautes capteur. Les auteurs, à partir d'une écriture linéaire de la valeur fournie par le capteur, présentent une structure originale de réseau Bayésien. Le réseau est composé de six noeuds dont quatre d'entre eux représentent des noeuds binaires de décision.

Chaque noeud de décision représente une faute donnée (biais, changement de variance, etc.). Une faute donnée est donc présente lorsque la probabilité a posteriori de sa présence dans le noeud de décision lui correspondant est la plus grande.

Dans (Verron et al., 2010a), un réseau Bayésien pour l'identification des variables impliquées dans un dysfonctionnement du système à un instant donné a été proposé. Il est basé sur la modélisation (Verron, 2007) et la décomposition (Li et al., 2008) de la statistique T^2 sous réseau Bayésien. Dans (Zhang et al., 2011), les auteurs proposent un schéma de détection dont un réseau Bayésien dynamique fait partie parmi d'autres méthodes de modélisation (théorie de Bond graphe) et d'analyse de données (analyse en composantes principales et transformée en ondelettes discrète). Dans (Najafi et al., 2012), les auteurs proposent une méthode de détection de fautes basée sur les réseaux Bayésiens en présence de données de fautes. Leur méthode a été proposée dans le cadre des systèmes de traitement d'air. Les réseaux proposés peuvent être vus comme des réseaux Bayésiens classifieurs.

Dans (Gonzalez et al., 2012), les auteurs proposent une méthode de détection en ligne basée sur un réseau Bayésien dynamique, à la manière d'un Switched Kalman Filter, pour la détection des fautes grossières dans des systèmes d'extraction de sable pétrolifère. Dans (Yu and Rashid, 2013), les auteurs se basent sur un réseau Bayésien dynamique appris par le biais d'un historique de données collectées lors du fonctionnement normal du système, et d'un diagramme de flux. Pour la détection, les auteurs ont proposé un indice appelé indice de vraisemblance anormale (abnormality likelihood index). Le réseau Bayésien proposé a été testé sur le Tennessee Eastman Process.

La plupart de ces méthodes lors de la détection ne permettent pas d'assurer un niveau de signification α lors de leur prise de décision en fonction des probabilités a posteriori calculées. Dans ce chapitre, nous allons nous intéresser aux statistiques quadratiques. Nous allons voir comment un réseau Bayésien peut être un cadre probabiliste permettant de généraliser ces statistiques. Ensuite, nous nous intéresserons au modèle linéaire Gaussien. Nous illustrerons que sous cette écriture, nous pouvons généraliser et unifier énormément de schémas de détection sous des réseaux Bayésiens capables de donner une décision sur l'état de fonctionnement du système. En d'autres termes, nous proposons un réseau Bayésien permettant de généraliser les schémas de détection se basant sur une transformation linéaire des observations. Spécifiquement, nous mettons l'accent

sur celui associé à l'analyse en composantes principales.

II Détection par calculs des limites de surveillance

Précédemment, nous avons vu que pour détecter une faute, étant donnée une nouvelle observation, il fallait décider entre deux hypothèses : l'hypothèse de fonctionnement normal et l'hypothèse alternative (traduisant une faute). Pour cela, il fallait définir une limite permettant d'établir un périmètre de décisions. Cette limite peut être définie empiriquement, mais celle-ci peut ne pas garantir un risque de fausse alarme fixée. Pour cela, des tests statistiques s'imposent. Dans la plupart des cas, les méthodes utilisées pour la détection de fautes sont associées à des tests statistiques se basant sur des indices quadratiques. Vu leur importance et leur nécessité pour décider de l'état de fonctionnement d'un système donné, nous proposons de les unifier sous un cadre probabiliste.

Ce cadre que nous allons présenter va permettre de gérer des indices (statistiques) quadratiques univariées/multivariées (e.g. SPE, T^2 , etc.). Ces statistiques de formes quadratiques, regroupées sous la même notation Δ , sont largement utilisées pour la détection de fautes et sont également à la base des tests statistiques associés à l'analyse en composantes principales (Ding, 2010). Pour une nouvelle observation x de \mathbf{x} , elles sont calculées et comparées à leurs seuils prédéfinis (limite de contrôle (LC_{Δ}) permettant de respecter un seuil de signification α) pour discriminer entre deux états : CFN (Classe de Fonctionnement Normal) et F (classe représentant les Fautes). Si la statistique Δ est supérieure à LC_{Δ} alors le système est déclaré hors contrôle (F pour Fautes), sinon le système est sous contrôle (CFN).

En se basant sur un réseau conditionnel Gaussien classifieur binaire (discriminant entre deux classes CFN et F , deux états de la variable discrète Δ) ayant la même structure que le réseau présenté dans la figure 4.1, nous allons chercher une représentation probabiliste à ces statistiques. Ceci dit, étant donné une statistique Δ , nous avons besoin de définir les paramètres du réseau lui correspondant ainsi que la limite de contrôle probabiliste ζ_{Δ}^F qui lui est associée, de sorte à ce que : si $p(\Delta = F|\mathbf{x} = x) \geq \zeta_{\Delta}^F$ le système soit déclaré sous la classe F , avec $p(F|x)$ la probabilité a posteriori de la classe F étant

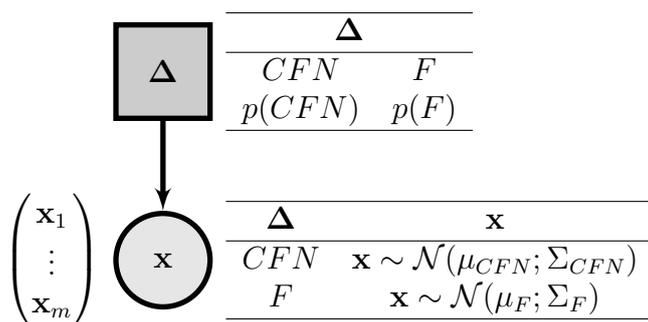


FIGURE 4.1 – Un réseau Bayésien classifieur pour la détection de fautes

donné x , et où $p(\Delta = F | \mathbf{x} = x) = 1 - p(\Delta = CFN | \mathbf{x} = x)$. Dans ce qui suit, nous nous référons à $\mathbf{x} = x$ par simplement x .

Les paramètres de la classe CFN (la distribution Gaussienne multivariée $p(\mathbf{x} | \Delta = CFN)$) sont estimés à partir des données de fonctionnement normal disponibles, utilisant par exemple l'estimation par maximum de vraisemblance. La classe F (la distribution Gaussienne multivariée $p(\mathbf{x} | \Delta = F)$), est quant à elle considérée comme une classe virtuelle représentant l'ensemble des observations ne pouvant pas être attribuées à la classe sous contrôle CFN . Ses paramètres sont définis de sorte à ce que $\mu_F = \mu_{CFN}$ et que Σ_F exprime plus de variabilité que Σ_{CFN} : $\Sigma_F = \mathbf{c}_\Delta \times \Sigma_{CFN}$, où $\mathbf{c}_\Delta > 1$ (si $\mathbf{c}_\Delta = 1$, les deux classes sont identiques, ce qui n'a pas de sens). Pour plus de simplicité, on note $\mu_\Delta = \mu_{CFN}$ et $\Sigma_\Delta = \Sigma_{CFN}$. Ainsi, le réseau Bayésien défini précédemment peut être redéfini comme le montre la figure 4.2.

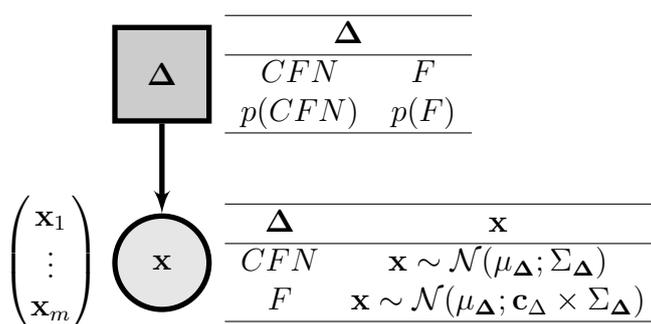


FIGURE 4.2 – Un exemple d'un réseau Bayésien pour la détection de fautes

Notons que cette différence entre les deux classes (le \mathbf{c}_Δ rendant la variance de F plus grande que celle de CFN) est très utilisée dans la littérature de la détection de fautes utilisant les réseaux Bayésiens (comme dans (Verron

et al., 2008b; Kawahara et al., 2005; Schwall and Gerdes, 2002; Lerner et al., 2000)), où la classe (CFN ou F) avec la plus grande probabilité a posteriori est choisie. Cependant, la valeur de \mathbf{c}_Δ est déterminée en utilisant les données de fautes qui sont généralement non disponibles ou alors en nombre insuffisant pour l'estimer précisément et éviter les fausses alarmes et mauvaises détections. Dans ce qui suit, au lieu de se préoccuper de la valeur exacte de \mathbf{c}_Δ , on cherche les limites de contrôle probabilistes ζ_Δ^{CFN} ou ζ_Δ^F , étant donné un $\mathbf{c}_\Delta > 1$, de sorte à ce que la règle de décision suivante soit maintenue :

$$x \in CFN : \text{ si } \Delta \leq LC_\Delta \quad (4.1)$$

sous une de ces règles de décisions :

$$x \in CFN : \text{ si } p(\Delta = CFN|x) \geq \zeta_\Delta^{CFN} \quad (4.2)$$

$$x \in CFN : \text{ si } p(\Delta = F|x) < \zeta_\Delta^F \quad (4.3)$$

Considérons (4.2) pour une observation x donnée, de sorte à ce que :

$$p(\Delta = CFN|x) = \zeta_\Delta^{CFN} \quad (4.4)$$

$$= \zeta_\Delta^{CFN} [p(\Delta = CFN|x) + p(\Delta = F|x)] \quad (4.5)$$

où $\zeta_\Delta^F = 1 - \zeta_\Delta^{CFN}$ et $p(\Delta = CFN|x) + p(\Delta = F|x) = 1$.

En se basant sur la formule de Bayes, la probabilité a posteriori de chaque classe (CFN ou F) peut être écrite comme ci-dessous :

$$p(\Delta|x = x) = \frac{p(\Delta)p(x|\Delta)}{p(x)}, \Delta \in \{CFN, F\} \quad (4.6)$$

A partir de (4.4) et (4.6) nous obtenons :

$$\frac{1}{p(x)} p(\Delta = CFN)p(x|\Delta = CFN) = \zeta_\Delta^{CFN} \left[\frac{p(\Delta = CFN)p(x|\Delta = CFN) + p(\Delta = F)p(x|\Delta = F)}{p(x)} \right] \quad (4.7)$$

$$\zeta_\Delta^{CFN} p(\Delta = F)p(x|\Delta = F) = p(\Delta = CFN)[p(x|\Delta = CFN) - \zeta_\Delta^{CFN} p(x|\Delta = CFN)] \quad (4.8)$$

Soit $\omega = \frac{p(\Delta=F)}{p(\Delta=CFN)}$, alors nous avons :

$$\begin{aligned}\zeta_{\Delta}^{CFN} \omega p(x|\Delta = F) &= p(x|\Delta = CFN) - \zeta_{\Delta}^{CFN} p(x|\Delta = CFN) \\ p(x|\Delta = CFN) &= \zeta_{\Delta}^{CFN} [p(x|\Delta = CFN) + \omega p(x|\Delta = F)] \\ \zeta_{\Delta}^{CFN} &= \frac{p(x|\Delta = CFN)}{p(x|\Delta = CFN) + \omega p(x|\Delta = F)}\end{aligned}\quad (4.9)$$

Le réseau Bayésien classifieur proposé correspond à une analyse discriminante quadratique dont chaque classe suit une distribution Gaussienne. Les probabilités conditionnelles des deux classes, CFN et F , peuvent être respectivement écrites sous les équations (4.10) et (4.11), où p représente une distribution Gaussienne multivariée de dimension m :

$$p(x|\Delta = CFN) = \frac{1}{2\pi^{\frac{m}{2}} |\Sigma_{\Delta}|^{\frac{1}{2}}} e^{-\frac{(x-\mu_{\Delta})^T \Sigma_{\Delta}^{-1} (x-\mu_{\Delta})}{2}} \quad (4.10)$$

$$p(x|\Delta = F) = \frac{1}{2\pi^{\frac{m}{2}} |\Sigma_{\Delta}|^{\frac{1}{2}} \mathbf{c}_{\Delta}^{\frac{m}{2}}} e^{-\frac{(x-\mu_{\Delta})^T \Sigma_{\Delta}^{-1} (x-\mu_{\Delta})}{2c_{\Delta}}} \quad (4.11)$$

Soit LC_{Δ} la forme carrée de Mahalanobis de x , où $p(\Delta = CFN|\mathbf{x} = x) = \zeta_{\Delta}^{CFN}$, avec $LC_{\Delta} = (x - \mu_{\Delta})^T \Sigma_{\Delta}^{-1} (x - \mu_{\Delta})$, alors :

$$\begin{aligned}\zeta_{\Delta}^{CFN} &= \frac{\frac{1}{2\pi^{\frac{m}{2}} |\Sigma_{\Delta}|^{\frac{1}{2}}} e^{-\frac{LC_{\Delta}}{2}}}{\frac{1}{2\pi^{\frac{m}{2}} |\Sigma_{\Delta}|^{\frac{1}{2}}} e^{-\frac{LC_{\Delta}}{2}} + \frac{\omega}{2\pi^{\frac{m}{2}} |\Sigma_{\Delta}|^{\frac{1}{2}} \mathbf{c}_{\Delta}^{\frac{m}{2}}} e^{-\frac{LC_{\Delta}}{2c_{\Delta}}}} \\ &= \frac{\mathbf{c}_{\Delta}^{\frac{m}{2}} e^{-\frac{LC_{\Delta}}{2}}}{\mathbf{c}_{\Delta}^{\frac{m}{2}} e^{-\frac{LC_{\Delta}}{2}} + \omega e^{-\frac{LC_{\Delta}}{2c_{\Delta}}}} \\ &= \frac{\mathbf{c}_{\Delta}^{\frac{m}{2}}}{\mathbf{c}_{\Delta}^{\frac{m}{2}} + \omega e^{\frac{(c_{\Delta}-1) LC_{\Delta}}{2c_{\Delta}}}} \\ &= \frac{1}{1 + \omega e^{\frac{(c_{\Delta}-1) LC_{\Delta}}{2c_{\Delta}}}} \frac{\mathbf{c}_{\Delta}^{\frac{m}{2}}}{\mathbf{c}_{\Delta}^{\frac{m}{2}}}\end{aligned}\quad (4.12)$$

Soit $\gamma = \frac{\omega}{\mathbf{c}_\Delta} \mathbf{e}^{\frac{(\mathbf{c}_\Delta - 1)}{2\mathbf{c}_\Delta} LC_\Delta}$, finalement nous obtenons :

$$\zeta_\Delta^{CFN} = \frac{1}{1 + \gamma} \text{ et } \zeta_\Delta^F = 1 - \zeta_\Delta^{CFN} \quad (4.13)$$

En se basant sur (4.13), avec un $\mathbf{c}_\Delta > 1$, nous sommes capables de modéliser sur un réseau conditionnel Gaussien un test basé sur la statistique quadratique Δ . Dans ce qui suit, nous allons développer une extension de ces tests statistiques sous réseau Bayésien permettant ainsi de généraliser et d'unifier certaines méthodes pour la détection de fautes.

III Schémas de détection

Afin d'illustrer notre cadre probabiliste sous réseau Bayésien pour des méthodes utilisées pour la détection de fautes et associées aux statistiques quadratiques (Atoui et al., 2015c, 2014a,b) nous nous concentrons dans un premier temps sur le schéma de détection de fautes par analyse en composantes principales. Par la suite, nous généraliserons notre cadre pour d'autres méthodes, qu'elles soient à base de données ou de modèle.

III.1 Schéma de détection par analyse en composantes principales sous réseau Bayésien

En se basant sur un modèle linéaire Gaussien (modèle probabiliste et génératif), (Tipping and Bishop, 1999) ont proposé l'analyse en composantes principales probabiliste. Ceci est un cas spécial de l'analyse factorielle statistique et une généralisation de l'analyse en composantes principales (Kim and Lee, 2003). Sous un tel modèle (dans le cas Gaussien), les m variables du système (mesures) sont considérées comme une combinaison linéaire de $a < m$ variables cachées mutuellement non-corrélées plus un bruit additif. La solution d'un tel modèle maximisant la vraisemblance correspond naturellement à l'analyse en composantes principales. En utilisant les étapes suivantes, nous pouvons obtenir la même partie systématique \hat{X} de X obtenue par l'analyse en composantes principales traditionnelle.

étape I : considérer et utiliser un modèle probabiliste d'entrées et sorties

où les distributions conditionnelles et marginales sont Gaussiennes :

$$\begin{aligned} \mathbf{x} &= \hat{A}\hat{\mathbf{t}} + \epsilon, \quad \mathbf{x} \in \mathbb{R}^m, \quad \hat{A} \in \mathbb{R}^{m \times a}, \quad \hat{\mathbf{t}} \in \mathbb{R}^a \\ \mathbf{x}|\hat{\mathbf{t}} &= \mathcal{N}(\hat{A}\hat{\mathbf{t}}; \nu I), \quad \hat{\mathbf{t}} \sim \mathcal{N}(0; I), \quad \epsilon \sim \mathcal{N}(0; \nu I), \quad \nu \approx 0 \end{aligned} \quad (4.14)$$

où I est une matrice d'identité, ν est un scalaire, et $\hat{\mathbf{t}}$ et ϵ sont des variables aléatoires indépendantes, où la variable cachée $\hat{\mathbf{t}}$ est supposée expliquer toute la variabilité systématique de \mathbf{x} .

Notons que, selon (Bishop, 2006), il n'y a aucune perte de généralité en posant la variable cachée $\hat{\mathbf{t}}$ avec une moyenne nulle et une matrice de variance-covariance unité, car une distribution Gaussienne plus générale donnera lieu à un modèle probabiliste équivalent.

étape II : étant donné X avec un vecteur ligne $(x^n)^T = [x_1^n, \dots, x_m^n]$, poser, par exemple, la matrice de projection \hat{A} égale à la matrice \hat{P} obtenue précédemment. D'autres méthodes peuvent être également utilisées pour déterminer a et \hat{A} (voir Bishop (2006)).

étape III : en se basant sur (4.14), déduire \hat{t}^n selon la distribution de probabilité a posteriori $p(\hat{\mathbf{t}}|\mathbf{x} = x^n)$ donnée par :

$$\begin{aligned} p(\hat{\mathbf{t}}|\mathbf{x} = x^n) &= p(\hat{\mathbf{t}}|\mathbf{x}_1 = x_1^n, \dots, \mathbf{x}_m = x_m^n) \\ &= \mathcal{N}(M\hat{A}^T[x_1^n, \dots, x_m^n]^T; \nu M), \quad M = (\nu I + \hat{A}^T\hat{A})^{-1} \end{aligned} \quad (4.15)$$

où on peut noter que contrairement à la covariance a posteriori, la moyenne a posteriori de $p(\hat{\mathbf{t}}|\mathbf{x} = x^n)$ dépend de \mathbf{x} , ainsi :

$$\hat{t}^n = \mathbb{E}[\hat{\mathbf{t}}|\mathbf{x} = x^n] = M\hat{A}^T[x_1^n, \dots, x_m^n]^T \quad (4.16)$$

étape IV : étant donné \hat{t}^n , calculer $\hat{x}^n \in \hat{X}^T$, avec $n \in 1, \dots, N$:

$$\begin{aligned} \hat{x}^n &= \mathbb{E}[\mathbf{x}|\mathbf{t} = \hat{t}^n] = [\mathbb{E}[\mathbf{x}_1|\mathbf{t} = \hat{t}^n] \dots \mathbb{E}[\mathbf{x}_m|\mathbf{t} = \hat{t}^n]]^T \\ &= [\hat{B}_1\hat{t}^n, \dots, \hat{B}_m\hat{t}^n]^T = \hat{A}\hat{t}^n \end{aligned} \quad (4.17)$$

Ainsi, l'analyse en composantes principales peut être représentée par un modèle génératif, où toutes les variables sont Gaussiennes, mais peut également être représentée sous un modèle graphique simple et illustratif : un réseau Bayésien. En effet, à partir de la définition donnée d'un noeud Gaussien, le

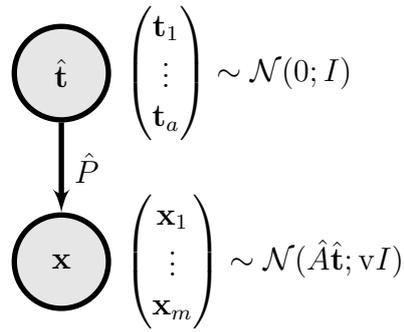


FIGURE 4.3 – Analyse en composantes principales sur réseau Bayésien, projection de \mathbf{x} dans l’espace principal, forme multivariée

modèle génératif représentant l’analyse en composantes principales peut être facilement implémenté sous un réseau conditionnel Gaussien. La figure (4.3) représente un noeud linéaire Gaussien \mathbf{x} et son parent Gaussien $\hat{\mathbf{t}}$, qui peut naturellement automatiser et généraliser les calculs probabilistes nécessaires au modèle Gaussien représentant l’analyse en composantes principales. Dans ce graphe, \mathbf{x} comme dans (4.14) correspond à une variable observée de m -dimension et qui suit une distribution Gaussienne conditionnelle $\mathcal{N}(\hat{A}\hat{\mathbf{t}}; \nu I)$, où $\nu \approx 0$, et $\hat{\mathbf{t}}$ correspond à une variable score de a -dimension et \hat{A} est la matrice de régression correspondant à la matrice \hat{P} comportant des vecteurs propres.

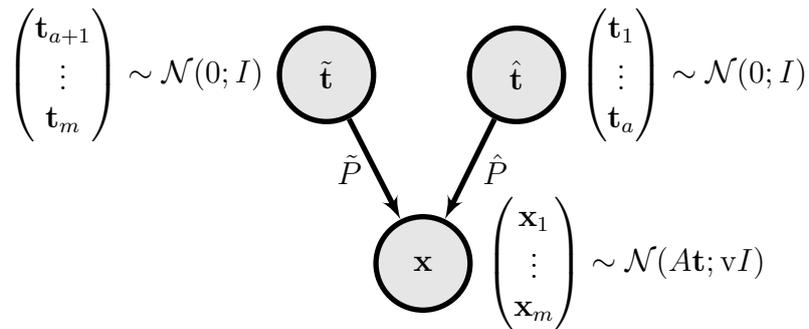


FIGURE 4.4 – Analyse en composantes principales sous un réseau Bayésien, projection de \mathbf{x} dans l’espace principal et résiduel, forme multivariée

Ce réseau peut être utilisé pour la réduction de données comme l’analyse en composantes principales standard mais aussi pour la détection de fautes en surveillant la partie principale. Cependant, il est judicieux de lui associer également la partie résiduelle permettant de détecter différents types de fautes. Pour cela, un autre noeud est ajouté, un noeud Gaussien $\tilde{\mathbf{t}}$ avec la même

distribution a priori que $\hat{\mathbf{t}}$, représentant une variable de dimension $(m - a)$. Ainsi, nous sommes capables de reconstruire entièrement \mathbf{x} et de gérer la partie systématique de l'analyse en composantes principales et les dernières $m - a$ composantes $\tilde{\mathbf{t}}$ de \mathbf{x} rendant ainsi possible la séparation de l'espace original en deux sous espaces. La nouvelle structure du réseau, montré par la figure 4.4, représente un modèle probabiliste pouvant être présenté comme suit :

$$\mathbf{x} = \hat{A}\hat{\mathbf{t}} + \tilde{A}\tilde{\mathbf{t}} + \epsilon \quad (4.18)$$

où $\mathbf{x} \in \mathbb{R}^m$, $\hat{\mathbf{t}} \in \mathbb{R}^a$, $\tilde{\mathbf{t}} \in \mathbb{R}^{m-a}$, $\hat{A} = \hat{P}$, $\tilde{A} = \tilde{P}$, $\epsilon \in \mathbb{R}^m$, $\epsilon \sim \mathcal{N}(0; \sigma^2 I)$, $v \approx 0$, $\hat{\mathbf{t}} \sim \mathcal{N}(0; I)$, $\tilde{\mathbf{t}} \sim \mathcal{N}(0; I)$, $\mathbf{x}|\hat{\mathbf{t}}, \tilde{\mathbf{t}} \sim \mathcal{N}(\hat{A}\hat{\mathbf{t}} + \tilde{A}\tilde{\mathbf{t}}; vI)$, avec $\tilde{\mathbf{t}}$ et $\hat{\mathbf{t}}$ indépendants.

La distribution marginale $p(\mathbf{x})$ de la variable observée \mathbf{x} peut être exprimée, grâce à des règles de probabilités (somme et produit), comme ci-dessous :

$$\begin{aligned} p(\mathbf{x}) &= \int \int p(\mathbf{x}, \hat{\mathbf{t}}, \tilde{\mathbf{t}}) d\hat{\mathbf{t}} d\tilde{\mathbf{t}} \\ &= \int p(\hat{\mathbf{t}}) \int p(\mathbf{x}|\hat{\mathbf{t}}, \tilde{\mathbf{t}}) p(\tilde{\mathbf{t}}) d\tilde{\mathbf{t}} d\hat{\mathbf{t}} \end{aligned} \quad (4.19)$$

L'équation (4.19) correspond à un modèle linéaire Gaussien et ainsi $p(\mathbf{x})$ suit une distribution Gaussienne donnée par ceci :

$$p(\mathbf{x}) = \mathcal{N}(\mu^{\mathbf{x}}; \Sigma^{\mathbf{x}}) \quad (4.20)$$

où les paramètres $\mu^{\mathbf{x}}$ et $\Sigma^{\mathbf{x}}$ peuvent être obtenus en manipulant les distributions Gaussienne (voir (4.19)) ou alors déduits directement à partir de (4.18) comme ceci :

$$\begin{aligned} \mu^{\mathbf{x}} &= \mathbb{E}[\mathbf{x}] = \hat{A}\mathbb{E}[\hat{\mathbf{t}}] + \tilde{A}\mathbb{E}[\tilde{\mathbf{t}}] + \mathbb{E}[\epsilon], \\ \Sigma^{\mathbf{x}} &= \text{cov}[\mathbf{x}] = \hat{A}\mathbb{E}[\hat{\mathbf{t}}\hat{\mathbf{t}}^T]\hat{A}^T + \tilde{A}\mathbb{E}[\tilde{\mathbf{t}}\tilde{\mathbf{t}}^T]\tilde{A}^T + \mathbb{E}[\epsilon\epsilon^T] \end{aligned}$$

De plus, la distribution de probabilité conditionnelle $p(\hat{\mathbf{t}}|\mathbf{x})$ du noeud caché

$\hat{\mathbf{t}}$ peut être obtenue comme suit :

$$p(\hat{\mathbf{t}}|\mathbf{x}) \propto p(\hat{\mathbf{t}})p(\mathbf{x}|\hat{\mathbf{t}}) \quad (4.21)$$

$$= \frac{p(\hat{\mathbf{t}}) \int p(\mathbf{x}|\hat{\mathbf{t}}, \tilde{\mathbf{t}})p(\tilde{\mathbf{t}})d\tilde{\mathbf{t}}}{p(\mathbf{x})} \quad (4.22)$$

$$= \frac{p(\hat{\mathbf{t}})p(\mathbf{x}|\hat{\mathbf{t}})}{p(\mathbf{x})} \quad (4.23)$$

$$= \mathcal{N}(\mu^{\hat{\mathbf{t}}|\mathbf{x}}; \Sigma^{\hat{\mathbf{t}}|\mathbf{x}}) \quad (4.24)$$

avec :

$$\begin{aligned} \mu^{\hat{\mathbf{t}}|\mathbf{x}} &= \mathbb{E}[\hat{\mathbf{t}}|\mathbf{x}] = \mathbb{E}[\hat{\mathbf{t}}] + \Sigma^{\hat{\mathbf{t}}\mathbf{x}}\Sigma^{(\mathbf{x}\mathbf{x})}^{-1}(\mathbf{x} - \mu^{\mathbf{x}}), \\ \Sigma^{\hat{\mathbf{t}}|\mathbf{x}} &= \text{cov}[\hat{\mathbf{t}}|\mathbf{x}] = \text{cov}[\hat{\mathbf{t}}] - \Sigma^{\hat{\mathbf{t}}\mathbf{x}}\Sigma^{\mathbf{x}\mathbf{x}}^{-1}\Sigma^{\mathbf{x}\hat{\mathbf{t}}} \end{aligned}$$

où les matrices $\Sigma^{\mathbf{x}\mathbf{x}}$, $\Sigma^{\mathbf{x}\hat{\mathbf{t}}}$ and $\Sigma^{\hat{\mathbf{t}}\mathbf{x}}$ appartiennent à l'inverse de la matrice de variance-covariance de la variable jointe $[\hat{\mathbf{t}} \ \mathbf{x}]^T$. Ainsi, pour une observation donnée x^n , $\hat{t}^n = \mathbb{E}[\hat{\mathbf{t}}|\mathbf{x}]$.

De plus, étant donnée l'hypothèse prise par le modèle probabiliste (4.18), où les éléments $\mathbf{x}_i, i \in 1, \dots, m$ de \mathbf{x} sont non corrélés (les éléments $\epsilon_i, i \in 1, \dots, m$ de ϵ sont mutuellement indépendants), le réseau proposé dans la figure 4.5 et celui présenté précédemment dans la figure 4.4 sont exactement les mêmes, différant uniquement dans la représentation de \mathbf{x} (figure 4.5 : représentation univariée ; figure 4.4 : représentation multivariée). Ainsi, chaque noeud Gaussien \mathbf{x}_i dans la figure 4.5 représente le modèle probabiliste suivant :

$$\mathbf{x}_i = \hat{B}_i\hat{\mathbf{t}} + \tilde{B}_i\tilde{\mathbf{t}} + \epsilon_i \quad (4.25)$$

où $\mathbf{x}_i \in \mathbb{R}$, $\hat{\mathbf{t}} \in \mathbb{R}^a$, $\tilde{\mathbf{t}} \in \mathbb{R}^{m-a}$, $\epsilon_i \in \mathbb{R}$, $\epsilon_i \sim \mathcal{N}(0; v)$, $v \approx 0$, $\hat{B}_i \in \mathbb{R}^{1 \times a}$, $\tilde{B}_i \in \mathbb{R}^{1 \times (m-a)}$, $\hat{\mathbf{t}} = \mathcal{N}(0; I)$, $\tilde{\mathbf{t}} = \mathcal{N}(0; I)$, $\mathbf{x}_i|\hat{\mathbf{t}}$ et $\tilde{\mathbf{t}} \sim \mathcal{N}(\hat{B}_i\hat{\mathbf{t}} + \tilde{B}_i\tilde{\mathbf{t}}; v)$

A partir de la figure 4.4, nous pouvons déduire la distribution marginale $p(\mathbf{x})$ de chaque variable observée comme suit :

$$\begin{aligned} p(\mathbf{x}_i) &= \int p(\hat{\mathbf{t}}) \int p(\tilde{\mathbf{t}}) \int_{\mathbf{x} \setminus \{\mathbf{x}_i\}} \prod_{j \neq i}^m p(\mathbf{x}_j|\hat{\mathbf{t}}, \tilde{\mathbf{t}})d(\mathbf{x} \setminus \{\mathbf{x}_i\})d\tilde{\mathbf{t}}d\hat{\mathbf{t}} \\ &= \int p(\hat{\mathbf{t}}) \int p(\tilde{\mathbf{t}})p(\mathbf{x}_i|\hat{\mathbf{t}}, \tilde{\mathbf{t}})d\tilde{\mathbf{t}}d\hat{\mathbf{t}} \end{aligned} \quad (4.26)$$

où $\int_{\mathbf{x}_j} \prod_{j \neq i}^m p(\mathbf{x}_j | \hat{\mathbf{t}}, \tilde{\mathbf{t}}) = 1$ et $\mathbf{x} \setminus \{\mathbf{x}_i\}$ signifie toutes les variables $\mathbf{x}_1, \dots, \mathbf{x}_m$ excepté \mathbf{x}_i .

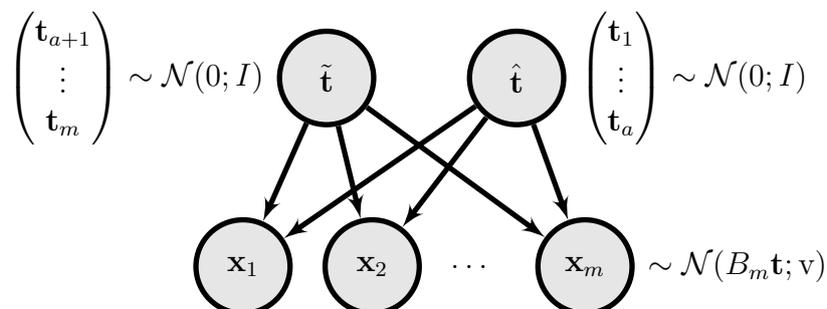


FIGURE 4.5 – Analyse en composantes principales sous un réseau Bayésien, forme univariée

La distribution de probabilité marginale de chaque variable \mathbf{x}_i , $p(\mathbf{x}_i)$, est une Gaussienne, de la forme :

$$p(\mathbf{x}_i) = \mathcal{N}(\mu^{\mathbf{x}_i}; \Sigma^{\mathbf{x}_i}) \quad (4.27)$$

où ces paramètres peuvent être déduits à partir de (4.25) comme suit :

$$\begin{aligned} \mu^{\mathbf{x}_i} &= \mathbb{E}[\mathbf{x}_i] = \hat{B}_i \mathbb{E}[\hat{\mathbf{t}}] + \tilde{B}_i \mathbb{E}[\tilde{\mathbf{t}}] + \mathbb{E}[\epsilon], \\ \Sigma^{\mathbf{x}_i} &= \text{cov}(\mathbf{x}_i) = \hat{B}_i \mathbb{E}[\hat{\mathbf{t}} \hat{\mathbf{t}}^T] \hat{B}_i^T + \tilde{B}_i \mathbb{E}[\tilde{\mathbf{t}} \tilde{\mathbf{t}}^T] \tilde{B}_i^T + \mathbb{E}[\epsilon \epsilon^T] \end{aligned}$$

La représentation univariée, où chaque variable a ses propres noeuds, peut être vue comme un exemple du modèle naïf de Bayes et peut être un bon choix quand certaines observations sont manquantes. En effet, de façon naturelle, il peut gérer les variables non-observées en considérant leurs noeuds comme cachés en effectuant l'inférence (calculs) étant données les observations disponibles, en marginalisant sur les variables non-observées. En se basant sur cette propriété, nous allons voir dans ce qui suit que l'outil proposé pour la détection de fautes, dans le cas de l'analyse en composantes principales par exemple, est capable de répondre à une requête lorsque certaines observations sont manquantes, contrairement au schéma traditionnel de l'analyse en composantes principales pour la détection de fautes.

Nous avons vu précédemment comment une analyse en composante principale peut être généralisée par un modèle Gaussien génératif. De plus, nous

avons également vu qu'il est possible de l'intégrer et la modéliser sous un réseau Bayésien, plus exactement un réseau conditionnel Gaussien. Dans ce qui suit, nous allons voir comment étendre ce dernier afin de pouvoir modéliser un schéma de détection basé sur l'analyse en composantes principales.

Pour cela, nous proposons d'augmenter le réseau conditionnel Gaussien représentant l'analyse en composantes principales par d'autres réseaux conditionnels Gaussiens permettant de modéliser les statistiques quadratiques associées à l'analyse en composantes principales. En d'autres termes, nous allons rajouter des noeuds discrets à chaque noeud caché considéré (représentation de l'espace principal et/ou résiduel) que l'on veut surveiller étant donnée une nouvelle observation. Toutefois, à titre d'exemple, nous considérons le schéma de détection usuellement utilisé, surveillant les deux espaces : principal et résiduel. Ceci donne lieu au deux réseaux Bayésiens des Figures (4.6), (4.7), où les noeuds \mathbf{T}^2 et \mathbf{SPE} sont des noeuds discrets, chacun associé à deux états CFN et F . Chacun de ces deux noeuds est associé à une limite probabiliste permettant de prendre une décision équivalente au test basé sur la statistique de forme quadratique de même nom, étant donné la probabilité a posteriori de chacun de ses états. Chaque noeud discret de décision est parent d'un noeud latent du réseau Bayésien précédent.

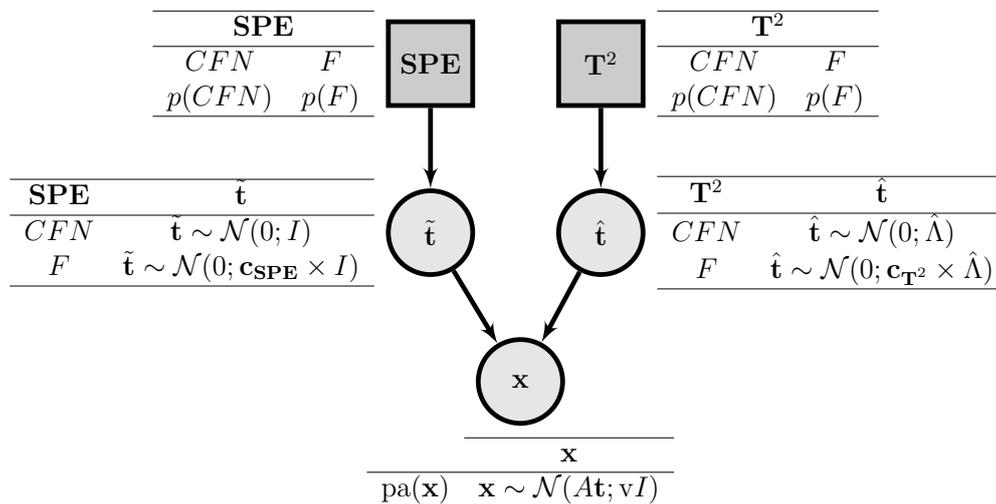


FIGURE 4.6 – Schéma de détection par analyse en composantes principales sous réseau Bayésien, forme multivariée

Ainsi, chaque noeud caché possède comme parent un noeud discret. Dans notre exemple, de façon similaire, les noeuds $\hat{\mathbf{t}}$ et $\tilde{\mathbf{t}}$, sont chacun descendants

directs d'un noeud discret respectivement \mathbf{T}^2 et \mathbf{SPE} , et pour chaque valeur de leurs parents, CFN et F , ils suivent une distribution Gaussienne.

