



HAL
open science

Détection de Fautes par Réseaux Bayésiens dans les Procédés Multivariés

Sylvain Verron, Teodor Tiplica, Abdessamad Kobi

► **To cite this version:**

Sylvain Verron, Teodor Tiplica, Abdessamad Kobi. Détection de Fautes par Réseaux Bayésiens dans les Procédés Multivariés. D. Lefebvre, H. Chafouk, A. El Hami, M. Bennouna. Diagnostic des systèmes complexes, D. Lefebvre, H. Chafouk, A. El Hami, M. Bennouna, 2008, 9954 - 8992 - 0 - 0. inria-00517065

HAL Id: inria-00517065

<https://inria.hal.science/inria-00517065>

Submitted on 13 Sep 2010

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.

Détection de Fautes par Réseaux Bayésiens dans les Procédés Multivariés

Sylvain Verron — Teodor Tiplica — Abdessamad Kobi

LASQUO/ISTIA
Université d'Angers
62, Avenue Notre Dame du Lac
49000 Angers
{sylvain.verron,tiplica,kobi}@istia.univ-angers.fr

RÉSUMÉ. L'objectif de cet article est de présenter une méthode permettant la détection de fautes d'un procédé multivarié, au moyen d'un réseau bayésien. Pour ce faire, la détection est assimilée à une tâche de classification telle que l'analyse discriminante, cette dernière étant aisément transposable en réseau bayésien. Nous prouvons mathématiquement, dans cet article, l'équivalence entre les méthodes de détection usuelles que sont les cartes de contrôle multivariées (cartes T^2 de Hotelling et MEWMA) et l'analyse discriminante quadratique (modélisée par réseau bayésien), rendant ainsi possible la détection de fautes directement par le biais du réseau bayésien.

ABSTRACT. The purpose of this article is to present a method for the fault detection in multivariate process, with a bayesian network. In this context, the detection is viewed as a classification task like the discriminant analysis, which can be transposed in a bayesian network. We prove mathematically the equivalence between the usual detection methods that are the multivariate control charts (Hotelling's T^2 , MEWMA) and the quadratic discriminant analysis (in a bayesian network). So, this makes possible the fault detection with a bayesian network.

MOTS-CLÉS : MSP, multivariés, détection, réseau bayésien, T^2 , MEWMA, Analyse Discriminante
KEYWORDS: SPC, multivariate, detection, bayesian network, T^2 , MEWMA, Discriminant Analysis.

1. Introduction

Dans le domaine de la surveillance des procédés, on distingue généralement deux grandes phases : la détection et le diagnostic. La détection a pour but de statuer sur la présence ou non d'une faute (anomalie) dans le procédé, alors que le diagnostic a pour but d'établir la ou les causes de cette anomalie. Cependant, la littérature assimile bien souvent le terme diagnostic à la combinaison de la détection et du diagnostic. Dans la littérature, on retrouve principalement deux grandes familles de méthodes (Zwingelstein, 1995) : les méthodes internes de diagnostic (se basant sur un modèle physique du procédé) et les méthodes externes de diagnostic (se basant uniquement sur un historique des données du procédé). Parmi les méthodes de diagnostic externe, spécialisées dans la phase de détection, on trouve généralement les cartes de contrôles. Celles-ci sont très utilisées en industrie et ont pour but de surveiller le centrage ou la dispersion du procédé. Cependant, les cartes de contrôle classiques : \bar{X} , R, S, EWMA (Exponentially Weighted Moving Average), CUSUM (CUmulative SUM), ne peuvent suivre qu'une caractéristique à la fois (Montgomery, 1997). Ce sont des cartes de contrôle univariées. Or, de nos jours la qualité des produits est devenue un élément clé de la réussite des entreprises. Celles-ci ne peuvent donc plus se permettre d'approcher la qualité d'un produit par une seule de ses caractéristiques. C'est pourquoi des cartes de contrôle multivariées ont été mises au point, la plus connue étant la carte T^2 de Hotelling (Hotelling, 1947). Ces cartes sont relativement efficaces pour détecter une faute dans un procédé multivarié. La faute peut être un changement de moyenne au niveau d'une ou plusieurs variables. Par exemple, on peut avoir une faute de type saut en échelon (défaillance d'un capteur, changement de matières premières, changement d'opérateur, décalibrage d'un outil de mesure, etc) ou de type dérive (usure d'outil, variation de conditions environnementales, etc). L'inconvénient des cartes de contrôle multivariées est l'absence d'indication concernant la cause d'une faute détectée, ce qui les rend difficilement exploitables en milieu industriel. Afin de diagnostiquer la cause d'une faute, beaucoup d'approches ont été mises au point (Tiplica *et al.*, 2001). Mais, il semble intéressant d'avoir sur un seul et même outil la possibilité de détecter les fautes et de les diagnostiquer. Diverses approches ont été proposées afin de réaliser le diagnostic de situation hors contrôle, on peut par exemple citer les réseaux de neurones, ou bien les approches par reconnaissance de forme (Dubuisson, 2001). Dans cette dernière catégorie, l'utilisation de réseaux bayésiens pour le diagnostic semble une approche intéressante (Verron *et al.*, 2006; Verron *et al.*, 2007). En effet, l'apprentissage de tels réseaux est beaucoup plus rapide que celui de réseaux neuronaux et permet ainsi d'inclure l'apprentissage du système en temps réel. L'utilisation de réseaux bayésiens peut également permettre d'estimer la fiabilité au cours du temps de systèmes dynamiques complexes (Weber *et al.*, 2003; Boudali *et al.*, 2005).

Les réseaux bayésiens semblent donc un outil idéal pour le diagnostic des procédés. Cependant, leur utilisation serait rendue plus intéressante encore si ceux-ci pouvaient effectuer la phase de détection en plus de la phase de diagnostic. On pourrait alors obtenir une synthèse de la détection et du diagnostic de faute dans un seul et même outil : un réseau bayésien. Dans ce but, il est nécessaire de proposer une mé-

thode permettant la détection de faute par l'intermédiaire d'un réseau bayésien. Cet article va permettre de démontrer qu'il est possible de modéliser des cartes de contrôle multivariées directement dans un réseau bayésien.

L'article est structuré de la manière suivante : dans le second paragraphe nous présentons la détection par cartes de contrôle multivariées, l'analyse discriminante, et l'analogie entre ces deux outils ; le troisième paragraphe décrit, suite à une présentation des réseaux bayésiens, la réalisation de l'analyse discriminante par réseau bayésien ; le quatrième paragraphe présente la modélisation des cartes de contrôle multivariées dans un réseau bayésien ; les conclusions et les perspectives sont énoncées dans le dernier paragraphe.

2. Détection et classification

La détection de faute peut s'effectuer au moyen de cartes de contrôle qui permettent de suivre les caractéristiques clés du procédé. Avant de décrire la méthode de détection par réseaux bayésiens, nous allons présenter les cartes les plus connues.

