

# Diagnostic et surveillance des processus complexes par réseaux bayésiens

Sylvain VERRON

LABoratoire en Sûreté de fonctionnement, QUALité et Organisation

Jeudi 13 décembre 2007



Directeur de thèse : M. Abdessamad KOBİ  
Co-encadrant de thèse : M. Teodor TIPLICA

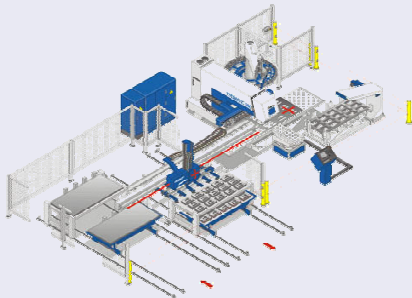


# Plan

- 1 Contexte des travaux
- 2 Détection
- 3 Diagnostic supervisé
- 4 Caractérisation de nouvelles fautes
- 5 Application sur le Tennessee Eastman Process
- 6 Conclusions et perspectives

# Procédé complexe

## Exemple

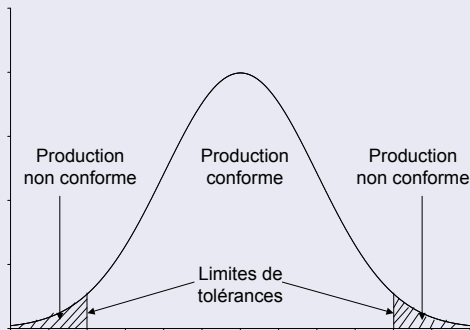


## Caractérisé par :

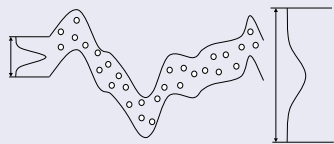
- nombreuses variables (continues et/ou discrètes)
- nombreuses corrélations entre ces variables
- base de données incomplète
- nombre important de fautes possibles

# Variations du procédé

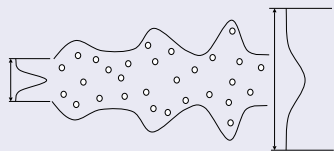
## Impacts sur la qualité



## Variations de centrage



## Variations de dispersion



# Les différentes approches

Trois principales approches pour la surveillance

## Méthodes à base de modèles analytiques

- Estimation de paramètres,
- Observateurs,
- Équations de parités,
- etc.

## Méthodes à base de connaissances

- Systèmes experts,
- AMDEC,
- Arbres de défaillances,
- etc.

## Méthodes basées sur les données

- Cartes de contrôle, ACP,
- Réseaux de neurones,
- Réseaux bayésiens,
- etc.

# Les différentes approches

Trois principales approches pour la surveillance

## Méthodes à base de modèles analytiques

- Estimation de paramètres,
- Observateurs,
- Équations de parités,
- etc.

## Méthodes à base de connaissances

- Systèmes experts,
- AMDEC,
- Arbres de défaillances,
- etc.

## Méthodes basées sur les données

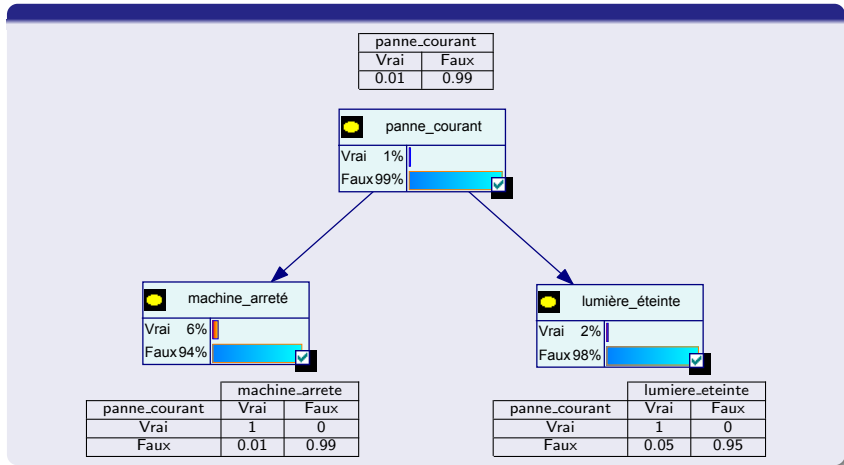
- Cartes de contrôle, ACP,
- Réseaux de neurones,
- Réseaux bayésiens,
- etc.