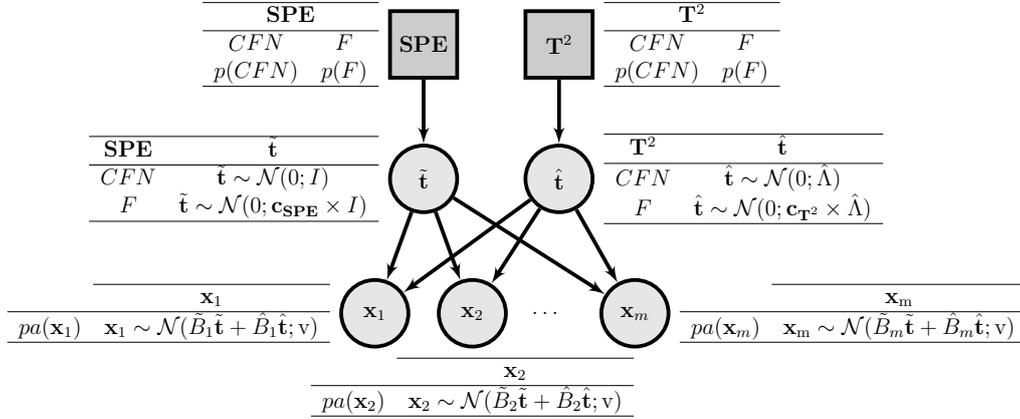


FIGURE 4.7 – Schéma de détection par analyse en composantes principales sous réseau Bayésien : forme univariée

Afin de pouvoir atteindre les mêmes décisions prises par les deux statistiques quadratiques \mathbf{T}^2 et \mathbf{SPE} , dans le cas de l'analyse en composantes principales, il faut redéfinir les distributions Gaussiennes des noeuds $\hat{\mathbf{t}}$ et $\tilde{\mathbf{t}}$. Précédemment, nous avons modélisé les statistiques quadratiques sous réseau Bayésien sous l'hypothèse que la classe F est fonction de la classe CFN . De la même manière, nous proposons de redéfinir les paramètres des distributions de chaque noeud caché. Ainsi, la distribution de chaque état dans ces noeuds est définie avec une moyenne nulle commune et des matrices de variance-covariance exprimant une variabilité plus importante, à travers le coefficient \mathbf{c}_Δ , par rapport à celle associée à l'état CFN . Dans notre exemple, les matrices de variance-covariance des Gaussiennes correspondantes à la classe CFN des noeuds \mathbf{T}^2 et \mathbf{SPE} sont définies respectivement égales à $\hat{\Lambda}$ (une matrice diagonale représentant les a plus grandes valeurs propres de la matrice de variance-covariance des observation \mathbf{x}) et I (une matrice diagonale, voir le schéma de détection par analyse en composantes principales dans le chapitre précédent). Nous rappelons que la matrice $\hat{\Lambda}$ est obtenue en utilisant uniquement les données de fonctionnement normal.

Concernant les autres noeuds des réseaux proposés, les noeuds discrets \mathbf{T}^2 et \mathbf{SPE} (les noeuds de décisions avec comme classes CFN , F) sont définis étant données les probabilités a priori de leurs valeurs (e.g. le nombre d'observation par état). Finalement, les noeuds continus restant suivent un modèle

Gaussien linéaire de régression étant donnés leurs noeuds parents Gaussiens cachés $\hat{\mathbf{t}}$ et $\tilde{\mathbf{t}}$. Dans les Figures (4.6), (4.7), nous donnons la table de probabilité conditionnelle définie et associée à chaque noeud.

A chaque instant, en utilisant le modèle probabiliste graphique proposé, nous sommes capables de prendre une décision concernant l'état d'un système qui est semblable à celle prise par le schéma traditionnel de détection par analyse en composantes principales. Une fois que les paramètres de ces noeuds sont définis, il est suffisant de calculer la probabilité a posteriori d'une classe de chaque noeud discret et la comparer à la limite probabiliste lui étant associée (équation (4.13)).

Les probabilités a posteriori du noeud \mathbf{T}^2 (de la même façon pour le noeud **SPE**) peuvent être déduites à partir du réseau proposé dans la figure 4.6 étant donnée une nouvelle observation x de la variable multivariée (noeud) \mathbf{x} , comme ci-dessous :

$$\begin{aligned}
 p(\mathbf{T}^2 = CFN | \mathbf{x} = x) &\propto p(\mathbf{T}^2 = CFN) p(x | \mathbf{T}^2 = CFN) \\
 &= \eta p(\mathbf{T}^2 = CFN) p(x | \mathbf{T}^2 = CFN) \\
 &= \eta p(\mathbf{T}^2 = CFN) \sum_{\mathbf{SPE}} p(\mathbf{SPE}) p(x | \mathbf{T}^2 = CFN, \mathbf{SPE})
 \end{aligned} \tag{4.28}$$

où $\eta = \frac{1}{\sum_{\mathbf{T}^2} p(x | \mathbf{T}^2) p(\mathbf{T}^2)}$ et $p(\mathbf{T}^2 = F | \mathbf{x} = x) = 1 - p(\mathbf{T}^2 = CFN | \mathbf{x} = x)$

Les paramètres de $p(x | \mathbf{T}^2 = CFN, \mathbf{SPE})$ pour chaque valeur (CFN, F) du noeud **SPE** peuvent être obtenus comme suit :

$$\begin{aligned}
 p(x | \mathbf{T}^2 = CFN, \mathbf{SPE} = CFN) &= \int p(\hat{\mathbf{t}} | \mathbf{T}^2 = CFN) \int p(x | \hat{\mathbf{t}}, \tilde{\mathbf{t}}) p(\tilde{\mathbf{t}} | \mathbf{SPE} = CFN) d\tilde{\mathbf{t}} d\hat{\mathbf{t}} \\
 &= \mathcal{N}(0; \mathbf{v}I + \tilde{A}\tilde{A}^T + \hat{A}\hat{A}^T)
 \end{aligned} \tag{4.29}$$

De la même manière :

$$p(x | \mathbf{T}^2 = CFN, \mathbf{SPE} = F) = \mathcal{N}(0; \mathbf{v}I + \tilde{A}(\mathbf{c}_{\mathbf{SPE}}I)\tilde{A}^T + \hat{A}\hat{A}^T) \tag{4.30}$$

Les réseaux présentés dans les Figures (4.6) et (4.7), malgré le fait qu'ils possèdent deux structures différentes, donnent des résultats identiques. Dans

ce qui suit, nous donnons les probabilités a posteriori de chaque état du noeud \mathbf{T}^2 (respectivement \mathbf{SPE}), dans la figure 4.7, étant donnée une nouvelle observation x_1, \dots, x_m des variables univariées (noeuds) $\mathbf{x}_1, \dots, \mathbf{x}_m$.

$$\begin{aligned}
p(\mathbf{T}^2 = CFN | \mathbf{x}_1 = x_1, \dots, \mathbf{x}_m = x_m) \\
&\propto p(\mathbf{T}^2 = CFN) \prod_{i=1}^m p(x_i | \mathbf{T}^2 = CFN) \\
&= \eta p(\mathbf{T}^2 = CFN) \prod_{i=1}^m p(x_i | \mathbf{T}^2 = CFN) \\
&= \eta p(\mathbf{T}^2 = CFN) \sum_{\mathbf{SPE}} p(\mathbf{SPE}) \prod_{i=1}^m p(x_i | \mathbf{T}^2 = CFN, \mathbf{SPE})
\end{aligned} \tag{4.31}$$

où $\eta = \frac{1}{\sum_{\mathbf{T}^2} p(x_1, \dots, x_m | \mathbf{T}^2) p(\mathbf{T}^2)}$ et $p(\mathbf{T}^2 = F | \mathbf{x}_1 = x_1, \dots, \mathbf{x}_m = x_m) = 1 - p(\mathbf{T}^2 = CFN | \mathbf{x}_1 = x_1, \dots, \mathbf{x}_m = x_m)$ et pour chaque état de \mathbf{SPE} , les paramètres de $p(x_i | \mathbf{T}^2 = CFN, \mathbf{SPE})$ sont donnés par :

$$\begin{aligned}
p(x_i | \mathbf{T}^2 = CFN, \mathbf{SPE} = CFN) \\
&= \int p(\hat{\mathbf{t}} | \mathbf{T}^2 = CFN) \int p(x_i | \hat{\mathbf{t}}, \tilde{\mathbf{t}}) p(\tilde{\mathbf{t}} | \mathbf{SPE} = CFN) d\tilde{\mathbf{t}} d\hat{\mathbf{t}} \\
&= \mathcal{N}(0; \mathbf{v} + \tilde{B}_i \tilde{B}_i^T + \hat{B}_i \hat{\Lambda} \hat{B}_i^T)
\end{aligned} \tag{4.32}$$

et

$$p(x_i | \mathbf{T}^2 = CFN, \mathbf{SPE} = F) = \mathcal{N}(0; \mathbf{v} + \tilde{B}_i \mathbf{c}_{\mathbf{SPE}} I \tilde{B}_i^T + \hat{B}_i \hat{\Lambda} \hat{B}_i^T) \tag{4.33}$$

Le réseau conditionnel Gaussien proposé sous forme univariée (voir figure 4.7) contrairement à l'analyse en composantes principales et la plupart des méthodes de détection, est capable de donner une réponse même si les valeurs de certaines variables sont manquantes (les noeuds leurs correspondant sont considérés comme étant cachés) et cela sans délai.

Soit \mathbf{x} un ensemble de m variables (noeuds). Parmi ces m variables, un ensemble de variables \mathbf{x}^+ sont observées tandis que d'autres \mathbf{x}^- ne le sont pas (elles sont manquantes). A partir des variables observées, les probabilités a posteriori des états des noeuds discrets peuvent être calculées en marginalisant sur \mathbf{x}^- . Ci-dessous, nous donnons un exemple de l'inférence de la probabilité a pos-

teriori de l'état CFN de \mathbf{T}^2 lorsque certaines observations sont manquantes. Notons que ces calculs sont exécutés automatiquement en utilisant un réseau Bayésien.

$$\begin{aligned}
p(\mathbf{T}^2 = CFN | \mathbf{x}^+ = x^+) & \\
& \propto p(\mathbf{T}^2 = CFN) \sum_{\mathbf{SPE}} \int p(\mathbf{SPE}) p(x^+, \mathbf{x}^- | \mathbf{T}^2 = CFN, \mathbf{SPE}) d\mathbf{x}^- \\
& = \eta p(\mathbf{T}^2 = CFN) \sum_{\mathbf{SPE}} p(\mathbf{SPE}) \int p(x^+, \mathbf{x}^- | \mathbf{T}^2 = CFN, \mathbf{SPE}) d\mathbf{x}^- \\
& = \eta p(\mathbf{T}^2 = CFN) \sum_{\mathbf{SPE}} p(\mathbf{SPE}) p(x^+ | \mathbf{T}^2 = CFN, \mathbf{SPE}) \\
& = \eta p(\mathbf{T}^2 = CFN) \sum_{\mathbf{SPE}} \prod_{x_i \in x^+} p(x_i | \mathbf{SPE}, \mathbf{T}^2 = CFN) \quad (4.34)
\end{aligned}$$

où $\int p(\mathbf{x}^- | \mathbf{T}^2 = CFN, \mathbf{SPE}) d\mathbf{x}^- = 1$, $\eta = \frac{1}{\sum_{\mathbf{T}^2} p(x^+ | \mathbf{T}^2) p(\mathbf{T}^2)}$ et $p(\mathbf{T}^2 = F | x^+) = 1 - p(\mathbf{T}^2 = CFN | x^+)$.

III.2 D'autres schémas de détection sous réseaux Bayésiens

Avant d'illustrer comment d'autres schémas de détection peuvent être déclinés sous réseaux Bayésiens, nous trouvons nécessaire d'illustrer dans un premier temps, comment une représentation d'état permet de généraliser l'analyse en composantes principales. L'analyse en composantes principales permet de modéliser et résumer les données en les projetant dans un autre espace de dimension réduite. Ainsi, une variable multidimensionnelle \mathbf{x} , supposée suivre une loi normale centrée et/ou réduite, est projetée dans un espace réduit \mathbf{z} comme ceci :

$$\mathbf{z} = P^T \mathbf{x} \quad (4.35)$$

où $P \in \mathcal{R}^{m \times a}$ est la matrice de projection avec $P^T P = I$.

Formellement cette écriture peut être représentée autrement à l'aide d'un modèle Gaussien, comme ceci :

$$\mathbf{x} = P\mathbf{z} + w, \quad \mathbf{z} \sim \mathcal{N}(0; \hat{\Lambda}) \quad (4.36)$$

où $\hat{\Lambda}$ est la matrice représentant les premières valeurs propres de la matrice de variance-covariance de \mathbf{x} et w est une variable Gaussienne. Cette dernière représente un bruit Gaussien avec une moyenne nulle et une matrice de variance-covariance vI où v est considéré comme un nombre réel très petit, tendant vers zéro.

Nous rappelons que pour modéliser le schéma de détection par analyse en composantes principales, nous avons augmenté le réseau Bayésien se basant sur ce modèle Gaussien par d'autres variables discrètes afin de décider d'une éventuelle présence de fautes dans le système.

Le modèle linéaire Gaussien donné dans (4.36) peut être généralisé et réécrit comme ceci :

$$\mathbf{x} = C\mathbf{z} + v, \quad \mathbf{z} \sim \mathcal{N}(0; Q), \quad v \sim \mathcal{N}(0; R) \quad (4.37)$$

en posant $C = P$, $R = vI$, $Q = \hat{\Lambda}$.

L'écriture (4.37) représente un cas spécial, le cas statique avec des variables cachées continues, de la représentation d'états proposée par (Roweis and Ghahramani, 1999). Cette représentation permet de générer et d'unifier des méthodes statistiques pouvant être employées pour la détection de fautes. Cette représentation d'états (formalisme répandu dans le domaine de l'automatique pour modéliser un système) décrite par les auteurs correspond à l'écriture suivante :

$$\begin{cases} \mathbf{z}(t+1) = A\mathbf{z}(t) + w \\ \mathbf{x}(t) = C\mathbf{z}(t) + v \end{cases} \quad (4.38)$$

où $\mathbf{z}(\cdot) \in \mathbb{R}^a$ est une variable non observable pouvant être continue ou discrète. Elle correspond à une projection informative de dimension réduite ou bien à un résumé des observations et évolue selon une dynamique de Markov. $\mathbf{x}(\cdot) \in \mathbb{R}^m$ est une variable observée dite de sortie, générée en fonction de $\mathbf{z}(\cdot)$. $A \in \mathbb{R}^{a \times a}$ est la matrice de transition, $C \in \mathbb{R}^{m \times a}$ est la matrice d'observations. $w \in \mathbb{R}^a$ et $v \in \mathbb{R}^m$ sont deux sources de bruits blancs (non corrélées au temps, mutuellement indépendantes et non observées), essentielles pour le modèle (régulant le vecteur d'état et lui permettant d'être un vecteur caché), suivant chacune

une distribution Gaussienne comme ceci :

$$w \sim \mathcal{N}(0; Q), \tag{4.39}$$

$$v \sim \mathcal{N}(0; R) \tag{4.40}$$

avec Q et R les matrices de variance et covariance de rang plein, par définition symétrique positive semi-définie, de respectivement w et v . Contrairement à R , il n'y a aucune perte d'information en permutant la structure de Q (pouvant être diagonalisée étant donnée la décomposition en valeurs propres) dans A ou C tout en la rendant diagonale ou égale à la matrice d'identité. Ceci dit, deux modèles peuvent être équivalents avec deux matrices Q non similaires. Suivant ce modèle nous pouvons générer des données statiques et d'autres dynamiques.

Dans un premier temps, nous considérons des modèles linéaires Gaussiens à états continus avec un bruit Gaussien générant des données n'ayant pas de dépendances temporelles et dont l'ordre d'apparition n'est pas pris en compte (tout comme l'analyse en composantes principales). En d'autres termes, nous considérons que les données générées sont indépendantes et identiquement distribuées.

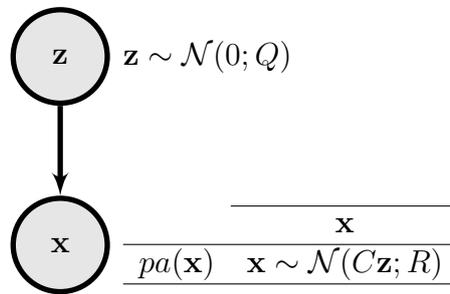


FIGURE 4.8 – Réseau Bayésien Gaussien représentant un modèle linéaire Gaussien avec un noeud continu caché : forme multivariée

Selon (Roweis and Ghahramani, 1999), la modélisation sous représentation d'états de ce type de données se fait en considérant le vecteur d'état $\mathbf{z}(\cdot)$, dans (4.38), comme une constante pourvue de dynamique corrompue par du bruit et associée à une matrice A nulle. Ainsi, la représentation d'état donnée dans (4.37) peut être écrite comme ceci :

$$\begin{cases} \mathbf{z} = w \\ \mathbf{x} = C\mathbf{z} + v \end{cases} \tag{4.41}$$

où la distribution marginale de la variable \mathbf{x} s'écrit comme ceci :

$$p(\mathbf{x}) = \mathcal{N}(0; C^T Q C + R) \quad (4.42)$$

Cette écriture peut être représentée sous un réseau Bayésien. Le réseau Bayésien lui correspondant et généralisant est donné dans la figure 4.8.

Plusieurs déclinaisons de cette représentation sont possibles en émettant des hypothèses sur les paramètres des différentes variables étant donnée la dégénérescence du modèle. En effet, sous cette représentation générative, plusieurs méthodes statistiques pour la modélisation des ensembles de données multidimensionnelles stationnaires (classiquement utilisées dans le cadre de la détection de fautes) peuvent être généralisées. Ces méthodes peuvent être vues comme un résultat de l'apprentissage du modèle sous certaines hypothèses, mais aussi comme étant tout simplement des méthodes pouvant être modélisées sous cette représentation sans pour autant être un résultat d'un apprentissage (e.g. en s'appuyant entre autres sur un $R = \nu I$ et un C représentant la pseudo-inverse de la matrice de projection/transformation associée à la méthode considérée). Parmi ces méthodes, on peut citer :

- l'Analyse Factorielle (AF), avec $Q = I$, $R = \text{diag}(\nu_1, \dots, \nu_m)$ et ν_i, C sont apprises à partir des données (Tipping and Bishop, 1999).
- l'Analyse en Composantes Principales Sensibles (ACPS), avec $Q = I$, $R = \epsilon I$, ϵ un scalaire, et ϵ, C sont identifiés à partir des données (Roweis, 1998).
- Projection dans les Structures Latentes (PLS), $Q = I$, $R = \text{diag}(\nu_1 I, \nu_2 I)$ et ν_1, ν_2, C à partir des données (Li et al., 2011).
- etc.

Dans le même esprit, les extensions de ces dernières ou d'autres transformations linéaires des observations peuvent être représentées sous un réseau Bayésien. Ces méthodes peuvent être utilisées pour la détection de fautes en les associant par exemple à des statistiques quadratiques. Cependant, ceci serait non optimal dans le sens où l'on passe d'un formalisme probabiliste à un autre déterministe. Nous proposons alors, comme précédemment pour l'analyse en composantes principales, de modéliser leurs schémas de détection sous réseaux Bayésiens. En d'autres termes, nous proposons sur un réseau Bayésien de généraliser les schémas de détection, se basant sur des statistiques quadratiques Δ , associés à des méthodes de projection/transformation linéaire des observations

dans un seul ou plusieurs espaces. Cette généralisation peut être traduite par le réseau Bayésien donné dans la figure 4.9. Ce réseau permet de détecter une faute, en surveillant les observations via leurs projections/transformations correspondant à la méthode choisie, en comparant la probabilité à posteriori d'un état de la variable Δ étant donnée une observation x à la limite probabiliste lui correspondant.

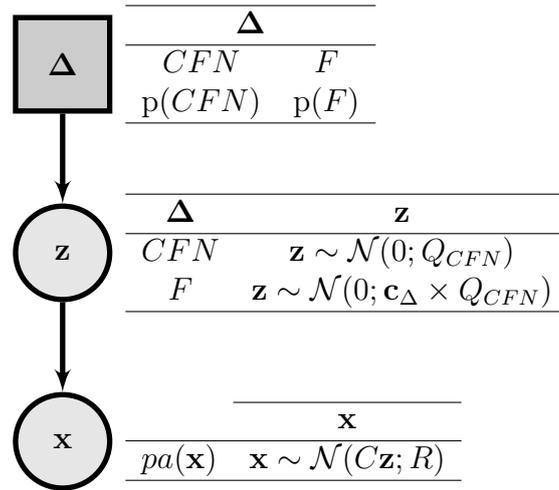


FIGURE 4.9 – Schéma de détection par modèle linéaire Gaussien sous réseau Bayésien : forme multivariée

Le réseau proposé dans la figure 4.9 peut être étendu selon la méthode linéaire statique utilisé, les variables impliquées et les matrices de variance-covariance leurs étant associées. Notons qu'étant donnée la nature divergente du noeud représentant la variable observée \mathbf{x} , ce dernier peut être augmenté par un noeud discret parent afin d'être surveillé à son tour par une statistique quadratique. Aussi, les noeuds représentant une variable multivariée dans le réseau peuvent être décomposés en plusieurs noeuds (univariés ou plusieurs noeuds joints) afin de, par exemple, gérer les observations manquantes et/ou surveiller plusieurs sous-espaces.

Nous avons jusqu'à présent considéré uniquement le cas statique, alors que d'autres méthodes dynamiques peuvent être utilisées pour la détection de fautes. Parmi ces méthodes, on peut citer celles se basant sur des modèles ARMA (Autoregressive Moving Average Mode, [Basseville et al. \(1993\)](#)). Ces méthodes peuvent être également représentées par un réseau Bayésien dynamique. Par exemple, une méthode se basant sur de tels modèles, plus précisément sur un modèle AR (Auto régressif), est la carte de contrôle MEWMA

(Multivariate Exponentially Weighted Moving Average, [Lowry et al. \(1992\)](#)). Cette statistique permet d'ailleurs de généraliser la statistique T^2 . Elle se base sur l'écriture récursive suivante :

$$\begin{cases} \mathbf{z}(t) = \lambda \mathbf{x}(t) + (I - \lambda) \mathbf{z}(t - 1) \\ \mathbf{z}(0) = \mu^{\mathbf{x}} \end{cases} \quad (4.43)$$

où μ est la moyenne des observations et λ est une matrice diagonale de pondération pouvant être régulée selon le besoin de détecter des sauts de faibles amplitudes ou des sauts de fortes amplitudes ([Testik and Borrer, 2004](#)). Étant donné ceci, la statistique MEWMA permettant de surveiller la variable \mathbf{x} à un instant t peut s'écrire comme ceci :

$$\Delta = MEWMA = \mathbf{z}(t)^T \Sigma^{\mathbf{z}(t)} \mathbf{z}(t) \quad (4.44)$$

où $\Sigma^{\mathbf{z}(t)}$ est la matrice de variance-covariance de \mathbf{z} à un instant t .

Nous proposons de réécrire (4.43) de façon probabiliste et ainsi modéliser (4.44) sous l'écriture suivante :

$$\begin{cases} \mathbf{x}(t) = \lambda^+ \mathbf{z}(t) - \lambda^+ (I - \lambda) \mathbf{z}(t - 1) + v \\ \mathbf{z}(0) = \mu \end{cases} \quad (4.45)$$

avec $\mathbf{z}(\cdot) \sim \mathcal{N}(0; \Sigma_k^{\mathbf{z}})$, $k \in \{CFN, F\}$, $v \sim \mathcal{N}(0; vI)$, $v \approx 0$, $\mathbf{x}(t) | \mathbf{z}(t), \mathbf{z}(t - 1) \sim \mathcal{N}(\lambda^+ \mathbf{z}(t) - \lambda^+ (I - \lambda) \mathbf{z}(t - 1); vI)$. Nous utilisons la pseudo-inverse de Moore-Penrose λ^+ de λ pour gérer le cas où un élément dans sa diagonale est nulle (la matrice λ peut être définie de diverses manières selon le besoin affiché). La matrice de variance-covariance de $\mathbf{z}(\cdot)$ est définie comme suit :

$$\begin{cases} k = CFN : \Sigma_{CFN}^{\mathbf{z}} = I \\ k = F : \Sigma_F^{\mathbf{z}} = \mathbf{c}_\Delta I \end{cases} \quad (4.46)$$

Le Réseau Bayésien Dynamique (RBD) représentant le modèle donné dans (4.45) est donné dans la figure (4.10). Ce réseau, représentant et modélisant une carte de contrôle MEWMA, peut être augmenté par des noeuds discrets permettant de surveiller les noeuds observés (ceci revient à l'association de la carte T^2 et la carte MEWMA). Toutefois, la structure de ce réseau pose problème en termes d'inférence. En effet, il n'est pas possible d'avoir une inférence exacte sur ce type de réseau. Ceci est dû au fait que plus on déroule le

réseau Bayésien plus on se retrouve avec des calculs impliquant des mélanges d'un nombre croissant de Gaussiennes. Nous verrons par la suite comment modéliser une carte MEWMA sur un RBD avec une possible inférence exacte.

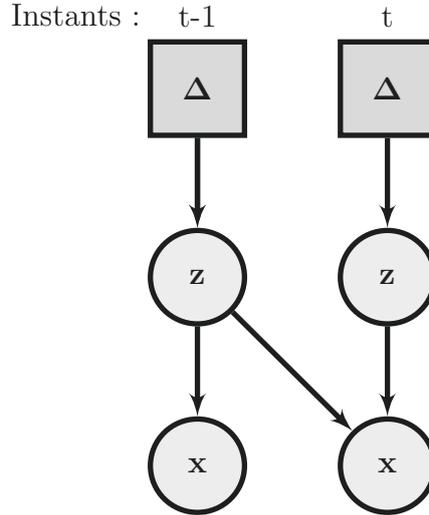


FIGURE 4.10 – Un réseau Bayésien représentant une carte MEWMA

Nous proposons également de modéliser d'autres méthodes dynamiques de détection sous réseau Bayésien. Précédemment dans la section II, nous avons abordé brièvement l'espace de parité qui est une méthode à base de modèle. Elle génère des résidus, en se basant sur une équation d'entrées et sorties, étant donnée une fenêtre de temps $(h + 1)$, selon l'équation suivante :

$$\mathbf{r}_h(t) = W(\mathbf{y}_h(t) - H_u \mathbf{u}_h(t)), W\Gamma = 0 \quad (4.47)$$

où

$$\mathbf{y}_h(t) = \begin{pmatrix} \mathbf{y}(t-h) \\ \mathbf{y}(t-h+1) \\ \vdots \\ \mathbf{y}(t) \end{pmatrix}, \quad \mathbf{u}_h(t) = \begin{pmatrix} \mathbf{u}(t-h) \\ \mathbf{u}(t-h+1) \\ \vdots \\ \mathbf{u}(t) \end{pmatrix} \quad (4.48)$$

Les variables $\mathbf{u}(\cdot)$ et $\mathbf{y}(\cdot)$ correspondent respectivement aux entrées et sorties temporelles passées du système. W correspond à une matrice choisie de sorte à ce qu'elle soit insensible aux perturbations et à la variable d'état inconnue,

et sensible aux fautes.

L'équation (4.47) peut être réécrite comme ceci :

$$\mathbf{r}_h(t) = \begin{pmatrix} W & -WH_u \end{pmatrix} \begin{pmatrix} \mathbf{y}_h(t) \\ \mathbf{u}_h(t) \end{pmatrix} \quad (4.49)$$

en posant :

$$\mathbf{x}_h(t) = \begin{pmatrix} \mathbf{y}_h(t) \\ \mathbf{u}_h(t) \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_p \end{pmatrix}, \quad (4.50)$$

$$O = \begin{pmatrix} W & -WH_u \end{pmatrix} \quad (4.51)$$

où $p = 2(h + 1)$.

alors (4.47) peut s'écrire comme ceci :

$$\mathbf{r}_h(t) = O\mathbf{x}_h(t) \quad (4.52)$$

Les résidus générés peuvent être utilisés pour la détection de fautes. Des statistiques quadratiques comme SPE peuvent être utilisées pour les surveiller. En se basant sur (4.52) nous proposons une représentation de ce générateur de résidus, dans le cas où les matrices du modèle sont invariantes dans le temps, sous réseau Bayésien. Pour cela, nous réécrivons de façon probabiliste (4.52) comme ci-dessous :

$$\mathbf{x}_h(t) = O^+\mathbf{r}_h(t) + v \quad (4.53)$$

où O^+ est la pseudo-inverse de O et

$$p(\mathbf{r}_h(\cdot)) = \mathcal{N}(0, I), \quad (4.54)$$

$$p(\mathbf{x}_h(\cdot)|\mathbf{r}_h(\cdot)) = \mathcal{N}(O^+\mathbf{r}_h(\cdot); vI), v \approx 0, \quad (4.55)$$

$$p(\mathbf{r}_h(\cdot)|\mathbf{x}_h(\cdot)) = \mathcal{N}(M(O^+)^T\mathbf{x}_h(\cdot); vM), M \approx ((O^+)^T O^+)^{-1} \quad (4.56)$$

En fonction de cela, nous proposons un réseau Bayésien permettant de modéliser un schéma de détection se basant sur un générateur d'espace de parité. Le réseau Bayésien proposé est fourni dans la figure 4.11.

Nous pouvons constater que lorsque W existe pour un $h = 0$, le schéma

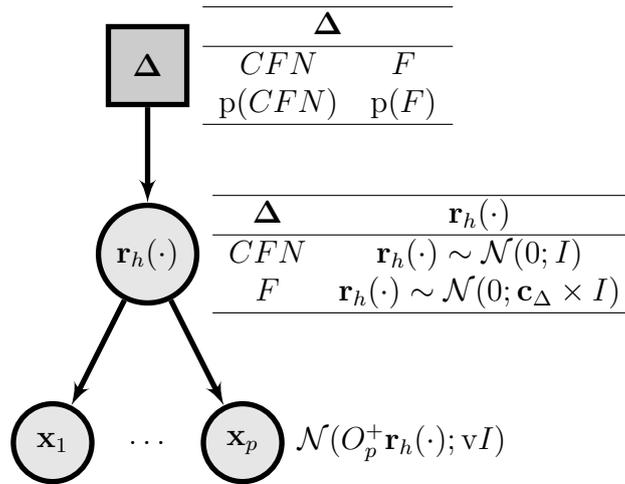


FIGURE 4.11 – Un réseau Bayésien représentant un générateur de résidus : espace de parité, avec une fenêtre h donnée et où O_p^+ correspond à la ligne p dans la matrice O^+

de détection se basant sur l'écriture (4.53) peut être présenté par le réseau Bayésien donné dans la figure 4.12. De plus, sous un réseau Bayésien comme celui présenté dans (4.7), nous pouvons associer un schéma de détection par analyse en composantes principales et un autre se basant sur l'espace de parité. En effet, il s'agit de surveiller le sous espace principal généré par l'analyse en composantes principales par une statistique comme celle du T^2 et surveiller les résidus générés par l'espace de parité (de même dimension que ceux générés par l'ACP en projetant les observations dans l'espace résiduel) par SPE ou une autre statistique quadratique. Cela est possible étant donnée l'équivalence prouvée entre l'espace résiduel et les résidus générés par l'espace de parité (Hagenblad et al., 2004; Gertler et al., 1999). Ceci est intéressant, car on vient associer deux méthodes de nature différente sous un même outil. D'ailleurs cela pourrait être d'une grande utilité pour détecter une faute en utilisant à la fois les données lorsqu'elles sont disponibles et/ou le modèle quand il est présent (permettant, par exemple, de déterminer le nombre de composantes principales).

Le réseau présenté dans 4.11 pourrait être également utilisé pour des fins d'isolation et du diagnostic de fautes. Il permet d'ailleurs de généraliser les réseaux proposés dans (Verron et al., 2010a; Weber et al., 2008; Pernestål et al., 2006; Schwall and Gerdes, 2002) permettant un traitement des résidus/symptômes à partir d'une modélisation d'une matrice d'incidence. En ef-

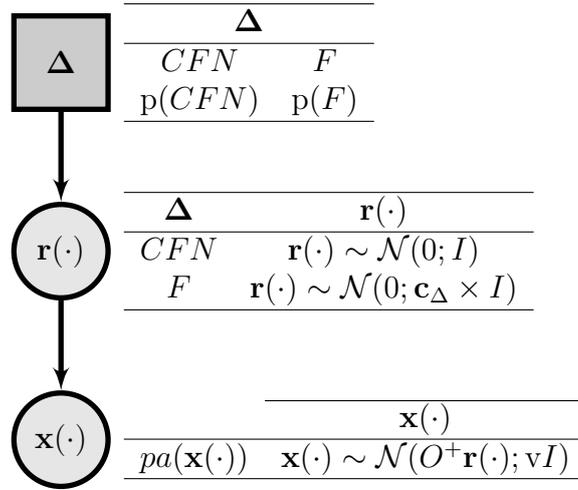


FIGURE 4.12 – Un réseau Bayésien représentant un générateur de résidus : espace de parité, où $h = 1$

fet, nous pourrions très bien augmenter ce réseau de sorte à modéliser une méthode à base de modèle entière (génération, évaluation des résidus et prise de décision). Pour cela, étant donnée l'hypothèse émise d'indépendance des résidus (matrice de variance-covariance diagonale), il suffit d'associer à chacun la ou les fautes dont il est sensible selon la matrice d'incidence. De plus, le réseau proposé dans 4.11 permettant de discriminer entre deux classes peut être facilement étendu pour discriminer entre $K, K > 2$ classes, et donc diagnostiquer une faute parmi les différentes fautes pouvant influencer les résidus. Plus encore, les noeuds $\mathbf{x}_1, \dots, \mathbf{x}_p$ sous forme multivariées peuvent être représentés par d'autres noeuds univariés, $((s \times (h + 1)) + (l \times (h + 1)))$ noeuds, où s et l sont respectivement les dimensions des variables $\mathbf{y}(\cdot)$ et $\mathbf{u}(\cdot)$, par définition, permettant ainsi de gérer les observations manquantes que ce soit lors de la détection, isolation et diagnostic de fautes.

Par ailleurs, d'autres générateurs de résidus linéaires invariants dans le temps (LTI, Linear Time-Invariant) peuvent être modélisés sous réseau Bayésien. Ces générateurs peuvent être généralisés, selon (Ding, 2012, 2008), par l'écriture ci-dessous :

$$\mathbf{r}(z) = R(z)\hat{N}(z)\mathbf{y}(z) - R(z)\hat{M}(z)\mathbf{u}(z) \quad (4.57)$$

De la même manière que l'espace de parité nous proposons de l'écrire

comme ceci :

$$\mathbf{x}(z) = O_g \mathbf{r}(z) + v \quad (4.58)$$

avec $O_g = [R(z)\hat{N}(z) - R(z)\hat{M}(z)]^+$. Bien entendu $R(z)$, $\hat{N}(z)$ et $\hat{M}(z)$ varient selon le générateur employé et permettent de généraliser certains générateurs de résidus pouvant être représentés sur un réseau Bayésien. Sous cette écriture, nous pourrions également tenir compte et intégrer des résultats obtenus par des méthodes de sous espaces notamment ceux proposés par (Ding, 2014; Ding et al., 2009). Par exemple, dans (Ding, 2014), les auteurs proposent d’identifier des générateurs de résidus comme celui de l’espace de parité directement à partir des données.

De même, d’autres méthodes issues du traitement de signal peuvent être également représentées sous réseau Bayésien. Par exemple, la décomposition discrète en ondelettes (Mallat, 1989) peut l’être après avoir été transcrite sous l’écriture suivante :

$$\mathbf{x} = O\mathbf{r} + v \quad (4.59)$$

où O représente l’inverse de la matrice orthogonale correspondante à la transformée discrète en ondelettes, \mathbf{r} représente les caractéristiques de \mathbf{x} , des observations d’une fonction donnée, sur J échelles.

Pour conclure, en se basant sur un réseau Bayésien comme celui présenté dans 4.11, nous pourrions représenter des méthodes à base de données et des méthodes à base de modèle à des fins de détection selon le problème considéré.

IV Applications

Nous proposons de confirmer la validité et la faisabilité de notre cadre probabiliste pour la détection de fautes. Pour cela, nous comparons un schéma de détection par analyse en composantes principales et le réseau Bayésien lui correspondant. Cette comparaison est effectuée sur les données du Tennessee Eastman Process (TEP). Ensuite, dans un exemple illustratif, nous illustrons la faculté de ce cadre probabiliste, comparativement aux schémas standards de détection, de gérer des observations manquantes. Cette faculté est valable sur les réseaux Bayésiens dont les variables observées sont considérées comme étant

indépendantes étant donnés les noeuds cachés, avec une matrice R diagonale.

IV.1 Tennessee Eastman Process

Le TEP est un système chimique industriel existant réellement dont la simulation fournie par la compagnie Eastman Chemical (le code FORTRAN est disponible sur internet) est largement utilisée comme un Benchmark pour les techniques de contrôle ainsi que pour les méthodes de détection et de diagnostic de fautes. La figure 4.13 décrit le diagramme de flux du système.

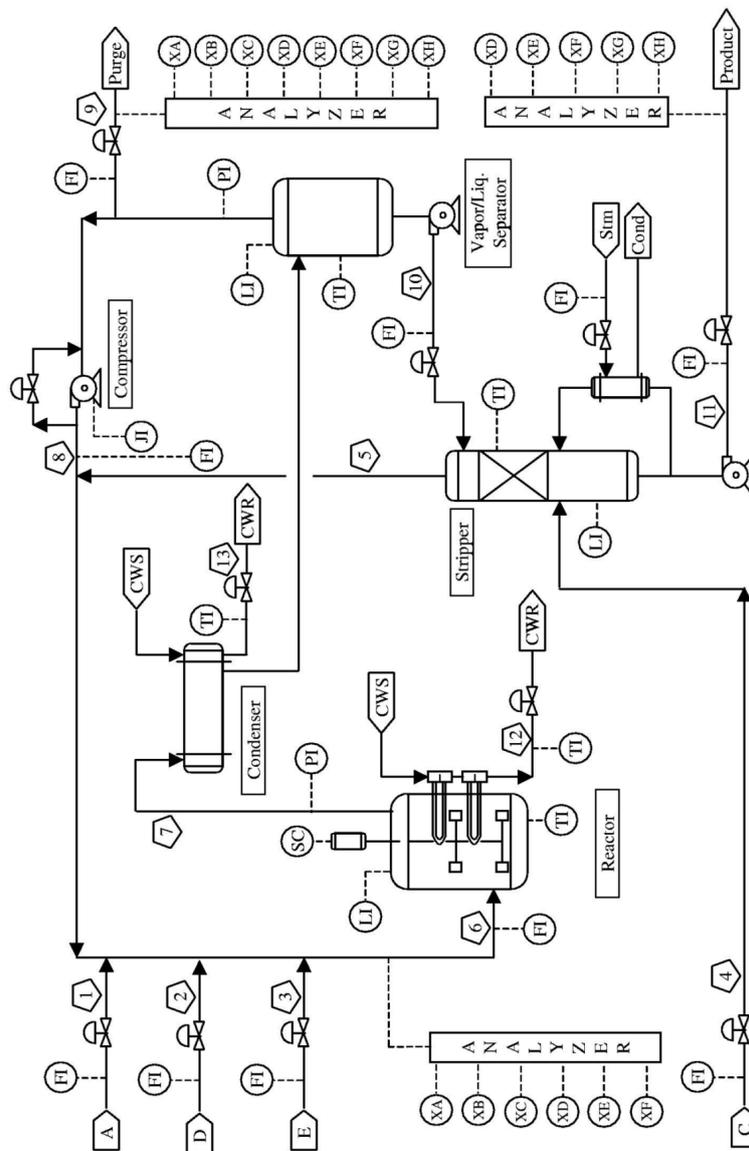


FIGURE 4.13 – Tennessee Eastman Process

Comme décrit dans (Downs and Vogel, 1993), il est composé de cinq composantes majeures : un réacteur, un condenseur, un compresseur, un décapeur, un séparateur. Toutefois, pour des raisons de confidentialité, il n'est pas la copie exacte du système réel. Le TEP se base sur quatre gaz réactifs et un autre non réactif pour générer deux composants liquides ainsi qu'un dérivé de production. Ceci donne lieu à 8 composants dénotés A, B, C, D, E, F, G, H. Bien entendu, le modèle mathématique du TEP n'est pas fourni. On peut dire que ce Benchmark est dédié au test des méthodes de surveillance à base de données.

Dans ce système, 52 variables sont généralement considérées. Parmi elles 41 sont mesurables (observées, voir les figures 4.14, 4.15) et 11 autres manipulées (variables d'asservissement, voir la figure 4.16). Les 41 variables mesurables ne sont pas toutes observables en continu. Uniquement 22 variables sont mesurables en continu (les 22 premières variables dans la figure 4.14) et les autres variables sont échantillonnées. Ces dernières correspondent aux dernières variables décrites dans la figure 4.15.