2.1. Cartes de contrôle multivariées

2.1.1. La carte T^2 de Hotelling

Les premiers travaux concernant la carte T^2 datent de 1947 (Hotelling, 1947). Hotelling a été le premier à mettre au point un concept de surveillance de procédé multivarié. Hawkins (Hawkins, 1991) a montré que le principal avantage de cette carte est qu'elle représente le meilleur test statistique pour détecter un dérèglement de la moyenne du procédé. Pour un procédé à p variables, et en utilisant les notations matricielles, on peut écrire le T^2 sous la forme de la distance statistique suivante :

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

où : \mathbf{x} est le vecteur d'observations à un instant donné de taille $1 \times p$, $\boldsymbol{\mu}$ est le vecteur cible du procédé, de taille $1 \times p$, $\boldsymbol{\Sigma}$ est la matrice de variance-covariance du procédé de taille $p \times p$, et le symbole T représente la transposée d'un vecteur ou d'une matrice.

Le calcul du T^2 nous donne un scalaire. On va donc pouvoir représenter cette mesure dans le temps sous la forme d'une carte de contrôle. Bien entendu, à chaque instant, la mesure du T^2 va être comparée à une limite de contrôle supérieure (la limite de contrôle inférieure étant fixée à 0). Le calcul des limites va dépendre de l'estimation des paramètres du procédé. Ainsi, suivant les cas, les limites seront différentes. Pour le détail des calculs des limites de contrôle, nous conseillons la lecture de Montgomery (Montgomery, 1997).

Ainsi, une fois les paramètres ($\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$) et la limite estimés, nous obtenons, pour chaque instant d'échantillonnage, le vecteur d'observations \mathbf{x} . On peut alors calculer la valeur du T^2 et la comparer à la Limite Supérieure de Contrôle (LSC). Si la valeur

du T^2 est inférieure à la LSC, alors le procédé sera déclaré sous contrôle, sinon il sera déclaré hors-contrôle.

2.1.2. La carte MEWMA

La carte MEWMA développée par Lowry (Lowry *et al.*, 1992) est l'extension multivariée de la carte univariée EWMA proposée par Roberts en 1959 (Roberts, 1959). Cette carte est bien adaptée pour le suivi de valeurs individuelles (taille d'échantillon $n = 1$), mais elle est surtout utile pour la détection d'écarts de faibles amplitudes par rapport à la cible. Cette carte est largement employée, mais pas toujours sous cette appellation. En effet, les automaticiens parlent de modèle AR (auto régressif), alors que les électroniciens parle de filtre à réponse impulsionnelle infinie (filtre Ri).

Pour la carte MEWMA (Lowry *et al.*, 1992), il faut calculer de manière récursive pour chaque échantillon la variable \mathbf{y}_t (voir équation 2), avec $\mathbf{y}_0 = \boldsymbol{\mu}$.

$$\mathbf{y}_t = \lambda \mathbf{x}_t + (\mathbf{I} - \lambda) \mathbf{y}_{t-1} \quad [2]$$

Dans l'équation 2, \mathbf{x}_t est le vecteur observation à l'instant t , \mathbf{I} est la matrice identité et λ est la matrice diagonale de pondération dont les éléments $\lambda_1, \lambda_2, \dots, \lambda_p$ seront compris entre 0 et 1 ($0 < \lambda_i < 1$).

Lowry (Lowry *et al.*, 1992) propose d'utiliser $\lambda_i = \lambda$ pour $i = 1, 2, \dots, p$ si il n'y pas de raison particulière de pondérer les variables différemment. Dans le cas inverse, la carte MEWMA est dite directionnelle puisqu'elle privilégie la détection dans certaines directions de l'espace, mais le traitement de la carte perd alors de sa simplicité. Ainsi, par la suite, nous considérerons le cas de la carte MEWMA non-directionnelle ($\lambda_i = \lambda$ pour tout i). On pourra donc réécrire l'équation 2 sous la forme :

$$\mathbf{y}_t = \lambda \mathbf{x}_t + (1 - \lambda) \mathbf{y}_{t-1} \quad [3]$$

Sur la carte de contrôle MEWMA, on tracera alors la statistique suivante :

$$T_t^2 = \mathbf{y}_t^T \boldsymbol{\Sigma}_{\mathbf{y}_t}^{-1} \mathbf{y}_t \quad [4]$$

où $\boldsymbol{\Sigma}_{\mathbf{y}_t}$ est la matrice de variance-covariance de la variable \mathbf{y} pour l'instant t . Elle est définie ainsi :

$$\boldsymbol{\Sigma}_{\mathbf{y}_t} = \left\{ \frac{\lambda[1 - (1 - \lambda)^{2t}]}{2 - \lambda} \right\} \boldsymbol{\Sigma} \quad [5]$$

Mais, dans le cas où λ n'est pas trop petit ($\lambda > 0.1$), cette matrice approche très rapidement (4 ou 5 échantillons) sa valeur asymptotique définie par :

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \left\{ \frac{\lambda}{2 - \lambda} \right\} \boldsymbol{\Sigma} \quad [6]$$

Le procédé est déclaré hors contrôle dès que T_t^2 dépasse la limite de contrôle h_M . Cette limite de contrôle h_M est calculée en fonction de p et de λ , afin de respecter une certaine fréquence donnée de fausses alertes (Lowry *et al.*, 1992). Il faut préciser que les performances de la carte MEWMA vont être fonction de λ . En effet, un λ faible permettra une détection performante des sauts de faibles amplitudes, alors qu'un λ élevé sera plus adapté à la détection de saut de fortes amplitudes. Il faudra donc choisir λ en fonction de l'amplitude du saut que l'on souhaite détecter. Il est également à signaler qu'un cas particulier de la carte MEWMA est la carte avec $\lambda = 1$. En effet, dans ce cas, nous obtenons une carte T^2 de Hotelling. La carte MEWMA possède donc, dans tous les cas, des performances équivalentes ($\lambda = 1$) ou supérieures ($\lambda < 1$) à la carte T^2 pour la détection de faibles déréglements. Par contre, dans le cas d'un saut d'une forte amplitude, la carte T^2 détectera plus rapidement que la carte MEWMA, car cette dernière possède une inertie que ne possède pas la carte T^2 .