# Étude de différents classifieurs

## Tableau comparatif

Classifieurs	SVM	kNN	ArbD	RN	AD	MMG	RB
Variables discrètes multinomiales	-	-	+	-	-	-	+
Variables continues (normales)	+	+	-	+	+	+	+
Corrélations entre variables	+	+	+	+	+	+	+
Temps d'apprentissage admissible	-	+	+	-	+	+	+
Temps de calcul admissible	+	-	+	+	+	+	+
Pas de limitations sur le nombre de classes	-	+	+	+	+	+	+
Gestion des données manquantes	-	-	-	-	-	-	+
Représentation visuelle et adaptabilité	-	-	-	-	-	-	+
Gestion de non-linéarité	+	+	-	+	-	+	-

## Exemple RB



A

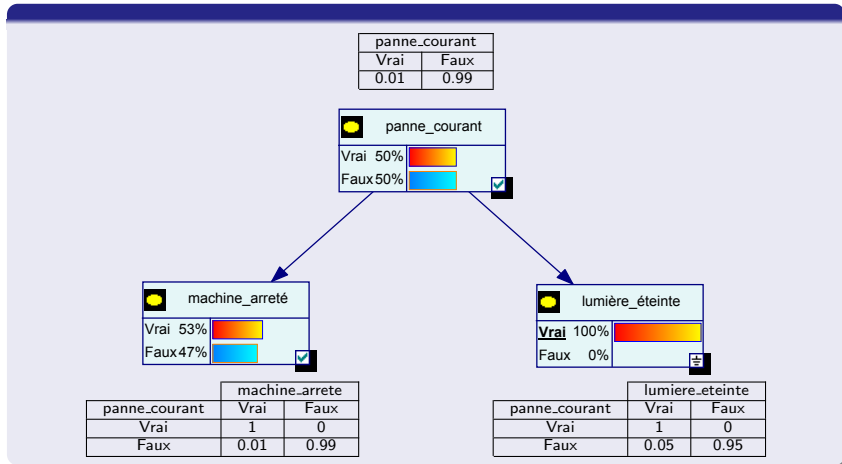
Nœud Discret Multimodal

B

Nœud Continu Gaussien



## Exemple RB



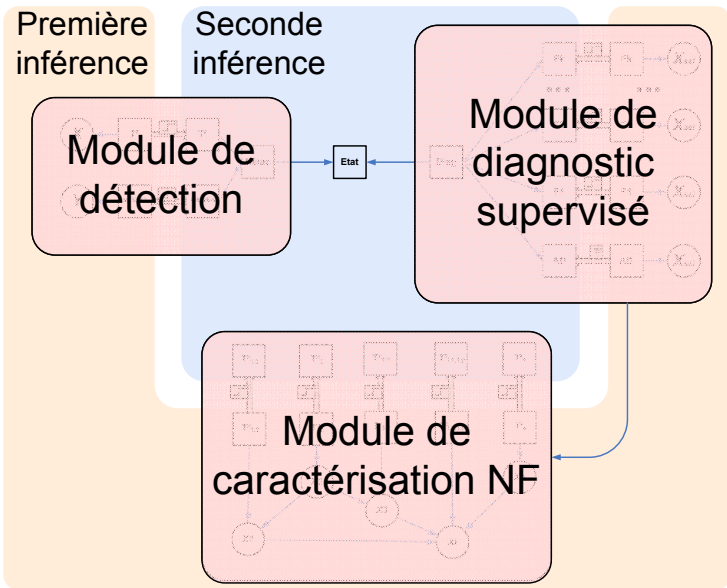
A

Nœud Discret Multimodal

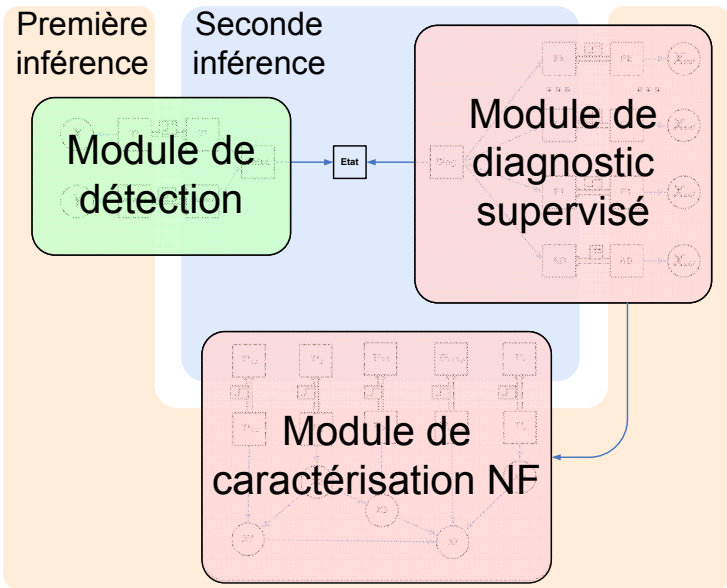
B

Nœud Continu Gaussien

# Surveillance et diagnostic par RB

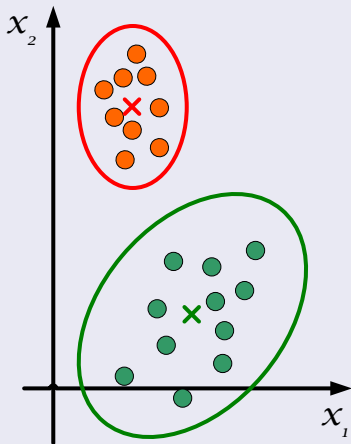


# Surveillance et diagnostic par RB



# Situation hors-contrôle

## Un espace bivarié



## Avec :

- : point sous contrôle SC
- : point hors-contrôle HC

Modèle gaussien des données :