Le TEP permet de simuler 21 fautes de nature différentes. Nous les décrivons dans la figure 4.17. Dans notre cas d'étude, comme dans (Yin et al., 2012), nous considérons uniquement les 22 variables observées et les 11 variables manipulées ($m = 33$). Bien évidemment, ces variables sont sujettes au bruit et même parfois à des dynamiques étant donnée leur sensibilité/robustesse aux différentes fautes pouvant être simulées sur le TEP. Notons que le choix de tester notre méthode sur 33 variables n'est pas dû à une contrainte quelconque.

Nous nous appuyons également sur la base de données proposée par (Chiang et al., 2001) pour la surveillance et elle est à disposition sur internet pour des éventuelles comparaisons entre méthodes de détection et de diagnostic. Celle-ci est divisée en deux parties : l'ensemble de données d'apprentissage et l'ensemble des données de test. Ces deux ensembles sont à leur tour composés en 22 sous-parties. Ces sous-parties représentent les 21 classes de fautes, décrites dans la figure 4.17, en plus de la classe de fonctionnement normal (*CFN*). Ces sous-parties comportent un nombre précis d'observations. Dans l'ensemble d'apprentissage, chacune de ces sous-parties fait référence à 480 observations issues de la simulation du TEP pour chacune des fautes sur une période de 24 heures avec une période d'échantillonnage de 3 minutes commune à toutes les variables. En ce qui concerne les 800 observations correspondant aux sous-parties de l'ensemble de test, elles sont obtenues sur une période de 40 heures.

Variable	Description	Unité
XMEAS(1)	Débit d'alimentation en A	kscmh
XMEAS(2)	Débit d'alimentation en D	kg/hr
XMEAS(3)	Débit d'alimentation en E	kg/hr
XMEAS(4)	Débit d'alimentation total	kscmh
XMEAS(5)	Débit de recyclage	kscmh
XMEAS(6)	Débit d'alimentation du réacteur	kscmh
XMEAS(7)	Précision du réacteur	kPa
XMEAS(8)	Niveau du réacteur	%
XMEAS(9)	Température du réacteur	C
XMEAS(10)	Débit de purge	kscmh
XMEAS(11)	Température du séparateur	°C
XMEAS(12)	Niveau du séparateur	%
XMEAS(13)	Pression du séparateur	kPa
XMEAS(14)	Débit du séparateur	m ³ /hr
XMEAS(15)	Niveau du séparateur	%
XMEAS(16)	Pression du décapeur	kPa
XMEAS(17)	Débit du décapeur	m ³ /hr
XMEAS(18)	Température du séparateur	°C
XMEAS(19)	Débit de gaz au séparateur	kg/hr
XMEAS(20)	Puissance du compresseur	kW
XMEAS(21)	Température de ref. liq. en sortie de réacteur	°C
XMEAS(22)	Température de ref. liq. en sortie de séparateur	°C

FIGURE 4.14 – La liste des variables mesurables en continu du TEP

Afin de comparer le réseau Bayésien proposé au schéma de détection par analyse en composantes principales, nous avons retenu $a = 9$ composantes principales en se basant sur la statistique PRESS (Predicted Residual Sum of Squares) et l'ensemble d'apprentissage du fonctionnement normal (480 échantillons). Les deux approches sont comparées en utilisant les deux indexes : TFA (Taux de Fausses Alarmes, voir équation 4.61) et TMD (Taux de Mauvaises Détections, voir équation 4.60) sur 21 classes de données de test, chacune avec 800 échantillons (classe 1 représente la classe de fonctionnement normal alors que les autres 20 classes représentent les 20 premières fautes).

$$TMD = 100 \times \frac{\text{Nbr. d'échantillons} \in F_j \text{ considérés comme } CFN}{\text{Nbr. Tot. d'échantillons} \in F_j} \quad (4.60)$$

$$TFA = 100 \times \frac{\text{Nbr. Tot. d'échantillons} \in CFN \text{ considérés comme } F}{\text{Nbr. Tot. d'échantillons} \in CFN} \quad (4.61)$$

Dans l'annexe A, nous présentons les résultats du traditionnel schéma

Variable	Description	Unité
XMEAS(23)	Analyse du débit d'alimentation du réacteur : composant A	mol%
XMEAS(24)	Analyse du débit d'alimentation du réacteur : composant B	mol%
XMEAS(25)	Analyse du débit d'alimentation du réacteur : composant C	mol%
XMEAS(26)	Analyse du débit d'alimentation du réacteur : composant D	mol%
XMEAS(27)	Analyse du débit d'alimentation du réacteur : composant E	mol%
XMEAS(28)	Analyse du débit d'alimentation du réacteur : composant F	mol%
XMEAS(29)	Analyse de la purge du gaz : composant A	mol%
XMEAS(30)	Analyse de la purge du gaz : composant B	mol%
XMEAS(31)	Analyse de la purge du gaz : composant C	mol%
XMEAS(32)	Analyse de la purge du gaz : composant D	mol%
XMEAS(33)	Analyse de la purge du gaz : composant E	mol%
XMEAS(34)	Analyse de la purge du gaz : composant F	mol%
XMEAS(35)	Analyse de la purge du gaz : composant G	mol%
XMEAS(36)	Analyse de la purge du gaz : composant H	mol%
XMEAS(37)	Analyse du produit : composant D	mol%
XMEAS(38)	Analyse du produit : composant E	mol%
XMEAS(39)	Analyse du produit : composant F	mol%
XMEAS(40)	Analyse du produit : composant G	mol%
XMEAS(41)	Analyse du produit : composant H	mol%

FIGURE 4.15 – La liste des variables mesurables échantillonnées du TEP

Variable	Description	Unité
XMV(1)	Débit d'alimentation en D	kg/hr
XMV(2)	Débit d'alimentation en E	kg/hr
XMV(3)	Débit d'alimentation en A	kscmh
XMV(4)	Débit d'alimentation en A et C	kscmh
XMV(5)	Valve de recyclage du compresseur	%
XMV(6)	Valve de purge	%
XMV(7)	Débit d'alimentation du séparateur	m ³ /hr
XMV(8)	Débit d'alimentation du décapeur	m ³ /hr
XMV(9)	Valve du décapeur	%
XMV(10)	Débit du refroidissement liquide au réacteur	m ³ /hr
XMV(11)	Débit du refroidissement liquide au condenseur	m ³ /hr

FIGURE 4.16 – La liste des variables manipulées du TEP

d'analyse en composantes principales pour la détection de fautes en utilisant les statistiques T^2 et SPE, et son équivalent sur réseau Bayésien sur les 21 classes de données ($CFN + 20$ fautes). Rappelons que pour les statistiques T^2 et SPE, un dépassement de leurs limites LC_{T^2} et LC_{SPE} signifie qu'une faute est survenue dans le système. De la même manière, concernant le réseau Bayésien proposé, lorsque la probabilité a posteriori de l'état F du noeud \mathbf{T}^2

Numéro	Description	Nature
F_1	Ratio d'alimentation A/C	Saut
F_2	Composition en B	Saut
F_3	Température d'alimentation en D	Saut
F_4	Température d'entrée du ref. liq. au réacteur	Saut
F_5	Température d'entrée du ref. liq. au condenseur	Saut
F_6	Baisse d'alimentation	Saut
F_7	Perte de pression de l'alimentation en C	Saut
F_8	Composition d'alimentation en A, B et C	Variation aléatoire
F_9	Température d'alimentation en D	Variation aléatoire
F_{10}	Température d'alimentation en C	Variation aléatoire
F_{11}	Température d'entrée du ref. liq. au réacteur	Variation aléatoire
F_{12}	Température d'entrée du ref. liq. au condenseur	Variation aléatoire
F_{13}	Cinétiques des réactions	Dérive lente
F_{14}	Valve du ref. liq. au réacteur	Bloquée
F_{15}	Valve du ref. liq. au condenseur	Bloquée
F_{16}	Inconnue	Inconnue
F_{17}	Inconnue	Inconnue
F_{18}	Inconnue	Inconnue
F_{19}	Inconnue	Inconnue
F_{20}	Inconnue	Inconnue
F_{21}	Vanne fixée dans une position stable	Constante

FIGURE 4.17 – La liste des différentes fautes impliquées dans le TEP

(resp. **SPE**) est supérieure ou égale à sa limite de contrôle probabiliste $\zeta_{T^2}^F$ (resp. ζ_{SPE}^F). Les résultats obtenus (les graphiques présentés dans l'annexe A) montrent que les deux méthodes sont identiques, elles fournissent les mêmes décisions à chaque instant (génèrent les mêmes TFAs et TMDs).

IV.2 Hot Forming Process

L'habilité du cadre proposé à gérer les observations manquantes est illustré sur le Hot Forming Process (HFP) présenté dans la figure 4.19. Il représente une illustration physique en deux dimensions d'un système de façonnage à chaud. Ceci est un simple cas d'étude (Li et al., 2008) avec 5 variables illustrées dans la figure 4.18.

Ces variables sont causalement liées entre elles. Par exemple, la variable \mathbf{x}_2 et \mathbf{x}_3 influent directement sur la variable \mathbf{x}_5 . Les variables \mathbf{x}_1 et \mathbf{x}_4 influent indirectement sur la variable \mathbf{x}_5 . En effet, ces deux variables affectent respectivement les variables \mathbf{x}_2 et \mathbf{x}_3 . Ces relations peuvent être facilement re-

Variable	Description
\mathbf{x}_1	Variable qualité
\mathbf{x}_2	Stress des flux de matières
\mathbf{x}_3	Tension dans la pièce
\mathbf{x}_4	Force du serre flan
\mathbf{x}_5	Variable qualité : la dimension final de la pièce

FIGURE 4.18 – La liste des variables du HFP

présentées par un modèle de régression ou un réseau conditionnel Gaussien en supposant que les 5 variables suivent une distribution Gaussienne. En considérant un changement de moyenne avec une amplitude ms dans ces variables, 31 scénarios sont alors envisageables : 5 fautes simples (concernant une seule variable) et 26 fautes multiples (correspondant aux différentes combinaisons des 5 fautes simples).

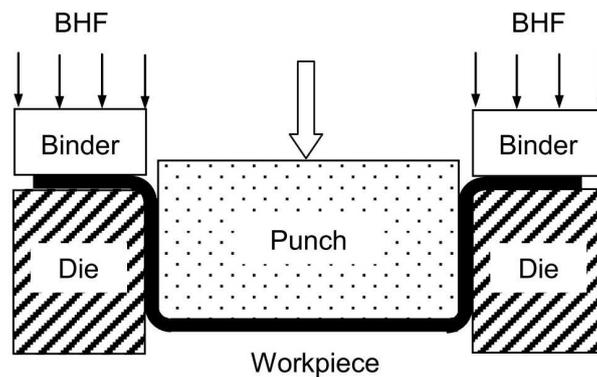


FIGURE 4.19 – Hot Forming Process

Dans le même contexte que précédemment, nous mettons l'accent sur le schéma de détection par analyse en composantes principales. Pour cela, nous avons simulé 100 observations de fonctionnement normal pour chaque variable, à partir d'elles les axes orthogonaux P et les a principaux sont définis.

Nous introduisons une faute avec $ms = 5$ dans \mathbf{x}_1 . La figure 4.20 présente le réseau Bayésien proposé représentant un schéma de détection par analyse en composantes principales (avec $c_{SPE} = c_{T^2} = 2$, $m=5$ et $\alpha = 1\%$) implémenté dans deux situations. Dans la figure 4.20.(a), l'échantillon de test est complet (toutes les variables sont observées) et dans la figure 4.20.(b) la valeur de la variable \mathbf{x}_2 est manquante (noeud blanc dans le graphe, $\mathbf{x}^- = \mathbf{x}_2$ et $\mathbf{x}^+ = [\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5]^T$).

Nous montrons que le réseau Bayésien proposé reste toujours opérationnel

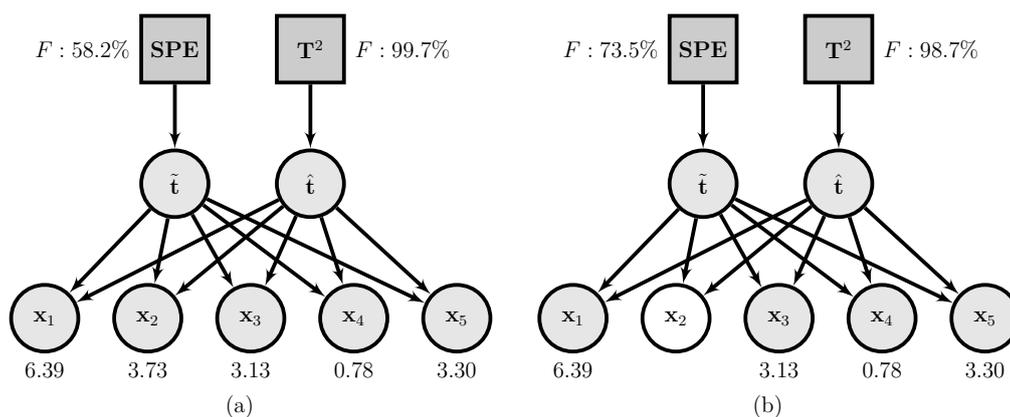


FIGURE 4.20 – Un exemple illustratif lorsque une faute est apparue dans x_1 (avec $ms=5$) avec $\zeta_{\text{T}^2}^F = 88.5\%$ et $\zeta_{\text{SPE}}^F = 47.9\%$. (a). Schéma de détection de fautes par analyse en composantes principales sous un réseau Bayésien lorsque toutes les variables sont disponibles. (b). Schéma de détection de fautes par analyse en composantes principales sous un réseau Bayésien lorsque la variable x_2 est manquante.

lorsque certaines variables sont non-observées et reste capable, de façon naturelle, de prendre une décision (donnent les probabilités de CFN avec F de chaque noeud discret en marginalisant sur x_2) en temps réel.

V Conclusion

Dans ce chapitre, nous avons proposé et introduit des réseaux Bayésiens originaux permettant de faire de la détection tout en tenant compte d'un niveau de signification α . Nous nous sommes concentrés, pour des raisons de clarté, sur le schéma de détection de fautes par analyse en composantes principales. Nous avons intégré l'analyse en composantes principales ainsi que les statistiques quadratiques qui lui sont associés (T^2 , SPE) sous un seul réseau Bayésien. Le cadre proposé projette une nouvelle observation dans un nouvel espace, identique à l'ACP, et fournit des probabilités sur l'état de fonctionnement du système et cela même dans le cas où les données observées sont incomplètes. Nous avons validé l'équivalence entre les deux approches sous le Tennessee Eastman Process (TEP). Pour conclure, nous avons proposé un cadre probabiliste pour les schémas de détection basés sur une transformation/projection linéaire des observations.

Diagnostic par réseau Bayésien

I Contexte et objectifs

Plusieurs méthodes ont été proposées pour le diagnostic de fautes à l'aide de réseaux Bayésiens. Dans ([Heckerman and Shwe, 1993](#)), un état de l'art et une comparaison des premières tentatives de diagnostic avec réseaux Bayésiens (les modèles noisy-or, multimembership Bayes, simple Bayes), majoritairement dédiées au diagnostic des maladies dans le domaine médical mais toutefois utilisables dans le cas du diagnostic de fautes dans les systèmes, est proposé par les auteurs. Dans ([Santoso et al., 1999](#)), un schéma de détection a été proposé pour une installation nucléaire. Les auteurs utilisent un RB discret dont la structure est définie par des experts, et les tables de probabilités conditionnelles apprises à partir des données d'apprentissage disponibles. Une procédure de création de schémas de diagnostic basée sur des réseaux Bayésiens discrets est proposée dans [Przytula and Thompson \(2000\)](#). Leur stratégie a été appliquée sur des locomotives diesel, des systèmes de communication satellite et des équipements satellite d'essai.

([Lerner et al., 2000](#)) proposent un réseau Bayésien dynamique hybride déduit d'un graphe causal temporel. Ce réseau représentant le comportement normal du système est utilisé comme un observateur. Pour le diagnostic, il est augmenté par des noeuds discrets afin de venir prendre une décision sur

l'état du système. Pour de tels réseaux Bayésiens, les auteurs introduisent un algorithme permettant de faire de l'inférence en évitant une explosion exponentielle dû à la présence de variables discrètes. Dans (Schwall and Gerdes, 2002) une évaluation des résidus en utilisant un réseau Bayésien dynamique, préalablement généré étant donné une matrice d'incidence et un historique de données de fautes, a été proposée.

Dans (Chien et al., 2002) un réseau Bayésien appris à partir d'un historique de données de fonctionnement normal et de fautes en plus des connaissances d'expert a été présenté par les auteurs. Il est utilisé pour le diagnostic des fautes dans un réseau de distribution d'électricité. (Mehranbod et al., 2005) propose un réseau Bayésien pour l'identification des différents types de fautes capteurs. Ces fautes sont représentées dans le réseau par des noeuds discrets de décision (composés chacun de plusieurs modalités). Une décision est prise, pour chaque noeud de décision, en surveillant dans le temps les variations des probabilités a posteriori par rapport à leurs valeurs lors du fonctionnement normal.

Dans (Dey and Stori, 2005), les auteurs abordent l'identification des causes racines des variations survenues dans un procédé d'usinage en utilisant une méthodologie originale basée sur les réseaux Bayésiens. Les auteurs s'appuient sur l'utilisation de plans d'expériences pour établir la structure et les paramètres du réseau. (Sheppard et al., 2005) proposent un réseau Bayésien pour le diagnostic de fautes. Ce réseau Bayésien est composé de noeuds discrets et sur les probabilités de mauvaises détections et de fausses alarmes estimées. Un autre réseau composé de noeuds discrets permettant de déterminer les capteurs en fautes est proposé dans (Matsuura et al., 2005). Ce réseau, comme vu précédemment, peut être utilisé pour la détection des fautes capteurs en examinant sa probabilité jointe. En ce qui concerne l'isolation, les auteurs proposent d'appliquer plusieurs fois la même stratégie que lors de la détection en ignorant à chaque fois un capteur donné. Le réseau Bayésien dynamique proposé par (Kawahara et al., 2005) et présenté précédemment dans la section I peut être utilisé également pour le diagnostic de fautes. Dans ce réseau, chaque faute est représentée par un noeud qui vient surveiller dans le réseau les variables concernées par cette faute selon les connaissances d'experts.

Dans (Roychoudhury et al., 2006), les auteurs s'appuient sur un Bond Graph pour définir la structure d'un réseau Bayésien dynamique mais aussi pour la prise de décision. Ils proposent une isolation qualitative qui vient op-

timiser une autre quantitative. Le schéma proposé par les auteurs permet de gérer les fautes naissantes et multiples. Il est utilisé sur le two tank benchmark. Dans (Pernestål et al., 2006) une comparaison de quatre méthodes d'isolation de fautes est fournie. Ces quatre méthodes correspondent à des réseaux Bayésiens ayant des structures différentes. Chaque structure représente une matrice d'incidence modélisée sous réseau Bayésien. La comparaison est effectuée dans le cadre du diagnostic du flux de gaz dans un moteur diesel.

Dans (Li and Shi, 2007), un réseau Bayésien composé de noeuds discrets est proposé. La structure de ce réseau est apprise, une fois qu'une sélection de variables est effectuée, à partir d'un historique de données ainsi que de connaissances d'experts. Le réseau permet d'identifier les variables responsables d'une non-conformité observée sur la caractéristique qualité considérée. Les auteurs donnent un exemple sur un système de laminage de barres. (Riascos et al., 2007) proposent un réseau Bayésien pour le diagnostic de fautes dans une pile à combustible. La structure du réseau a été apprise grâce à des méthodes probabilistes d'apprentissage sous contraintes en se basant sur une base de données de fautes ainsi que sur les connaissances d'experts. (Mechraoui et al., 2008) proposent un réseau Bayésien pour le diagnostic de fautes dans les moteurs électriques. L'identification des variables, leurs modes, les relations causales et la quantification des probabilités sont établis en fonction des connaissances d'experts. Dans (Huang, 2008) plusieurs autres réseaux Bayésiens dédiés au diagnostic de fautes simples et multiples dans le contexte des véhicules ont été proposés.

Dans (Weber et al., 2008) les auteurs proposent une architecture originale de réseaux Bayésiens permettant l'isolation de fautes. Les auteurs, à partir d'une matrice d'incidence, discriminent entre les différentes fautes sachant les nouveaux symptômes (résultat de l'évaluation des résidus) et la fiabilité des composants. (Pernestål, 2009) proposent une nouvelle représentation de la matrice d'incidence pouvant être apprise automatiquement à partir des données et des connaissances d'experts. Dans (Lo et al., 2011, 2003), une procédure générale pour la construction d'un réseau Bayésien à partir d'un Bond graphe a été proposée. Les relations causales entre variables sont générées à partir d'équations qualitatives et les tables de probabilités conditionnelles sont apprises à partir des données d'historiques et les connaissances d'experts. Dans (Zaidi et al., 2012), les auteurs s'appuient sur les relations de redondances tirées d'un Bond graphe pour aboutir à des résidus qui seront par la suite évalués

par un réseau Bayésien dont les décisions sont augmentées par les probabilités de fiabilité des composantes. Cette approche généralise celle proposée par (Weber et al., 2008). Les auteurs proposent dans (Yang and Lee, 2012) un réseau Bayésien pour l'identification des causes racines conduisant à des fautes dans le cadre de la production de micro-processeurs. Dans (Jiao et al., 2012), un autre réseau a été proposé dans le cadre des réseaux analogiques étant donnée l'hypothèse qu'un ensemble de données de fautes est disponible. Dans (Yu, 2013), un réseau Bayésien dynamique est utilisé. A ce réseau sont associés des indices permettant d'isoler et d'identifier les variables responsables d'un dérèglement du système ainsi qu'une stratégie permettant de déterminer efficacement les causes racines.

Dans ce qui suit, nous allons exposer une famille de réseaux Bayésiens hybrides qui sont, contrairement à ceux présentés précédemment, capable de réaliser de façon simultanée de la détection et du diagnostic tout en intégrant le rejet de distance. En d'autres termes, nous allons illustrer comment des méthodes assez répandues pour le diagnostic de fautes peuvent être modélisées sur réseau Bayésien et comment elles peuvent être utilisées en même temps pour la détection de fautes tout en gérant le rejet de distance.

II Limites spécifiques de surveillance

Une fois une faute détectée, des méthodes de classification supervisée peuvent être utilisées pour le diagnostic. En effet, le diagnostic peut être vu comme la solution d'un problème de classification entre différentes classes de fautes. La tâche de classification nécessite alors un classifieur. Dans la classification supervisée, le nombre de classes et la classe d'appartenance de chaque observation de l'ensemble d'apprentissage sont connus et utilisés pour apprendre le classifieur, permettant alors de classer une nouvelle observation dans une des classes connues.

Par exemple, étant donné un ensemble de données d'apprentissage d'un système bivarié avec deux fautes différentes connues, comme illustré dans la figure 5.1, nous pouvons facilement utiliser une classification supervisée pour classer une nouvelle observation en faute.

D'autres approches incluent la classe de fonctionnement normal *CFN* dans le classifieur. La figure 5.2 représente cette configuration : la classe de fonctionnement normal (cercles verts). Ces approches sont intéressantes car la détection

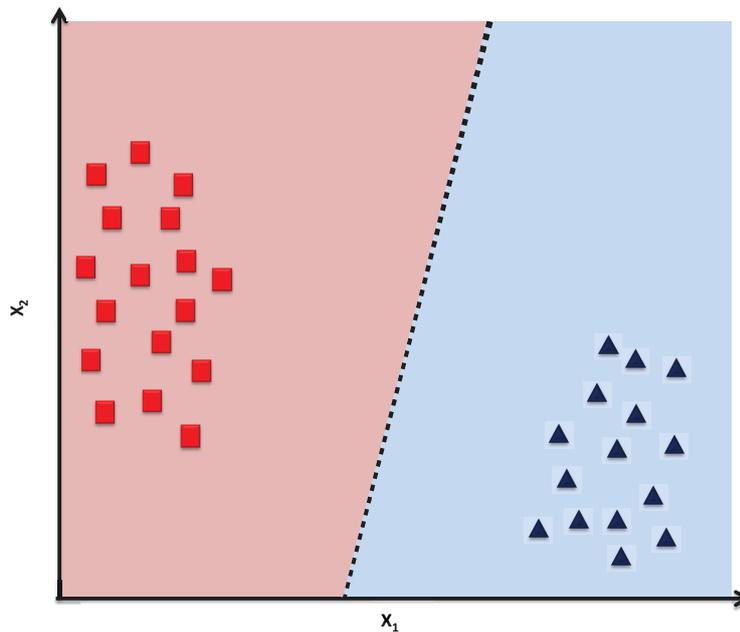


FIGURE 5.1 – Classification entre les classes de fautes

et le diagnostic de fautes peuvent être réalisés avec seulement un unique outil, un classifieur, dans une seule phase. Cependant, cette façon d'agir comporte deux inconvénients majeurs. Premièrement, le taux de fausses alarmes peut ne pas être assuré (Chiang et al., 2001). Comme montré dans la figure 5.2, la zone représentant la classe *CFN* n'est pas une ellipsoïde classique, il dépend d'autres classes (classes de fautes). Le second inconvénient apparaît lorsqu'une observation d'une faute donnée est classifiée dans la classe de fonctionnement normal car elle est la plus proche (statistiquement parlant) de celle-ci. Un troisième inconvénient, pouvant concerné également le diagnostic, est l'apparition d'une nouvelle observation très éloignée des différentes classes, voir figure 5.3. Dans ce cas, il faut considérer le fait que d'autres classes non considérées peuvent exister ou que de nouvelles classes viennent de faire leurs apparitions.

Ainsi, pour plusieurs cas pratiques, cette approche ne peut pas être appliquée car elle conduit à l'arrêt fréquent du système (plusieurs fausses alarmes), ou alors à ne pas détecter des fautes non désirables et ainsi les laisser évoluer dans le système.

Nous proposons, sur de tels classifieurs, de détecter une faute étant donné un niveau de signification α , tout en diagnostiquant les différentes fautes. De

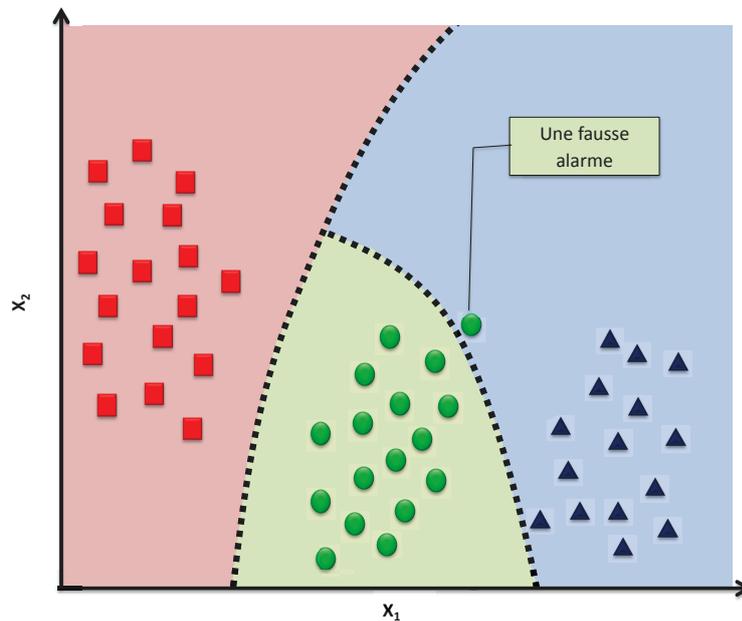


FIGURE 5.2 – Classification entre les classes de fautes et la classe de fonctionnement normal : exemple de fausse alarme

plus, l'approche proposée permet également d'intégrer le rejet de distance (Veron et al., 2010b). Pour effectuer cela, une limite probabiliste LP spécifique à chaque classe va être définie.

II.1 Démonstration du calcul de LP

Le problème du diagnostic de fautes, peut être vu comme un problème de discrimination entre plusieurs classes. Plusieurs méthodes pour la classification ont été proposées. Ces méthodes peuvent être discriminées en 3 catégories : supervisée, non-supervisée, semi-supervisée. Les réseaux Bayésiens appartiennent à la première catégorie, bien que quelques propositions dans la littérature les ont utilisé pour la classification non-supervisée. Ces modèles probabilistes généralisant l'analyse discriminante peuvent être utilisés pour le diagnostic. En effet, ils peuvent être utilisés pour le diagnostic en discriminant entre les différentes fautes du système. Ainsi, la faute ayant la plus grande probabilité d'occurrence est prise en compte. Sachant que chaque faute se voit attribuer une probabilité d'occurrence, elles peuvent être classées de façon décroissante permettant ainsi d'aider l'ingénieur ou l'opérateur à prendre une décision.

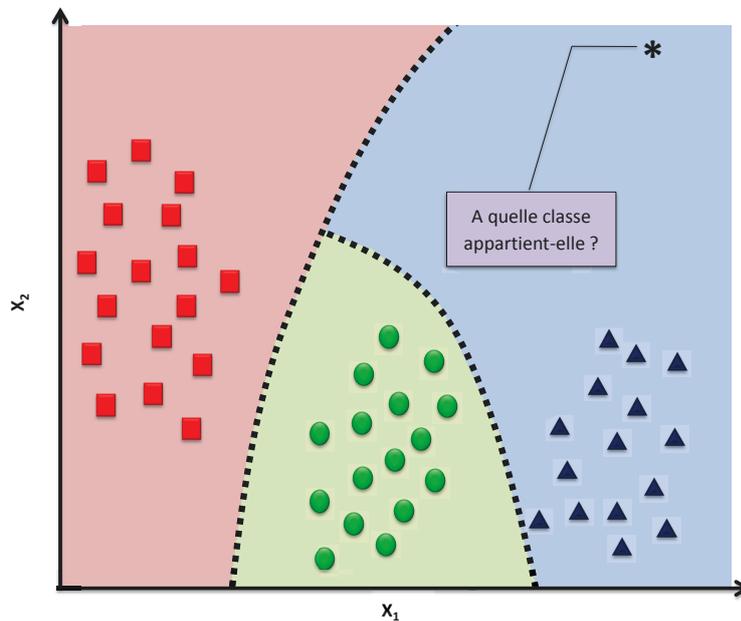


FIGURE 5.3 – Classification entre les classes de fautes et la classe de fonctionnement normal : exemple d'un rejet de distance

Nous présentons dans ce travail comment, sous un réseau Bayésien, nous pouvons prendre une décision pour détecter et diagnostiquer une faute tout en prenant en compte le rejet de distance. Pour ce faire, nous allons utiliser un réseau Bayésien discriminant entre les différentes fautes du système tout en intégrant la classe de fonctionnement normal afin d'obtenir une méthode complète pour la détection et le diagnostic de fautes. Pour prendre une décision dans ce type de réseau, nous présentons une généralisation de la règle de décision présentée précédemment dans le cadre de la détection (voir sous-section II), mais aussi d'autres règles susceptibles d'être également appliquées selon les différentes configurations considérées par la moyenne et la matrice de variance-covariance de chaque classe. Bien entendu, en plus de la décision prise par les réseaux Bayésiens que nous allons présenter s'ajoute une règle de décision permettant de garantir statistiquement un taux de fausse alarme, de détecter et de diagnostiquer les différentes fautes du système tout en prenant en compte le rejet de distance.

Afin de pouvoir tenir compte de cela sur un réseau Bayésien classifieur comme celui présenté dans la figure 5.4, nous allons augmenter la règle argmax

(attribution d'une observation à la classe avec la plus grande probabilité a posteriori) en une règle se basant sur des limites de probabilités respectant un niveau de signification α .

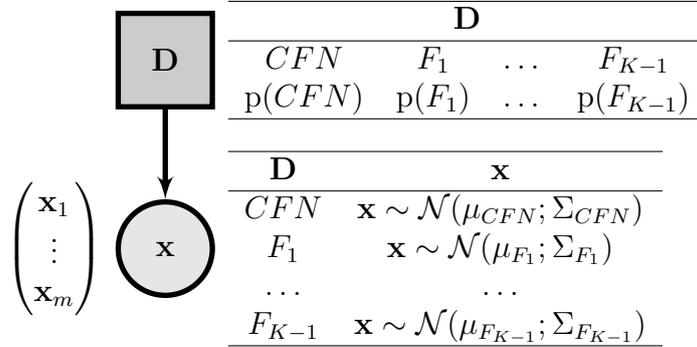


FIGURE 5.4 – Un exemple d'un réseau Bayésien pour la détection et le diagnostic de fautes (forme multivariée)

Bien évidemment, l'application de cette règle réside dans le calcul des limites probabilistes permettant de décider de manière statistique pour chaque classe. Dans ce qui suit nous allons démontrer, dans le cas général, la limite de probabilité $LP_{\Delta}^{C_k}$ associée à la probabilité a posteriori $p(\mathbf{D} = C_k|x)$ d'une classe C_k sachant une observation donnée x . Dans un premier temps et pour des raisons de clarté, nous allons noter l'état de fonctionnement normal *CFN* par classe C_1 et les $K - 1$ fautes respectivement par C_2, \dots, C_K . Dans un second temps, dans le but de respecter un niveau de signification α , nous faisons l'analogie entre un test se basant sur une statistique quadratique donnée (associée à une limite statistique) et l'analyse discriminante se basant sur la règle de Bayes. Ainsi, nous proposons de généraliser ces statistiques quadratiques Δ et leurs limites sous une représentation probabiliste permettant de prendre une décision sur la probabilité a posteriori de chaque classe étant données les nouvelles observations.

Rappelons qu'une statistique quadratique permet de prendre une décision en la comparant à la limite statistique lui étant associée. Ainsi, un test se basant sur une statistique Δ , pour une classe donnée C_k , peut s'écrire comme suit :

$$x \in C_k : \text{si } \Delta_{C_k} \leq LC_{\Delta}^{C_k} \quad (5.1)$$

où $LC_{\Delta}^{C_k}$ est une limite statistique définie étant donné la statistique quadra-

tique et les paramètres de la distribution de la classe CFN .

Nous voulons reproduire ou obtenir le même résultat que cette règle sous une autre règle équivalente se basant sur la probabilité a posteriori de la classe en question comme ceci :

$$x \in C_k : \text{ si } p(\mathbf{D} = C_k|x) \geq LP_{\Delta}^{C_k} \quad (5.2)$$

Sous cette règle nous sommes en mesure de prendre une décision sur le fait qu'une nouvelle observation x de \mathbf{x} appartienne ou non à la classe C_k étant donné la probabilité a posteriori de cette dernière.

Pour cela, il faut définir la limite $LP_{\Delta}^{C_k}$ correspondant à la classe $\mathbf{D} = C_k$. Soit x^* de \mathbf{x} , une observation vérifiant l'égalité suivante $\Delta_{C_k} = LC_{\Delta}^{C_k}$ et nous permettant d'écrire :

$$p(\mathbf{D} = C_k|x^*) = LP_{\Delta}^{C_k} \quad (5.3)$$

$$= LP_{\Delta}^{C_k} \sum_{j=1}^K p(\mathbf{D} = C_j|x) \quad (5.4)$$

où $\sum_{j=1}^K p(\mathbf{D} = C_j|x) = 1$ et x est une observation quelconque de \mathbf{x} .

Notons que, selon la règle de Bayes, la probabilité a posteriori de chaque classe C_k peut être écrite :

$$p(\mathbf{D} = C_k|x) = \frac{p(\mathbf{D} = C_k)p(x|\mathbf{D} = C_k)}{p(x)}, C_k \in \{C_1, \dots, C_K\} \quad (5.5)$$

En remplaçant (5.5) dans (5.4) nous obtenons :

$$LP_{\Delta}^{C_k} \left[\frac{\sum_{k=1}^K p(\mathbf{D} = C_k)p(x|\mathbf{D} = C_k)}{p(x)} \right] = \frac{p(\mathbf{D} = C_k)p(x^*|\mathbf{D} = C_k)}{p(x^*)}$$

$$LP_{\Delta}^{C_k} = \frac{p(x)}{p(x^*)} \frac{p(\mathbf{D} = C_k)p(x^*|\mathbf{D} = C_k)}{\sum_{k=1}^K p(x|\mathbf{D} = C_k)p(\mathbf{D} = C_k)} \quad (5.6)$$

avec $p(x) = \sum_{j=1}^K p(x|\mathbf{D} = C_j)p(\mathbf{D} = C_j)$.

En utilisant (5.6) nous pouvons réécrire (5.2) comme ceci :

$$x \in C_k, \text{ si } p(\mathbf{D} = C_k|x) \geq \frac{p(x)}{p(x^*)} \frac{p(\mathbf{D} = C_k)p(x^*|\mathbf{D} = C_k)}{p(x)} \quad (5.7)$$

Toutefois, on peut constater que le terme $p(x^*)$ dans (5.7) n'est pas défini. En effet, $p(x^*)$ est égale à $\sum_{j=1}^K p(\mathbf{D} = C_j)p(x^*|\mathbf{D} = C_j)$ où les vraisemblances des classes C_j sont différentes de la classe C_k considérée. Ceci est dû au fait que x^* n'est pas connu. Ainsi, pour gérer ce terme $p(x^*)$ dans (5.7) il suffit de simplifier (5.6) en éliminant le ratio $\frac{p(x)}{p(x^*)}$ de sorte à ce que :

$$x \in C_k, \text{ si } p(\mathbf{D} = C_k|x) \geq \frac{p(\mathbf{D} = C_k)p(x^*|\mathbf{D} = C_k)}{p(x)}, C_k \in CFN, F \quad (5.8)$$

avec $p(\mathbf{D} = C_k|x) = p(x)^{-1}p(x|\mathbf{D} = C_k)p(\mathbf{D} = C_k)$

La suppression de ce ratio ne modifie pas la prise de décision. Afin de prouver ceci, il suffit de démontrer que la condition (5.8) est vérifiée pour une observation x donnée appartenant à la classe C_k et vérifiant cette condition :

$$x \in C_k, \text{ si } \Delta_{C_k} \leq LC_{\Delta}^{C_k} \quad (5.9)$$

En se basant sur (5.9), nous allons démontrer, ci-dessous, que nous pouvons écrire (5.8) :

$x \in C_k$, si

$$\begin{aligned} p(\mathbf{D} = C_k|x) &\geq \frac{p(x^*|\mathbf{D} = C_k)p(\mathbf{D} = C_k)}{p(x)} \\ \frac{p(x|\mathbf{D} = C_k)p(\mathbf{D} = C_k)}{p(x)} &\geq \frac{p(x^*|\mathbf{D} = C_k)p(\mathbf{D} = C_k)}{p(x)} \\ p(x|\mathbf{D} = C_k)p(\mathbf{D} = C_k) &\geq p(x^*|\mathbf{D} = C_k)p(\mathbf{D} = C_k) \\ p(x|\mathbf{D} = C_k) &\geq p(x^*|\mathbf{D} = C_k) \\ \frac{1}{2\pi^{\frac{m}{2}}|\Sigma_{C_k}|^{\frac{1}{2}}} e^{-\frac{(x-\mu_{C_k})^T \Sigma_{C_k}^{-1} (x-\mu_{C_k})}{2}} &\geq \frac{1}{2\pi^{\frac{m}{2}}|\Sigma_{C_k}|^{\frac{1}{2}}} e^{-\frac{-LC_{\Delta}^{C_k}}{2}} \\ e^{-\frac{\Delta_{C_k}}{2}} &\geq e^{-\frac{-LC_{\Delta}^{C_k}}{2}} \\ \frac{-\Delta_{C_k}}{2} &\geq \frac{-LC_{\Delta}^{C_k}}{2} \\ \Delta_{C_k} &\leq LC_{\Delta}^{C_k} \end{aligned}$$

où $\Delta_{C_k} = (x - \mu_{C_k})^T \Sigma_{C_k}^{-1} (x - \mu_{C_k})$

Ainsi, on peut poser $LP_{\Delta}^{C_k} = p(x)^{-1}p(x^*|\mathbf{D} = C_k)p(\mathbf{D} = C_k)$.