2.2. L'analyse discriminante

En dehors de son caractère descriptif, l'analyse discriminante peut également être vue comme une technique statistique de classification se basant sur la règle de Bayes. En effet, elle affecte à un nouvel individu \mathbf{x} la classe C_i qui possède la probabilité a posteriori $P(C_i|\mathbf{x})$ maximale d'appartenance sachant la valeur de tous les descripteurs :

$$\mathbf{x} \in C_i, \text{ si } i = \underset{i=1, \dots, k}{\operatorname{argmax}} \{P(C_i|\mathbf{x})\} \quad [7]$$

On appelle cette règle "règle de Bayes", elle permet d'obtenir la valeur de $P(C_i|\mathbf{x})$ par la formule de l'équation 8, où $P(C_i)$ est la probabilité a priori d'appartenance à la classe C_i .

$$P(C_i|\mathbf{x}) = \frac{P(C_i)P(\mathbf{x}|C_i)}{P(\mathbf{x})} \quad [8]$$

On voit que pour chaque classe, le dénominateur de l'équation 8 est le même, il ne va donc pas intervenir dans la fonction discriminante. L'équation 7 peut ainsi se réécrire sous la forme :

$$\mathbf{x} \in C_i, \text{ si } i = \underset{i=1, \dots, k}{\operatorname{argmax}} \{P(C_i)P(\mathbf{x}|C_i)\} \quad [9]$$

Mais, pour plus de facilité nous allons écrire cette règle de décision sous forme de fonction de coût K telle que donnée par :

$$K_i(\mathbf{x}) = -2\log(P(C_i)P(\mathbf{x}|C_i)) \quad [10]$$

On peut alors écrire la règle d'attribution d'un nouvel individu \mathbf{x} à une classe C_i par la règle de l'équation 11 suivante :

$$\mathbf{x} \in C_i, \text{ si } i = \underset{i=1, \dots, k}{\operatorname{argmin}} \{K_i(\mathbf{x})\} \quad [11]$$

Classiquement, cette règle est utilisée avec des lois paramétriques, la plus connue étant la loi normale multivariée.

2.2.1. Application à la loi normale multivariée

La loi normale multivariée est l'extension de la loi normale (ou loi de Gauss) au domaine multivarié. Sa fonction de densité ϕ conditionnellement à une classe C_i peut s'écrire comme indiquée par l'équation 12, où $\boldsymbol{\mu}_i$ représente le vecteur des moyennes de la classe C_i et $\boldsymbol{\Sigma}_i$ représente la matrice de variance-covariance de la classe C_i .

$$\phi(\mathbf{x}|C_i) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right) \quad [12]$$

Comme les valeurs exactes des paramètres de loi : $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \dots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ ne sont généralement pas connus, il faut les estimer à partir des données. Pour cela, on peut utiliser la méthode d'Estimation par Maximum de Vraisemblance (EMV). L'intérêt premier d'utiliser l'EMV est qu'elle possède de bonnes propriétés de convergence lorsque la taille d'échantillon utilisée pour l'estimation s'accroît. De plus, elle reste la méthode la plus simple d'estimation de paramètres de loi. Dans le cas d'une loi normale multivariée, on peut montrer que l'EMV du vecteur des moyennes $\boldsymbol{\mu}$ revient à l'équation 13. Cette estimation est sans biais.

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad [13]$$

De même, une estimation non-biaisée de $\boldsymbol{\Sigma}$ est :

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \quad [14]$$

Pour une justification approfondie de ces estimations, on peut notamment consulter (Duda *et al.*, 2001). Il faut également noter que dans certains cas (présence de valeurs aberrantes dans les données, trop de peu d'exemples), ces estimateurs peuvent se révéler imprécis. Dans ces cas, d'autres estimateurs, appelés estimateurs robustes, peuvent être utilisés (Huber, 1981; Kobi, 1994).

2.2.2. Analyse discriminante quadratique

Dans le cas de la règle de décision de l'analyse discriminante appliquée à une loi normale multivariée, l'équation 10 peut alors s'écrire :

$$K_i(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - 2\log(P(C_i)) + \log(|\boldsymbol{\Sigma}_i|) + p\log(2\pi) \quad [15]$$

Dans l'équation 15, on peut voir que le dernier terme $p \log(2\pi)$ est constant à chaque K_i et n'interviendra donc pas pour la discrimination. Cette règle se nomme "Analyse Discriminante Quadratique". Elle réalise des séparations quadratiques entre chaque classe. On peut remarquer que l'expression $(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$ est le T^2 de \mathbf{x} pour la classe C_i . Si l'on fait l'hypothèse d'indépendance des variables, alors $\boldsymbol{\Sigma}_i$ est diagonale (toutes les covariances sont nulles) : $\text{diag}(\boldsymbol{\Sigma}_i) = (\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$. Cette règle de décision est également connue sous le nom de classifieur de Bayes ou réseau bayésien naïf. En continuité à l'hypothèse d'indépendance des variables, on peut également faire l'hypothèse, pour chaque classe, d'égalité des variances σ^2 des p variables. Alors, chaque classe aura une forme sphérique mais de taille différente (taille dépendante de σ_i^2). Le problème de l'analyse discriminante quadratique est qu'elle exige l'estimation de beaucoup de paramètres. Or, pour estimer convenablement de nombreux paramètres il faut avoir beaucoup de données. Les problèmes d'estimation venant principalement des différentes matrices de variance-covariance, une solution pourrait être l'analyse discriminante linéaire.

2.2.3. L'analyse discriminante linéaire

Pour réaliser une analyse discriminante linéaire, nous allons faire l'hypothèse d'égalité des matrices de variance-covariance. C'est à dire que pour toute classe C_i , $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$, avec $\boldsymbol{\Sigma}$ la matrice de variance-covariance commune à toutes les classes. Cette matrice peut être obtenue facilement par l'équation 16 dans laquelle les n_i représentent les nombre de cas des classes C_i que contient l'ensemble d'apprentissage et où n est le nombre de cas total (en d'autres termes $n = n_1 + n_2 + \dots + n_k$).

$$\boldsymbol{\Sigma} = \frac{(n_1 - 1)\boldsymbol{\Sigma}_1 + (n_2 - 1)\boldsymbol{\Sigma}_2 + \dots + (n_k - 1)\boldsymbol{\Sigma}_k}{n - k} \quad [16]$$

Ainsi, nous allons de nouveau pouvoir simplifier la discrimination de l'équation 15 puisque le terme $\log(|\boldsymbol{\Sigma}_i|)$ va devenir $\log(|\boldsymbol{\Sigma}|)$ et sera donc constant pour chaque classe. Ainsi, en posant $\log(|\boldsymbol{\Sigma}|) + p \log(2\pi) = Cste$ nous obtenons la fonction coût de l'équation 17.

$$K_i(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - 2 \log(P(C_i)) + Cste \quad [17]$$

On peut remarquer que si les probabilités marginales de chaque classe sont égales (même nombre d'exemples pour chaque classe), alors la règle de décision revient à calculer les distances de Mahalanobis pour chaque classe et à attribuer au nouvel individu la classe pour laquelle cette distance est la plus faible. Cette fonction coût va réaliser des séparations linéaires (droites, plans, hyperplans) entre les classes. Mais, comme dans le cas de l'analyse discriminante quadratique, nous pourrions également faire l'hypothèse supplémentaire que $\boldsymbol{\Sigma}$ est diagonale ou bien sphérique. L'analyse discriminante linéaire est plutôt robuste face aux fluctuations sur les hypothèses de normalité des classes et d'égalité des matrices de variance-covariance. De ce fait, elle est fréquemment utilisée et doit être considérée comme une méthode de référence.

Cependant, bien souvent, l'analyse discriminante linéaire est confondue avec l'analyse factorielle discriminante. Or, l'analyse factorielle discriminante n'est pas une méthode de classification, mais une étape de réduction dimensionnelle.