- ✕ :  $\mu$  position de l'ellipse
- :  $\Sigma$  forme de l'ellipse

# Cartes de contrôle

## Carte du $T^2$

Distance statistique :

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

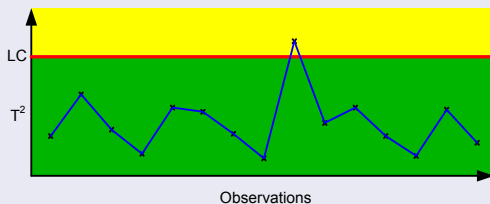
Si  $\boldsymbol{\mu}$  et  $\boldsymbol{\Sigma}$  sont connus, alors

$$T^2 \sim \chi_{\alpha, p}^2$$

## Carte MEWMA

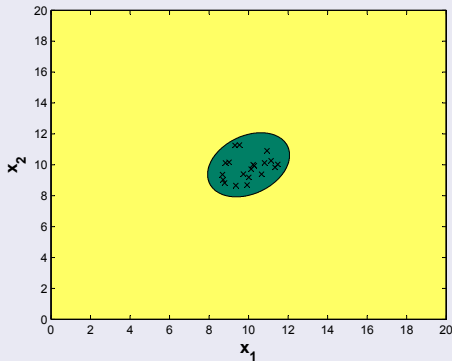
- Filtrage des données par :  
 $\mathbf{y}_t = \lambda \mathbf{x}_t + (1 - \lambda) \mathbf{y}_{t-1}$
- Carte  $T^2$  sur  $\mathbf{y}$
- Adaptée aux sauts de faibles amplitudes
- Inertie, moins performante pour les fortes amplitudes

## La carte



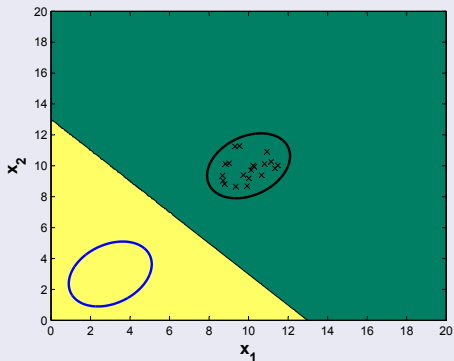
# Analogie détection - classification

## Graphes



# Analogie détection - classification

## Graphes

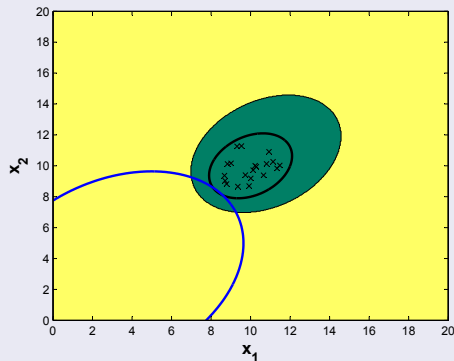


## Solution

- Création d'une "classe virtuelle HC" de paramètres  $\mu_{HC}$  et  $\Sigma_{HC}$

# Analogie détection - classification

## Graphes



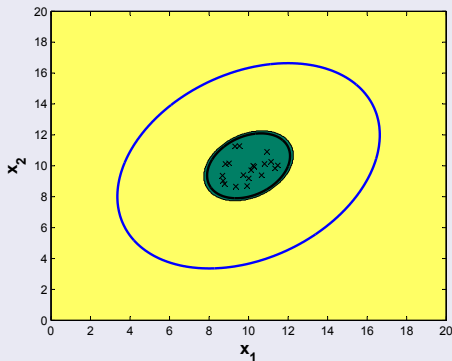
## Solution

- Création d'une "classe virtuelle HC" de paramètres  $\mu_{HC}$  et  $\Sigma_{HC}$
- Augmentation de la variabilité de HC :  $\Sigma_{HC} = c\Sigma_{SC}$



# Analogie détection - classification

## Graphes

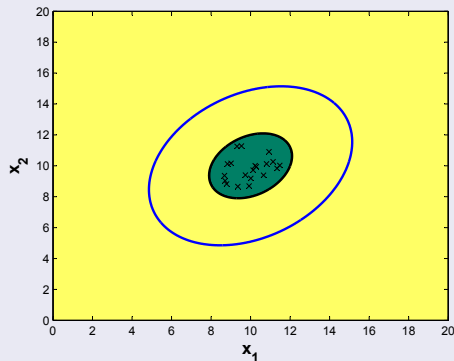


## Solution

- Création d'une "classe virtuelle HC" de paramètres  $\mu_{HC}$  et  $\Sigma_{HC}$
- Augmentation de la variabilité de HC :  $\Sigma_{HC} = c\Sigma_{SC}$
- Forme parfaitement centrée et ellipsoïdale pour  $\mu_{HC} = \mu_{SC}$

# Analogie détection - classification

## Graphes



## Solution

- Création d'une "classe virtuelle HC" de paramètres  $\mu_{HC}$  et  $\Sigma_{HC}$
- Augmentation de la variabilité de HC :  $\Sigma_{HC} = c\Sigma_{SC}$
- Forme parfaitement centrée et ellipsoïdale pour  $\mu_{HC} = \mu_{SC}$
- Ajustement de  $c$  pour respecter un taux de fausses alertes

# Le coefficient $c$

## Problème

Comment fixer  $c$  pour obtenir le taux voulu de fausses alertes  $\alpha$  ?