Cette limite simplifiée doit cependant être calculée sous cette forme pour chaque nouvelle observation x de \mathbf{x} introduite dans le réseau Bayésien utilisé.

Cela est dû au terme $p(x)$. Ceci peut être évité, par exemple, dans le cas d'une analyse discriminante quadratique contrairement aux réseaux Bayésiens classifieurs où les probabilités a posteriori sont automatiquement calculées. En effet, la limite peut alors n'être définie qu'une seule fois, sans avoir besoin d'être redéfinie par la suite. Il suffit d'enlever le terme $p(x)$ dans les deux parties de la règle (5.7) et décider en fonction de la vraisemblance de chaque observation au lieu de $p(\mathbf{D} = C_k|x)$ comme ceci :

$$x \in C_k, \text{ si } p(x|\mathbf{D} = C_k) \geq p(x^*|\mathbf{D} = C_k) \quad (5.10)$$

où $p(x^*|\mathbf{D} = C_k)$ est calculé hors ligne.

Comme mentionné précédemment, dans notre contexte de détection et diagnostic de faute sous réseaux Bayésiens, C_k représente les différents états de fonctionnement du système : CFN, F_1, \dots, F_{K-1} . Chaque classe est supposée suivre une distribution normale avec des paramètres qui lui sont propres (connus ou estimés). Étant donnée cela, nous pouvons, par exemple, développer et écrire la limite probabiliste LP_{Δ}^{CFN} de la classe CFN comme ceci :

$$\begin{aligned} LP_{\Delta}^{CFN} &= \frac{p(\mathbf{D} = CFN)p(x^*|\mathbf{D} = CFN)}{p(x)} \\ &= \frac{p(\mathbf{D} = CFN)p(x^*|\mathbf{D} = CFN)}{p(\mathbf{D} = CFN)p(x|\mathbf{D} = CFN) + \dots + p(x|\mathbf{D} = F_{K-1})p(\mathbf{D} = F_{K-1})} \\ &= \frac{e^{-\frac{LC_{\Delta}^{CFN}}{2}}}{e^{-\frac{\Delta_{CFN}}{2}} + \omega_1 \frac{|\Sigma_{CFN}|^{\frac{1}{2}}}{|\Sigma_{F_1}|^{\frac{1}{2}}} e^{-\frac{\Delta_{F_1}}{2}} + \dots + \omega_{(K-1)} \frac{|\Sigma_{CFN}|^{\frac{1}{2}}}{|\Sigma_{F_{K-1}}|^{\frac{1}{2}}} e^{-\frac{\Delta_{F_{K-1}}}{2}}} \\ &= \frac{1}{e^{\frac{1}{2}\Delta_1}} \cdot \frac{1}{1 + \omega_1 \frac{|\Sigma_{CFN}|^{\frac{1}{2}}}{|\Sigma_{F_1}|^{\frac{1}{2}}} e^{-\frac{1}{2}\Delta_2} + \dots + \omega_{(K-1)} \frac{|\Sigma_{CFN}|^{\frac{1}{2}}}{|\Sigma_{F_{K-1}}|^{\frac{1}{2}}} e^{-\frac{1}{2}\Delta_K}} \end{aligned} \quad (5.11)$$

avec

$$\Delta_1 = LC_{\Delta}^{CFN} - \Delta_{CFN}, \quad (5.12)$$

$$\Delta_{j+1} = \Delta_{F_j} - \Delta_{CFN}, \quad (5.13)$$

$$\omega_j = \frac{p(\mathbf{D} = F_j)}{p(\mathbf{D} = CFN)} \quad (5.14)$$

où $j = 1, \dots, K - 1$ et $\Delta_{CFN} = (x - \mu_{CFN})^T \Sigma_{CFN}^{-1} (x - \mu_{CFN})$.

De la même manière, les autres limites $LP_{\Delta}^{F_1}, \dots, LP_{\Delta}^{F_{K-1}}$ correspondant

respectivement aux différentes fautes F_1, \dots, F_{K-1} peuvent être déterminées. Notons que les limites des différentes classes sont égales lorsque un niveau de signification α est fixé pour toutes les classes.

II.2 Détection de fautes

L'écriture (5.11) permet de généraliser la limite proposée précédemment dans le cas de la détection de fautes. Elle permet également de proposer une famille de cas de figures non présentés ou considérés dans le chapitre précédent (différentes hypothèses émises sur les paramètres de classes).

Considérant un réseau Bayésien classifieur discriminant entre la classe de fonctionnement normal CFN et une autre classe représentant tout ce qui ne fait pas partie du fonctionnement normal, notée F et dont les paramètres restent à définir. Dans ce cadre, la limite LP_{Δ}^{CFN} , selon (5.11), correspond à :

$$\begin{aligned}
 LP_{\Delta}^{CFN} &= \frac{p(\mathbf{D} = CFN)p(x^*|\mathbf{D} = CFN)}{p(x)} \\
 LP_{\Delta}^{CFN} &= \frac{p(\mathbf{D} = CFN)p(x^*|\mathbf{D} = CFN)}{p(\mathbf{D} = CFN)p(x|\mathbf{D} = CFN) + p(x|\mathbf{D} = F)p(\mathbf{D} = F)} \\
 &= \frac{\frac{1}{2\pi^{\frac{m}{2}}|\Sigma_{CFN}|^{\frac{1}{2}}}e^{-\frac{LC_{\Delta}^{CFN}}{2}}}{\frac{1}{2\pi^{\frac{m}{2}}|\Sigma_{CFN}|^{\frac{1}{2}}}e^{-\frac{\Delta_{CFN}}{2}} + \frac{\omega}{2\pi^{\frac{m}{2}}|\Sigma_F|^{\frac{1}{2}}}e^{-\frac{\Delta_F}{2}}} \quad (5.15)
 \end{aligned}$$

où $\Delta_{CFN} = (x - \mu_{CFN})^T \Sigma_{CFN}^{-1} (x - \mu_{CFN})$ et $\Delta_F = (x - \mu_F)^T \Sigma_F^{-1} (x - \mu_F)$.

L'équation (5.15) peut être représentée sous l'écriture suivante :

$$\begin{aligned}
 LP_{\Delta}^{CFN} &= \frac{1}{e^{\frac{LC_{\Delta}^{CFN}}{2}} e^{-\frac{\Delta_{CFN}}{2}}} \cdot \frac{1}{1 + \omega \frac{|\Sigma_{CFN}|^{\frac{1}{2}}}{|\Sigma_F|^{\frac{1}{2}}} e^{-\frac{1}{2}\Delta_2}} \\
 &= \tau \cdot \frac{1}{1 + \beta e^{-\gamma\varphi}} \quad (5.16)
 \end{aligned}$$

avec $\gamma = \frac{1}{2}$,

$$\varphi = \Delta_2 = \Delta_F - \Delta_{CFN}, \quad (5.17)$$

$$\tau = \frac{1}{e^{(LC_{\Delta}^{CFN} - \Delta_{CFN})}} = \frac{1}{e^{\frac{1}{2}\Delta_1}}, \quad (5.18)$$

$$\beta = \frac{\omega}{\nu_2}, \quad (5.19)$$

$$\nu_2 = \frac{|\Sigma_F|^{\frac{1}{2}}}{|\Sigma_{CFN}|^{\frac{1}{2}}} \quad (5.20)$$

et $\omega = \frac{p(\mathbf{D}=CFN)}{p(\mathbf{D}=F)}$.

Ainsi, nous pouvons constater que LP_{Δ}^{CFN} correspond à une fonction logistique $f(\varphi)$.

Ceci se justifie par le fait que τ est un réel positif (un exponentiel est par définition positif), ainsi que β dont le rapport des probabilités a priori non nuls et le rapport des déterminants (les matrices de variance-covariance sont par définition des matrices symétriques semi-définies positives non dégénérées) sont positifs.

La courbe représentative de cette fonction logistique (courbe en S) a pour asymptote $LP_{\Delta}^{CFN} = 0$, $LP_{\Delta}^{CFN} = \tau$, et centre de symétrie le point d'inflexion de coordonnées suivant :

$$I = \left(\frac{\ln(\beta)}{\gamma}, \frac{\tau}{2} \right) = \left(\ln(\beta), \frac{\tau}{2} \right) \quad (5.21)$$

Il est évident que les quatre variables $\tau, \gamma, \varphi, \beta$ influencent, à leur manière, la forme et l'évolution que la courbe de la fonction peut prendre. Ainsi, il serait judicieux de les étudier afin de déterminer quelles valeurs celles-ci peuvent prendre et comment elles influencent la limite LP_{Δ}^{CFN} . Pour établir ceci, il est nécessaire d'étudier la fonction (5.16). Cette fonction nous permet de définir ou de généraliser différentes configurations permettant de discriminer entre les deux classes CFN et F .

Précédemment, nous avons vu que pour détecter des fautes sous réseaux Bayésiens, il fallait discriminer entre deux classes Gaussiennes décrivant l'état de fonctionnement du système : l'une décrivant son fonctionnement nominal et l'autre son état de fonctionnement anormal. Toutefois, il est généralement difficile ou impossible de donner une distribution permettant de décrire de façon globale les différentes fautes pouvant survenir dans le système. Dans le

chapitre précédent, nous avons choisi, au lieu de chercher à faire une approximation par une distribution Gaussienne donnée, de l'ajuster de sorte à ce que nous puissions prendre une décision. En effet, nous l'avons restreint à une distribution Gaussienne définie avec une moyenne semblable à celle de la classe de fonctionnement normal mais avec une plus grande variabilité. Cette différence de variabilité est due à un coefficient \mathbf{c} introduit venant augmenter la matrice de variance-covariance de la classe de fonctionnement normal. En fonction de l'écriture donnée dans (5.16) nous pouvons en déduire la limite de probabilité obtenue dans le chapitre précédent.

En effet, en considérant que la distribution de la classe F est déduite à partir des paramètres de la distribution de classe CFN estimée à partir des données, alors $p(x^*)$ dans (5.6) peut être calculée. Étant donné cela, (5.7) s'écrit comme ceci :

$$\begin{aligned} x \in C_k, \text{ si } p(\mathbf{D} = C_k|x) &\geq \frac{p(\mathbf{D} = C_k)p(x^*|\mathbf{D} = C_k)}{p(x^*)} \\ &\geq LP_{\Delta}^{C_k} \end{aligned} \quad (5.22)$$

où $C_k \in \{CFN, F\}$.

Ainsi, en faisant l'hypothèse que la matrice de variance-covariance de la classe fautive est fonction de celle de la classe de fonctionnement normal, nous pouvons écrire le terme quadratique de la distribution de la classe fautive en fonction de celui de la distribution de la classe de fonctionnement normal. Ainsi, de (5.16), nous pouvons écrire :

$$\begin{aligned} \varphi &= \Delta_F - \Delta_{CFN} = \nu_1 LC_{\Delta}^{CFN} - LC_{\Delta}^{CFN} \\ &= (\nu_1 - 1)LC_{\Delta}^{CFN} \end{aligned} \quad (5.23)$$

où $\Delta_{CFN} = LC_{\Delta}^{CFN}$ pour tout $\mathbf{x} = x^*$. Δ_F est alors une fonction linéaire en LC_{Δ}^{CFN} (une fois calculée, elle est fixée), ν_1 est un coefficient (un nombre réel) différenciant les deux classes de fonctionnement. Notons également que pour toute observation x^* on a :

$$\Delta_1 = \Delta_{CFN} - LC_{\Delta}^{CFN} = 0 \quad (5.24)$$

A partir de (5.23) et (5.24), nous obtenons l'écriture (5.25) donnée ci-

dessous de (5.16) :

$$\begin{aligned} LP_{\Delta}^{CFN} &= \frac{1}{1 + \beta e^{-\frac{1}{2}(\nu_1 - 1)LC_{\Delta}^{CFN}}} \\ &= \frac{1}{1 + \beta e^{-\gamma_1 \varphi_1}} \end{aligned} \quad (5.25)$$

avec $\tau = 1$ dans (5.18), et

$$\gamma_1 = \frac{1}{2}(\nu_1 - 1), \quad (5.26)$$

$$\varphi_1 = LC_{\Delta}^{CFN} \quad (5.27)$$

Il est évident qu'à partir de (5.25) nous pouvons en déduire la limite de probabilité ζ_{Δ}^{CFN} (voir (4.13)) en posant $\nu_1 = \frac{1}{c}$. En considérant cela, (5.16) correspond à une fonction logistique $f(\varphi_1)$ avec comme asymptotes $LP_{\Delta}^{CFN} = 0$, $LP_{\Delta}^{CFN} = 1$ et un centre de symétrie :

$$I = \left(-\frac{m\mathbf{c}}{1 - \mathbf{c}} \ln(\omega\mathbf{c}), \frac{1}{2} \right) \quad (5.28)$$

où m est la dimension de la variable \mathbf{x} .

Cette configuration peut poser des problèmes numériques liés à la statistique quadratique mise en place et sa limite de contrôle LC_{Δ}^{CFN} lui correspondant, nombre de variables, nombre d'échantillons disponibles de fonctionnement normal, niveau de signification α , et ainsi que le choix du \mathbf{c} .

Ce dernier paramètre, non lié au système, peut être varié à notre guise, et influe sur le calcul des probabilités a posteriori ainsi que sur la limite probabiliste calculée. Il peut alors être posé afin de réguler ces dernières. En effet, il doit être choisi de sorte à éviter que la limite et les probabilités à posteriori n'atteignent les asymptotes de la fonction générée, pouvant conduire à des problèmes numériques rendant la prise décision peu précise. En d'autres termes, la configuration de la classe de fautes doit tenir compte des différents paramètres pour éviter tout problème numérique. Par exemple, on peut poser la limite probabiliste LP_{Δ}^{CFN} et rechercher le \mathbf{c} lui correspondant.

Toutefois, d'autres hypothèses peuvent être émises afin de développer des limites nous permettant également de réaliser la détection. En effet, en déve-

loppant le terme φ dans (5.16), nous pouvons le réécrire comme ceci :

$$\varphi = \Delta_F - \Delta_{CFN} = f(x) - a \quad (5.29)$$

On peut voir qu'il correspond à une fonction quadratique en x , où :

$$f(x) = x^T \Sigma x + 2x^T h, \quad (5.30)$$

$$a = \mu_{CFN}^T \Sigma_{CFN} \mu_{CFN} - \mu_F^T \Sigma_F \mu_F \quad (5.31)$$

avec

$$\Sigma = \Sigma_F^{-1} - \Sigma_{CFN}^{-1}, \quad (5.32)$$

$$h = \Sigma_{CFN}^{-1} \mu_{CFN} - \Sigma_F^{-1} \mu_F \quad (5.33)$$

En fonction de (5.30), on peut écrire (5.16) comme ceci :

$$\begin{aligned} LP_{\Delta}^{CFN} &= \tau \cdot \frac{1}{1 + \frac{\omega}{\nu_2} \mathbf{e}^{-\frac{1}{2}(f(x)-a)}} \\ &= \tau \cdot \frac{1}{1 + \frac{\omega}{\nu_2} \mathbf{e}^{\frac{1}{2}a} \mathbf{e}^{-\frac{1}{2}f(x)}} \\ &= \tau \cdot \frac{1}{1 + \beta_2 \mathbf{e}^{-\frac{1}{2}\varphi_2}} \end{aligned} \quad (5.34)$$

avec

$$\varphi_2 = f(x), \quad (5.35)$$

$$\beta_2 = \mathbf{e}^{\ln \beta + \frac{1}{2}a} \quad (5.36)$$

Cette fonction logistique a pour asymptotes $LP_{\Delta}^{CFN} = 0$, $LP_{\Delta}^{CFN} = \tau$, et un centre de symétrie : le point d'inflexion avec comme coordonnées :

$$I = \left(\ln(\beta) + \frac{1}{2}a, \frac{\tau}{2} \right) \quad (5.37)$$

À partir de (5.34) plusieurs configurations possibles peuvent être envisagées pour définir les paramètres Σ_F et μ_F , et ainsi que les limites de probabilités nous permettant de faire de la détection. Notons qu'en fonction des paramètres choisis, les termes dans $f(x)$ non dépendants de x sont amenés vers a tout en faisant évoluer $f(x)$. Ceci conduit à différentes fonctions logistiques ayant les

mêmes asymptotes.

Par ailleurs, l'écriture (5.34) de LP_{Δ}^{CFN} peut nous permettre d'utiliser un autre type de réseau Bayésien pour la détection de fautes. Pour voir cela, nous posons $\Sigma_{CFN} = \Sigma_F$. Ainsi, la matrice Σ devient nulle et $f(x)$ linéaire en x . L'équation (5.34) peut alors s'écrire :

$$\begin{aligned} LP_{\Delta}^{CFN} &= \tau \cdot \frac{1}{1 + \frac{\omega}{\nu_2} e^{\frac{1}{2}a} e^{-x^T h}} \\ &= \tau \cdot \frac{1}{1 + \beta_2 e^{-\varphi_3}} \end{aligned} \quad (5.38)$$

où

$$\varphi_3 = x^T h = x^T \Sigma_{CFN}^{-1} (\mu_{CFN} - \mu_F) \quad (5.39)$$

Sous cette configuration, une prise de décision peut être établie en comparant la limite LP_{Δ}^{CFN} , étant donnée une observation x , à la probabilité a posteriori $p(\bar{\mathbf{D}} = CFN|x)$ où $\bar{\mathbf{D}}$ prend les mêmes valeurs, CFN , F , que la variable \mathbf{D} utilisée précédemment. $p(\bar{\mathbf{D}} = CFN|x)$ peut s'écrire comme ci-dessous :

$$\begin{aligned} p(\bar{\mathbf{D}} = CFN|x) &= \frac{p(\mathbf{D} = CFN)p(x|\mathbf{D} = CFN)}{p(\mathbf{D} = CFN)p(x|\mathbf{D} = CFN) + p(\mathbf{D} = F)p(x|\mathbf{D} = F)} \\ &= \frac{1}{1 + e^{-x^T \Sigma \mu + \ln(\beta_2)}} \end{aligned} \quad (5.40)$$

Cette écriture de $p(\bar{\mathbf{D}} = CFN|x)$ peut être vue comme une fonction softmax pouvant s'écrire comme ceci :

$$p(\bar{\mathbf{D}} = CFN|x) = \sigma_{CFN}(x) = \frac{e^{w_1 x^T + b_1}}{e^{w_1 x + b_1} + e^{w_2 x + b_2}} \quad (5.41)$$

où w_1 et b_1 sont nulles et

$$w_2 = -(\Sigma_{CFN}^{-1} (\mu_{CFN} - \mu_F))^T, \quad (5.42)$$

$$b_2 = \ln(\beta_2) \quad (5.43)$$

avec

$$p(\bar{\mathbf{D}} = F|x) = \sigma_F(x) = 1 - \sigma_{CFN}(x) \quad (5.44)$$

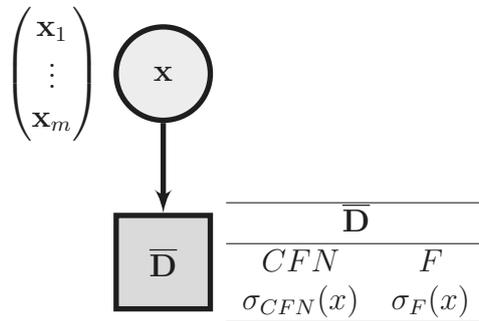


FIGURE 5.5 – Détection dans un réseau Bayésien ayant un noeud softmax

Ainsi, sur un réseau comme celui présenté dans la figure (5.5), une décision peut être prise sur la présence d'une faute, étant donnée une nouvelle observation, selon la règle suivante :

$$x \in C_k, \text{ si } p(\bar{\mathbf{D}} = C_k | x) \geq LP_{\Delta}^{C_k}, C_k \in CFN, F \quad (5.45)$$

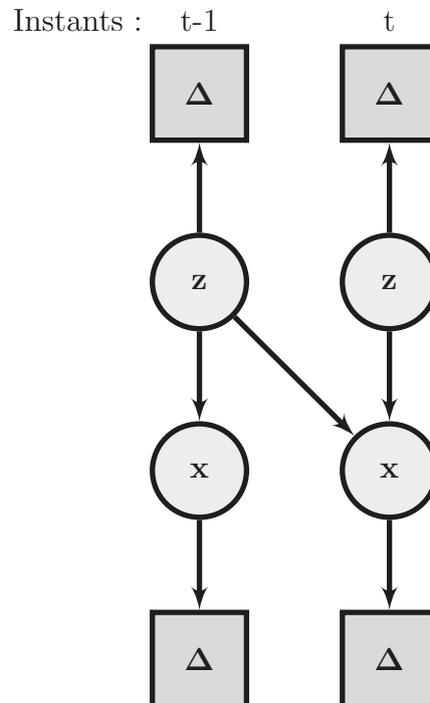


FIGURE 5.6 – Un réseau Bayésien combinant une carte MEWMA et d'autres statistiques quadratiques

Le réseau Bayésien proposé dans la figure (5.5) pour la détection de fautes est composé d'un noeud discret softmax $\overline{\mathbf{D}}$ et un noeud continu ayant la particularité d'être non caché (\mathbf{x} , toujours observé). Ce noeud discret softmax peut être une solution au problème présenté précédemment dans le cas de la carte MEWMA (voir section III.2). En effet, en remplaçant les noeuds de décision par ces noeuds dans le réseau Bayésien dynamique proposé précédemment, nous pourrions éviter donc d'avoir des mélanges de Gaussiennes et donc obtenir une inférence exacte. Aussi, dans ce sens, nous pourrions tenter également, par exemple, la combinaison sur un seul réseau d'une carte MEWMA et, par exemple, la statistique T^2 . Le réseau Bayésien dynamique correspondant à de tels combinaisons de statistiques quadratiques est donné dans la figure 5.6.

II.3 Détection et diagnostic simultanés de fautes

De façon générale, sous l'écriture (5.11), nous proposons une limite de probabilité nous permettant de faire de la détection et du diagnostic sous un réseau Bayésien (Atoui et al., 2015b) comme celui présenté dans la figure 5.4. Afin de montrer les avantages de la méthode proposée, nous l'illustrons dans un exemple de simulation d'un système bivariée, avec trois classes : une classe représentant l'état de fonctionnement normal CFN , et deux autres classes pour deux fautes systèmes : F_1 et F_2 . Les paramètres de chaque classe obtenus après apprentissage sont définis de la sorte :

$$\begin{aligned}\mu_{CFN} &= \begin{pmatrix} 12 \\ 5 \end{pmatrix}, & \Sigma_{CFN} &= \begin{pmatrix} 1 & 0.2 \\ 0.2 & 2 \end{pmatrix}, \\ \mu_{F_1} &= \begin{pmatrix} 5 \\ 5 \end{pmatrix}, & \Sigma_{F_1} &= \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}, \\ \mu_{F_2} &= \begin{pmatrix} 19 \\ 5 \end{pmatrix}, & \Sigma_{F_2} &= \begin{pmatrix} 2 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}\end{aligned}$$

Dans cet exemple, nous considérons la statistique quadratique de Hotelling T^2 . La limite statistique lui correspondant est égale à χ_α^2 lorsque un nombre suffisant de données est disponible. A partir des données CFN , et un niveau de signification ($\alpha = 1\%$) choisi, $LP_{T^2}^{CFN}$ peut être déduite.

Ainsi, pour une observation x de \mathbf{x} , nous proposons de faire de la détection et du diagnostic sous un seul réseau Bayésien suivant cette règle :

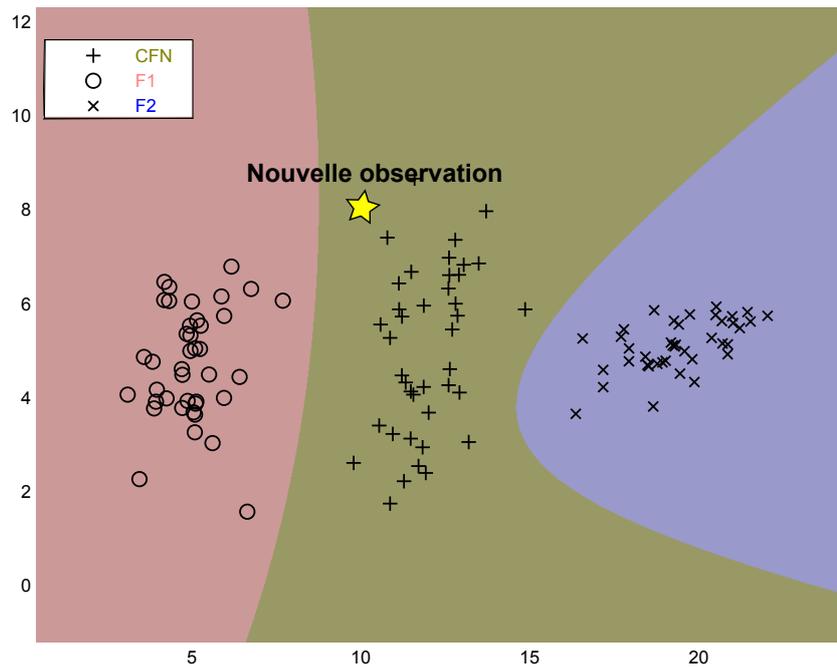


FIGURE 5.7 – Détection et diagnostic classique

si $p(\mathbf{D} = CFN|x) \geq LP_{\Delta}^{CFN}$ alors x appartient à la classe CFN , sinon parmi les autres classes, la classe avec la probabilité a posteriori maximale est prise en considération.

La figure 5.7 représente la décision du classifieur n'intégrant pas la limite LP_{Δ}^{CFN} dans la classe CFN . La figure 5.8 illustre les limites de décisions contournant chaque classe définie par le classifieur précédent augmenté par la règle proposée. Dans cette dernière figure 5.8, nous pouvons observer la décision du classifieur proposé avec les limites probabilistes. Nous remarquons clairement que la classe CFN est maintenant représentée par une ellipse. La taille de cette ellipse dépend de la valeur de α . Ainsi, nous pouvons voir que la limite LP_{Δ}^{CFN} permet de respecter un taux de fausse alarme fixé, et au classifieur de faire de la détection et du diagnostic.

Afin de tester les performances de la méthode en termes de diagnostic, nous proposons de la tester sur un système complexe : le Tennessee Eastman Process (TEP). Nous avons pris en compte 3 types de fautes : les fautes 2, 10 et 14 (voir la figure 5.9). Évidemment, la classe de fonctionnement normal (CFN) est aussi prise en compte. Comme indiqué dans la figure 5.9, chaque classe est associée à deux ensembles de données : un ensemble d'apprentissage et un

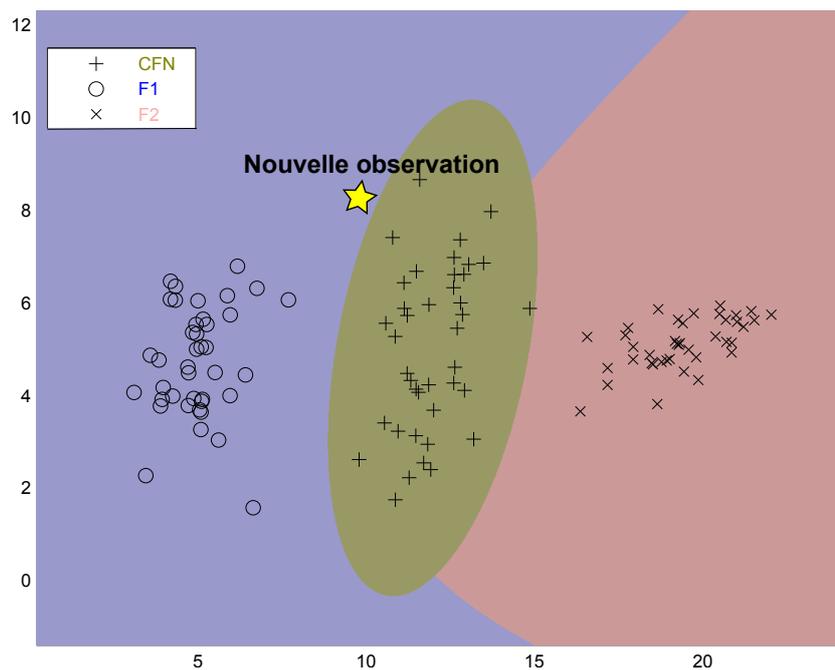


FIGURE 5.8 – Détection et diagnostic de fautes intégrant la limite probabiliste de CFN

ensemble de test, comportant chacun respectivement 480 et 800 observations.

Classe	Type de faute	Données d'apprentissage	Données de test
CFN	classe de fonctionnement normal	480	800
F_2	saut dans la composition en B	480	800
F_{10}	variation aléatoire dans la température d'alimentation en C	480	800
F_{14}	blocage de la valve du liquide de refroidissement du réacteur	480	800

FIGURE 5.9 – Description de l'ensemble de données

Premièrement, nous avons appliqué le réseau Bayésien classifieur classique

sans les limites probabilistes. Dans ce cas, les résultats de classification des 3200 observations de test sont présentés dans la matrice de confusion de la figure 5.10.

La figure 5.10 nous donne la possibilité de voir comment la discrimination entre les différentes fautes est réalisée. Chaque colonne de la matrice représente les observations dans la classe prédite, alors que chaque ligne représente les observations dans la classe actuelle. Par exemple, pour 800 observations de test de la faute 2, la procédure de diagnostic donne 7 observations comme appartenant à la classe *CFN*, et 3 observations comme appartenant à la faute 10, ainsi 1.25% (10/800) d'observations mal classées en faute 2. Concernant les quatre classes, le taux de mauvaise classification est de 9.28% ($\frac{188+71+7+3+18+10}{3200}$).

Classe	<i>CFN</i>	F_2	F_{10}	F_{14}	Total
<i>CFN</i>	541	0	188	71	800
F_2	7	790	3	0	800
F_{10}	18	0	772	10	800
F_{14}	0	0	0	800	800
Total	566	790	963	881	3200

FIGURE 5.10 – Matrice de confusion du réseau Bayésien classifieur

Dans ce cas, le taux de fausse alarme peut ne pas être pris en compte. En effet, on peut voir que 259 (188 + 71) observations de *CFN* ont été classées comme des observations en faute par le réseau Bayésien, et cela avec un taux de fausse alarme de 32.37%. Un tel taux de fausse alarme n'est pas acceptable en pratique, car il mène à stopper le système chaque 3 observations (en moyenne).

Dans un second temps, nous associons notre limite probabiliste développée à la probabilité a posteriori de la classe de fonctionnement normal. Le taux de fausse alarme théorique a été fixé à $\alpha = 1\%$. La figure 5.11 représente la matrice de confusion de ce cas. Précédemment, nous avons montré que la limite probabiliste associée à la classe *CFN* permet d'augmenter la précision de la classification. Maintenant, nous allons étudier si la même conclusion peut être obtenue pour d'autres combinaisons des trois fautes. Afin de comparer correctement les deux approches, nous allons calculer le taux de d'erreurs de classification pour la *CFN* additionnée à chaque combinaison possible de 3 fautes parmi les 15 fautes connues du TEP, comme présenté dans la figure 5.12.

Nous obtenons 455 erreurs pour le réseau Bayésien classifieur, et 455 pour

Deux résultats significatifs nécessitent d'être remarqués. Premièrement, le taux de mauvaise classification a diminué : 3.63% (au lieu de 9.28% précédemment). Ainsi, globalement, les décisions données par le classifieur ont été améliorées. Deuxièmement, dans ce cas, nous remarquons clairement que la classe *CFN* a été contrôlée. Nous avons réduits considérablement le taux de fausses alarmes de 32.37 % à 6.38%. En effet, avec notre méthode seulement 51 observations (au lieu de 259 observations dans le cas précédent) ont été classées comme étant des observations en fautes (42 comme faute 10 et 9 comme faute 14).

Classe	<i>CFN</i>	F_2	F_{10}	F_{14}	Total
<i>CFN</i>	749	0	42	9	800
F_2	9	790	1	0	800
F_{10}	75	0	724	1	800
F_{14}	0	0	0	800	800
Total	833	790	767	810	3200

FIGURE 5.11 – Matrice de confusion de réseau Bayésien classifieur avec une limite probabiliste

Error	Classe 1	Classe 2	Classe 3	Classe 4	RB	RB avec <i>LP</i>
1	<i>CFN</i>	F_1	F_2	F_3	0,2119	0,2469
2	<i>CFN</i>	F_1	F_2	F_4	0,0038	0,0194
⋮	⋮	⋮	⋮	⋮	⋮	⋮
454	<i>CFN</i>	F_{12}	F_{13}	F_{15}	0,2969	0,2778
453	<i>CFN</i>	F_{12}	F_{14}	F_{15}	0,2722	0,2406
455	<i>CFN</i>	F_{13}	F_{14}	F_{15}	0,2794	0,2478

FIGURE 5.12 – Erreurs de classification pour les différentes combinaisons du *CFN* + 3 fautes

le réseau Bayésien classifieur intégrant la limite probabiliste à la classe *CFN*. Les résultats sont donnés dans la figure 5.13.

La figure 5.13 donne la moyenne et l'écart-type de l'erreur des deux classifieurs, un réseau Bayésien classifieur classique et un réseau Bayésien classique avec une limite probabiliste (*LP*), calculées pour 455 combinaisons de 3 fautes. Nous pouvons voir que l'erreur est moindre avec le réseau Bayésien intégrant la limite probabiliste à la classe *CFN*.

Classifieur	Moyenne des erreurs	Écart-type des erreurs
RB	0.1841	0.1464
RB avec <i>LP</i>	0.1782	0.1452

FIGURE 5.13 – Moyennes et écart-types des erreurs du classifieur

Afin de confirmer ce résultat, nous appliquons un test d'hypothèse comparant les deux populations d'erreurs (comparaison des deux moyennes). Nous avons utilisé le test de Wilcoxon à rang signé. Pour ce test, étant donné l'hypothèse nulle que les moyennes des 2 populations sont considérées comme égales, nous avons trouvé une p-value de $7.23E^{-09}$. Nous concluons que la différence de performance entre les deux méthodes est significative. En moyenne, la différence est de 0.59%. Ainsi, nous pouvons dire que pour l'application TEP, intégrer la limite probabiliste permet de prendre de meilleurs décisions.

Ainsi, nous avons présenté le calcul d'une limite probabiliste (pour garantir un certain taux de fausses alarmes) qui peut être utilisé dans un réseau Bayésien classifieur intégrant les différentes fautes du système, mais aussi la classe *CFN*. L'intérêt de la limite probabiliste est de rendre le réseau Bayésien classifieur capable de gérer un test basé sur une statistique Δ . Les résultats obtenus sur le TEP montrent toute l'importance et la plus-value d'utiliser l'approche proposée.

II.4 Diagnostic de fautes augmenté par un rejet de distance

Nous avons vu qu'avec un réseau Bayésien classifieur, nous sommes capables de faire de la détection ou du diagnostic de fautes ou les deux simultanément. Cependant, en pratique, les différentes classes de fautes peuvent ne pas décrire tous les états possibles de dysfonctionnement qu'un système donné peut avoir. De plus, dans certain cas, il n'est pas possible d'obtenir des données sur un état de fonctionnement car il est rare ou bien cet état ne s'est jamais produit auparavant.

De même, il n'est pas toujours possible d'identifier le nombre exact de défaillances possibles pouvant altérer l'état de fonctionnement du système à un moment donné. Ainsi, selon ([Dubuisson, 2001](#)) cela justifie le fait de devoir décider, dans le contexte de la reconnaissance de formes, sans avoir une vue exhaustive du nombre d'états. Ainsi, nous pouvons prendre une décision et dire que l'origine ou la cause d'une nouvelle observation est due à une classe de faute alors qu'il se peut que cette observation ne lui appartienne pas.

Il serait intéressant dans de tels classifieurs Bayésiens de pouvoir décider lorsqu'une observation (pouvant être aberrante) n'appartient pas vraiment à la classe ayant la plus grande probabilité a posteriori. Pour remédier à cela,

nous augmentons le classifieur par une nouvelle classe de fautes représentant les différentes fautes inconnues ou non classées (nouveau type de fautes survenu). Nous nommerons une telle classe : classe de Fautes Non Connues *FNC*.

Ainsi sur un réseau Bayésien dédié au diagnostic, comme celui présenté dans la figure 5.4, nous proposons une nouvelle règle de décision afin de tenir compte de cela. En d'autres termes, nous voudrions diviser l'espace de décision de sorte à avoir un sous espace supplémentaire à ceux consacrés aux différentes classes définies dans le réseau Bayésien classifieur. Ce nouvel espace est consacré à la classe *FNC*. Ainsi, tout ce qui n'appartient pas aux différentes classes définies appartient à celui-ci. Une observation donnée appartiendra à la classe *FNC* si elle est éloignée ou n'appartient pas aux zones de décisions dédiées aux autres classes obtenues lors de l'apprentissage. Pour avoir une idée encore plus précise du rejet de distance, considérons une classe de faute définie étant donné un ensemble de données d'apprentissage lui correspondant. En construisant une carte de contrôle à partir de cet ensemble, nous décidons comme ceci pour une observation donnée : si la statistique ou l'indice est inférieur à la limite de contrôle, alors l'observation appartient à la faute sinon elle appartient à l'hypothèse alternative car elle est éloignée statistiquement de la zone définie pour la classe de faute. La dernière décision menant à l'hypothèse alternative peut être vue comme un rejet de distance.

Rappelons que sous ce type de réseau Bayésien, nous pouvons modéliser l'analyse discriminante comme on peut définir des cartes de contrôle (test statistique) sur chaque classe. Ainsi, nous disposons des probabilités a posteriori de chaque classe après classification ainsi que deux règles de décision permettant de différencier la classification et le test statistique. En fonction de cela, nous allons voir comment tenir compte du rejet de distance sur ce type de réseau.

Dans un premier temps, pour une nouvelle observation, nous calculons la probabilité a posteriori correspondant à chaque classe. Ensuite, nous calculons pour chaque classe la limite probabiliste lui correspondant, ainsi qu'au test statistique établi avec un niveau de signification α . Nous décidons, étant donnée une nouvelle observation, suivant les règles mentionnées ci-dessous :

- détection : si la probabilité a posteriori de la classe *CFN* respecte la limite lui correspondant alors l'observation décrit un fonctionnement normal.
- diagnostic : sinon,

- si l'observation respecte la limite associée à la classe de fautes ayant la plus grande probabilité a posteriori alors l'observation décrit un fonctionnement anormal et appartient à cette classe.
- sinon, l'observation est considérée comme faisant partie d'une nouvelle classe non définie CND.

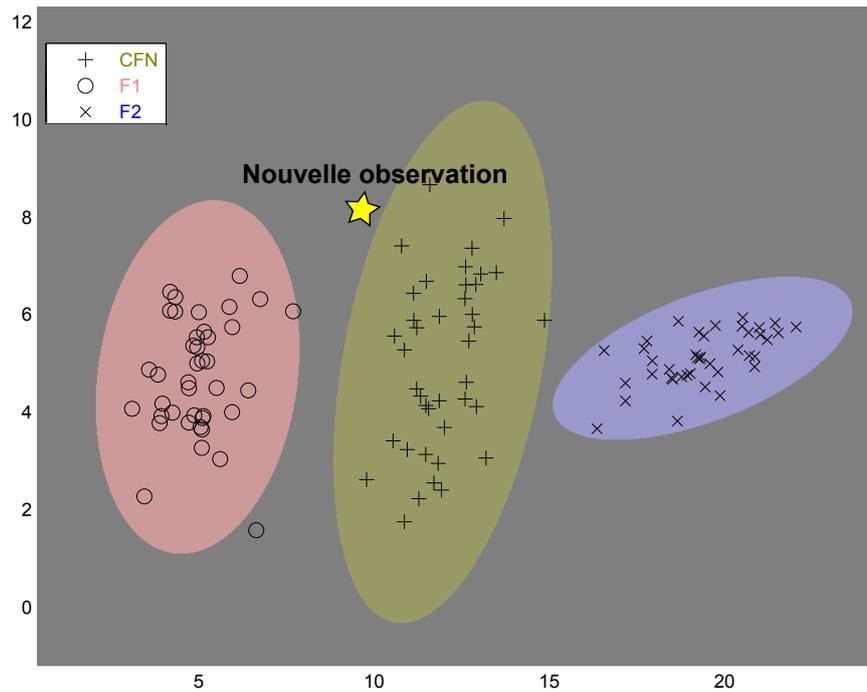


FIGURE 5.14 – Détection et diagnostic de fautes augmentés par le rejet de distance

Ainsi, en intégrant le rejet de distance, nous pouvons constater le changement des limites de décisions discriminantes entre les différentes classes. Chaque classe est isolée statistiquement des autres classes étant donné son ensemble d'apprentissage. Sous l'hypothèse de normalité les observations de chaque classe appartiennent à une ellipsoïde, voir la Figure 5.14.