2.3. Analogie entre détection et classification

La détection, comme nous l'avons déjà définie, consiste à déceler la présence de fautes dans le procédé. Elle a pour but de détecter si celui-ci est soumis à l'effet d'une cause spéciale qui impliquera un accroissement de la variabilité à plus ou moins long terme. La détection peut être vue comme une classification en deux classes : fonctionnement normal du procédé, et respectivement fonctionnement anormal du procédé. Cependant, la phase de détection ne peut pas être directement assimilée aux types de classification classique (supervisée ou non-supervisée). En effet, dans le cas de la détection, bien que l'on suppose posséder des exemples d'individus de la classe de fonctionnement normal, nous devons pouvoir effectuer la détection même si aucun exemple de fonctionnement anormal n'est disponible. Ce type de classification est appelé classification monoclasse ("one-class classification") (Tax *et al.*, 2001). Le but de la classification monoclasse est de décrire une classe d'individus, et de pouvoir distinguer si un nouvel individu appartient ou non à cette classe. Dans le cas de la détection, la classe d'intérêt est celle décrite par des exemples d'individus supposés sous-contrôle : la classe décrivant le fonctionnement normal du procédé. Cependant, lorsque l'on parle de classification, il est évident que l'on cherche à faire la distinction entre plusieurs classes. Or, pour traiter le cas de la classification monoclasse, nous n'avons qu'une seule classe à disposition. Nous devons donc créer, au minimum, une seconde classe que nous appellerons classe virtuelle. La première classe représentant le fonctionnement normal du procédé (classe "Sous-Contrôle" ou classe *SC*), nous dénommerons cette seconde classe (virtuelle) : classe "Hors-Contrôle" ou classe *HC*. Ainsi, cette seconde classe va représenter l'ensemble des individus qui ne peuvent pas appartenir à la première classe *SC*.

Un exemple typique de classification monoclasse est la carte du T^2 de Hotelling. En effet, un nouvel individu est déclaré comme appartenant à la classe *SC* si la mesure de son T^2 ne dépasse pas une certaine limite de contrôle *LC*, alors que cet individu sera déclaré comme appartenant à la classe *HC* si son T^2 dépasse la limite de contrôle *LC*. Dans ce cas, il est clair que la frontière de décision entre les deux classes *SC* et *HC* est représentée par la limite de contrôle *LC*. Dans le cas de cette carte de contrôle, il est possible de représenter, pour un exemple en deux dimensions, la frontière de décision induite par l'application de la limite de contrôle *LC*. Sur la figure 1, on s'aperçoit que cette frontière est une ellipse entourant alors la classe de fonctionnement normal du procédé (classe *SC*), et que tout individu à l'extérieur de cette ellipse est considéré comme un individu appartenant à la classe hors-contrôle *HC*.

La figure 1 montre une chose très importante, la classification engendrée ne peut pas être une classification linéaire. Ainsi, il va être impossible d'utiliser une Analyse Discriminante Linéaire (voir section 2.2.3) pour arriver à une frontière de classifica-

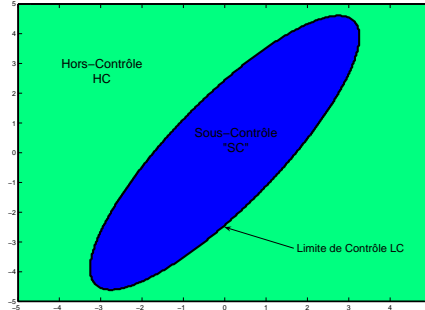


Figure 1. Frontière de décision de la carte T^2 dans l'espace bivarié

tion comme celle décrite sur la figure 1. En effet la frontière de classification n'est pas un hyperplan (linéaire). Cependant, ce type de frontière est typique d'une Analyse Discriminante Quadratique. Nous allons donc nous orienter vers ce type de classification, afin de réaliser la classification monoclasse (détection). Mais, puisque notre but est de réaliser cette détection dans un réseau bayésien, nous allons tout d'abord voir comment réaliser une analyse discriminante quadratique par un réseau bayésien.

3. Analogie entre RB et AD

3.1. Réseaux Bayésiens

Un Réseau Bayésien (RB) (Jensen, 1996; Naim *et al.*, 2004) est un modèle graphique dans lequel les connaissances sont représentées sous forme de variable. Chaque variable est un nœud du graphe et prend ses valeurs dans un ensemble discret ou continu. Le graphe est toujours dirigé et acyclique. Les arcs dirigés représentent un lien de dépendance directe (la plupart du temps il s'agit de causalité). Ainsi un arc allant de la variable X à la variable Y exprimera le fait que X dépend directement de Y . L'absence d'arc ne renseigne alors que sur la non-existence d'une dépendance directe. Les paramètres expriment les poids donnés à ces relations et sont les probabilités conditionnelles des variables sachant leurs parents (exemple : $P(Y|X)$) ou les probabilités a priori si la variable n'a pas de parents. On peut définir formellement un réseau bayésien comme étant un triplet $\{\mathbf{G}, \mathbf{E}, \mathbf{D}\}$ où :

$\{\mathbf{G}\}$ est un graphe acyclique orienté, $\mathbf{G} = (V, A)$, où V est l'ensemble des nœuds de \mathbf{G} , et A est l'ensemble des arcs de \mathbf{G} ,

$\{\mathbf{E}\}$ est un espace probabilisé fini (Ω, Z, P) , avec Ω un espace non vide, Z un ensemble de sous-espaces de Ω , et P une mesure de probabilité sur Z avec $P(\Omega) = 1$,

$\{\mathbf{D}\}$ est un ensemble de variables aléatoires associées aux nœuds de \mathbf{G} et défini sur \mathbf{E} , tel que :

$$P(V_1, V_2, \dots, V_n) = \prod_{i=1}^n P(V_i | C(V_i)) \quad [18]$$

où $C(V_i)$ est l'ensemble des causes (parents) de V_i dans le graphe \mathbf{G} .

Il est possible de réaliser des classifieurs performants grâce aux réseaux bayésiens (Langley *et al.*, 1994; Friedman *et al.*, 1997; Pernkopf, 2005). Nous présentons ici les principaux types de structures pour employer les réseaux bayésiens comme classifieurs. Un classifieur bayésien d'un problème à p variables a pour particularité de posséder $p + 1$ nœuds. En effet, tous les classifieurs bayésiens modélisent le fait d'appartenance à une classe par un nœud discret. Nous nommons ce nœud "nœud de classe", et nous le notons C . Ce nœud est un nœud discret multinomial à k modalités, où k représente le nombre de classe de notre problème (C_1, C_2, \dots, C_k). Ce nœud de classe a pour particularité de ne pas posséder de nœud parent. Les autres variables, au nombre de p , que nous nommons variables descriptives, sont notées X_i (i de 1 à p).