## Règles de décision

Hors-contrôle si :

- $T^2 > LC$  pour les cartes de contrôle
- $P(SC|x) < p(SC)$  pour les réseaux bayésiens

## Equivalence

L'équivalence entre les deux règles s'obtient en résolvant :

$$1 - c + \frac{pc}{LC(\alpha, p)} \ln(c) = 0$$

Deux racines, mais l'une est un non-sens (solution évidente  $c=1$ ).

# Cartes par RB (1)

## Carte $T^2$



Classe C	
SC	HC
$1 - \alpha$	$\alpha$



C	X
SC	$X \sim \mathcal{N}(\mu, \Sigma)$
HC	$X \sim \mathcal{N}(\mu, c \times \Sigma)$

## Carte MEWMA



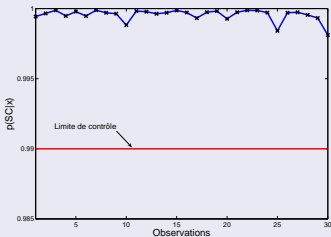
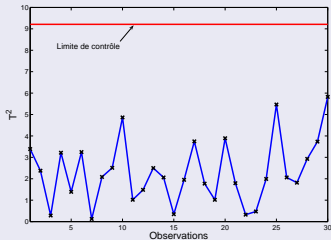
Classe C	
SC	HC
$1 - \alpha$	$\alpha$



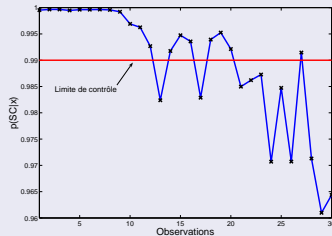
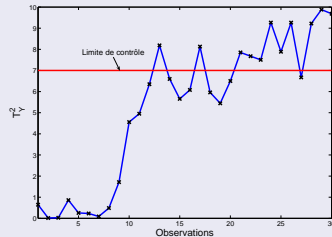
C	Y
SC	$Y \sim \mathcal{N}(\mu, (\frac{\lambda}{2-\lambda}) \Sigma)$
HC	$Y \sim \mathcal{N}(\mu, c \times (\frac{\lambda}{2-\lambda}) \Sigma)$