III Propositions de structures de réseau Bayésien

Dans cette section nous allons présenter une famille de réseaux Bayésiens pouvant être utilisés pour la détection et le diagnostic de fautes. Ensuite, nous

allons développer un cadre probabiliste permettant de les associer.

III.1 Modèles Gaussiens pour le diagnostic de fautes

Certains réseaux Bayésiens peuvent être utilisés pour la classification mais aussi pour la détection et le diagnostic de fautes (avec ou sans rejet de distance). Ces réseaux Bayésiens sont vus comme une représentation graphique de méthodes de classification à base de modèle probabiliste.

Parmi ces réseaux, l'un des plus intuitif est le réseau Bayésien naïf semi condensé (voir figure 5.4). Ce réseau donne la probabilité a posteriori de l'appartenance d'une nouvelle observation de la variable examinée à une des classes connues considérées. Rappelons que ce réseau Bayésien correspond à l'analyse discriminante quadratique et peut être utilisé pour la détection et le diagnostic de fautes dans les systèmes.

Cependant, ce genre de réseau peut poser des problèmes par exemple lors de la présence d'un ensemble de données de grande dimension, d'un nombre limité de données par classe, etc. En effet, cela est dû en grande partie aux problèmes liées à l'inversion des matrices de variance-covariance pouvant être singulières ou mal estimés. Pour éviter ces problèmes un bon nombre de méthodes ont été adoptées. La plupart portent sur la modification des matrices de variance-covariance (e.g. réécriture, pénalisation, l'utilisation de la pseudo inverse de Moore-Penrose, etc.). Ces méthodes sont dites techniques de régularisation (Bouveyron, 2013; Thomaz et al., 2004; Mkhadri et al., 1997). Bien évidemment, le résultat de l'application de ces méthodes peut être représenté sous un réseau Bayésien comme celui présenté dans 5.4. Ce réseau Bayésien peut être couteux en termes de nombre de paramètres à estimer. Cela conduit parfois à son sur-apprentissage. Pour remédier à cela, on peut s'appuyer sur les différentes approches alternatives proposées dans le cadre de l'analyse discriminante quadratique. Comparativement à celle-ci, ces approches sont bien évidemment moins coûteuses en terme de nombre de paramètres à estimer mais elles sont, dans la plupart des cas, moins performantes. Toutefois, généralement, elles proposent un compromis entre coût et efficacité. Ces approches émettent des hypothèses (des contraintes, par exemple l'égalité des matrices de variance-covariance (voir e.g. l'équation (5.46)), donnant lieu à une Analyse Discriminante Linéaire, afin d'estimer le moins de paramètres possibles) sur

les paramètres des classes (Bouveyron et al., 2007).

$$\Sigma = \frac{(N_1 - 1)\Sigma_1 + \dots + (N_K - 1)\Sigma_K}{N - K} \quad (5.46)$$

où N_1, N_K sont respectivement le nombre d'échantillons dans la classe C_1 et C_K , avec $N = N_1 + \dots + N_K$.

Ces solutions peuvent être réécrites et généralisées à partir de la décomposition en valeurs propres des matrices de variance-covariance (Celeux and Govaert, 1995), correspondant à chaque classe C_k d'un réseau Bayésien naïf semi condensé, donnée ci-dessous :

$$\Sigma_{C_k} = \lambda_{C_k} D_{C_k} A_{C_k} D_{C_k}^T \quad (5.47)$$

où D_{C_k} correspond à la matrice des vecteurs propres, A_{C_k} est une matrice diagonale proportionnelle aux valeurs propres et λ_{C_k} correspond à un scalaire.

Dans le même contexte que ces approches, nous pouvons utiliser les réseaux Bayésiens présentés dans le chapitre 2, où on vient intervenir sur la structure du réseau en terme de relations d'indépendances conditionnelles (un réseaux Bayésiens naïf ou d'autres augmentés (Nielsen and Jensen, 2009)) pour réduire le nombre de paramètres à estimer.

Les réseaux Bayésiens présentés jusqu'à présent correspondent à des différentes déclinaisons du réseau fourni dans la figure 5.4 que ce soit en termes de paramètres ou de structures. Tous traitent directement les observations sans réduire leurs dimensions ou les transformer, c'est à dire sans avoir recours à une autre représentation des données observées. Toutefois, ces réseaux peuvent être utilisés pour la détection et le diagnostic de fautes (avec ou sans rejet de distance). Pour cela, il suffit de comparer les probabilités a posteriori des différentes classes, entre elles et/ou avec les limites probabilistes leurs correspondants déduites de (5.11).

D'autres solutions peuvent être utilisées afin de réduire le nombre de paramètres à estimer. Ces solutions correspondent à des méthodes de classification augmentées par d'autres méthodes de réduction de dimensions. Ces dernières essayent de trouver un espace de dimension réduit, conservant les informations capitales de l'espace original, sur lequel la discrimination est effectuée. Certaines de ces solutions peuvent être traduites et exprimées sous le modèle linéaire, générant des observations indépendantes et identiquement distribuées,

donné ci-dessous :

$$\mathbf{x}_{C_k} = A_{C_k} \mathbf{z} + \mu_{C_k} + \epsilon, \quad (5.48)$$

$$\epsilon \sim \mathcal{N}(0; \Psi_{C_k}), \quad (5.49)$$

$$\mathbf{z} \sim \mathcal{N}(0; I), \quad (5.50)$$

$$\mathbf{x}|\mathbf{z}, \mathbf{D} = C_k \sim \mathcal{N}(A_{C_k} \mathbf{z} + \mu_{C_k}; \Psi_{C_k}) \quad (5.51)$$

où A_{C_k} est la matrice de projection correspondant à chaque classe C_k et \mathbf{z} est une variable cachée représentant les projections de \mathbf{x} dans un nouvel espace de dimension a . ϵ est un bruit Gaussien centré avec une matrice de variance-covariance diagonale Ψ_{C_k} associée à chaque classe. La densité marginale de \mathbf{x} peut être donnée et déduite à partir de l'équation suivante :

$$p(\mathbf{x}) = \int \sum_{k=1}^K p(C_k) p(\mathbf{z}|\mathbf{D} = C_k) p(\mathbf{x}|\mathbf{z}, \mathbf{D} = C_k) d\mathbf{z} \quad (5.52)$$

En posant A_{C_k} comme l'ensemble des $a, a < m$, vecteurs propres associés à chaque classe C_k , et $\Psi_{C_k} = \Psi$, où Ψ est une matrice diagonale commune à toutes les classes, le modèle donné dans (5.48) correspond à une mixture d'analyse factorielle (Ghahramani et al., 1996). Cette mixture attribue un modèle d'analyse factorielle à chaque classe C_k . Une mixture d'analyse en composantes principales peut être obtenue en considérant $\Psi = \mathbf{v}I$, avec $\mathbf{v} \approx 0$. D'autres mixtures, mélange de modèles de projection linéaire, peuvent être également déclinés en redéfinissant les matrices A_{C_k} et Ψ_{C_k} ou en contraignant et paramétrant leurs structures (Bouveyron, 2013; Montanari and Viroli, 2010; Baek et al., 2010; Mcnicholas and Murphy, 2008; Bouveyron et al., 2007; Yoshida et al., 2006, 2004).

Le modèle fournie dans (5.48) correspond au réseau Bayésien présenté dans la figure (5.15). Ce réseau peut être utilisé pour la détection et le diagnostic simultanés de fautes (avec ou sans rejet de distance). Pour cela, il faut déterminer la limite probabiliste $LP_{\Delta}^{C_k}$ correspondant à la statistique quadratique Δ surveillant \mathbf{x} sachant la classe C_k . Cette dernière peut être développée et déduite à partir de l'écriture suivante de la même manière que l'équation (5.11) :

$$LP_{\Delta}^{CFN} = \frac{p(\mathbf{D} = CFN) p(\mathbf{z}) p(x^*|\mathbf{z}, \mathbf{D} = CFN)}{p(x)} \quad (5.53)$$

Ainsi, pour déterminer l'appartenance statistique d'une nouvelle observation x à une classe C_k on procède comme ceci :

$$\text{si } p(\mathbf{D} = C_k|x) \geq LP_{\Delta}^{C_k} \text{ Alors : } x \in C_k \text{ sinon } x \notin C_k \quad (5.54)$$

où $p(\mathbf{D} = C_k|x) \propto p(\mathbf{D} = C_k) \int p(\mathbf{z})(p(x|\mathbf{z}, \mathbf{D} = C_k))d\mathbf{z}$.

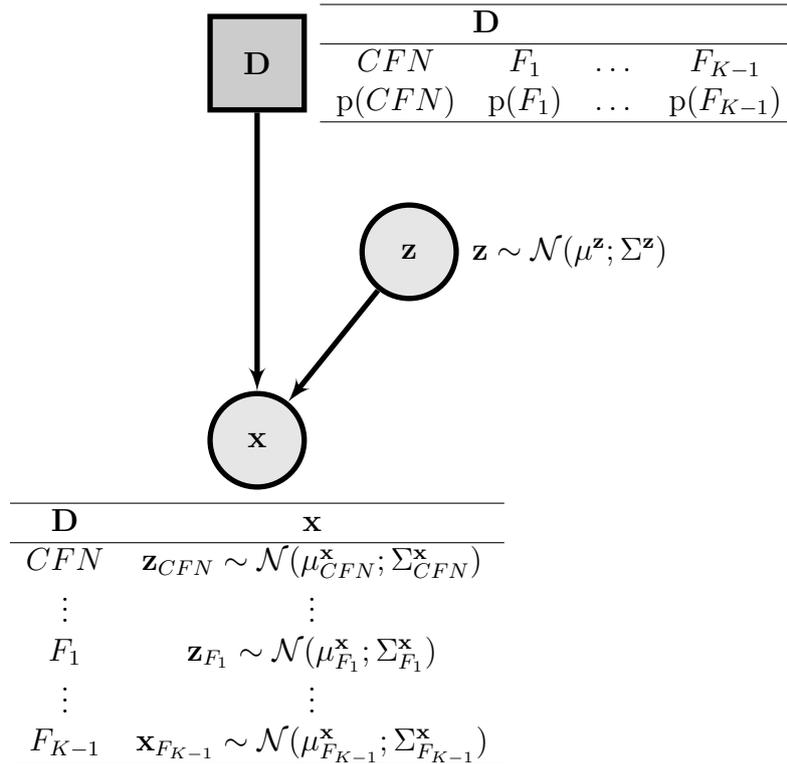


FIGURE 5.15 – Un réseau Bayésien classifieur considérant le noeud observé \mathbf{x} comme étant fonction d'une variable Gaussienne cachée \mathbf{z} , et un bruit Gaussien ϵ pour chaque classe C_k

La classification supervisée avec ce type de réseau et de modèle augmenté par une méthode de réduction de données peut ne pas être efficace. En effet, il se peut que les informations conservées (le nouvel espace de discrimination) ne soit pas toujours convenables, cohérentes et optimales pour la méthode de classification utilisée (Bouveyron, 2013). En effet, réduire la dimension de l'espace des observations peut parfois engendrer une perte d'informations pouvant être utile pour la classification et donc pour le diagnostic.

Une alternative aux méthodes de réduction de données est l'utilisation de méthodes de sélection de variables (Verron et al., 2008a). Ces méthodes ré-

duisent la dimension de l'espace original en maintenant les variables les plus discriminatoires sachant le classifieur utilisé (Dash and Liu, 1997; Kohavi and John, 1997).

D'autres substitutions aux méthodes de réduction de données peuvent être également utilisées. Parmi ces méthodes, les méthodes dites de sous espace pour la classification. Ces méthodes définissent un sous espace de discrimination de petite dimension comparativement à l'espace original sans pour autant réduire la dimension de départ. En d'autres termes, ils utilisent le sous espace et l'espace original pour faire de la classification. Parmi ces méthodes, l'analyse discriminante de Fisher (Duda et al., 2001). Dans (Bouveyron and Brunet, 2012), les auteurs proposent une représentation probabiliste de cette dernière. Le modèle proposé, contrairement aux modèles vues précédemment, représente l'espace original de chaque classe comme une fonction linéaire d'un espace caché commun à toutes les classes. Cette espace de dimension d est assumé comme étant le plus discriminatoire contrairement, par exemple, à celui de l'analyse en composantes principales maximisant la variance projetée. Le modèle lui correspondant peut être défini comme ceci, lorsque $K < m$:

$$\mathbf{x}_{C_k} = U\mathbf{z}_{C_k} + \epsilon \quad (5.55)$$

où $\mathbf{z} \in E$, $E \subset R^m$, E est un sous-espace considéré comme le plus discriminant de dimension $d \leq K - 1$ et U est une matrice orthogonale commune aux différentes classes considérées. ϵ suit la distribution normale $\mathcal{N}(0; \Gamma_{C_k})$ et modélise l'information non discriminante contrairement à \mathbf{z} qui suit lui aussi une distribution Gaussienne pour chaque valeur de ses parents :

$$\mathbf{z}|\mathbf{D} = C_k \sim \mathcal{N}(\mu_{C_k}^{\mathbf{z}}, \Sigma_{C_k}^{\mathbf{z}}) \quad (5.56)$$

Ainsi, la distribution marginale de \mathbf{x} , étant donnée la classe C_k , $p(\mathbf{x}|\mathbf{D} = C_k)$, correspond à une Gaussienne avec comme moyenne :

$$m_{C_k} = U\mu_{C_k}^{\mathbf{z}} \quad (5.57)$$

et matrice de variance-covariance :

$$S_{C_k} = U\Sigma_{C_k}^{\mathbf{z}}U^T + \Gamma_{C_k} \quad (5.58)$$

où

$$\mathbf{x}|\mathbf{z}, \mathbf{D} = C_k \sim \mathcal{N}(\mu_{C_k}^{\mathbf{x}}, \Sigma_{C_k}^{\mathbf{x}}), \quad (5.59)$$

$$\mu_{C_k}^{\mathbf{x}} = U\mathbf{z}, \quad (5.60)$$

$$\Sigma_{C_k}^{\mathbf{x}} = \Gamma_{C_k} \quad (5.61)$$

Ce modèle proposé par les auteurs est défini de sorte à ce que $U\Gamma_{C_k}U^T = 0$ et $V\Gamma_{C_k}V^T = \beta_{C_k}I$ où V est le complément orthogonal de U et $W = [U, V]$. Ceci donne lieu à $\Delta_{C_k} = W^T S_{C_k} W$, un bloc diagonal pouvant être écrit comme ceci :

$$\delta_{C_k} = \text{diag}(\Sigma_{C_k}^{\mathbf{z}}, \beta_{C_k}I) \quad (5.62)$$

où I est de dimension $R^{m-d \times m-d}$. Ceci signifie que les sous-espaces, discriminant et non-discriminant, sont orthogonaux.

A partir de ce modèle, plusieurs autres modèles peuvent être déduits en relaxant ou ajoutant des contraintes à ces paramètres (e.g. relaxer β_{C_k} , contraindre Σ_{C_k} , etc.). Ce modèle peut être représenté sous un réseau Bayésien comme celui présenté dans la figure 5.16.

Ce réseau Bayésien peut être à son tour être utilisé pour la détection et le diagnostic simultanés de fautes (avec ou sans rejet de distance). De la même manière que les réseaux précédents, pour décider statistiquement sur l'appartenance d'une nouvelle observation \mathbf{x} à une classe C_k il est nécessaire de définir la limite probabiliste lui correspondant comme ceci :

$$LP_{\Delta}^{C_k} = \frac{p(\mathbf{D} = C_k)p(\mathbf{z}|\mathbf{D} = C_k)p(x^*|\mathbf{z}, \mathbf{D} = C_k)}{p(x)} \quad (5.63)$$

Cette limite correspond à une statistique quadratique Δ surveillant la variable \mathbf{x} . A partir de LP_{Δ}^{CFN} une décision est prise suivant la règle ci-dessous :

$$\text{si } p(\mathbf{D} = C_k|x) \geq LP_{\Delta}^{C_k} \text{ Alors : } x \in C_k \text{ sinon } x \notin C_k \quad (5.64)$$

où $p(\mathbf{D} = C_k|x) \propto p(\mathbf{D} = C_k)p(x|\mathbf{D} = C_k)$.

Le réseau donné dans 5.16 permet de généraliser d'autres méthodes de projection/transformation linéaire de données en modifiant les matrices composant le réseau ou contraindre leurs structures. Par exemple, le modèle d'analyse en

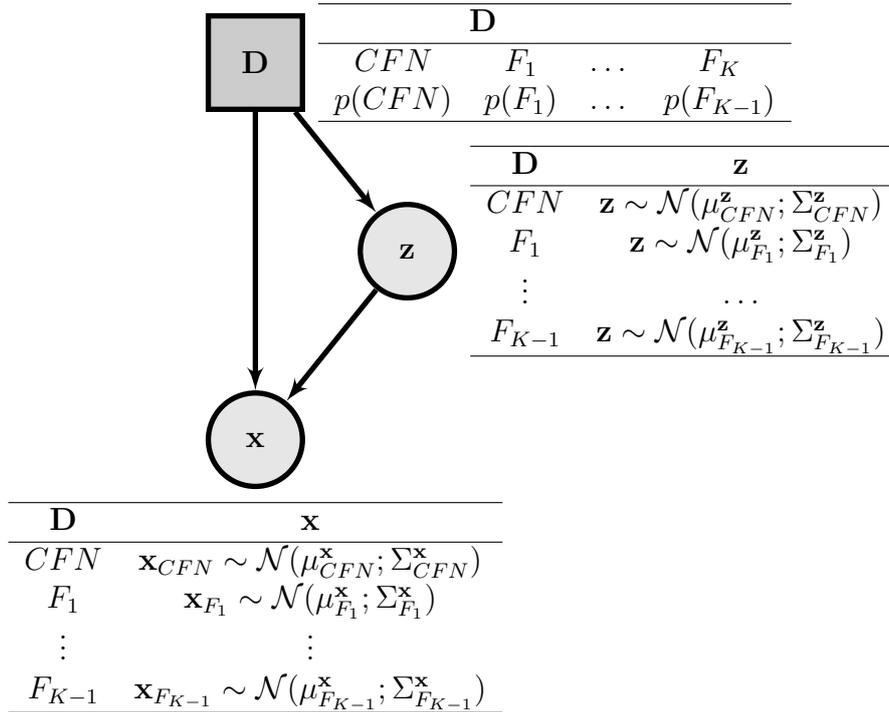


FIGURE 5.16 – Un réseau Bayésien classifieur traitant le noeud observé \mathbf{x} et le noeud \mathbf{z}

composantes principales utilisé dans (Chiang et al., 2001) pour le diagnostic de fautes (PCA1) peut être modélisé par ce réseau en posant $U = A$, où A est la matrice de vecteurs propres estimés à partir de toutes les données de fautes, $\Gamma = \mathbf{v}I$ et $\Sigma_{C_k}^z = A_{C_k}^T \Sigma_{C_k} A_{C_k}$.

Sous ce réseau, des méthodes à base de modèle dédiés à la surveillance des systèmes peuvent être utilisée. Prenons, le cas des résidus structurés. Ces résidus se basent une une matrice d’incidence. Cette matrice peut être représentée par un réseau Bayésien (Atoui et al., 2015a; Verron et al., 2010a; Pernestål, 2009; Weber et al., 2008; Schwall and Gerdes, 2002). Le réseau Bayésien lui correspondant exprime les relations de cause à effet entre fautes et résidus. Nous proposons de l’étendre afin de tenir compte de l’étape de génération des résidus. Ceci correspond à regrouper les résidus ainsi que les fautes dans de noeuds joints leurs correspondant. Le réseau présenté dans la figure 5.16 permet de modéliser entièrement une méthode à base de modèle (génération des résidus + évaluation + isolation).

D’autre stratégies à base de modèle pour l’isolation de fautes se basent sur l’utilisation d’un banc de générateurs de résidus (Isermann, 2006). Ainsi,

nous pouvons faire de même avec un banc constitué de réseaux Bayésiens modélisant chacun un générateur de résidu(s) (des générateurs de même ou de nature différente) sensible, selon la stratégie employée, à une faute ou un ensemble de fautes. Des bancs de réseaux Bayésiens représentant chacun une méthode à base de données dédiée à la détection d'une faute donnée, peuvent être également utilisés pour le diagnostic de fautes.

III.2 Combinaison de méthodes pour le diagnostic de fautes sous réseau Bayésien

Les différentes méthodes, que ce soit dans le cadre de la détection ou du diagnostic, modélisées sous des réseaux Bayésiens peuvent être facilement combinées. Dans ce but, nous proposons de les lier à un noeud discret de décision. En d'autres termes, en se basant sur la règle de Bayes on vient fusionner les différentes probabilités obtenues par chaque méthode ou leurs discrétisation (e.g. en les comparant à des seuils ou limites de contrôles) Nous pouvons constater la facilité de la manoeuvre étant donné le fait que les méthodes associées sont de même nature. De plus, un réseau Bayésien est un outil naturel de fusion.

Le cadre proposé peut être vu comme un réseau Bayésien hiérarchique à deux couches. Chaque couche est caractérisée par des entrées et sorties. Les sorties de la première couche sont les entrées de la seconde. Une couche correspond à un ou plusieurs réseaux Bayésiens dans lequel on fera l'inférence. En d'autres termes, dans notre cadre nous sommes amenés à faire de l'inférence en deux temps :

- première inférence : entrées (observations), sorties (probabilités a posteriori de chaque classe données par chaque méthode ou leurs transformées).
- seconde inférence : entrées (probabilités a posteriori de chaque classe ou leurs transformées (la comparaison des probabilités a posteriori à des seuils), sorties (probabilités a posteriori de chaque classe).

Il est évident que la première inférence est établie dans chaque réseau Bayésien correspondant à une méthode de diagnostic donnée. La seconde inférence peut être faite dans un réseau Bayésien naïf où les noeuds discrets observés sont supposés être indépendants étant donné le noeud de décision (le noeud racine discret). Nous proposons ainsi de fusionner les décisions d'un nombre \mathbf{b} de méthodes sous un réseau Bayésien classifieur composé uniquement de noeuds

discrets de même dimension. La table de probabilité conditionnelle d'un noeud enfant du noeud racine peut être simplement, par exemple, configuré suivant la règle suivante :

- si la valeur du noeud racine correspond à celle du noeud enfant, alors la probabilité conditionnelle du noeud enfant sachant le noeud racine est 1
- sinon 0

L'hypothèse d'indépendance entre les méthodes combinées peut être relaxée si certaines de ces méthodes sont identifiées ou assumées comme étant non conditionnellement indépendantes. Dans la figure 5.17 nous donnons un schéma illustratif du cadre proposé. Toutefois, il faut noter que fusionner plusieurs

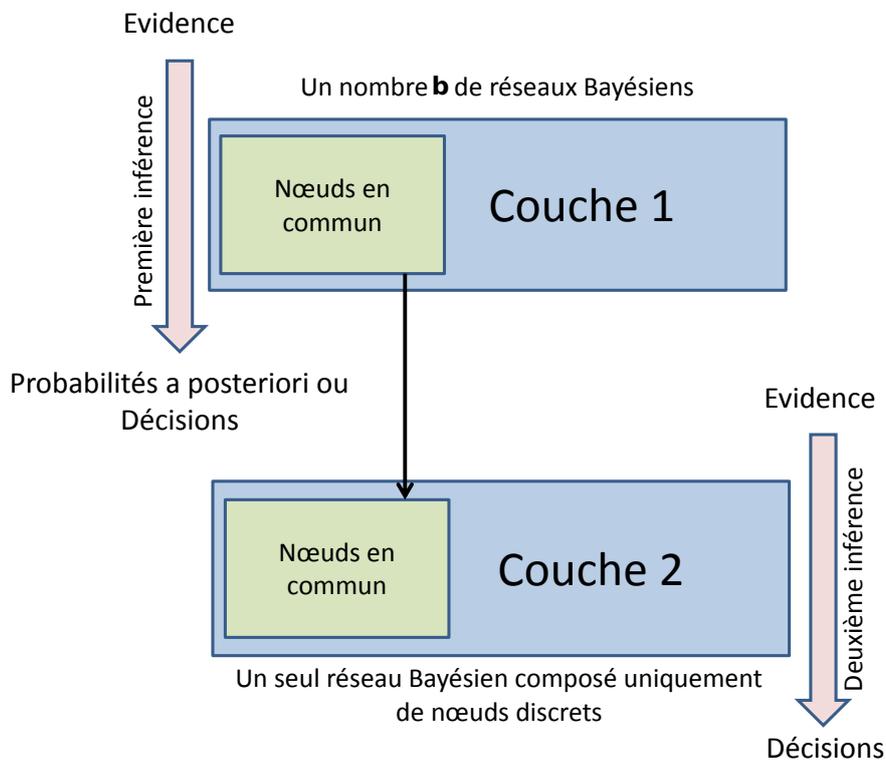


FIGURE 5.17 – Cadre proposé pour la surveillance des systèmes

méthodes ne permet pas forcément d'améliorer la décision. En effet, avant de fusionner ou combiner **b** méthodes il faut s'assurer de leurs hétérogénéités et complémentarités, et étudier l'intérêt de leur association pour plus de précision et de performance.

Par ailleurs, ce cadre probabiliste dans le cadre de la surveillance pourrait

nous permettre de bénéficier du maximum d'information disponible (données et modèle du système) sur le système. Précédemment, nous avons vu qu'un bon nombre de méthodes à base de données ou à base de modèle peuvent être modélisées sous des réseaux Bayésiens. Certaines d'entre elles peuvent même être utilisées lorsque certaines observations sont manquantes (étant donné l'aspect probabiliste des réseaux Bayésiens). Sous ce cadre, nous pourrions associer ces réseaux Bayésiens, fusionner leurs décisions et bénéficier donc de l'avantage de chacun.

Afin d'illustrer ceci nous prenons en exemple une combinaison d'une méthode à base de données, une analyse discriminante linéaire, et l'autre représentant une méthode à base de modèle, une multitude de tests statistiques (pour l'évaluation des résidus générés à partir d'un générateur de résidus adapté et adéquat au modèle du système traité) arrangés étant donnée une matrice d'incidence. Cette association peut être justifiée par la simple raison que les deux méthodes sont issues de deux classes de méthodes employant des informations de type différent.

La structure de la première méthode est présentée dans la figure 5.18. Il s'agit d'un réseau Bayésien classifieur discriminant entre les fautes et l'état *CFN* (état de fonctionnement normal). Le réseau proposé est alors composé d'un noeud discret \mathbf{S}_m avec K états et un noeud \mathbf{x} joignant les m variables du système. Pour chaque état $\mathbf{S}_{m_j=1,\dots,K} \in (CFN, F_1, F_2, \dots, F_{K-1})$, \mathbf{x} suit une distribution normale multivariée avec une moyenne $\mu_{\mathbf{S}_{m_j}}$ et une matrice de variance-covariance $\Sigma_{\mathbf{S}_{m_j}}$. Ses paramètres sont estimés étant donnée l'hypothèse de disponibilité d'un ensemble d'apprentissage. Comme le réseau Bayésien classifieur correspond à une analyse discriminante linéaire, une matrice de variance-covariance commune est définie pour l'ensemble des classes (voir (5.46)).

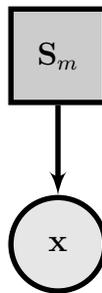


FIGURE 5.18 – Le réseau Bayésien à base de données pour le diagnostic de fautes

Les probabilités a priori des états du noeud \mathbf{S}_m sont supposées équiprobables. Ceci est justifié par le fait que les données de fautes sont collectées une fois implémentés. Ainsi, la probabilité a priori d'un état donné ne peut être exactement définie. Une fois le réseau mis à jour, étant donnée une nouvelle observation, les probabilités a posteriori correspondant à chaque état sont déterminées. Une décision peut être, par exemple, prise en considérant l'état ayant la plus grande probabilité a posteriori comme ceci :

$$\delta : x \in \mathbf{S}_{m_j}^*, \text{ si } j^* = \operatorname{argmax}_{j=1, \dots, K} \frac{p(\mathbf{S}_{m_j})p(\mathbf{x} = x | \mathbf{S}_{m_j})}{p(\mathbf{x} = x)} \quad (5.65)$$

où $p(\mathbf{S}_{m_j} | \mathbf{x} = x)$ est la probabilité a posteriori de la classe \mathbf{S}_{m_j} étant donné x , $p(\mathbf{x})$ est la fonction de densité de \mathbf{x} , $p(\mathbf{x} | M_j)$ est la vraisemblance et $p(\mathbf{S}_{m_j})$ est la probabilité a priori de \mathbf{S}_{m_j} .

\mathbf{S}_m			
CFN	F_1	\dots	F_{K-1}
$p(CFN)$	$p(F_1)$	\dots	$p(F_{K-1})$

FIGURE 5.19 – La table de probabilités conditionnelles de \mathbf{S}_m

\mathbf{S}_m	\mathbf{x}
CFN	$\mathbf{x} \sim \mathcal{N}(\mu_{CFN}; \Sigma_{CFN})$
F_1	$\mathbf{x} \sim \mathcal{N}(\mu_{F_1}; \Sigma_{F_1})$
\vdots	\vdots
F_{K-1}	$\mathbf{x} \sim \mathcal{N}(\mu_{F_{K-1}}; \Sigma_{F_{K-1}})$

FIGURE 5.20 – La table de probabilités conditionnelles du noeud \mathbf{x}

Afin d'assister le réseau Bayésien donné précédemment, une autre stratégie de diagnostic est considérée. Au lieu d'utiliser les données du système, elle se base sur un modèle analytique a priori disponible du système (une méthode à base de modèle). Le réseau Bayésien proposé lui correspondant est constitué de noeuds discrets $\mathbf{F}_1, \dots, \mathbf{F}_{K-1}$, représentant les différentes fautes et comportant chacun deux états (présence (*Oui*) et non présence (*Non*)), et de noeuds continus (des résidus sensibles à F_j avec $j = 1, \dots, K - 1$) représentant chacun une variable Gaussienne univariée. Chaque arc connectant ses noeuds correspond à un lien entre une faute \mathbf{F}_j et un résidu \mathbf{r}_i décrit par un "1" dans la matrice d'incidence correspondant au système traité. Un exemple d'une matrice d'incidence est donné par 5.21. La structure du réseau Bayésien lui correspondant

est donnée dans la figure 5.22.

	CFN	F_1	F_2	\dots	F_{K-1}
\mathbf{r}_1	0	1	0	\dots	0
\mathbf{r}_2	0	0	1	\dots	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
\mathbf{r}_n	0	0	1	\dots	1

FIGURE 5.21 – Un exemple d’une matrice d’incidence

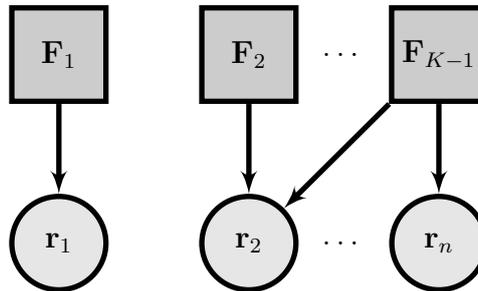


FIGURE 5.22 – Le réseau Bayésien à base de modèle pour le diagnostic de fautes

A chaque instant, ce réseau évalue les sorties d’un générateur de résidus (espace de parité, observateurs, etc.) avec un niveau de signification α associé à chaque classe, donnant la probabilité d’occurrence de chaque faute (la valeur d’état de chaque noeud \mathbf{F}_j). En fonction de ces probabilités a posteriori, nous pouvons détecter et isoler les différentes fautes décrites dans la matrice d’incidence (Verron et al., 2010a). Nous considérons uniquement les fautes simples, les fautes multiples ne sont pas prises en compte (pour des raisons de compatibilité avec la méthode à base de données considérée dans cet exemple). Par ailleurs, notons que comme l’état de fonctionnement normal est associé à un vecteur caractéristique nul $[0, \dots, 0]^T$ dans n’importe quelle matrice d’incidence (voir la figure 5.21), le noeud CFN n’est lié à aucun autre noeud dans le réseau Bayésien. Le système est déclaré implicitement sous contrôle (cas sans fautes) si et uniquement si aucune faute F_j n’est détectée. Nous proposons d’étendre ce réseau Bayésien afin qu’il puisse exprimer explicitement les probabilités d’appartenance à une des $K - 1$ fautes en plus de celle de la classe de fonctionnement normal (CFN). Pour cela, nous proposons d’introduire (voir figure 5.27) un nouveau noeud discret parent \mathbf{S}_r avec K états (CFN, F_1, \dots, F_{K-1}), connectant les $K - 1$ noeuds de fautes.

Les tables de probabilités conditionnelles du noeud \mathbf{S}_r et ses noeuds fils $\mathbf{F}_j, j \in \{1, \dots, K-1\}$ sont données dans les figures 5.23 et 5.24.

\mathbf{S}_r			
CFN	F_1	\dots	F_{K-1}
$p(CFN)$	$p(F_1)$	\dots	$p(F_{K-1})$

FIGURE 5.23 – La table de probabilités conditionnelles du noeud \mathbf{S}_r

\mathbf{S}_r	\mathbf{F}_j	
	<i>Oui</i>	<i>Non</i>
CFN	0	1
F_1	0.5	0.5
\vdots	\vdots	\vdots
F_j	1	0
\vdots	\vdots	\vdots
F_{K-1}	0.5	0.5

FIGURE 5.24 – La table de probabilités conditionnelles des noeuds \mathbf{F}_j

La table de probabilités conditionnelles de chaque noeud \mathbf{F}_j est configurée avec les règles intuitives suivantes :

- si $\mathbf{S}_r = F_j$, alors il est certain ($p(\mathbf{F}_j = \text{Oui} | \mathbf{S}_r = F_j) = 1$) que l'observation appartient à la faute F_j ;
- si $\mathbf{S}_r = CFN$, alors il est certain ($p(\mathbf{F}_j = \text{Oui} | \mathbf{S}_r = CFN) = 0$) que l'observation se différencie de la faute F_j ;
- si $\mathbf{S}_r = F_{o(o \neq j)}$, alors aucune connaissance n'est apprise à propos de l'appartenance de la faute F_j , nous fixons $p(\mathbf{F}_j = \text{Oui} | \mathbf{S}_r = F_{o(o \neq j)}) = p(\mathbf{F}_j = \text{Non} | \mathbf{S}_r = F_{o(o \neq j)}) = 0.5$, cette valeur peut être définie afin d'avantager certaines fautes.

Sous le cadre probabiliste proposé, nous allons voir comment associer ces deux méthodes issues de deux domaines différents et gérer à la fois les différentes variables impliquées (variables et résidus). Pour des raisons de clarté et de généralité nous avons préféré utiliser directement les résidus générés, sous l'hypothèse qu'ils sont issus d'un générateur adéquat au problème formulé.

Les deux méthodes probabilistes présentées précédemment utilisent deux différents types d'informations. Nous pouvons imaginer que leur association permettra aux deux méthodes probabilistes de bénéficier des avantages de chacune et limitera à un certain point leurs inconvénients (e.g. lorsque les

paramètres ne sont pas bien estimés, le modèle est peu précis et les résidus peu fiables). Dans (Atoui et al., 2015a), cela a été testé sur un exemple d’une simulation d’un chauffe-eau (Weber et al., 2008) où nous avons pu constater une amélioration concernant la prise de décision.

Le chauffe-eau se compose d’un réservoir équipé de deux résistances \mathbf{R}_1 et \mathbf{R}_2 . Les entrées sont le débit de flux d’eau \mathbf{Q}_i , la température de l’eau T_i et la puissance électrique de chauffage \mathbf{P} . Les sorties sont le taux de flux d’eau \mathbf{Q}_0 et la température \mathbf{T} régulée autour d’un point de fonctionnement. La température de l’eau entrante T_i est supposée constante.

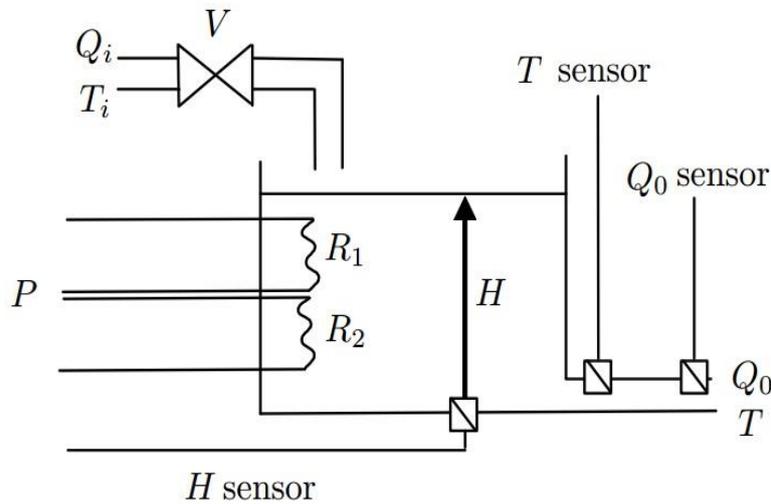


FIGURE 5.25 – Heating water system

L’objectif du système est de fournir un certain débit d’eau à une température donnée. Dans cette analyse, seule les fautes capteurs et composantes sont prises en compte ($K = 4$) : F_1 : niveau du liquide, F_2 : température sortante, F_3 : débit d’eau sortant. Le modèle mathématique détaillé du système de chauffe-eau est présenté dans (Weber et al., 2008). En s’appuyant sur ce dernier trois résidus sont générés et une matrice d’incidence est établi (voir figure 5.26).

Afin d’associer les deux méthodes, nous allons définir les deux couches composant le cadre probabiliste. La première couche contient les deux réseaux et permet d’obtenir, sachant de nouvelles observations, les probabilités a posteriori pour chaque valeur de chaque noeud discret de décision, notamment \mathbf{S}_m

	<i>CFN</i>	F_1	F_2	F_3
\mathbf{r}_1	0	1	0	0
\mathbf{r}_2	0	0	1	0
\mathbf{r}_3	0	0	1	1

FIGURE 5.26 – Matrice d’incidence du chauffe-eau

et $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$. Ces différentes probabilités obtenues sont ensuite injectées dans la seconde couche sous forme d’évidence. Nous avons choisi d’utiliser les différentes probabilités a posteriori de chaque classe comme observations (évidence virtuelle) dans le réseau Bayésien concerné par la seconde inférence. Rappelons que le réseau correspondant à cette couche n’est composé que de noeuds discrets. Ces noeuds discrets, dans notre exemple, représentent une copie des noeuds de sorties de la première couche (noeuds de décision du réseau Bayésien) dont certains $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3$ sont reliés à un noeud \mathbf{S}_r explicitant les différents états de fonctionnement du système. En plus de ces noeuds, un nouveau noeud discret $\mathbf{S}_{r\&m}$ est rajouté. Ce noeud permet de prendre une décision étant données les deux méthodes. Il vient lier naïvement les deux noeuds \mathbf{S}_r et \mathbf{S}_m . Dans cet exemple, il représente une variable discrète qui a les mêmes valeurs, K états (CFN, F_1, F_2, F_3), que les noeuds discrets \mathbf{S}_r et \mathbf{S}_m .