Le classifieur bayésien possédant la structure la plus simple est le Réseau Bayésien Naïf (RBN), également appelé classifieur de Bayes. On le qualifie de naïf car il fait l'hypothèse, très forte, que chaque variable descriptive est, conditionnellement à la classe, indépendante des autres. Lorsque toutes les variables descriptives sont incorporées au modèle, on parlera alors de structure naïve complète. Ce classifieur est extrêmement connu car ses performances (notamment dans le cas où toutes les variables sont discrètes) sont intéressantes dans certains domaines et dépassent des techniques beaucoup plus sophistiquées même lorsque l'hypothèse d'indépendance est violée (Domingos *et al.*, 1996). Au vu de l'hypothèse forte que ce classifieur implique, il est normal que certains chercheurs aient voulu améliorer ce classifieur. Friedman (Friedman *et al.*, 1997) propose d'ajouter des arcs entre les différentes variables descriptives du classifieur naïf. Pour cela, il décide de créer un arbre entre les variables descriptives, à la manière de Chow et Liu (Chow *et al.*, 1968), afin d'obtenir un TAN (Tree Augmented Naïve Bayes). L'algorithme part d'un réseau bayésien naïf et ajoute alors un arc entre les variables qui partagent la plus importante information mutuelle. Mais, pour respecter la topologie de l'arbre, l'algorithme interdit à chaque nœud d'avoir plus de 2 parents (soit un parent en plus du nœud classe). Afin de prendre en compte la corrélation entre les différents descripteurs, il a également été proposé le CSNBN (Condensed Semi-Naïve Bayesian Network) : un réseau bayésien semi naïf condensé (Kononenko, 1991; Pazzani, 1997). On le nomme condensé car il introduit un nouveau type de variable : les variables jointes. Ces nouvelles variables jointes représentent en fait un groupement de variables descriptives. Bien entendu, une variable descriptive ne pourra se trouver que dans une seule variable jointe. Le fait que deux variables se trouvent dans une variable jointe implique que ces deux variables sont corrélées. Un regroupement de p variables continues suivra une loi normale multivariée et sera donc représenté par un seul nœud continu de dimension p .

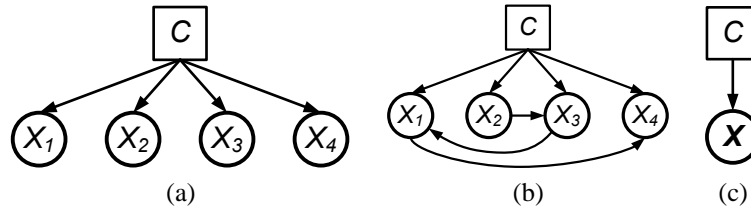


Figure 2. Différents classifieurs bayésiens : RBN (a), TAN (b) et CSNBN (c).

3.2. Analyse discriminante par Réseau Bayésien

Nous allons pouvoir réaliser aisément des analyses discriminantes par réseaux bayésiens. En effet, l'inférence dans un réseau bayésien est basée sur la règle de Bayes, et l'analyse discriminante est également basée sur cette règle de décision. La structure pour réaliser des analyses discriminantes classiques (quadratiques, linéaires) sur un système à p variables comprenant k types de fonctionnement peut se modéliser par un nœud continu normal multivarié de dimension p (variable \mathbf{X}), relié à un nœud discret de dimension k (variable C), comme indiqué sur la figure 3 (a). Nous avons déjà présenté rapidement ce type de réseau, il s'agit d'un réseau bayésien semi-naïf condensé. Ce réseau représente en fait une loi normale multivariée conditionnellement à la classe, tout comme l'est une analyse discriminante. Le choix entre les différentes possibilités d'analyse discriminante (quadratique, linéaire, diagonale, sphérique, régularisée, etc) se fera alors sur le choix des k matrices de variance-covariance qui vont être attribuées à \mathbf{X} . On peut également réaliser une analyse discriminante quadratique en faisant apparaître chaque variable univariée X_i composant la variable multivariée \mathbf{X} , puis en reliant toutes ces variables afin de prendre en compte toutes les relations pouvant exister entre les X_i . Un exemple de ceci pour quatre variables est décrit sur la figure 3 (b). Ce réseau représente en fait une série de $p - 1$ régressions linéaires. L'avantage de ce type de réseau est qu'il permet de donner un résultat même si nous n'avons pas les valeurs de chaque variable descriptive. Cependant, l'évaluation de tous les paramètres de régression ainsi que le temps d'inférence vont demander plus de temps calcul que dans le cas précédent. Supposons à présent que l'hypothèse de diagonalité des matrices de variance-covariance soit faite. Cela signifie que chaque variable X_i est indépendante des autres variables, excepté la variable de classe C . Dans ce cas, il est toujours possible d'exprimer ceci sous la forme générale vu précédemment (figure 3). Cependant, il est aussi possible de modéliser une analyse discriminante quadratique diagonale grâce au réseau de la figure 3 (c). Ce type de réseau se nomme également réseau bayésien naïf, ou classifieur de Bayes, car il fait l'hypothèse naïve que chaque variable X_i est indépendante. Là encore, l'avantage de cette modélisation est le fait de pouvoir obtenir un résultat même si nous n'avons pas les valeurs de chaque X_i .

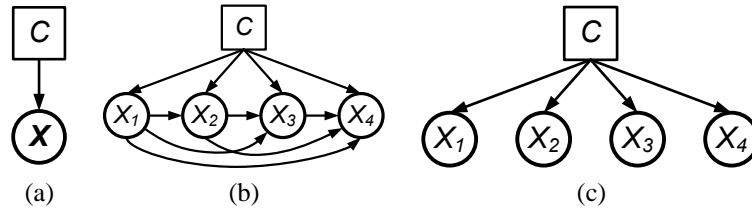


Figure 3. Différents type d'analyse discriminante (AD) : AD classique (a), AD pour données manquantes (b) et AD diagonale (c).

4. Cartes multivariées par réseau bayésien

Comme nous l'avons vu dans les deux sections précédentes, d'une part il est possible d'assimiler la détection (et notamment le principe des cartes multivariées) à une étape de classification, et d'autre part, il est simple de réaliser une analyse discriminante par réseau bayésien. Dans cette section, nous allons présenter mathématiquement la modélisation des cartes de contrôle multivariées dans un réseau bayésien.

4.1. Définition de la classe HC

Nous avons vu dans la section 2 que la détection par carte de contrôle multivariée peut être assimilée à une classification en deux classes : la classe de fonctionnement normal (classe SC) et la classe virtuelle représentant tous les individus n'appartenant pas à la première classe (classe HC).