# Cartes par RB (2)

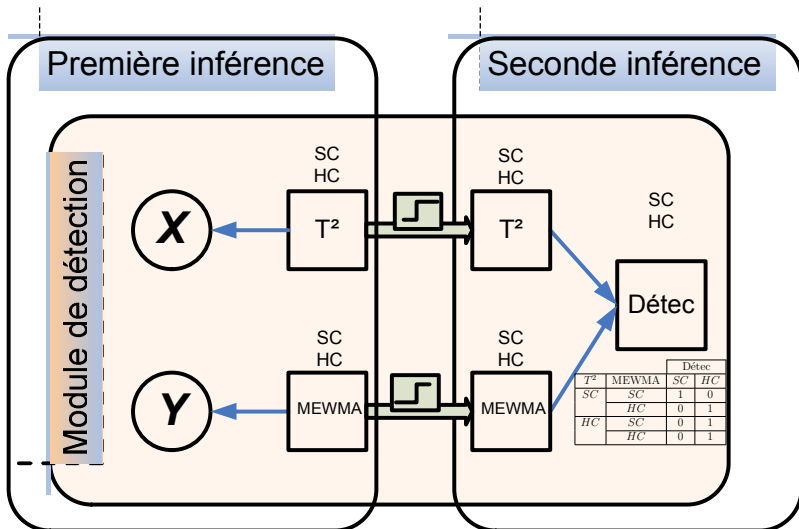
## Carte $T^2$



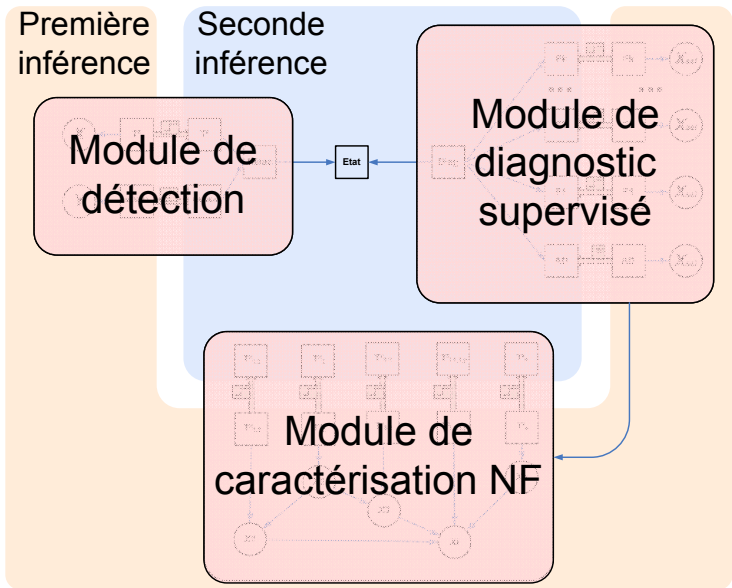
## Carte MEWMA



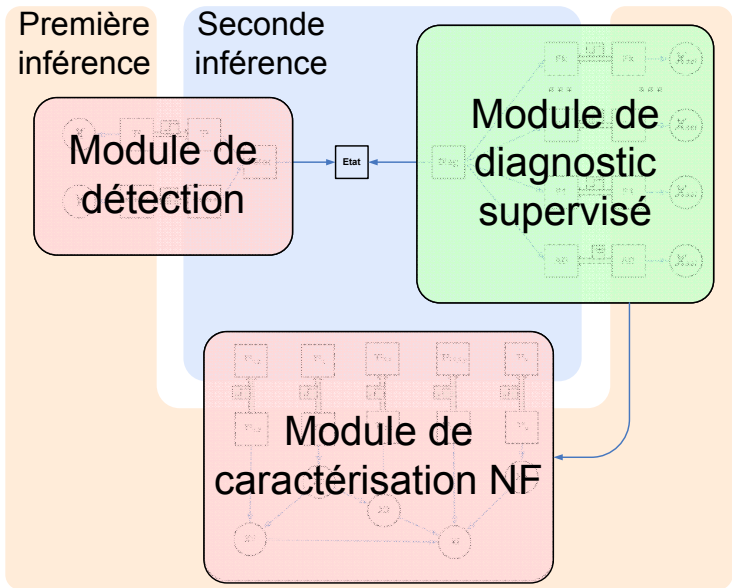
# Module de détection par RB



# Surveillance et diagnostic par RB



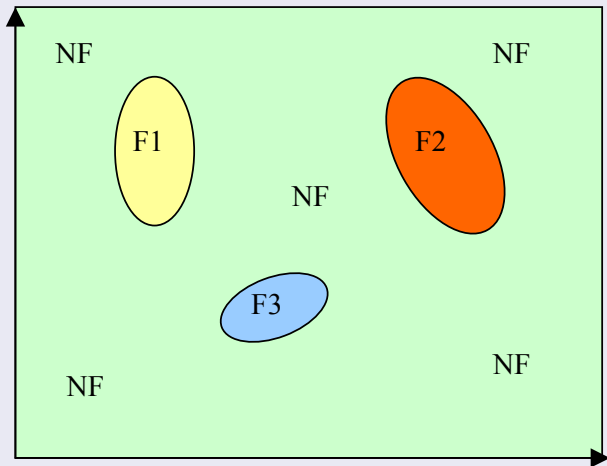
# Surveillance et diagnostic par RB





# Diagnostic supervisé

## Objectifs



# Analyse discriminante par RB

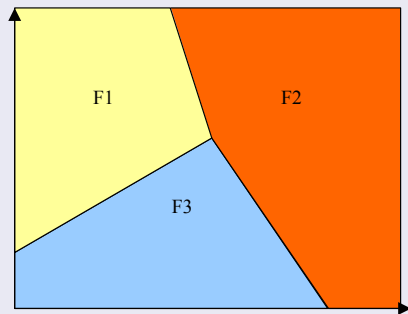
## Réseau bayésien



AD		
$F_1$	...	$F_k$
$\frac{1}{k}$	...	$\frac{1}{k}$

AD	$X$
$F_1$	$X \sim \mathcal{N}(\mu_1, \Sigma_1)$
...	...
$F_k$	$X \sim \mathcal{N}(\mu_k, \Sigma_k)$

## Zones de classification



# Analyse discriminante par RB

## Réseau bayésien

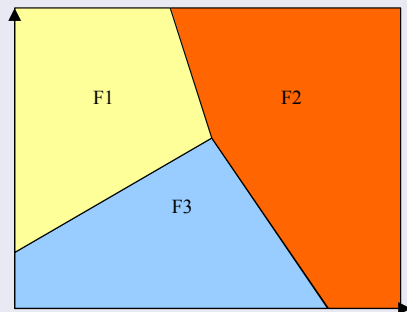


AD		
$F_1$	...	$F_k$
$\frac{1}{k}$	...	$\frac{1}{k}$



$AD$	$X$
$F_1$	$X \sim \mathcal{N}(\mu_1, \Sigma_1)$
...	...
$F_k$	$X \sim \mathcal{N}(\mu_k, \Sigma_k)$