La probabilité a posteriori de chaque état K du noeud $\mathbf{S}_{r\&m}$, étant donné une nouvelle évidence virtuelle (ve) : $e_F = (e_{F_1}, e_{F_2}, e_{F_3}), e_{S_m}$ respectivement les probabilités a posteriori des états des noeuds $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3), \mathbf{S}_r$ étant donnés les résidus et les observations des variables du système, peut être inféré comme suit (voir (Blimes, 2004; Pearl, 1986), pour plus de détails à propos des évidences virtuelles dans les réseaux Bayésiens) :

$$\begin{aligned}
 p(\mathbf{S}_{r\&m_j} = |ve) &= \sum_{\mathbf{S}_r} \sum_{\mathbf{F}} p(\mathbf{S}_{r\&m_j}, \mathbf{S}_r, \mathbf{S}_m, \mathbf{F}) \frac{p(\mathbf{S}_m, \mathbf{F} | ve)}{p(\mathbf{S}_m, \mathbf{F})} = \\
 &= \frac{p(\mathbf{S}_{r\&m_j}) \sum_{\mathbf{S}_m} p(\mathbf{S}_m | \mathbf{S}_{r\&m_j}) e_{S_m} \sum_{\mathbf{S}_r} p(\mathbf{S}_r | \mathbf{S}_{r\&m_j}) \sum_{\mathbf{F}} p(\mathbf{F} | \mathbf{S}_r) e_F}{\sum_{\mathbf{S}_m} \sum_{\mathbf{F}} p(\mathbf{S}_m, \mathbf{F}) e_{S_m} e_F} \quad (5.66)
 \end{aligned}$$

Ainsi, une combinaison de probabilités est réalisée sous l’hypothèse que la méthode à base de modèle et la méthode à base de données sont conditionnellement indépendantes sachant le noeud rajouté. La table de probabilité conditionnelle du noeud rajouté est donnée par la figure 5.28.

Une fois les réseaux construits, le cadre probabiliste (les deux couches fusionnant les décisions prises par les deux méthodes (voir la figure 5.27)), on

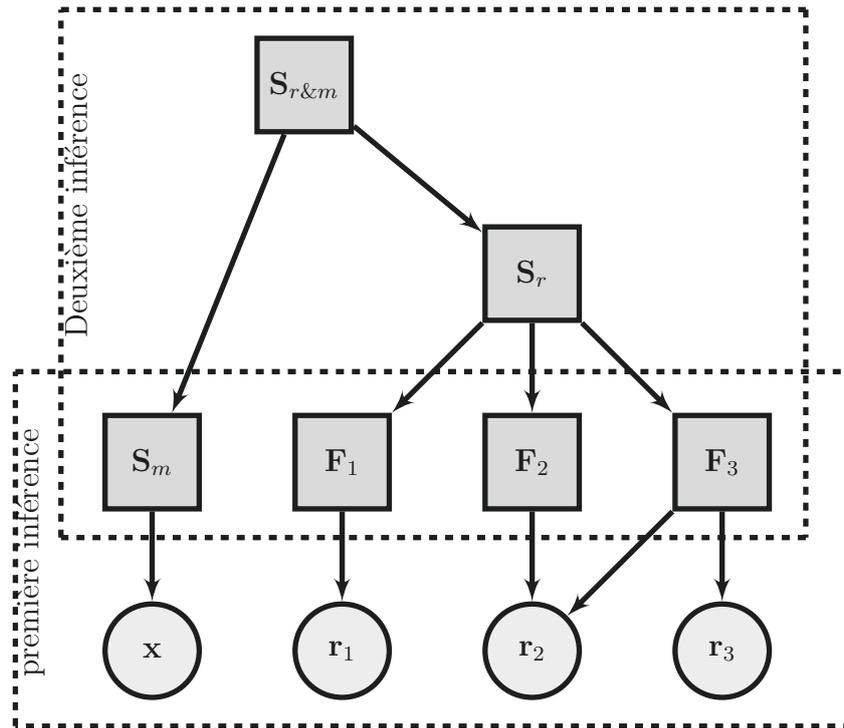


FIGURE 5.27 – Les différentes couches d’inférence pour la combinaison de deux méthodes pour l’isolation de fautes

$\mathbf{S}_{r\&m}$			
CFN	F_1	F_2	F_3
$p(CFN)$	$p(F_1)$	$p(F_2)$	$p(F_3)$

FIGURE 5.28 – La table de probabilités conditionnelles du nœud $\mathbf{S}_{r\&m}$

vient introduire dans chaque couche l’évidence lui correspondant. Nous avons simulé le système selon les scénarios décrits dans la figure 5.29. A la fin de chaque inférence dans la deuxième couche, le nœud $\mathbf{S}_{r\&m}$ indique, pour chaque classe, la probabilité de son occurrence. Parmi d’autres critères, nous choisissons celui du maximum a posteriori. Ainsi, pour chaque instant, la classe avec la plus grande probabilité a posteriori est choisie, comme ceci :

$$\delta : ve \in \mathbf{S}_{r\&m_j}^* = \operatorname{argmax}_{j=1,\dots,K} \frac{p(\mathbf{S}_{r\&m_j})p(ve|\mathbf{S}_{r\&m_j})}{p(ve)} \quad (5.67)$$

Les résultats obtenus en testant notre méthode sous les différents scénarios sont proposés dans la table de confusion présentée dans la figure 5.30. Nous pouvons remarquer que les deux méthodes combinées donnent un taux

Échantillons	1-30	31-60	61-90	91-120
Scénarios	CFN	F_1	F_2	F_3

FIGURE 5.29 – Scénarios simulés

de mauvaise erreurs inférieur à celui obtenu par les deux méthodes utilisées séparément. Le fait que les taux d’erreurs soient faibles se justifie par le fait que les résidus et les données issues du système ne sont pas difficile à isoler. Toutefois, à travers cette simple et illustrative combinaison, nous pouvons voir le potentiel du réseau Bayésien à modéliser et combiner de façon naturelle des approches de nature différente.

Méthode à base de données					
State	CFN	F_1	F_2	F_3	Total
CFN	30	0	0	0	30
F_1	1	29	0	0	30
F_2	1	0	29	0	30
F_3	0	0	0	30	30
Total	32	29	29	30	120

(a)

Méthode à base de modèle					
State	CFN	F_1	F_2	F_3	Total
CFN	29	1	0	0	30
F_1	0	30	0	0	30
F_2	0	0	30	0	30
F_3	0	0	0	30	30
Total	29	31	30	30	120

(b)

Combinaison					
State	CFN	F_1	F_2	F_3	Total
CFN	30	0	0	0	30
F_1	0	30	0	0	30
F_2	0	0	30	0	30
F_3	0	0	0	30	30
Total	30	30	30	30	120

(c)

FIGURE 5.30 – Matrices de confusion pour chaque approche proposée sur les données simulées à titre d’exemple

Pour conclure, nous avons proposé de modéliser et de représenter des méthodes de détection et de diagnostic de fautes à base de réseaux Bayésiens. La

sortie de ces réseaux Bayésiens correspond aux probabilités a posteriori des différents états de fonctionnement considérées. Ces sorties peuvent être considérées comme telles (évidences softs ou virtuelles) ou transformées (discrétisées après comparaison à des seuils ou limites probabilistes). En tant qu'évidence, elles sont par la suite combinées en utilisant la règle de Bayes. Le cadre proposé ainsi que les réseaux Bayésiens pris individuellement, grâce à leurs aspects probabilistes, peuvent être un moyen efficace pour assister les ingénieurs dans leur prise de décision à propos de l'état de fonctionnement du système surveillé (e.g. ordonner et/ou décider sur les différents états du système selon leurs probabilités a posteriori d'occurrence).

IV Conclusion

Ce chapitre nous a permis de présenter nos contributions en termes de détection et diagnostic de fautes sous réseaux Bayésiens. Ces contributions portent sur plusieurs points. Nous avons dans un premier temps généraliser les statistiques quadratiques et leurs limites probabilistes sur réseaux Bayésiens. Cette généralisation nous a permis de faire de la détection et du diagnostic de façon simultanée tout en intégrant la notion de rejet de distance. Nous avons présenté et identifié une famille de réseaux Bayésiens classifieurs intégrant cela. Finalement, nous avons proposé un cadre probabiliste permettant d'unifier et combiner des méthodes dédiées à la détection et d'autres au diagnostic.

Conclusion générale

De nos jours, les entreprises, cherchant à optimiser le rapport gain/perte d'un côté et le rapport qualité/prix d'un autre, sont de plus en plus amener à utiliser des moyens permettant de détecter rapidement la présence d'une faute (un dérèglement, une anomalie, etc.) et de bien l'identifier par la suite. Ces moyens sont nécessaires pour maintenir une production conforme aux attentes et éviter de lourdes conséquences humaines, matérielles et environnementales. Dans ce cadre, nous avons proposé une approche probabiliste pour la détection et le diagnostic de fautes. Plus précisément, nous avons utilisé un réseau Bayésien, un outil permettant la modélisation de plusieurs schémas de détection et de diagnostic basés sur une transformation linéaire des observations avec la possibilité d'intégrer un rejet de distance.

Dans le premier chapitre, nous avons commencé par introduire les réseaux Bayésiens. Nous avons également mentionné l'intérêt émanant de l'utilisation de variables Gaussiennes. Par la suite, nous avons décrit certaines propriétés principales propres aux réseaux Bayésiens. Après la définition de quelques noeuds pouvant composer un réseau Bayésien, nous avons présenté quelques structures de réseaux Bayésiens pouvant faire de la classification supervisée. Suite à cette présentation, nous avons décrit brièvement deux algorithmes populaires d'inférence dans les réseaux Bayésiens et utilisées dans le cadre de cette thèse.

Le second chapitre a fait l'objet d'un état de l'art se voulant non-exhaustif des méthodes proposées dans la littérature dans le contexte de la détection et du diagnostic de fautes. Une présentation de la famille des méthodes à base de modèle a été fournie suivie par celles des méthodes à base de données (statistiques). Le troisième et quatrième chapitres ont été consacrés aux contributions apportées au domaine de la détection et du diagnostic de fautes dans les systèmes par réseaux Bayésiens.

Dans le troisième chapitre, nous avons mis l'accent sur la détection de

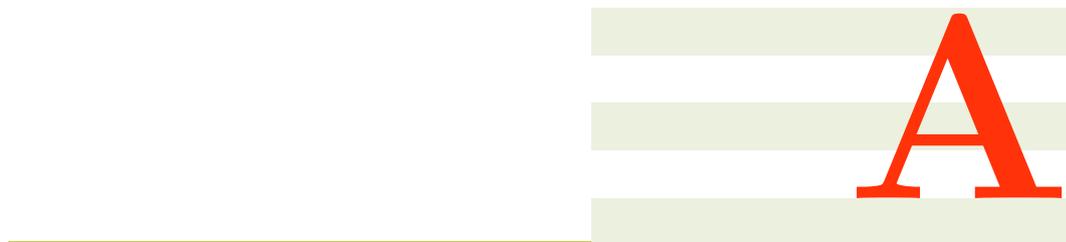
fautes. Les différentes méthodes proposées dans la littérature utilisant les réseaux Bayésiens pour la détection de fautes ont été illustrées suivies par nos contributions dans le même contexte. Dans un premier temps, nous avons mis l'accent sur l'analyse en composantes principales étant donné sa popularité. Nous avons proposé l'équivalent de son schéma de détection sous réseau Bayésien. Dans un second temps, nous avons présenté un cadre permettant de modéliser les schémas de détection se basant sur une transformé linéaire des observations.

Dans le quatrième chapitre, nous nous sommes placés spécialement dans le contexte du diagnostic de fautes. Après avoir présenté un état de l'art de réseaux Bayésiens proposés dans ce contexte, nous avons exposé nos travaux. Nous avons proposé une famille de réseaux Bayésiens permettant de faire de la détection et du diagnostic de façon simultanée, tout en tenant compte d'un rejet de distance. A la fin de ce chapitre un cadre probabiliste permettant d'unifier différents réseaux Bayésiens dédiées à la détection et au diagnostic de fautes a été présenté.

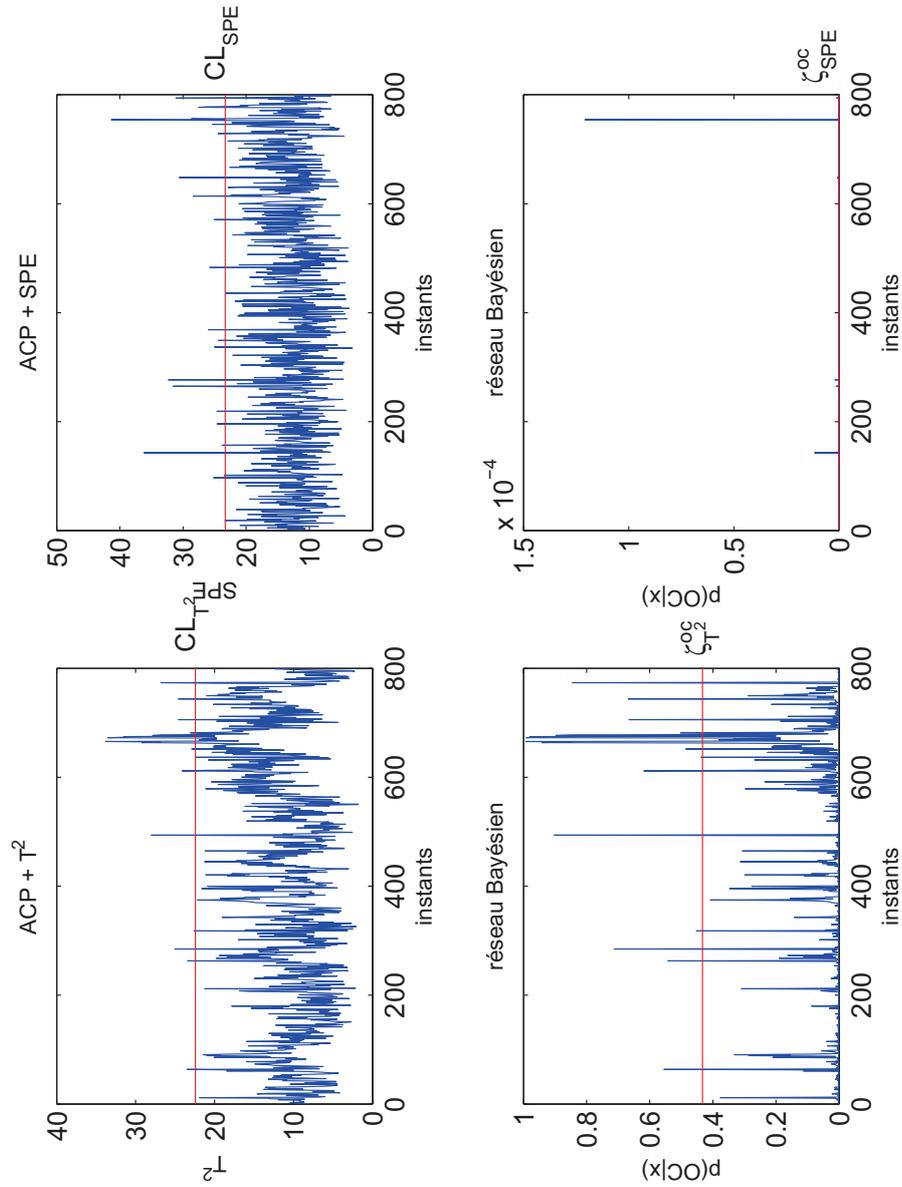
Les perspectives de ces travaux sont nombreuses. Concernant les réseaux Bayésiens proposés pour la détection et le diagnostic de fautes, il serait intéressant d'étudier leurs extensions que ce soit dans le cas non-gaussien ou non-linéaire. Aussi, nous avons vu que certains des réseaux Bayésiens proposés permettent de gérer les observations manquantes. Une comparaison plus approfondie de ces réseaux vis-à-vis d'autres méthodes gérant les observations manquantes (avec imputations ou sans imputations) serait d'un grand intérêt. De même, cette propriété caractérisant la plupart des réseaux Bayésiens proposés devrait être encore plus étudié. Nous pouvons par exemple s'en servir en ignorant à chaque fois une variable de sorte à identifier les variables responsables de la présence de fautes. Il serait également intéressant de comparer le rendu de ces méthodes par rapport à d'autres méthodes d'identification liées ou non à la méthode modélisée.

Une autre voie intéressante serait d'augmenter et d'intégrer à des réseaux Bayésiens modélisant les différentes relations entre les différentes variables du système (Kawahara et al., 2005; Schwall and Gerdes, 2002; Lerner et al., 2000) les réseaux Bayésiens proposés. Ceci permettra, par exemple, de surveiller les différents variables du modèle que ce soit de façon individuelle (une variable multivariée) ou de façon collective (plusieurs variables univariées). Nous avons vu dans ces travaux qu'il est théoriquement possible d'intégrer complètement

une carte MEWMA sur un réseau Bayésien dynamique. Pour cela, nous avons proposé l'utilisation de noeuds softmax. Ces noeuds pourraient être une solution lors de l'extension des réseaux proposés dans le cas dynamique. Enfin, une autre perspective serait d'augmenter les méthodes modélisées, comme l'analyse en composantes principales, par des réseaux Bayésiens prenant en compte d'autres informations, par exemple la fiabilité des composants (Zaidi et al., 2012; Weber et al., 2008). De même, ces méthodes peuvent être combinées par le cadre probabiliste proposé. Ce cadre nous a permis jusqu'à présent de combiner des méthodes à base de données et/ou d'autres à base de modèle. Il serait intéressant d'inclure à ce cadre d'autres méthodes à base de connaissances d'experts. Par ailleurs, on peut voir chaque méthode comme étant à part entière. Ainsi, nous pourrions créer une bibliothèque avec tous les réseaux Bayésiens pouvant être utilisés pour la détection et le diagnostic de fautes. Une étude de leur compatibilité serait intéressante. Nous pouvons également utiliser les réseaux Bayésiens orientés objet (Koller and Pfeffer, 1997) afin de créer une hiérarchie de réseaux et/ou des modules pouvant être combinés à notre guise.

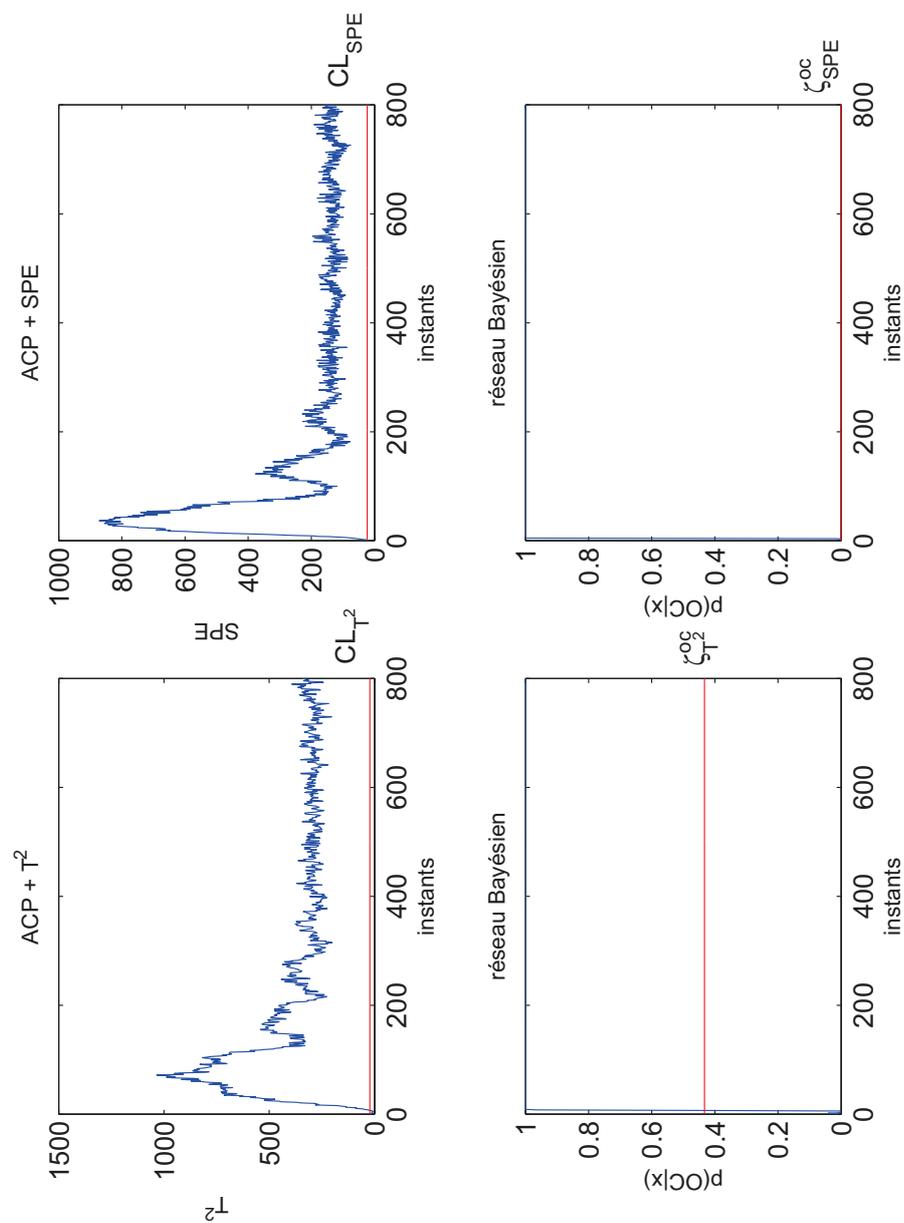


Comparaison des résultats



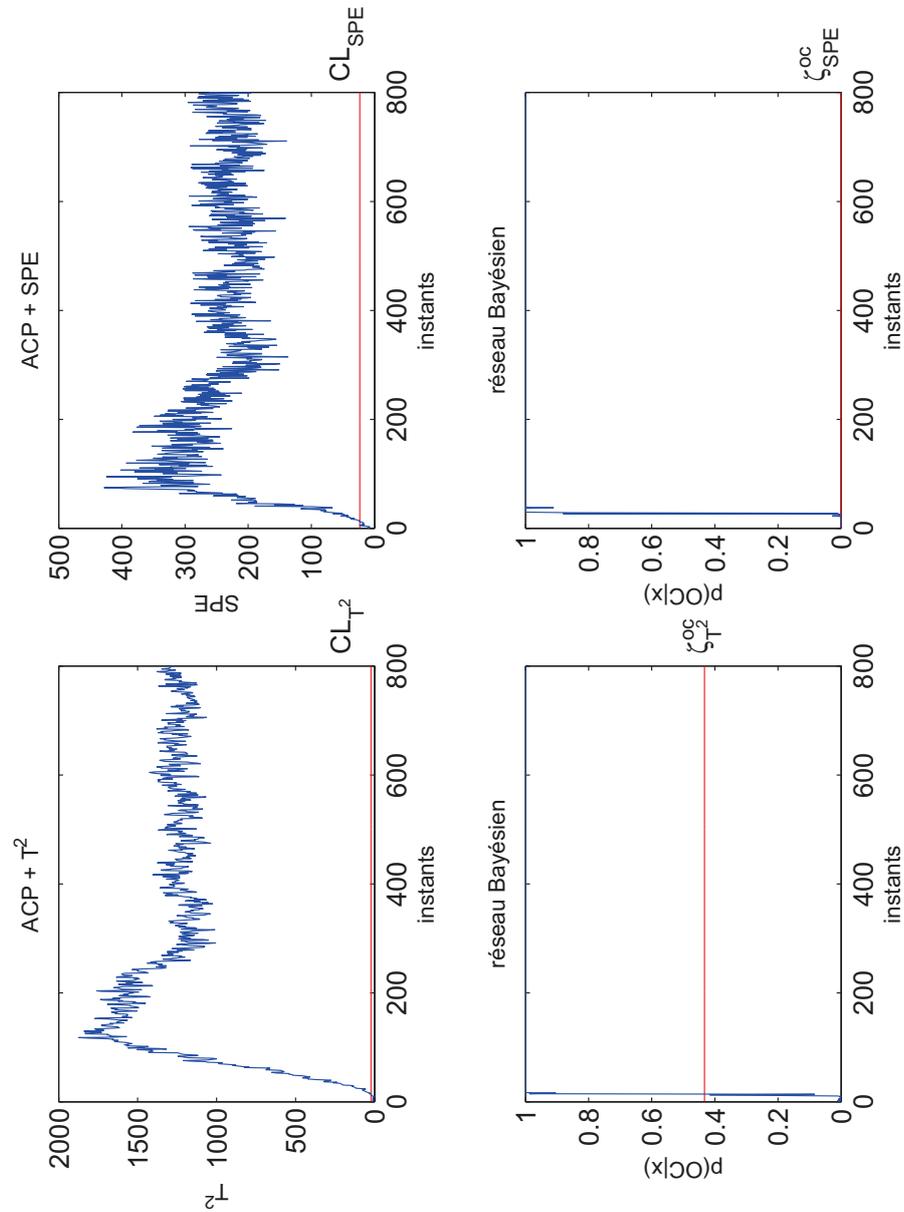
ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
2.5%	2.5%	3.125%	3.125%

FIGURE A.1 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe CFN (avec $c_{\Delta} = 10$) : représentation graphique + TFAs



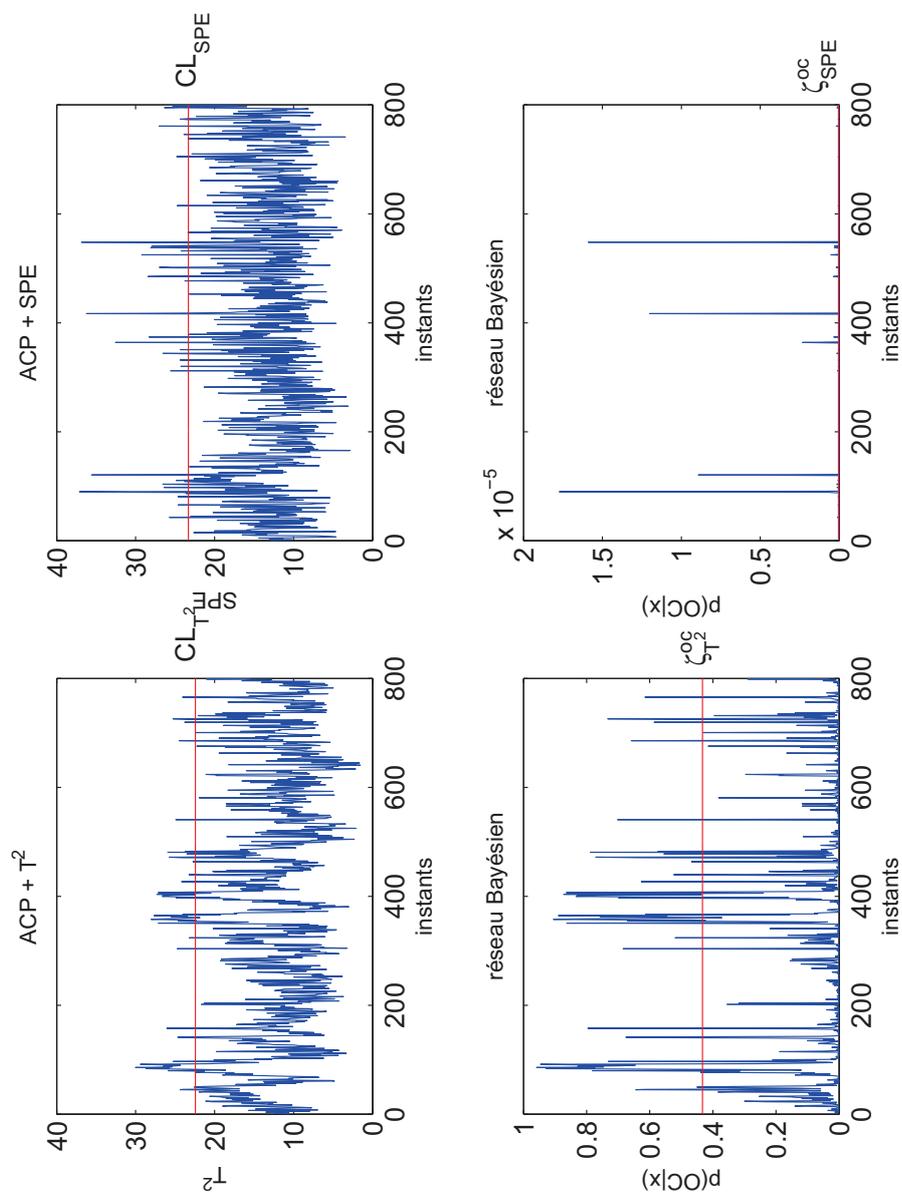
ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
0.75%	0.75%	0.125%	0.125%

FIGURE A.2 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_1 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs



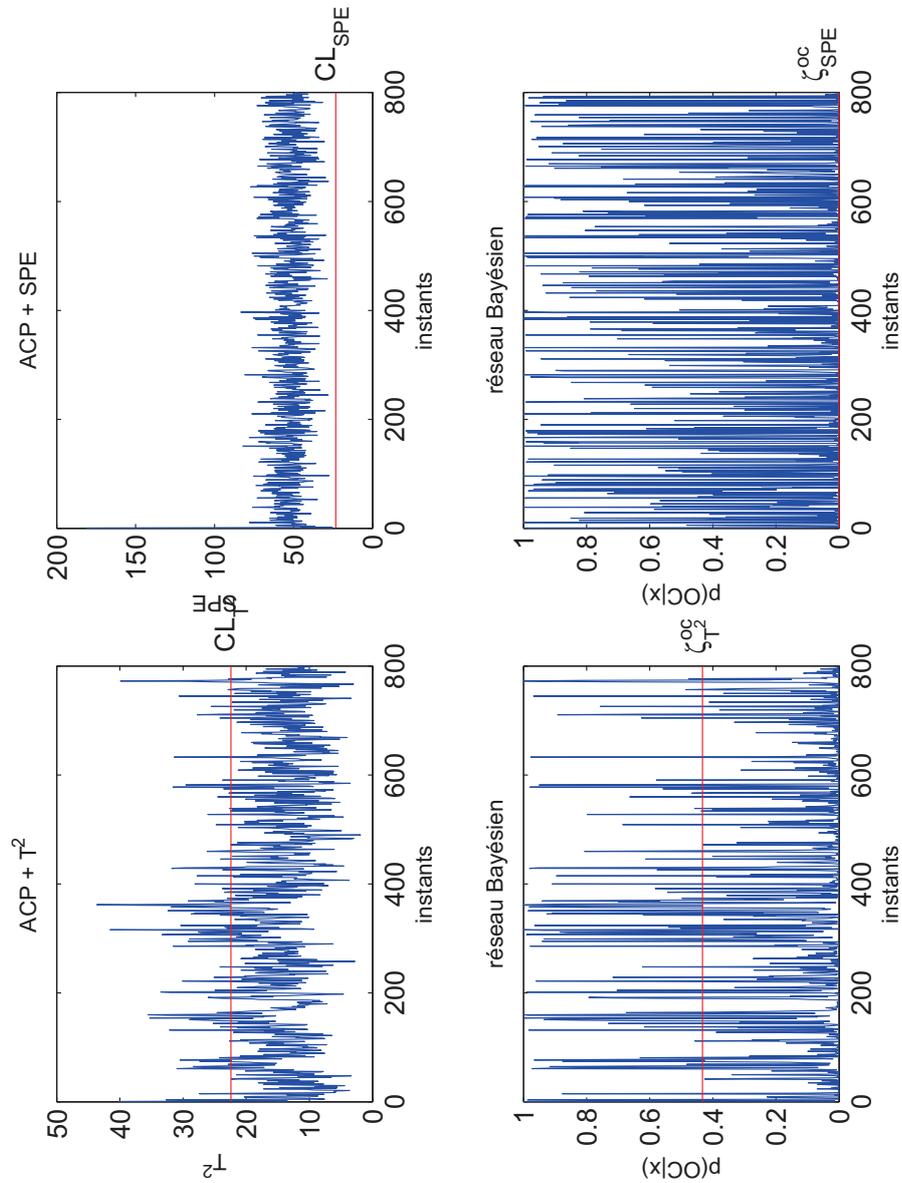
ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
1.75%	1.75%	1.375%	1.375%

FIGURE A.3 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_2 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs



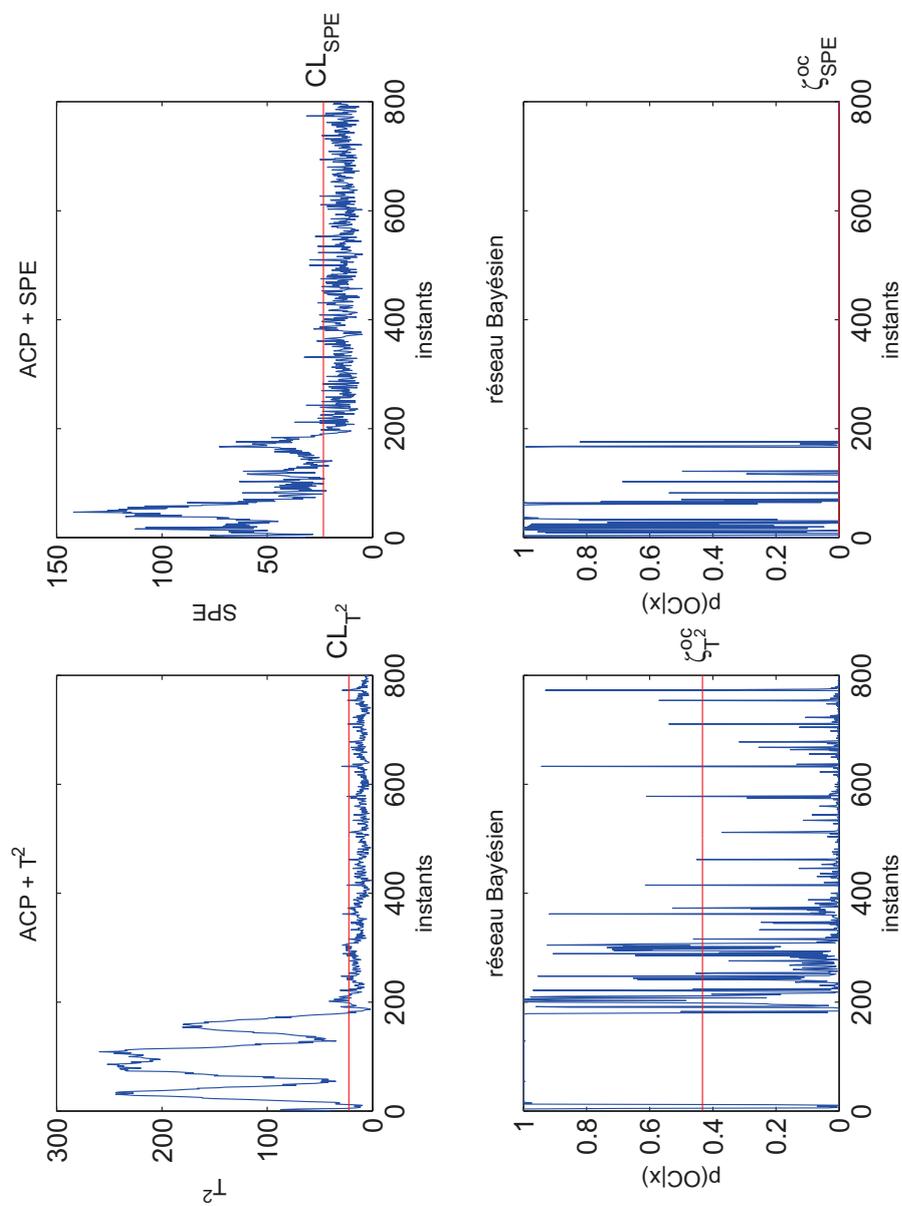
ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
93%	93%	94.75%	94.75%

FIGURE A.4 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_3 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
87.125%	87.125%	0%	0%

FIGURE A.5 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_4 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
72.75%	72.75%	72.625%	72.625%

FIGURE A.6 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_5 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs

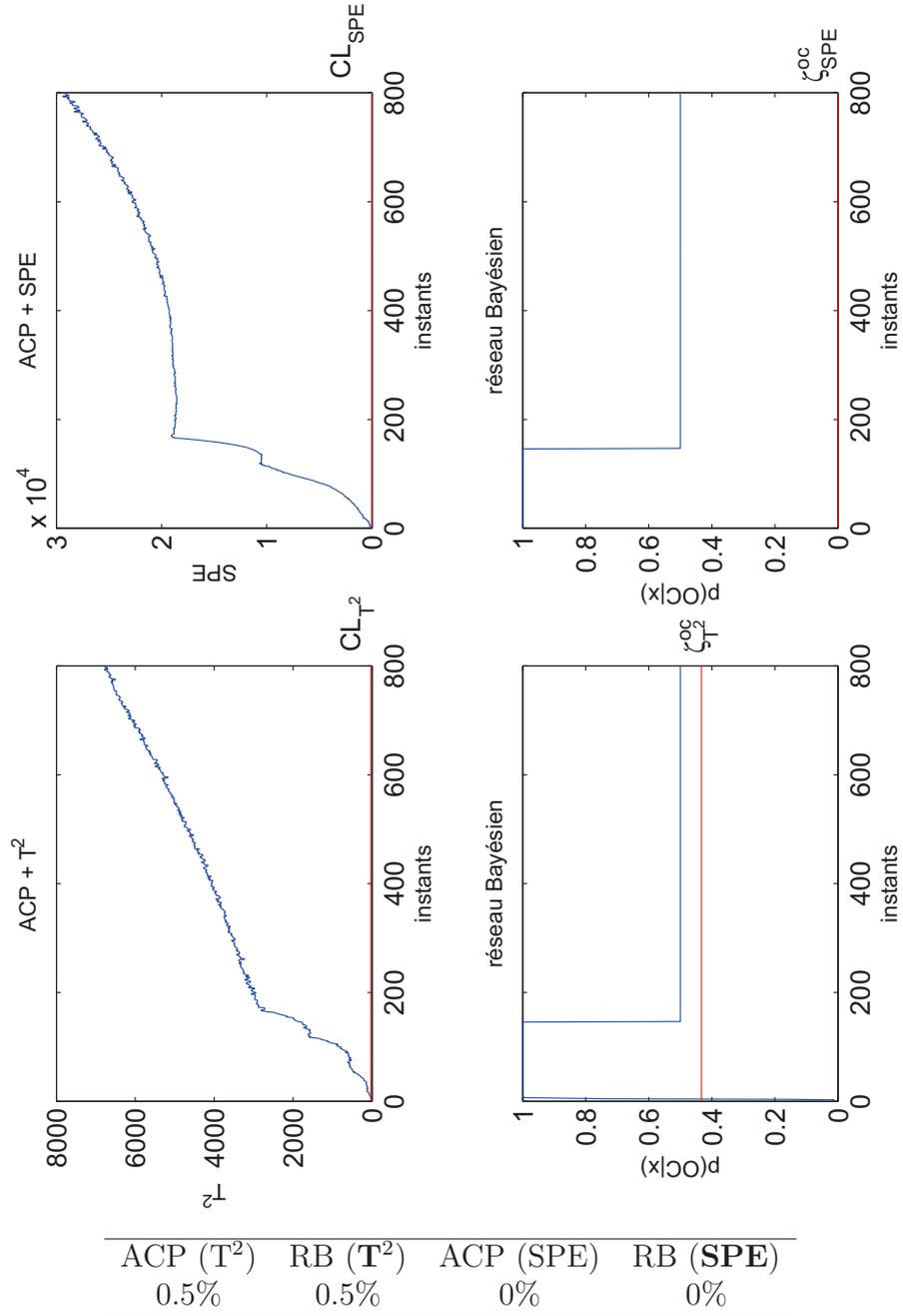
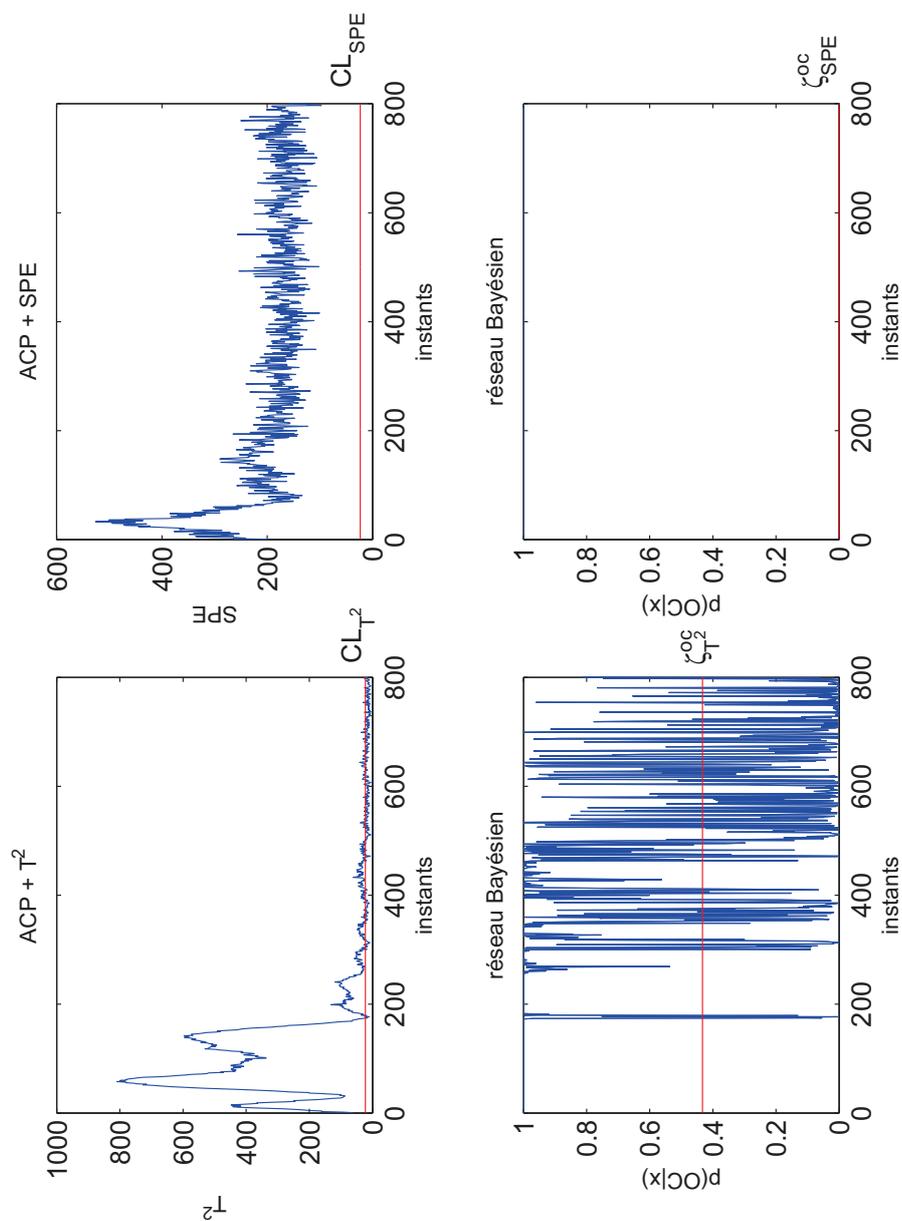


FIGURE A.7 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_6 (avec $c_\Delta = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
37.75%	37.75%	0%	0%

FIGURE A.8 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_7 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs

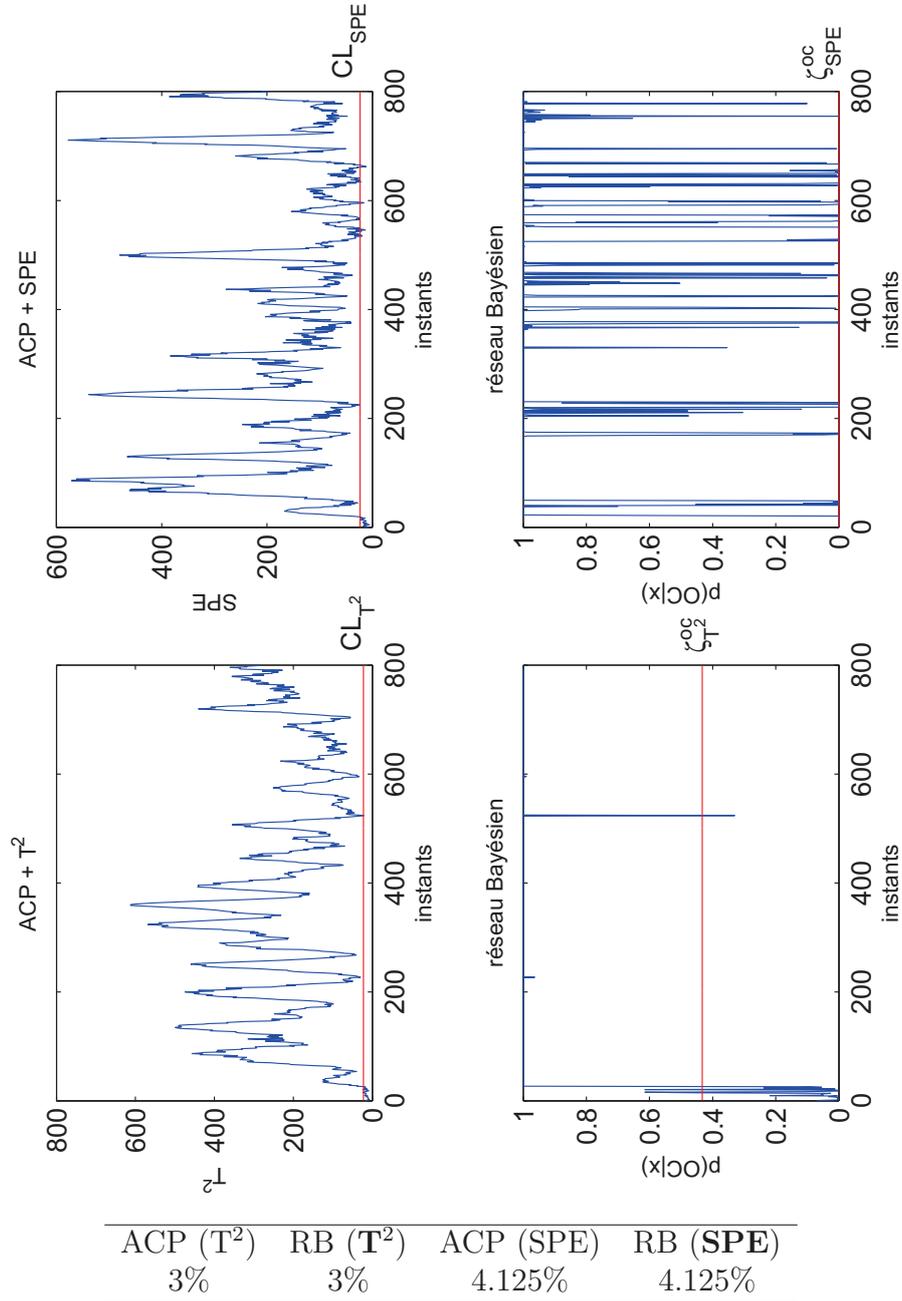
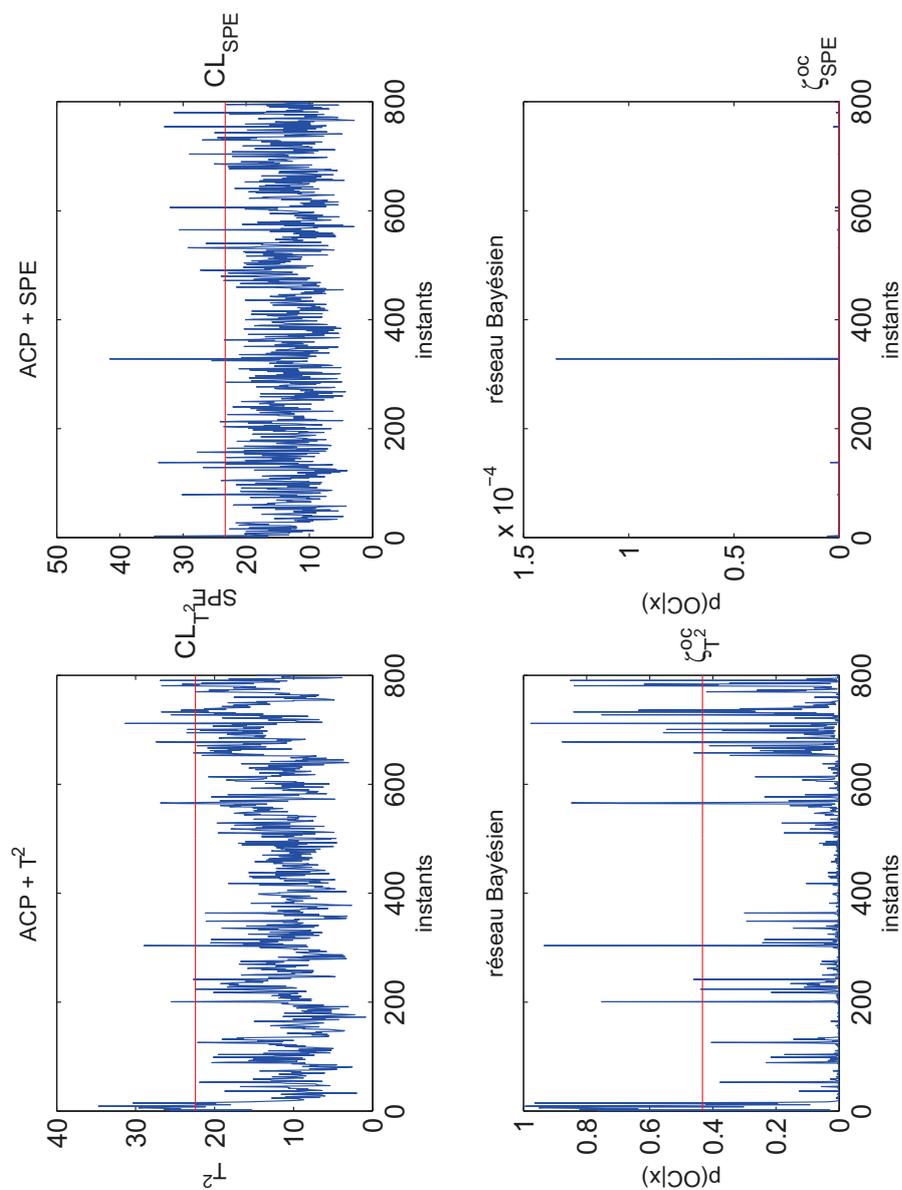
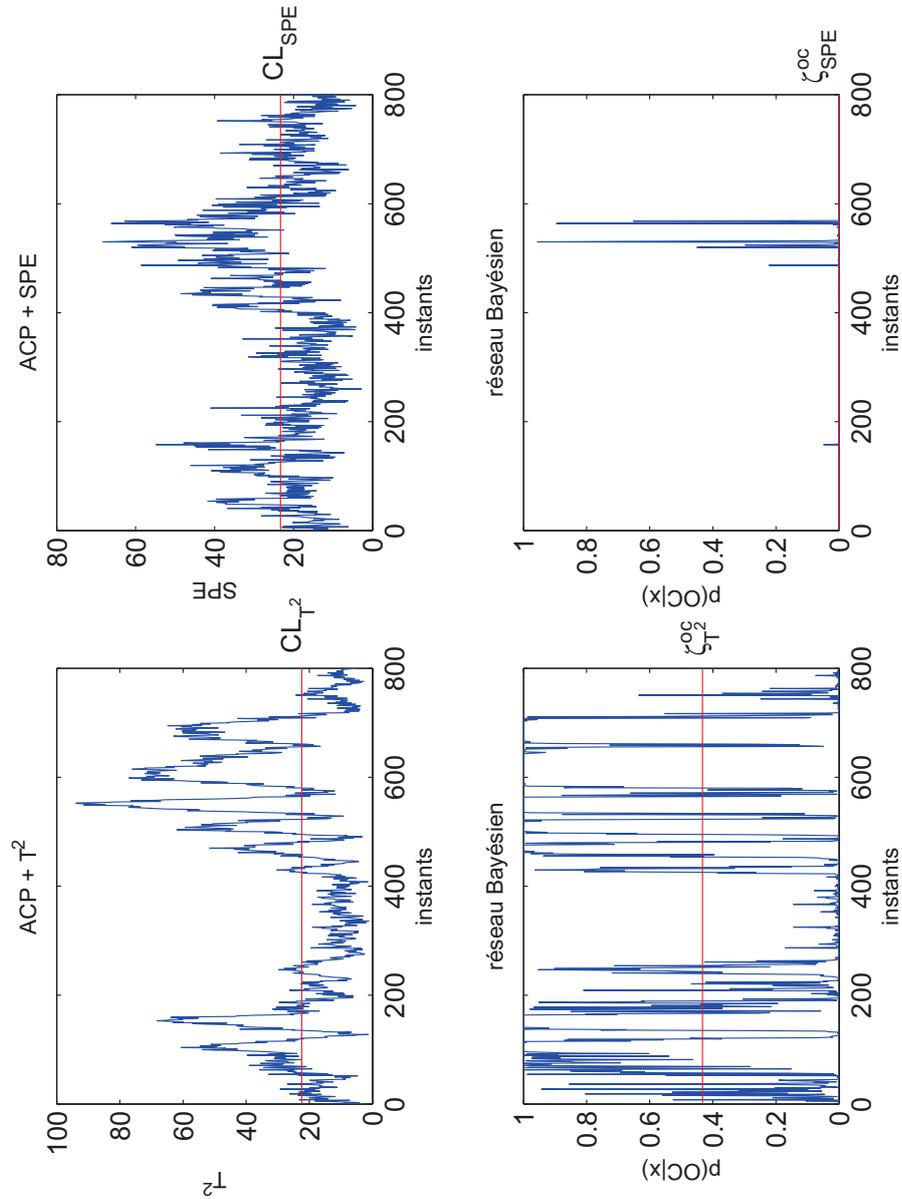


FIGURE A.9 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_8 (avec $c_\Delta = 10$) : représentation graphique + TMDs



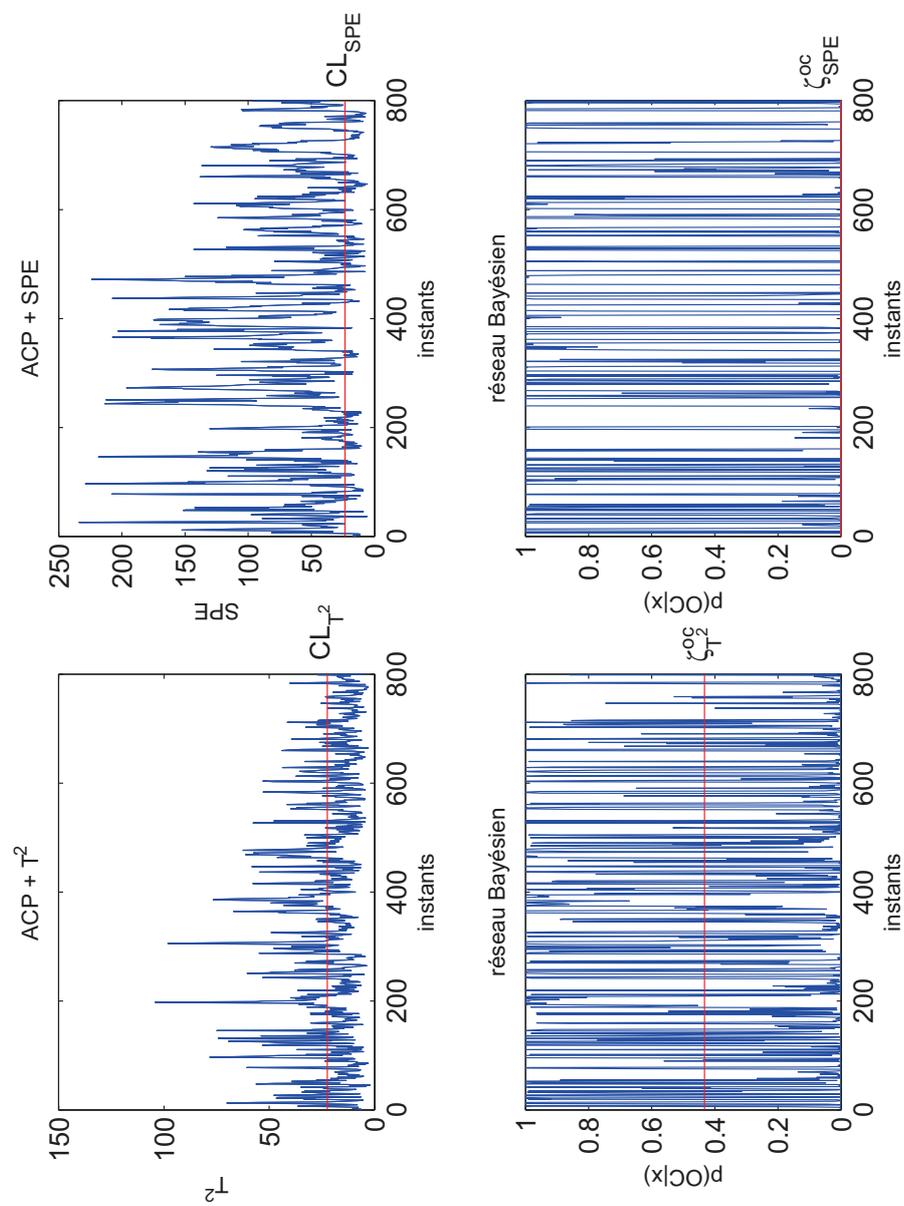
ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
96.375%	96.375%	96.375%	96.375%

FIGURE A.10 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_9 (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs



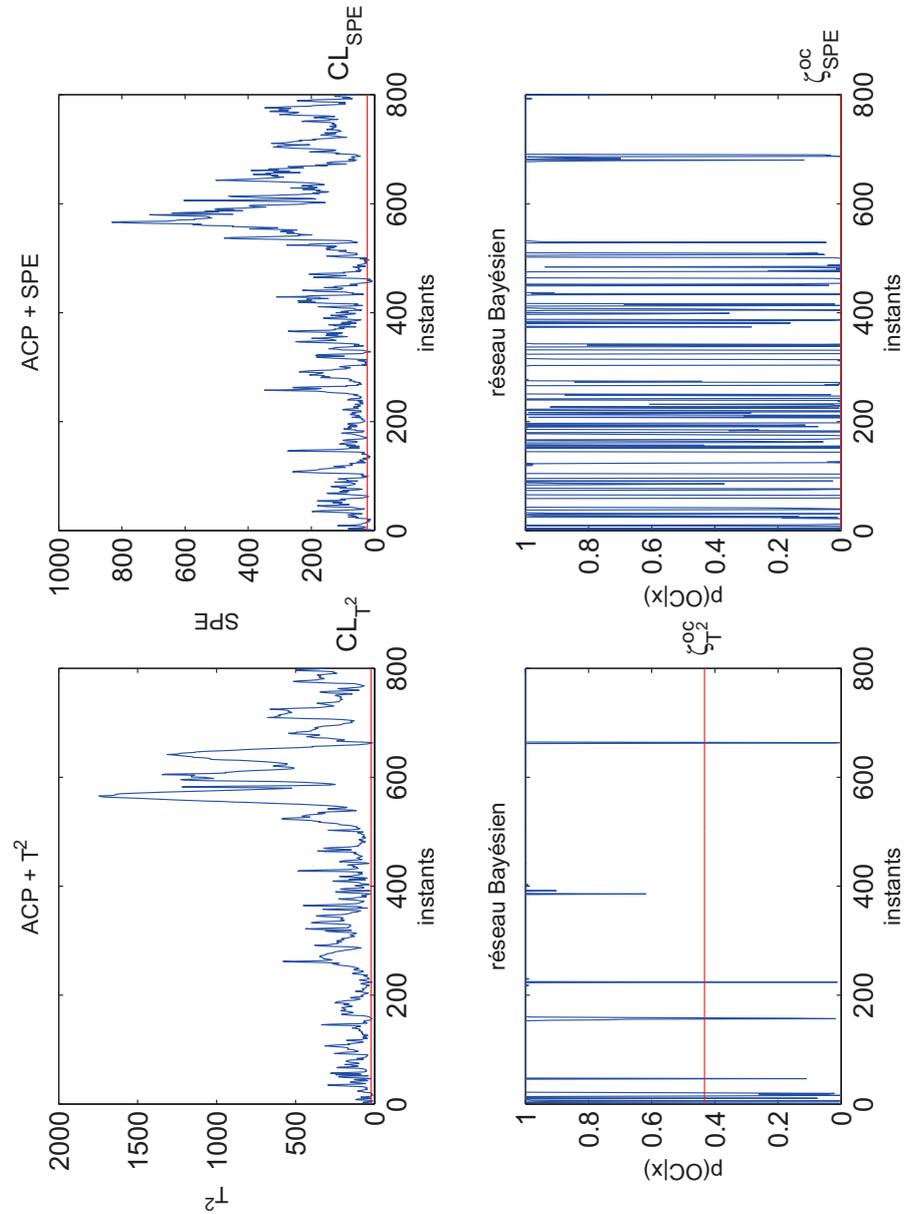
ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
55.875%	55.875%	63.75%	63.75%

FIGURE A.11 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{10} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
67.25%	67.25%	23.125%	23.125%

FIGURE A.12 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{11} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
2%	2%	3.875%	3.875%

FIGURE A.13 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{12} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs

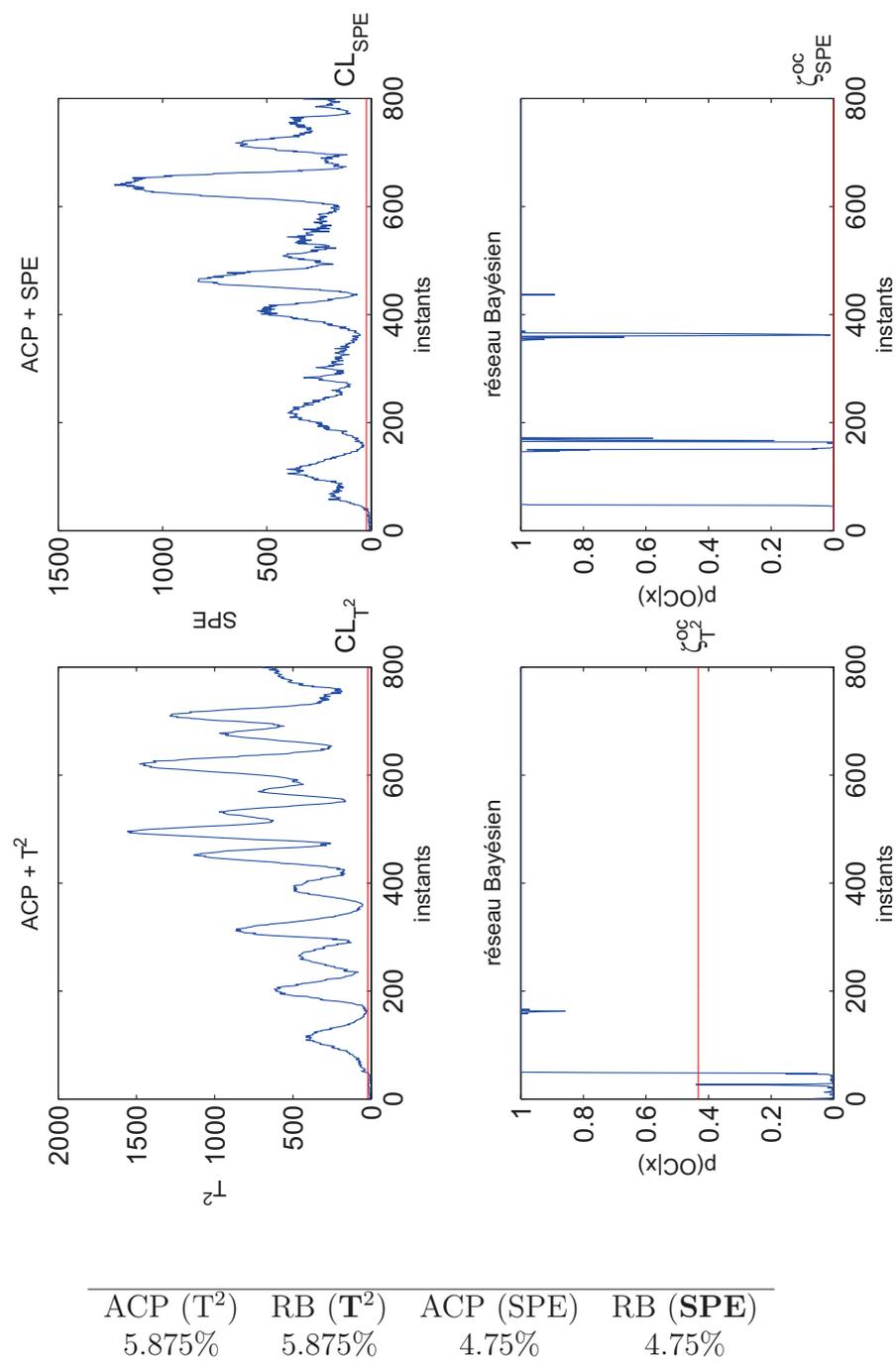


FIGURE A.14 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{13} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs

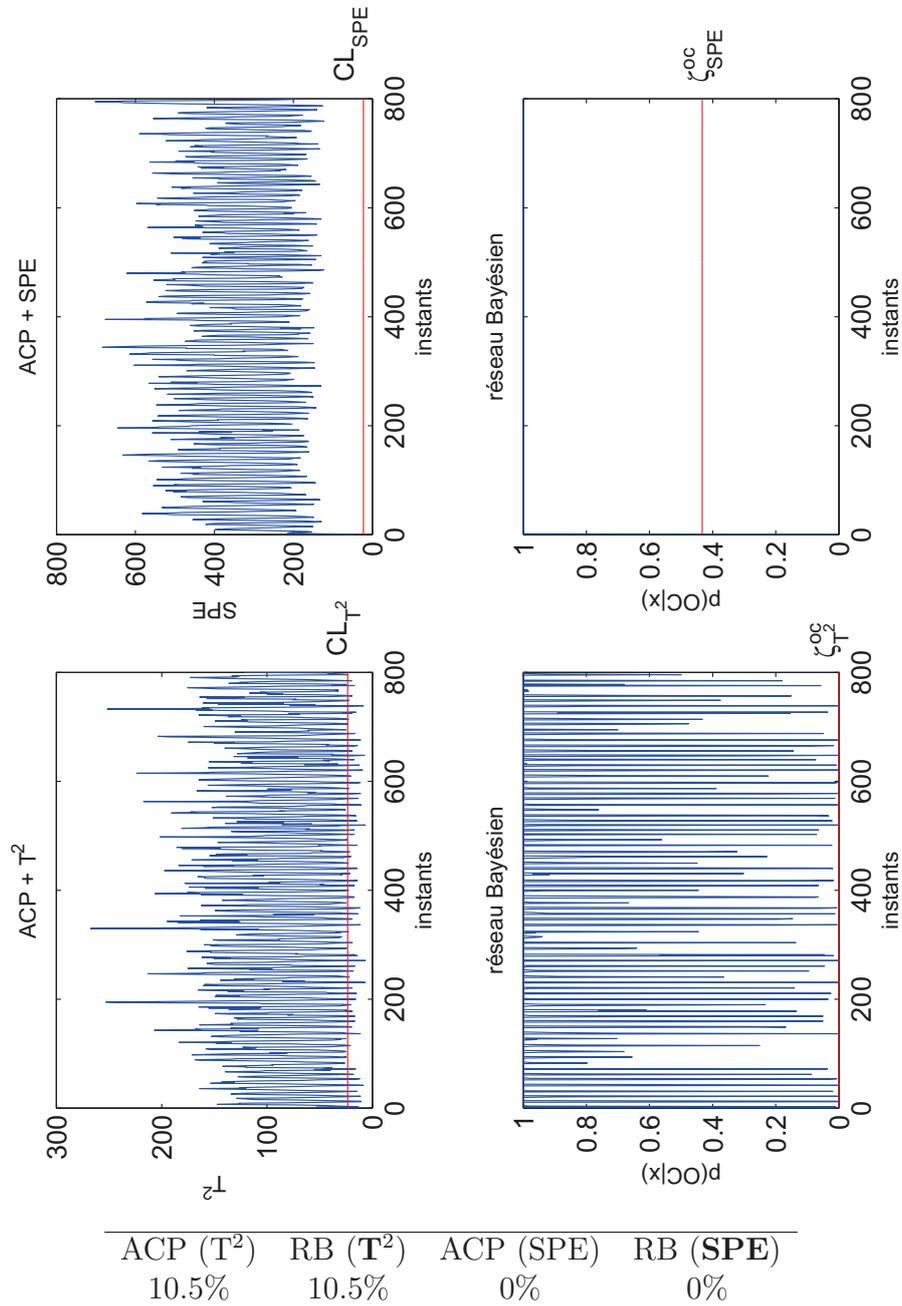
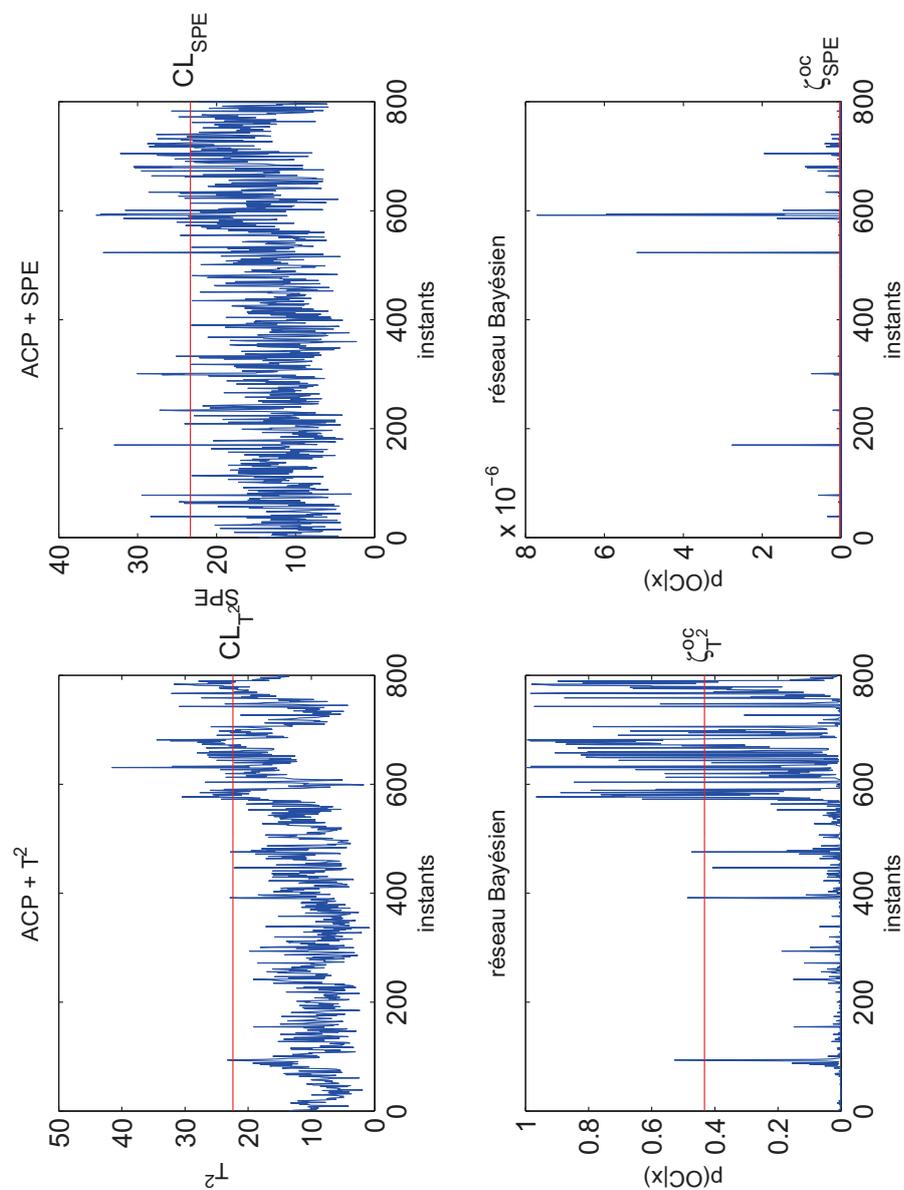
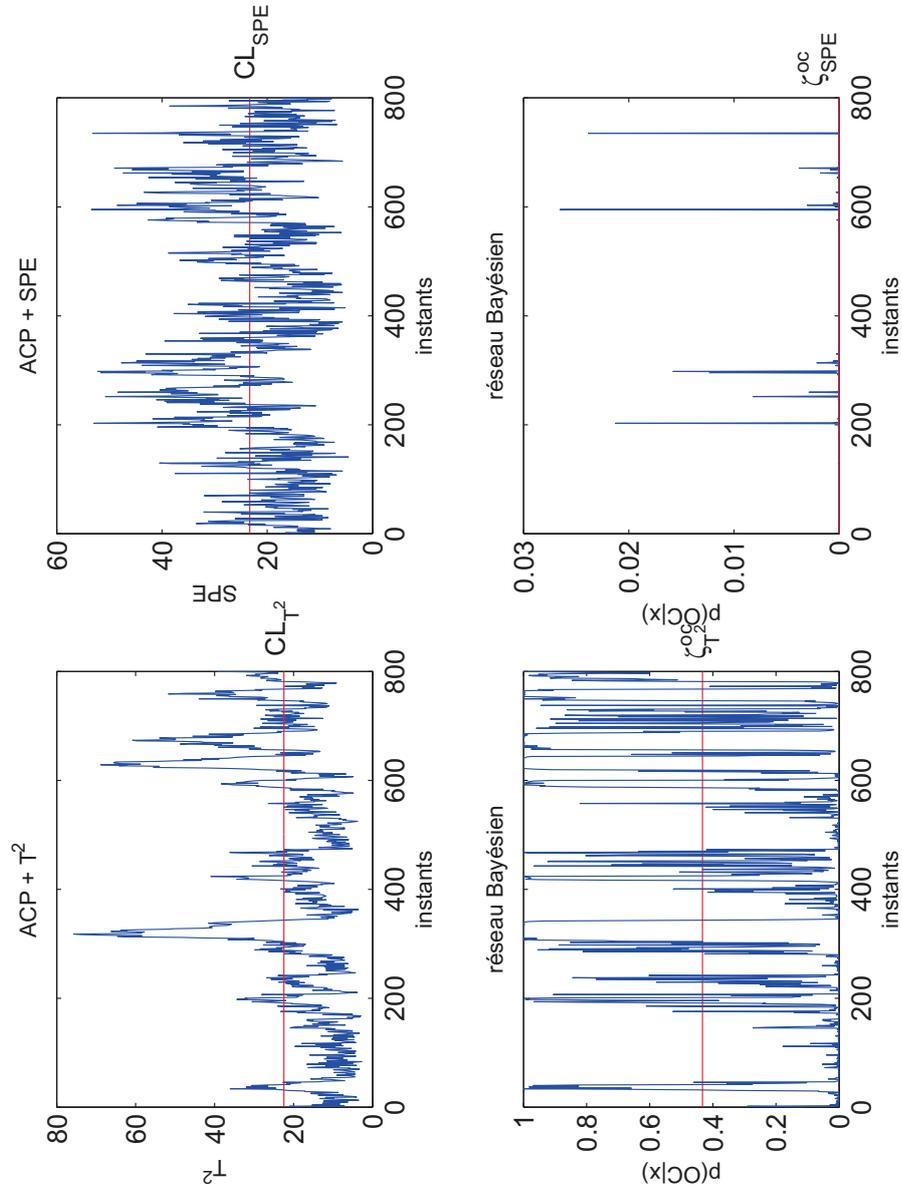


FIGURE A.15 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{14} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs



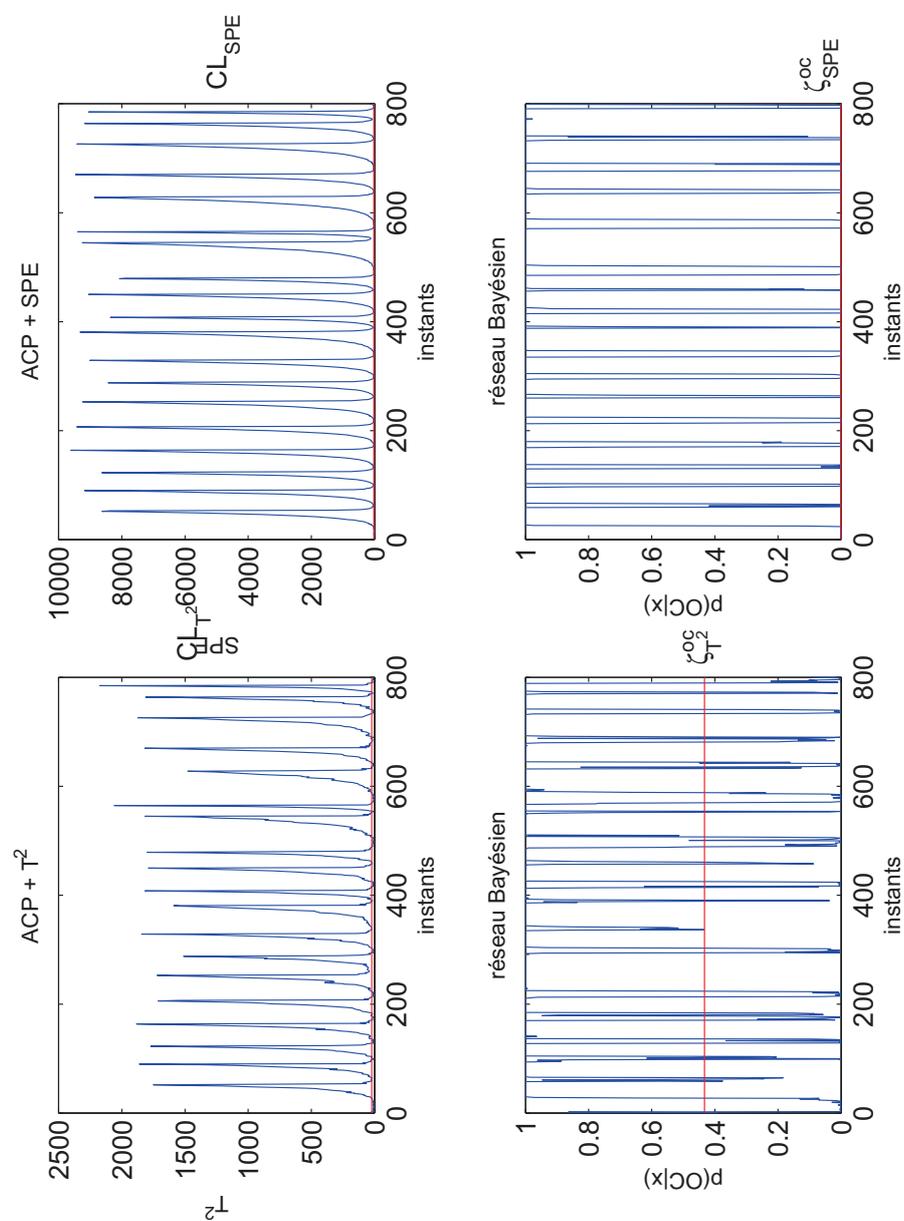
ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
93.375%	93.375%	92.75%	92.75%