Afin de définir la classe de fonctionnement normal (SC), nous supposons que plusieurs individus supposés sous-contrôle sont disponibles. De même que pour les cartes multivariées, ces individus nous permettent d'estimer le vecteur des moyennes μ ainsi que la matrice de variance-covariance Σ de notre procédé, lorsque celui-ci est en fonctionnement normal. Concernant la classe virtuelle HC , nous n'avons pas forcément d'individus à disposition. De plus, même si nous en avons, ceux-ci ne suffiraient jamais à couvrir toute la zone de l'espace pour laquelle nous voulons attribuer la classe hors-contrôle HC . En analysant un peu la différence entre les deux classes sur la figure 1, on voit qu'il est possible d'assimiler leur centre de classe au même point. Alors la seule chose différenciant les deux classes SC et HC est leur variabilité. En effet, la classe SC possède une variabilité plus faible que la classe HC . Cette remarque nous pousse donc à définir le vecteur des moyennes de la classe HC comme étant le même que celui de la classe SC , mais avec une matrice de variance-covariance exprimant plus de variabilité. Pour cela, nous définissons alors la matrice de variance-covariance de la classe HC comme étant $c \times \Sigma$ où Σ est la matrice de variance-covariance de la classe SC , et c est un coefficient strictement supérieur à 1, permettant ainsi d'aug-

menter la variabilité de la classe HC par rapport à celle de la classe SC . Nous avons donc les deux classes possédant les paramètres répertoriés dans la table 1.

Classe	Distribution
Sous contrôle (SC)	$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
Hors-contrôle (HC)	$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, c \times \boldsymbol{\Sigma})$

Tableau 1. Paramètres des classes pour la détection

Cependant, fixer les lois sous-jacentes au fonctionnement du procédé ne suffit pas pour tirer des conclusions sur son état. En effet, la définition de ces paramètres permet l'application d'une analyse discriminante directe ou par réseau bayésien. Lorsqu'un nouvel individu \mathbf{x} est présenté au classifieur, nous allons alors obtenir une probabilité $P(SC|\mathbf{x})$ que cet individu appartienne à la classe SC et une probabilité $P(HC|\mathbf{x})$ que cet individu appartienne à la classe HC (avec $P(SC|\mathbf{x}) + P(HC|\mathbf{x}) = 1$, puisque seulement deux classes sont présentes). Cependant, comment alors interpréter ces différentes probabilités, comment prendre une décision correcte en fonction de ces probabilités ? Plusieurs facteurs vont alors rentrer en ligne de compte. En effet, comme pour le cas des cartes multivariées, nous allons devoir nous fixer un certain taux α de fausses alertes. De plus, la valeur du paramètre c définissant la classe HC va jouer un rôle sur les valeurs des probabilités et donc de la décision à prendre. Nous allons donc dans la section suivante nous attacher à adapter les paramètres du réseau bayésien afin d'obtenir la même règle de décision que celle des cartes de contrôle multivariées.

4.2. Equivalence RB - cartes de contrôle

Après avoir montré l'équivalence qu'il peut y avoir entre un réseau bayésien et une analyse discriminante (voir section 3.2), nous allons prouver ici l'équivalence entre le principe des cartes de contrôle et un réseau bayésien, ou plus précisément définir les paramètres de celui-ci permettant d'obtenir l'équivalence avec les cartes de contrôle multivariées.

Dans un premier temps, comme pour le cas des cartes de contrôles multivariées et en plus de l'hypothèse de normalité, nous allons nous fixer un certain taux α de fausses alarmes. Ainsi, nous nous fixons des probabilités a priori d'être sous-contrôle $P(SC)$ et hors-contrôle $P(HC)$ données par :

$$P(SC) = 1 - \alpha$$

$$P(HC) = \alpha$$

De plus, comme pour les cartes de contrôle, nous allons nous fixer un seuil permettant de prendre une décision : "procédé hors-contrôle si $P(HC|\mathbf{x}) > \alpha$ ", ou de façon équivalente "procédé sous-contrôle si $P(SC|\mathbf{x}) > 1 - \alpha$ ", soit :

$$\mathbf{x} \in SC \text{ si } P(SC|\mathbf{x}) > 1 - \alpha \quad [19]$$

Puisque nous nous sommes fixés le paramètre α , ainsi que la règle de décision voulue, il ne nous reste plus qu'un seul paramètre à étudier : le coefficient c permettant l'augmentation de la variabilité de la classe HC . Dans ce but, nous allons développer l'inéquation de la règle de décision (équation 19) :

$$\begin{aligned} P(SC|\mathbf{x}) &> 1 - \alpha \\ P(SC|\mathbf{x}) &> (1 - \alpha)(P(SC|\mathbf{x}) + P(HC|\mathbf{x})) \\ P(SC|\mathbf{x}) &> (1 - \alpha)P(SC|\mathbf{x}) + (1 - \alpha)P(HC|\mathbf{x}) \\ P(SC|\mathbf{x}) &> \left(\frac{1 - \alpha}{\alpha}\right) P(HC|\mathbf{x}) \end{aligned}$$

Or, d'après la loi de Bayes, on a :

$$P(SC|\mathbf{x}) = \frac{P(SC)P(\mathbf{x}|SC)}{P(\mathbf{x})} \quad [20]$$

et

$$P(HC|\mathbf{x}) = \frac{P(HC)P(\mathbf{x}|HC)}{P(\mathbf{x})} \quad [21]$$

avec $P(SC) = 1 - \alpha$ et $P(HC) = \alpha$, on peut alors obtenir :

$$\begin{aligned} \frac{P(SC)P(\mathbf{x}|SC)}{P(\mathbf{x})} &> \left(\frac{1 - \alpha}{\alpha}\right) \frac{P(HC)P(\mathbf{x}|HC)}{P(\mathbf{x})} \\ \left(\frac{P(SC)}{P(HC)}\right) P(\mathbf{x}|SC) &> \left(\frac{1 - \alpha}{\alpha}\right) P(\mathbf{x}|HC) \\ P(\mathbf{x}|SC) &> P(\mathbf{x}|HC) \end{aligned} \quad [22]$$

Or, dans le cas d'une analyse discriminante avec normalité de k classes C_i , les probabilités conditionnelles (par rapport aux différentes classes) peuvent être calculées par l'équation 23, où ϕ représente la fonction de densité de probabilité de la loi normale multivariée correspondante à la classe.

$$P(\mathbf{x}|C_i) = \frac{\phi(\mathbf{x}|C_i)}{\sum_{j=1}^k P(C_j)\phi(\mathbf{x}|C_j)} \quad [23]$$

L'inéquation d'équivalence 22 peut alors s'exprimer comme :

$$\phi(\mathbf{x}|SC) > \phi(\mathbf{x}|HC) \quad [24]$$

On rappelle que la fonction de densité de la loi normale multivariée de dimension p , de paramètres $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$, d'un individu \boldsymbol{x} est donnée par :

$$\phi(\boldsymbol{x}) = \frac{e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \quad [25]$$

A présent, si les paramètres de la loi sont $\boldsymbol{\mu}$ et $c \times \boldsymbol{\Sigma}$, alors la fonction de densité va être :

$$\phi(\boldsymbol{x}) = \frac{e^{-\frac{1}{2c}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2} c^{p/2}} \quad [26]$$