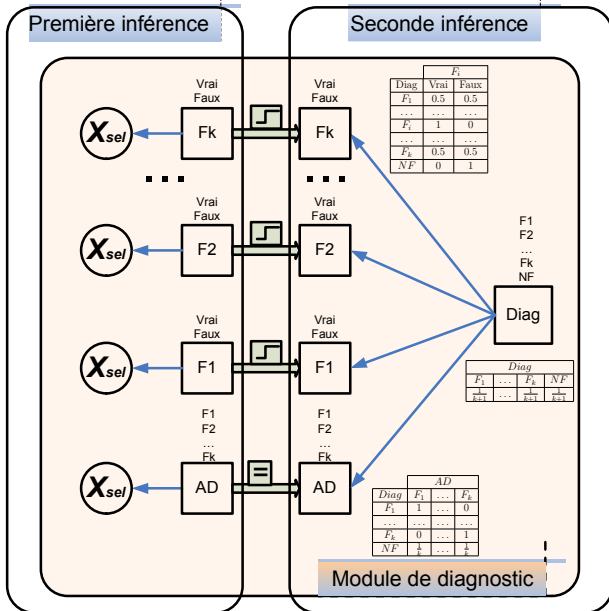
## Zones de classification



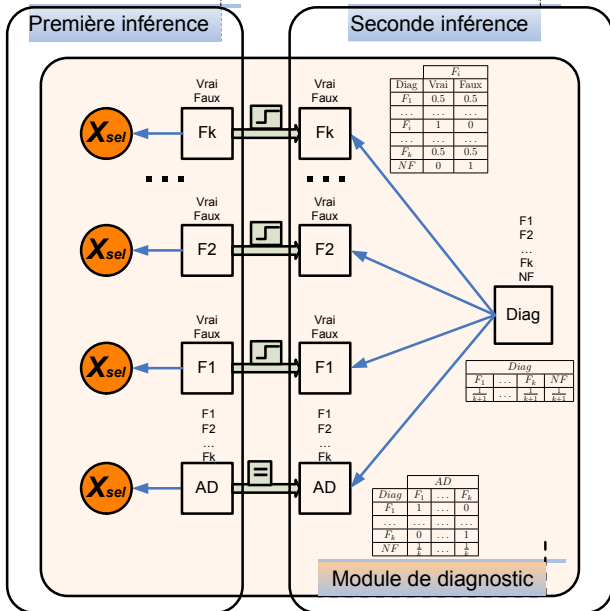
## Solution

Effectuer (en plus de la classification) une carte de contrôle sur chaque classe de faute.

# Module de diagnostic par RB



# Module de diagnostic par RB

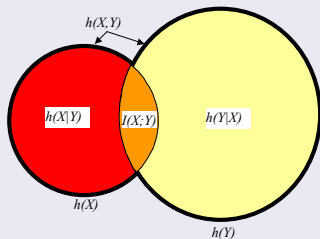


# Sélection de composantes (1)

## Entropie

Mesure de l'information contenue par une source d'information (par exemple, une variable aléatoire)

## Représentation graphique



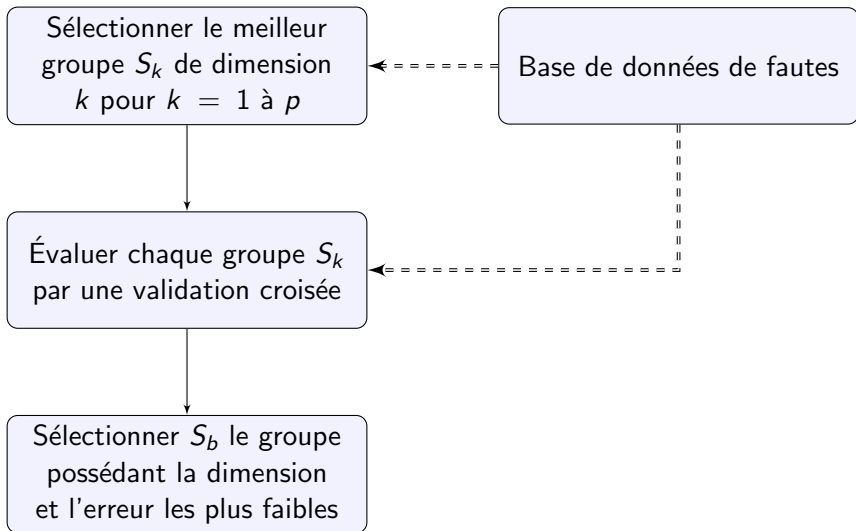
## Information Mutuelle

$$I(X; C) = \frac{1}{2} \left[ \log(\sigma^2) - \sum_{c=1}^r P(c) \log(\sigma_c^2) \right]$$