FIGURE A.16 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{15} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
70.75%	70.75%	61.125%	61.125%

FIGURE A.17 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{16} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
20.75%	20.75%	5.875%	5.875%

FIGURE A.18 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{17} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs

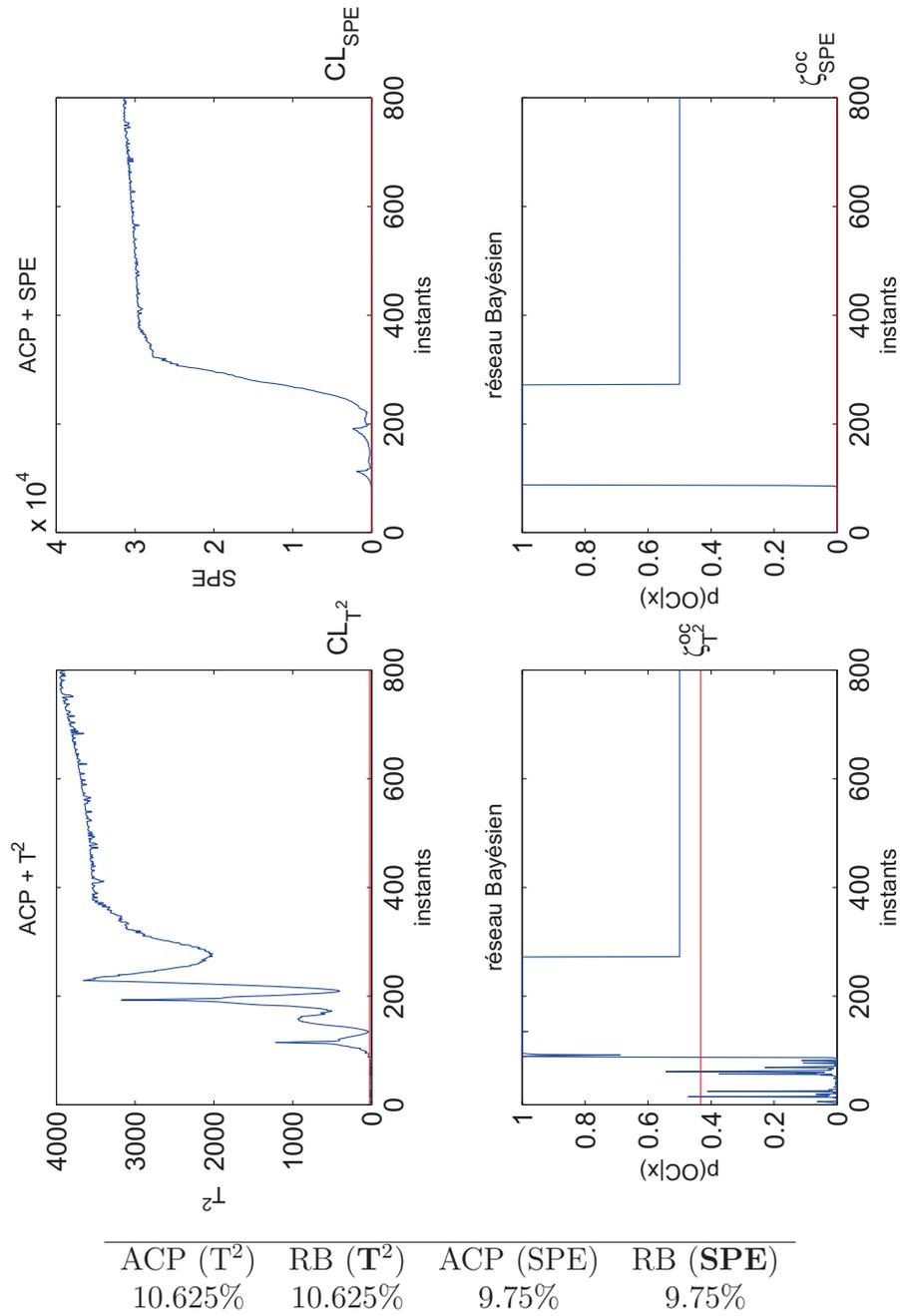
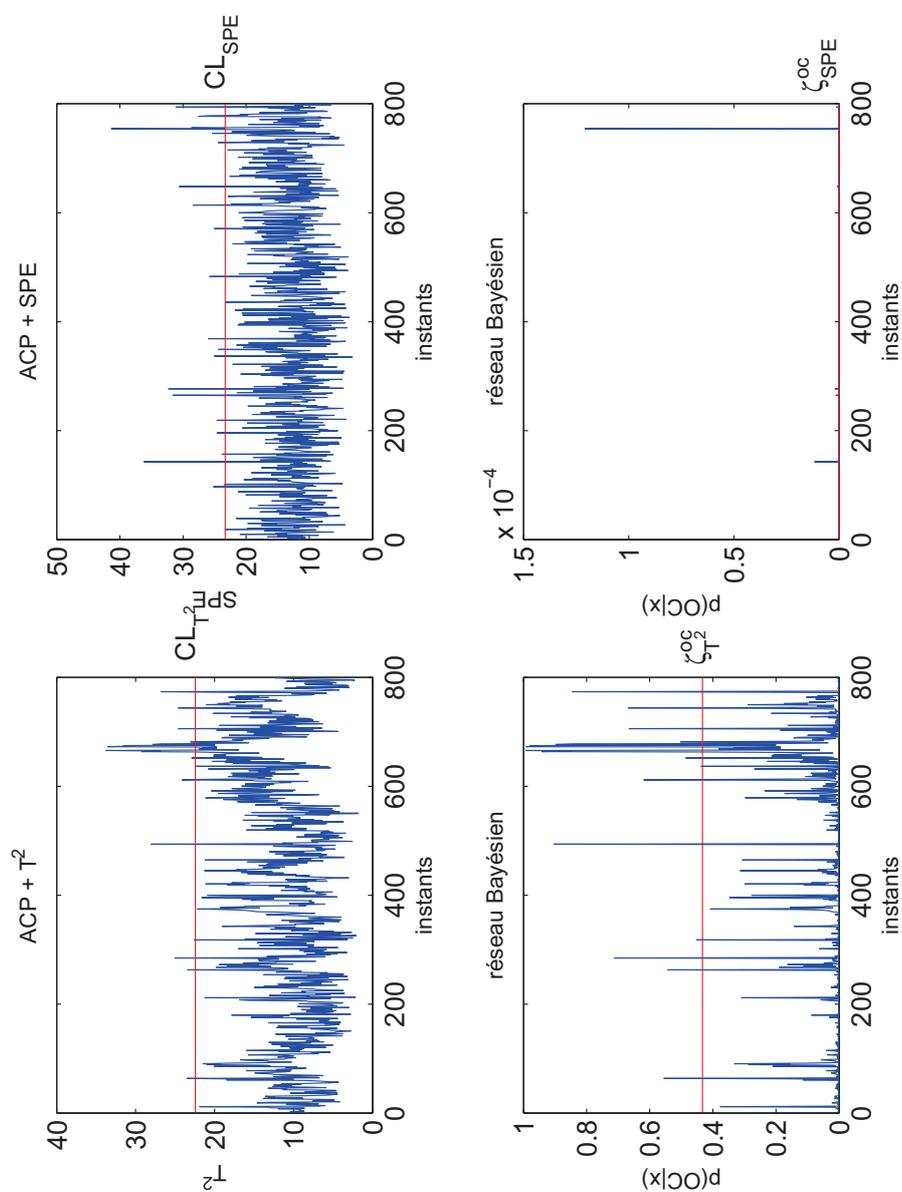
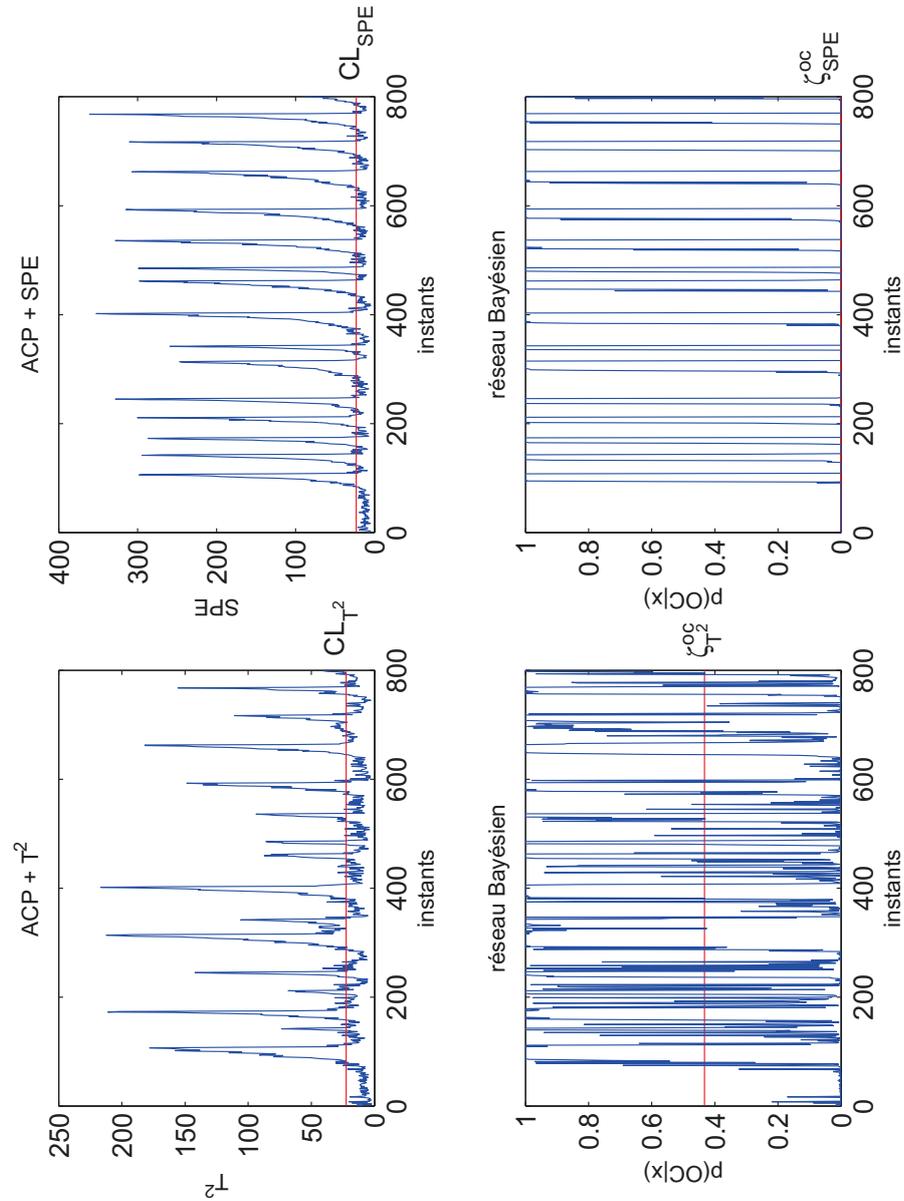


FIGURE A.19 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{18} (avec $\mathbf{c}_\Delta = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
98.125%	98.125%	58.875%	58.875%

FIGURE A.20 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{19} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs



ACP (T^2)	RB (T^2)	ACP (SPE)	RB (SPE)
59%	59%	45.125%	45.125%

FIGURE A.21 – Comparaison entre le schéma de détection par ACP et le réseau Bayésien proposé par rapport à la classe F_{20} (avec $c_{\Delta} = 10$) : représentation graphique + TMDs

Bibliographie

Bibliographie

- Ahrendt, P. (2005). The multivariate gaussian probability distribution. Technical report, Technical University of Denmark.
- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6) :716–723.
- Aradhye, H. B. (1997). *Sensor fault detection, isolation, and accommodation using neural networks, fuzzy logic, and bayesian belief networks*. M.S. Diss., University of New Mexico, Albuquerque, NM.
- Atoui, M. A., Verron, S., and Kobi, A. (2014a). Conditional gaussian network as pca for fault detection. In *World Congress*, pages 1935–1940.
- Atoui, M. A., Verron, S., and Kobi, A. (2014b). Pca in a bayesian framework for fault detection. In *Control & Automation (ICCA), 11th IEEE International Conference on*, pages 354–359. IEEE.
- Atoui, M. A., Verron, S., and Kobi, A. (2015a). A bayesian network dealing with measurements and residuals for system monitoring. *Transactions of the Institute of Measurement and Control*, page 0142331215581446.
- Atoui, M. A., Verron, S., and Kobi, A. (2015b). Fault detection and diagnosis in a bayesian network classifier incorporating probabilistic boundary 1. *IFAC-PapersOnLine*, 48(21) :670–675.
- Atoui, M. A., Verron, S., and Kobi, A. (2015c). Fault detection with conditional gaussian network. *Engineering Applications of Artificial Intelligence*, 45 :473–481.

- Baek, J., McLachlan, G. J., and Flack, L. K. (2010). Mixtures of factor analyzers with common factor loadings : Applications to the clustering and visualization of high-dimensional data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(7) :1298–1309.
- Baïkeche, H. (2007). *Diagnostic des systèmes linéaires en boucle fermée*. PhD thesis, Institut National Polytechnique de Lorraine-INPL.
- Barber, D. (2003). Probabilistic modelling and reasoning : The junction tree algorithm. *Course notes*, 2004.
- Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.
- Basseville, M., Nikiforov, I. V., et al. (1993). *Detection of abrupt changes : theory and application*, volume 104. Prentice Hall Englewood Cliffs.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.
- Blimes (2004). On soft evidence in bayesian networks. Technical report, Technical Report UWEE-TR-2004-0016, University of Washington, Dept. of EE.
- Bouveyron, C. (2013). Probabilistic model-based discriminant analysis and clustering methods in chemometrics. *Journal of Chemometrics*, 27(12) :433–446.
- Bouveyron, C. and Brunet, C. (2012). Probabilistic fisher discriminant analysis : A robust and flexible alternative to fisher discriminant analysis. *Neurocomputing*, 90 :12–22.
- Bouveyron, C., Girard, S., and Schmid, C. (2007). High-dimensional discriminant analysis. *Communications in Statistics-Theory and Methods*, 36(14) :2607–2623.
- Byun, H. and Lee, S.-W. (2002). Applications of support vector machines for pattern recognition : A survey. In *Pattern recognition with support vector machines*, pages 213–236. Springer.
- Celeux, G. and Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern recognition*, 28(5) :781–793.

- Chamroukhi, F. (2010). *Hidden process regression for curve modeling, classification and tracking*. PhD thesis, Université de Technologie de Compiègne.
- Chen, T. and Sun, Y. (2009). Probabilistic contribution analysis for statistical process monitoring : A missing variable approach. *Control Engineering Practice*, 17(4) :469–477.
- Chiang, L., Russel, E., and Braatz, R. D. (2001). *Fault Detection and Diagnosis in Industrial Systems*. Springer.
- Chiang, L. H., Kotanchek, M. E., and Kordon, A. K. (2004). Fault diagnosis based on fisher discriminant analysis and support vector machines. *Computers and Chemical Engineering*, 28(8) :1389 – 1401.
- Chien, C.-F., Chen, S.-L., and Lin, Y.-S. (2002). Using bayesian network for fault location on distribution feeder. *Power Delivery, IEEE Transactions on*, 17(3) :785–793.
- Cobb, B. R., Rumi, R., and Salmerón, A. (2007). Bayesian network models with discrete and continuous variables. In *Advances in probabilistic graphical models*, pages 81–102. Springer.
- Cover, T. M. and Hart, P. E. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1) :21–27.
- Cowell, R. G. (2005). Local propagation in conditional gaussian bayesian networks. *The Journal of Machine Learning Research*, 6 :1517–1550.
- Dash, M. and Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, 1(1) :131–156.
- Daszykowski, M. (2007). From projection pursuit to other unsupervised chemometric techniques. *Journal of Chemometrics*, 21(7-9) :270–279.
- Deventer, R. (2004). *Modeling and control of static and dynamic systems with Bayesian networks*. PhD thesis, Universitat Erlangen-Nurnberg.
- Dey, S. and Stori, J. (2005). A bayesian network approach to root cause diagnosis of process variations. *International Journal of Machine Tools and Manufacture*, 45(1) :75–91.

- Ding, S. (2011). A survey of the application of basic data-driven and model-based methods in process monitoring and fault diagnosis. *18th IFAC World Congress, Milano, Italy*.
- Ding, S. (2014). Data-driven design of monitoring and diagnosis systems for dynamic processes : A review of subspace technique based schemes and some recent results. *Journal of Process Control*, 24(2) :431–449.
- Ding, S., Zhang, P., Naik, A., Ding, E., and Huang, B. (2009). Subspace method aided data-driven design of fault detection and isolation systems. *Journal of process control*, 19(9) :1496–1510.
- Ding, S. X. (2008). *Model-Based Fault Diagnosis Techniques : Design Schemes, Algorithms and Tools*. Springer.
- Ding, S. X. (2012). Data-driven design of model-based fault diagnosis systems. In *Proceedings of IFAC ADCHEM*, pages 10–13.
- Ding, X., Guo, L., and Jeansch, T. (1999). A characterization of parity space and its application to robust fault detection. *IEEE Transactions on Automatic Control*, 44(2) :337–343.
- Ding, Y. (2010). Fault diagnosis in multivariate manufacturing processes. *Electronic Theses, Treatises and Dissertations*.
- Dong, J. and Verhaegen, M. (2009). Subspace based fault detection and identification for lti systems. In *Fault Detection, Supervision and Safety of Technical Processes*, pages 330–335.
- Downs, J. J. and Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 17(3) :245–255.
- Dubuisson, B. (2001). *Diagnostic, intelligence artificielle et reconnaissance des formes*. Hermès science publications.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *pattern classification 2nd edition*. Wiley.
- Favoreel, W., De Moor, B., and Van Overschee, P. (2000). Subspace state space system identification for industrial processes. *Journal of process control*, 10(2) :149–155.

- Fellouah, R. (2007). *Contribution au diagnostic de pannes pour les systèmes différentiellement plats*. PhD thesis, INSA de Toulouse.
- Fossard, A. and Biannic, J.-M. (2006). Représentations d'un système linéaire. *Techniques de l'ingénieur. Informatique industrielle*, 1(S7130).
- Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, 29(2) :131–163.
- Fukunaga, K. (1990). *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA.
- Gertler, J. (1997). Fault detection and isolation using parity relations. *Control engineering practice*, 5(5) :653–661.
- Gertler, J., Li, W., Huang, Y., and McAvoy, T. (1999). Isolation enhanced principal component analysis. *AIChE Journal*, 45(2) :323–334.
- Ghahramani, Z., Hinton, G. E., et al. (1996). The em algorithm for mixtures of factor analyzers. Technical report, Technical Report CRG-TR-96-1, University of Toronto.
- Ghosh, K., Ng, Y. S., and Srinivasan, R. (2011). Evaluation of decision fusion strategies for effective collaboration among heterogeneous fault diagnostic methods. *Computers & Chemical Engineering*, 35(2) :342 – 355.
- Gonzalez, R., Huang, B., Xu, F., and Espejo, A. (2012). Dynamic bayesian approach to gross error detection and compensation with application toward an oil sands process. *Chemical Engineering Science*, 67(1) :44 – 56.
- Hagenblad, A., Gustafsson, F., and Klein, I. (2004). A comparison of two methods for stochastic fault detection : the parity space approach and principal component analysis. In *Reglermöte 2004, Göteborg, Sweden, May, 2004*.
- Heckerman, D. and Shwe, M. (1993). Diagnosis of multiple faults : a sensitivity analysis. In *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*, pages 80–87. Morgan Kaufmann Publishers Inc.
- Heger, A. S. and Aradhye, H. B. (2002). Bayesian belief networks for industrial processes. US Patent 6,415,276.

- Hood, C. S. and Ji, C. (1997). Proactive network-fault detection [telecommunications]. *Reliability, IEEE Transactions on*, 46(3) :333–341.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6) :417.
- Huang, B. (2008). Bayesian methods for control loop monitoring and diagnosis. *Journal of process control*, 18(9) :829–838.
- Huang, Y., McMurrin, R., Dhadyalla, G., and Jones, R. P. (2008). Probability based vehicle fault diagnosis : Bayesian network method. *Journal of Intelligent Manufacturing*, 19(3) :301–311.
- Isermann, R. (2004). Model-based fault detection and diagnosis : status and applications. In *In Proceedings of the 16th IFAC Symposium on Automatic Control in Aerospace, St*, pages 71–85.
- Isermann, R. (2006). *Fault-diagnosis systems : an introduction from fault detection to fault tolerance*. Springer.
- Isermann, R. and Ballé, P. (1997). Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5(5) :709 – 719.
- Jackson, J. E. (2005). *A user's guide to principal components*, volume 587. Wiley. com.
- Jackson, J. E. and Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, 21(3) :341–349.
- Jaulin, L. (2005). *Représentation d'état pour la modélisation et la commande des systèmes*. Hermès Science publications.
- Jebara, T. (2001). *Discriminative, generative and imitative learning*. PhD thesis, Massachusetts Institute of Technology.
- Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Jensen, F. V. and Nielsen, T. D. (2007). *Bayesian networks and decision graphs*. Springer.

- Jiao, P., Wang, X., and Li, C. (2012). Diagnosis method for analog circuits fault using bayesian network. In *2nd International Conference on Electronic and Mechanical Engineering and Information Technology*.
- Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library.
- Kawahara, Y., Fujimaki, R., Yairi, T., and Machida, K. (2005). Diagnosis method for spacecraft using dynamic bayesian networks. In *'i-SAIRAS 2005'-The 8th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, volume 603, page 85.
- Kempowski, T. (2004). *Surveillance de procédés à base de méthodes de classification : conception d'un outil d'aide pour la détection et le diagnostic des défaillances*. PhD thesis, INSA de Toulouse.
- Kim, D. and Lee, I.-B. (2003). Process monitoring based on probabilistic pca. *Chemometrics and intelligent laboratory systems*, 67(2) :109–123.
- Kjærulff, U. (1990). Triangulation of graphs—algorithms giving small total state space. *Technical Report R-90-09, Dept. of Math. and Comp. Sci., Aalborg Univ., Denmark*.
- Kjærulff, U. B. and Madsen, A. L. (2005). Probabilistic networks—an introduction to bayesian networks and influence diagrams. *Aalborg University*.
- Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1) :273–324.
- Koller, D. and Pfeffer, A. (1997). Object-oriented bayesian networks. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 302–313. Morgan Kaufmann Publishers Inc.
- Kulcsar, B., Dong, J., and Verhaegen, M. (2009). Model-free fault tolerant control approach for linear parameter varying systems. In *Fault Detection, Supervision and Safety of Technical Processes*, pages 876–881.
- Larimore, W. E. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In *Decision and Control, 1990., Proceedings of the 29th IEEE Conference on*, pages 596–604. IEEE.

- Lauritzen, S. L. (1992). Propagation of probabilities, means, and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87(420) :1098–1108.
- Lauritzen, S. L. and Jensen, F. (2001). Stable local computation with conditional gaussian distributions. *Statistics and Computing*, 11(2) :191–203.
- Le Rohellec, F. (1998). *Identification des matrices physiques dans l'espace état. Recombinaison de sous-représentations*. PhD thesis, Montpellier 2.
- Leimer, H.-G. (1988). Triangulated graphs with marked vertices. *Annals of Discrete Mathematics*, 41 :311–324.
- Lerner, U., Parr, R., Koller, D., and Biswas, G. (2000). Bayesian fault detection and diagnosis in dynamic systems. In *AAAI/IAAI*, pages 531–537.
- Lerner, U., Segal, E., and Koller, D. (2001). Exact inference in networks with discrete children of continuous parents. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 319–328. Morgan Kaufmann Publishers Inc.
- Li, J., Jin, J., and Shi, J. (2008). Causation-based t2 decomposition for multivariate process monitoring and diagnosis. *Journal of Quality Technology*, 40(1) :46.
- Li, J. and Shi, J. (2007). Knowledge discovery from observational data for process control using causal bayesian networks. *IIE transactions*, 39(6) :681–690.
- Li, S., Gao, J., Nyagilo, J. O., and Dave, D. P. (2011). Probabilistic partial least square regression : a robust model for quantitative analysis of raman spectroscopy data. In *Bioinformatics and Biomedicine (BIBM), 2011 IEEE International Conference on*, pages 526–531. IEEE.
- Lieftucht, D., Völker, M., Sonntag, C., Kruger, U., Irwin, G. W., and Engell, S. (2009). Improved fault diagnosis in multivariate systems using regression-based reconstruction. *Control Engineering Practice*, 17(4) :478–493.
- Ljung, L. (1998). *System Identification : Theory for the User*. Pearson Education.

- Lo, C., Wong, Y., and Rad, A. (2003). Bayesian network for fault diagnosis. In *European Control Conference. IEEE*.
- Lo, C., Wong, Y., and Rad, A. B. (2011). Bond graph based bayesian network for fault diagnosis. *Applied soft computing*, 11(1) :1208–1212.
- Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1) :46–53.
- Luo, J., Namburu, S. M., Pattipati, K. R., Qiao, L., and Chigusa, S. (2010). Integrated model-based and data-driven diagnosis of automotive antilock braking systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 40(2) :321–336.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition : the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7) :674–693.
- Matsuura, J. P. and Yoneyama, T. (2004). Learning bayesian networks for fault detection. In *Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop*, pages 133–142. IEEE.
- Matsuura, J. P., Yoneyama, T., and Galvão, R. K. H. (2005). Learning bayesian networks for fault detection : application to the 747 longitudinal motion. *18th Int. Cong. Of Mechanical Engineering*.
- McKelvey, T. (1995). *Identification of state-space models from time and frequency data*. Department of Electrical Engineering, Linköping University.
- McLachlan, G. (2004). *Discriminant analysis and statistical pattern recognition*, volume 544. John Wiley & Sons.
- Menicholas, P. D. and Murphy, T. B. (2008). Parsimonious gaussian mixture models. *Statistics and Computing*, 18(3) :285–296.
- Mechraoui, A., Medjaher, K., and Zerhouni, N. (2008). Bayesian based fault diagnosis : application to an electrical motor. In *17th IFAC World Congress*, volume 400, pages 6–pages.

- Mehranbod, N., Soroush, M., and Panjapornpon, C. (2005). A method of sensor fault detection and identification. *Journal of Process Control*, 15(3) :321 – 339.
- Mkhadri, A., Celeux, G., and Nasroallah, A. (1997). Regularization in discriminant analysis : an overview. *Computational Statistics & Data Analysis*, 23(3) :403–423.
- Mnassri, B. (2012). *Analyse de données multivariées et surveillance des processus industriels par analyse en composantes principales*. PhD thesis, Aix-Marseille.
- Montanari, A. and Viroli, C. (2010). Heteroscedastic factor mixture analysis. *Statistical Modelling*, 10(4) :441–460.
- Murphy, K. P. (1998). *Inference and learning in hybrid bayesian networks*. University of California, Berkeley, Computer Science Division.
- Murphy, K. P. (2002). *Dynamic bayesian networks : representation, inference and learning*. PhD thesis, University of California.
- Najafi, M., Auslander, D. M., Bartlett, P. L., Haves, P., and Sohn, M. D. (2012). Application of machine learning in the fault diagnostics of air handling units. *Applied Energy*, 96 :347–358.
- Neil, M., Fenton, N., Forey, S., and Harris, R. (2001). Using bayesian belief networks to predict the reliability of military vehicles. *Computing & Control Engineering Journal*, 12(1) :11–20.
- Nielsen, T. D. and Jensen, F. V. (2009). *Bayesian networks and decision graphs*. Springer Science & Business Media.
- Nunnari, G., Cannavó, F., and Vrânceanu, R. (2006). Bayesian networks approach for a fault detection and isolation case study. In *Applied Soft Computing Technologies : The Challenge of Complexity*, pages 173–183. Springer.
- Odgaard, P. F., Stoustrup, J., and Kinnaert, M. (2013). Fault-tolerant control of wind turbines : A benchmark model. *Control Systems Technology, IEEE Transactions on*, 21(4) :1168–1182.

- Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial intelligence*, 29(3) :241–288.
- Pernestål, A. (2009). *Probabilistic fault diagnosis with automotive applications*. PhD thesis, Linköping University.
- Pernestål, A., Nyberg, M., and Wahlberg, B. (2006). A bayesian approach to fault isolation with application to diesel engine diagnosis. In *17th Workshop on Principles for Diagnosis, Peñaranda de Duero, Burgos, Spain, June 26-28, 2006*, pages 211–218.
- Przytula, K. W. and Thompson, D. (2000). Construction of bayesian networks for diagnostics. In *Aerospace Conference Proceedings, 2000 IEEE*, volume 5, pages 193–200. IEEE.
- Qin, S. J. (2006). An overview of subspace identification. *Computers & chemical engineering*, 30(10) :1502–1513.
- Qin, S. J. and Li, W. (2001). Detection and identification of faulty sensors in dynamic processes. *AIChE Journal*, 47(7) :1581–1593.
- Riascos, L. A. M., Simoes, M. G., and Miyagi, P. E. (2007). A bayesian network fault diagnostic system for proton exchange membrane fuel cells. *Journal of power sources*, 165(1) :267–278.
- Rojas-Guzman, C. and Kramer, M. A. (1993). Comparison of belief networks and rule-based expert systems for fault diagnosis of chemical processes. *Engineering Applications of Artificial Intelligence*, 6(3) :191–202.
- Roweis, S. and Ghahramani, Z. (1999). A unifying review of linear gaussian models. *Neural computation*, 11(2) :305–345.
- Roweis, S. T. (1998). Em algorithms for pca and spca. In *Advances in Neural Information Processing Systems*, pages 626–632.
- Roychoudhury, I., Biswas, G., and Koutsoukos, X. (2006). A bayesian approach to efficient diagnosis of incipient faults. In *17th International Workshop on Principles of Diagnosis DX*, pages 243–264. Citeseer.
- Ruz-Hernandez, J. A., Sanchez, E. N., and Suarez, D. A. (2007). Optimal training for associative memories : application to fault diagnosis in fossil electric power plants. In *Hybrid Intelligent Systems*, pages 329–356. Springer.

- Samantaray, A. K. and Bouamama, B. O. (2008). *Model-based process supervision : a bond graph approach*. Springer Science & Business Media.
- Santoso, N. I., Darken, C., Povh, G., and Erdmann, J. (1999). Nuclear plant fault diagnosis using probabilistic reasoning. In *Power Engineering Society Summer Meeting, 1999. IEEE*, volume 2, pages 714–719. IEEE.
- Schubert, U., Kruger, U., Arellano-Garcia, H., de Sá Feital, T., and Wozny, G. (2011). Unified model-based fault diagnosis for three industrial application studies. *Control Engineering Practice*, 19(5) :479–490.
- Schwall, M. L. and Gerdes, J. C. (2002). A probabilistic approach to residual processing for vehicle fault detection. In *American Control Conference, 2002. Proceedings of the 2002*, volume 3, pages 2552–2557.
- Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2) :461–464.
- Sheppard, J. W., Kaufman, M., et al. (2005). Bayesian diagnosis and prognosis using instrument uncertainty. In *Autotestcon, 2005. IEEE*, pages 417–423. IEEE.
- Simani, S., Fantuzzi, C., and Patton, R. J. (2013). *Model-based fault diagnosis in dynamic systems using identification techniques*. Springer Science & Business Media.
- Testik, M. C. and Borrór, C. M. (2004). Design strategies for the multivariate exponentially weighted moving average control chart. *Quality and Reliability Engineering International*, 20(6) :571–577.
- Theilliol, D. (2003). Contribution à l'étude et au développement des systèmes tolérants aux défauts : diagnostic et accommodation à base de modèles linéaires et au-delà. *Habilitation á Diriger des Recherches, Université Henri Poincaré, Nancy, 1*.
- Thomaz, C., Duncan, G. F., and Raul, F. Q. (2004). A new covariance estimate for bayesian classifiers in biometric recognition. *IEEE Transactions on Circuit and systemes for video technology*, 14.
- Tiplica, T. (2012). Contributions au contrôle des processus complexes. *Habilitation á Diriger des Recherches, Université d'Angers*.

- Tipping, M. E. and Bishop, C. M. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 61(3) :611–622.
- Tracy, N., Young, J., and Mason, R. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, 24(2).
- Treasure, R. J., Kruger, U., and Cooper, J. E. (2004). Dynamic multivariate statistical process control using subspace identification. *Journal of Process Control*, 14(3) :279 – 292.
- Valle, S., Li, W., and Qin, S. J. (1999). Selection of the number of principal components : the variance of the reconstruction error criterion with a comparison to other methods. *Industrial & Engineering Chemistry Research*, 38(11) :4389–4401.
- Van Overschee, P. and De Moor, B. (1996). *Subspace Identification for linear systems : Theory - Implementation - Applications*. Kluwer Academic Publishers.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K. (2003). A review of process fault detection and diagnosis : Part iii : Process history based methods. *Computers & Chemical Engineering*, 27(3) :327 – 346.
- Verhaegen, M. and Dewilde, P. (1992). Subspace model identification part 1. the output-error state-space model identification class of algorithms. *International journal of control*, 56(5) :1187–1210.
- Verron, S. (2007). *Diagnostic et surveillance des processus complexes par réseaux bayésiens*. PhD thesis, Université d’Angers.
- Verron, S., Tiplica, T., and Kobi, A. (2008a). Fault detection and identification with a new feature selection based on mutual information. *Journal of Process Control*, 18(5) :479–490.
- Verron, S., Tiplica, T., and Kobi, A. (2008b). Fault detection with bayesian network. In *Frontiers in Robotics, Automation and Control*, pages 341–356. InTech Education and Publishing.

- Verron, S., Tiplica, T., and Kobi, A. (2010a). Fault diagnosis of industrial systems by conditional gaussian network including a distance rejection criterion. *Engineering Applications of Artificial Intelligence*, 23(7) :1229 – 1235.
- Verron, S., Tiplica, T., and Kobi, A. (2010b). Fault diagnosis of industrial systems by conditional gaussian network including a distance rejection criterion. *Engineering applications of artificial intelligence*, 23(7) :1229–1235.
- Wang, J. and Qin, S. J. (2002). A new subspace identification approach based on principal component analysis. *Journal of Process Control*, 12(8) :841–855.
- Wang, Y., Ma, G., Ding, S. X., and Li, C. (2011). Subspace aided data-driven design of robust fault detection and isolation systems. *Automatica*, 47(11) :2474–2480.
- Weber, P., Theilliol, D., and Aubrun, C. (2008). Component reliability in fault diagnosis decision-making based on dynamic bayesian networks. *Proceedings of the Institution of Mechanical Engineers Part O Journal of Risk and Reliability*, 222(2) :161–172. SURFDIAG.
- Yang, L. and Lee, J. (2012). Bayesian belief network-based approach for diagnostics and prognostics of semiconductor manufacturing systems. *Robotics and Computer-Integrated Manufacturing*, 28(1) :66–74.
- Yin, S., Ding, S. X., Haghani, A., Hao, H., and Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of Process Control*, 22(9) :1567 – 1581.
- Yoshida, R., Higuchi, T., and Imoto, S. (2004). A mixed factors model for dimension reduction and extraction of a group structure in gene expression data. In *Computational Systems Bioinformatics Conference, 2004. CSB 2004. Proceedings. 2004 IEEE*, pages 161–172. IEEE.
- Yoshida, R., Higuchi, T., Imoto, S., and Miyano, S. (2006). Arraycluster : an analytic tool for clustering, data visualization and module finder on gene expression profiles. *Bioinformatics*, 22(12) :1538–1539.

- Yu, J. (2013). A new fault diagnosis method of multimode processes using bayesian inference based gaussian mixture contribution decomposition. *Engineering Applications of Artificial Intelligence*, 26(1) :456–466.
- Yu, J. and Rashid, M. M. (2013). A novel dynamic bayesian network-based networked process monitoring approach for fault detection, propagation identification, and root cause diagnosis. *AIChE Journal*.
- Zaidi, A., Bouamama, B. O., and Tagina, M. (2012). Bond graph and bayesian networks for reliable diagnosis. *International Journal of SIMULATION : Systems, Science & Technology*, 12(5) :1240–1253.
- Zhang, B., Sconyers, C., Byington, C., Patrick, R., Orchard, M. E., and Vachtsevanos, G. (2011). A probabilistic fault detection approach : application to bearing fault detection. *Industrial Electronics, IEEE Transactions on*, 58(5).
- Zhang, G. P. (2000). Neural networks for classification : a survey. *Systems, Man, and Cybernetics, Part C : Applications and Reviews, IEEE Transactions on*, 30(4) :451–462.
- Zwingelstein, G. (1995). *Diagnostic des défaillances, théorie et pratique pour les systèmes industriels*. Ed. HERMES.

Thèse de Doctorat

Mohamed Amine ATOUI

Contributions à la détection et au diagnostic de fautes dans les systèmes par réseaux Bayésiens

Contributions to fault detection and diagnosis in systems by Bayesian networks

Résumé

Les fautes systèmes peuvent conduire à des conséquences sérieuses pour l'humain, l'environnement et le matériel. Or, y remédier peut s'avérer coûteux voire même dangereux. Ainsi, afin d'éviter ces situations, il est devenu essentiel pour les systèmes complexes modernes de détecter et d'identifier tout changement dans leur fonctionnement nominal avant que cela ne devienne critique. De ce fait, plusieurs méthodes de détection et de diagnostic ont été proposées ou améliorées durant les dernières décennies. Parmi ces méthodes, celles présentant un fort intérêt se basent sur un outil statistique et probabiliste nommé réseau Bayésien. Toutefois, la majorité d'entre elles ne tiennent pas compte du risque de fausse alarme dans leur prise de décision. L'intérêt de cette thèse est alors d'introduire sous réseau Bayésien des limites probabilistes permettant le respect d'un niveau de signification considéré. Plus exactement, nous proposons une modélisation des statistiques quadratiques et les limites leurs correspondant sur réseau Bayésien. Ceci nous permet de généraliser sous réseau Bayésien des schémas de détection de fautes comme par exemple ceux basés sur l'analyse en composantes principale. Cette modélisation nous permet également de proposer une famille de réseaux Bayésiens permettant de faire de la détection et du diagnostic de façon simultanée, tout en tenant compte d'un rejet de distance. Enfin, nous proposons un cadre probabiliste permettant d'unifier les différents réseaux Bayésiens pouvant être utilisés pour la détection ou le diagnostic de fautes.

Mots clés

détection et diagnostic de fautes, réseaux Bayésiens, réseaux conditionnels Gaussiens, statistiques, classification supervisée

Abstract

Systems failures can potentially lead to serious consequences for human, environment and material, and sometimes fixing them could be expensive and even dangerous. Thus, in order to avoid these undesirable situations, it becomes very important and essential for modern complex systems to detect and identify any changes in their nominal operations before they become critical. To do so, several detection and diagnosis methods have been proposed or enhanced during the last decades. Among these methods, those with a great interest are based on a statistical and probabilistic tool named Bayesian network. However, the majority of these methods do not handle the risk of false alarm in their decision-making. The interest of this thesis is to introduce, under Bayesian network, probabilistic limits able to respect a given significance level. More precisely, we propose to model the quadratic statistics and their limits in Bayesian network. This allows us to generalize under Bayesian network fault detection schemes as those associated to the principal component analysis. This modeling allows us also to propose a family of Bayesian networks that can make detection and diagnosis simultaneously, while taking into account the distance rejection. Finally, we propose a probabilistic framework able to unify different BNs dedicated to the detection or diagnosis of systems faults.

Key Words

faults detection and diagnosis, Bayesian networks, conditional Gaussian networks, statistics, supervised classification