En identifiant l'expression $(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ comme étant le T^2 pour l'individu \boldsymbol{x} , on peut alors écrire les inéquations suivantes :

$$\begin{aligned} \phi(\boldsymbol{x}|SC) &> \phi(\boldsymbol{x}|HC) \\ \frac{e^{-\frac{T^2}{2}}}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} &> \frac{e^{-\frac{T^2}{2c}}}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2} c^{p/2}} \\ e^{-\frac{T^2}{2}} &> \frac{e^{-\frac{T^2}{2c}}}{c^{p/2}} \\ -\frac{T^2}{2} &> -\frac{T^2}{2c} - \frac{p \ln(c)}{2} \\ T^2 &< \frac{T^2}{c} + p \ln(c) \\ T^2 &< \frac{p \ln(c)}{1 - \frac{1}{c}} \end{aligned} \quad [27]$$

Or, nous recherchons les valeurs de c permettant d'obtenir une règle de décision équivalente aux cartes de contrôle multivariée, à savoir : $\boldsymbol{x} \in SC$ si $T^2 < LC$. Nous obtenons donc l'équation suivante pour c :

$$\frac{p \ln(c)}{1 - \frac{1}{c}} = LC \quad [28]$$

Ce qui revient à résoudre l'équation :

$$1 - c + \frac{pc}{LC} \ln(c) = 0 \quad [29]$$

L'équation 29 admet deux solutions. Cependant, l'une des solutions (évidente) est $c = 1$. Mais ceci n'est pas possible puisque nous avons précédemment défini le paramètre c comme étant strictement supérieur à 1. De plus, attribuer la valeur 1 à c reviendrait à réaliser une analyse discriminante entre deux même classes, ce qui n'aurait pas de sens. L'équation 29 admet également une autre racine calculable numériquement. Ainsi, c ne dépendant que de LC et de p , on peut établir des abaques en fonction de la dimension p du système à surveiller et du risque α fixé (en effet, LC est uniquement

fonction de p et de α). Nous avons établis ces abaques pour les cartes de contrôle du T^2 de Hotelling ainsi que pour la carte MEWMA (à différents paramètres λ), pour des valeurs de α de 1% et de 0.5%, et des valeurs de p allant de 1 à 50 variables. Les cartes de contrôle univariés n'étant que des cas particuliers des cartes de contrôle multivariées, la méthode développée ici (ainsi que le calcul du paramètre c) est également valable pour ce type de carte de contrôle. Les différents abaques sont disponibles en annexes (voir annexes A). Ainsi, suivant la carte que l'on veut utiliser, la dimension du système et le risque α choisi, ces abaques nous donnent la valeur du coefficient c à utiliser afin d'obtenir dans tous les cas une règle de décision comparable à celle de la carte de contrôle choisie.

4.3. Exemple des cartes T^2 et MEWMA

Nous proposons d'illustrer l'approche décrite sur un exemple très simple d'un système à deux dimensions. Ainsi, nous allons étudier sur ce système une carte T^2 et sa transposition par un réseau bayésien, ainsi qu'une carte MEWMA (avec $\lambda = 0.1$) et sa transposition par un réseau bayésien, les deux cartes possédant un risque de fausses alertes $\alpha = 1\%$. Lorsque ce système est sous-contrôle, il suit une loi normale multivariée de paramètres μ et Σ tels que :

$$\begin{aligned}\mu &= (5 \quad 10) \\ \Sigma &= \begin{pmatrix} 1 & 1.2 \\ 1.2 & 2 \end{pmatrix}\end{aligned}$$

Afin de surveiller ce procédé, nous appliquons alors la méthode proposée de détection par réseau bayésien. Ainsi, pour une carte du T^2 de Hotelling, nous obtenons le réseau bayésien de la figure 4, où sont également représentées les tables de probabilités conditionnelles de chaque nœud, et où le paramètre c est égal à 95,28.

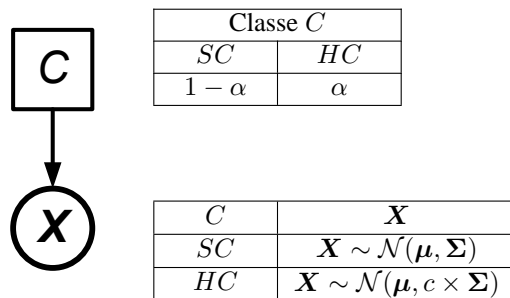


Figure 4. Réseau bayésien similaire à la carte T^2

De la même façon, on peut également surveiller le procédé par une carte MEWMA sous forme du réseau bayésien de la figure 5, où le paramètre c est égal à 90,29.

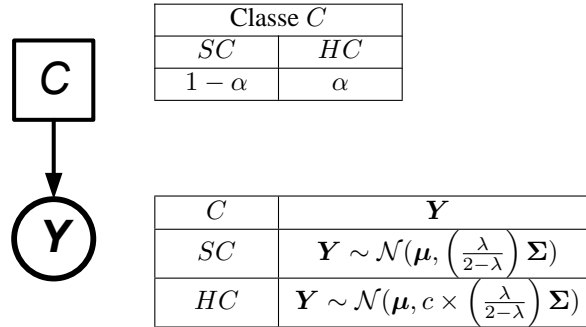


Figure 5. Réseau bayésien similaire à la carte MEWMA

Nous avons simulé ce système sur 30 observations, mais nous avons introduit un saut en échelon d'amplitude 0.5 à partir de l'observation 6 sur la première variable. La figure 6 présentent graphiquement les différentes conclusions obtenues, respectivement pour la carte T^2 (graphiques de gauche) et la carte MEWMA (graphiques de droite). Sur chaque figure, le graphique du haut représente le calcul de la distance statistique pour la carte de contrôle, alors que le graphique du bas représente la probabilité de la modalité SC du nœud classe pour la modélisation par réseau bayésien. La limite de contrôle est également représentée pour chaque graphique. Ainsi, on observe bien la limite sur le réseau bayésien, fixée à $1 - \alpha$, soit 99%.

La figure 6 permet de voir que les décisions prises à un instant t , entre une carte de contrôle multivariée et son équivalence par réseau bayésien, sont les mêmes. De plus, on peut remarquer qu'il existe une certaine ressemblance entre ces signaux, malgré une légère différence du fait de la prise en compte de la borne supérieure et inférieure (à savoir 0 et 1) pour le calcul de $P(SC|\mathbf{x})$. Désormais, il est ainsi possible de pratiquer la détection de faute dans un procédé multivarié, directement par un réseau bayésien.