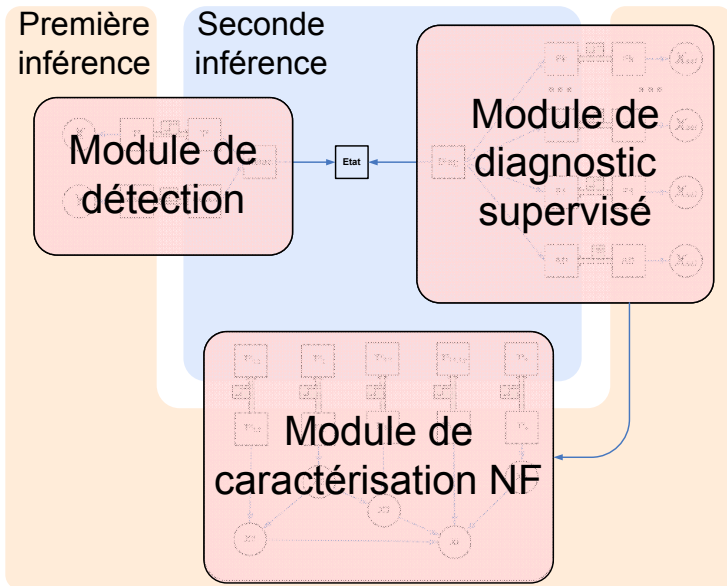
- $C$  est une variable multinomiale avec  $r$  modalités
- $\mathbf{X}$  est une variable normale multivariée de paramètres  $\boldsymbol{\mu}$  et  $\boldsymbol{\Sigma}$

$$I(\mathbf{X}; C) = \frac{1}{2} \left[ \log(|\boldsymbol{\Sigma}|) - \sum_{c=1}^r P(c) \log(|\boldsymbol{\Sigma}_c|) \right]$$

# Sélection de composantes (2)

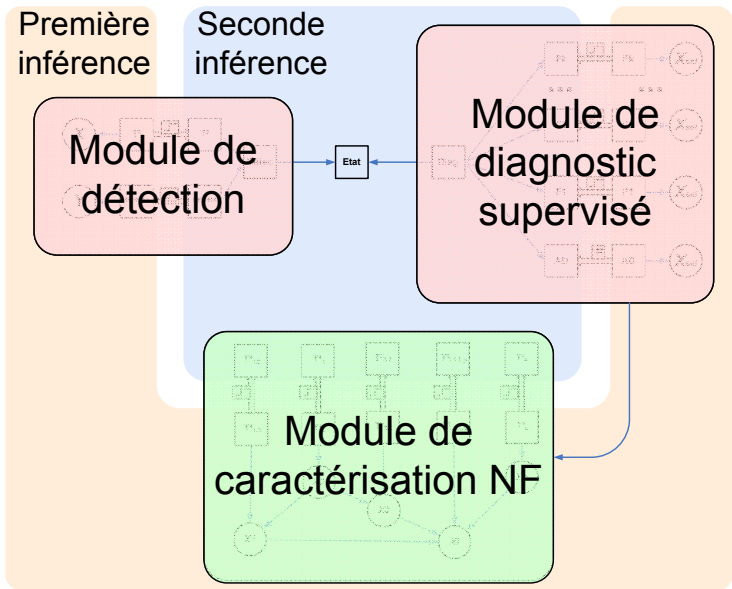


# Surveillance et diagnostic par RB





# Surveillance et diagnostic par RB



# Caractérisation de nouvelles fautes

## Objectifs

Identifier les variables responsables d'une situation hors-contrôle

## Décomposition causale du $T^2$

- Décomposition causale =  
décomposition MYT + modèle causal
- Dans ce cas, Li et al. prouvent que :

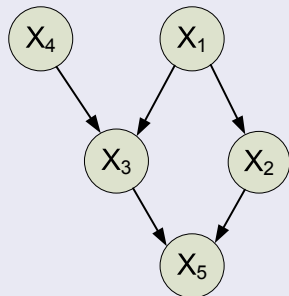
$$T^2 = \sum_{i=1}^p T_{i \bullet PA(X_i)}^2$$

- Exemple du graphe :

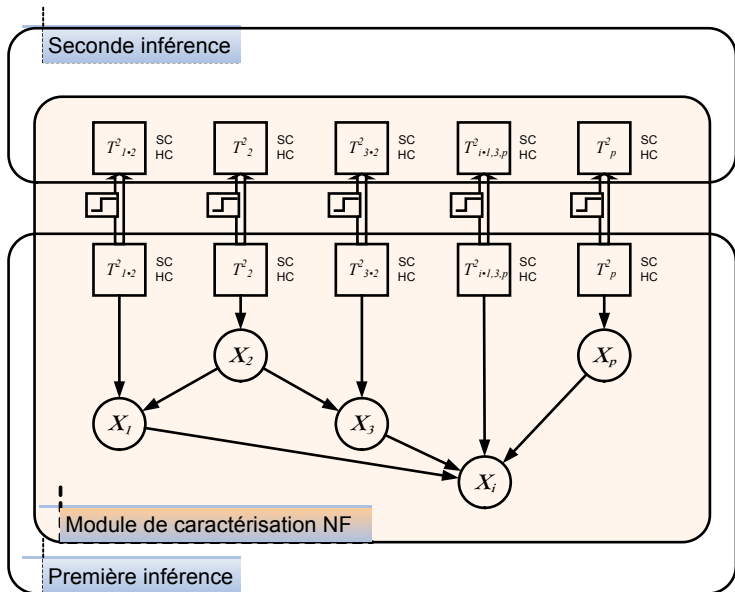
$$T^2 = T_4^2 + T_1^2 + T_{3 \bullet 1,4}^2 + T_{2 \bullet 1}^2 + T_{5 \bullet 2,3}^2$$