5. Conclusions et perspectives

Dans cet article nous avons présenté une approche permettant d'effectuer la détection de fautes dans un procédé multivarié. Cette approche se base sur un réseau bayésien. Un nœud de classe permet de représenter l'état du système (sous contrôle ou hors-contrôle), alors qu'un nœud gaussien multivarié représente les données du système à un instant donné. Nous avons alors prouvé que l'on pouvait obtenir une équivalence entre ce réseau et une carte de contrôle multivariée (T^2 ou MEWMA). L'intérêt principal de cette méthode est qu'elle va pouvoir être étendue puisque le diagnostic de faute par réseau bayésien a déjà démontré son efficacité (Verron *et al.*, 2006; Verron

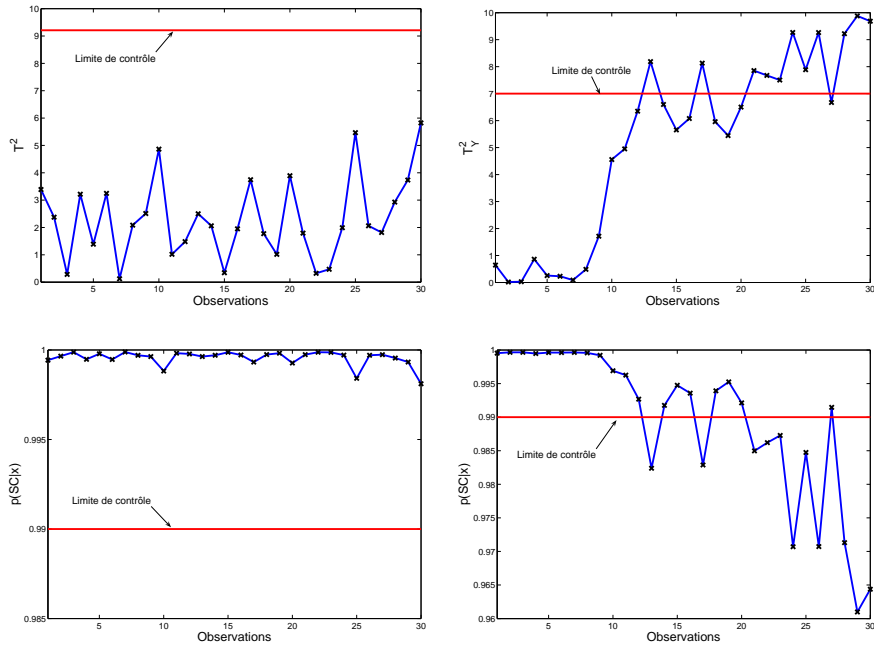


Figure 6. Résultats des cartes de contrôle T^2 et MEWMA, ainsi que leurs équivalences en réseau bayésien, pour l'exemple du système bivarié

et al., 2007). Ainsi, la perspective qu'entraîne cette méthode est la mise au point d'un réseau bayésien complet permettant à la fois la détection et le diagnostic de faute dans un procédé multivarié.

A. Abaques du coefficient c

La table 2 présente les valeurs du coefficient c afin de modéliser une carte MEWMA, pour des valeurs de fausses alertes α de 1% ou 0.5%, pour un nombre de variables p de 1 à 50, et pour un λ de 0.1 ou bien de 1 (équivalent à la carte T^2).

λ	p		1	2	5	10	20	50
	α							
1	1%		754.54	95.28	17.14	7.46	4.17	2.48
	0.5%		2634.5	194.63	24.91	9.53	4.92	2.75
0.1	1%		650.81	90.29	16.57	7.3	4.01	2.45
	0.5%		2228.61	111.49	23.42	9.37	4.81	2.72

Tableau 2. Coefficient c pour la carte MEWMA

B. Bibliographie

- Boudali H., Dugan J., « A discrete-time Bayesian network reliability modeling and analysis framework », *Reliability Engineering and System Safety*, vol. 87, n° 3, p. 337-349, 2005.
- Chow C., Liu C., « Approximating discrete probability distributions with dependence trees », *Information Theory, IEEE Transactions on*, vol. 14, n° 3, p. 462-467, 1968.
- Domingos P., Pazzani M. J., « Beyond Independence : Conditions for the Optimality of the Simple Bayesian Classifier », *International Conference on Machine Learning*, 1996.
- Dubuisson B., *Diagnostic, intelligence artificielle et reconnaissance des formes*, Traité IC2 information. Série productique, Hermès sciences publications, 2001.
- Duda R. O., Hart P. E., Stork D. G., *Pattern Classification 2nd edition*, Wiley, 2001.
- Friedman N., Geiger D., Goldszmidt M., « Bayesian Network Classifiers », *Machine Learning*, vol. 29, n° 2-3, p. 131-163, 1997.
- Hawkins D. M., « Multivariate quality control based on regression-adjusted variables », *Technometrics*, vol. 33, n° 1, p. 61-75, 1991.
- Hotelling H., « Multivariate Quality Control », *Techniques of Statistical Analysis*, vol. , p. 111-184, 1947.
- Huber P., *Robust Statistics*, Wiley, 1981.
- Jensen F. V., *An introduction to Bayesian Networks*, Taylor and Francis, London, United Kingdom, 1996.
- Kobi A., Diagnostic de processus continus : application à la détection de valeurs aberrantes dans les signaux d'entrée et de sortie de système, PhD thesis, Institut national polytechnique de Lorraine, 1994.
- Kononenko I., « Semi-naive Bayesian classifier », *EWSL-91 : Proceedings of the European working session on learning on Machine learning*, p. 206-219, 1991.
- Langley P., Sage S., « Induction of Selective Bayesian Classifiers », *In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, 1994.
- Lowry C. A., Woodall W. H., Champ C. W., Rigdon S. E., « A multivariate exponentially weighted moving average control chart », *Technometrics*, vol. 34, n° 1, p. 46-53, 1992.
- Montgomery D. C., *Introduction to Statistical Quality Control, Third Edition*, John Wiley and Sons, 1997.
- Naim P., Willemin P.-H., Leray P., Pourret O., Becker A., *Réseaux Bayésiens - 2ème édition*, Eyrolles, 2004.
- Pazzani M., « Searching for dependencies in Bayesian classifiers », *Learning from Data Artificial Intelligence and Statistics*, vol. 5, p. 239-248, 1997.
- Pernkopf F., « Bayesian network classifiers versus selective k-NN classifier », *Pattern Recognition*, vol. 38, n° 1, p. 1-10, 2005.
- Roberts S. W., « Control chart tests based on geometric moving averages », *Technometrics*, vol. 1, n° 3, p. 239-250, Août, 1959.
- Tax D. M. J., Duin R. P. W., « Combining One-Class Classifiers », *Lecture Notes in Computer Science*, 2001.

- Tiplica T., Kobi A., Barreau A., « Synthèse et Comparaison des Méthodes pour la Maîtrise Statistique des Processus Multivariés », *Actes du congrès QUALITA*, Annecy, France, p. 134-142, 2001.
- Verron S., Tiplica T., Kobi A., « A new procedure based on mutual information for fault diagnosis of industrial systems », *Workshop on Advanced Control and Diagnosis*, 2006.
- Verron S., Tiplica T., Kobi A., « Fault diagnosis of industrial systems with bayesian networks and mutual information », *European Control Conference*, 2007.
- Weber P., Jouffe L., « Reliability modelling with dynamic bayesian network », *SafeProcess 2003, 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Washington D.C., 2003.
- Zwingelstein G., *Diagnostic des défaillances, théorie et pratique pour les systèmes industriels*, Ed. HERMES, 1995.