- $p$  termes à calculer ( $p \times 2^{p-1}$  pour MYT)
- Chaque terme suit  $\chi_{1,\alpha}^2$ , donc  
identification par cartes de contrôle

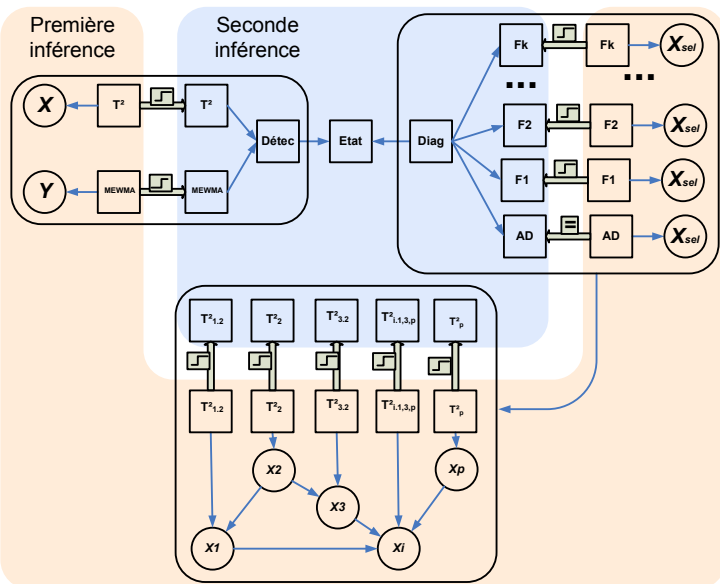
## Modèle causal linéaire gaussien



# Module de caractérisation NF par RB



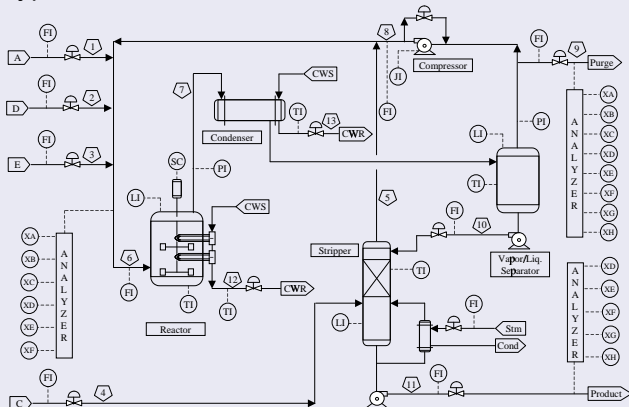
# Surveillance et diagnostic par RB



# Le Tennessee Eastman Process

## Procédé chimique complexe

- 12 variables d'entrée
- 41 variables de sortie
- 20 types de fautes identifiées



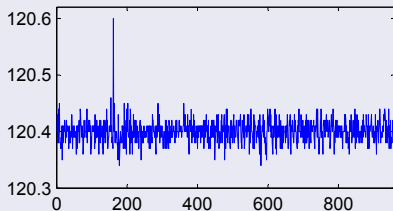
# Exemple de faute

## Faute F4

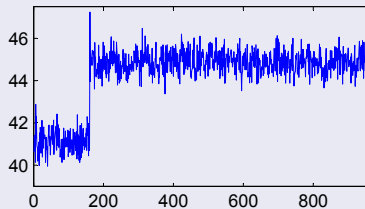
- Au réacteur : augmentation de la température d'entrée du liquide de refroidissement
- Variable 9 (sortie) : température du réacteur
- Variable 51 (entrée) : débit du refroidissement liquide

## Graphes

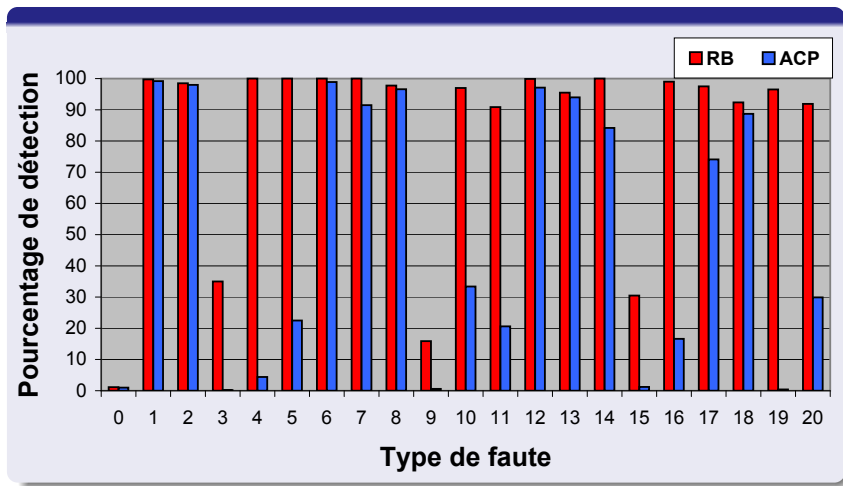
Variable 9 : Faute F4



Variable 51 : Faute F4



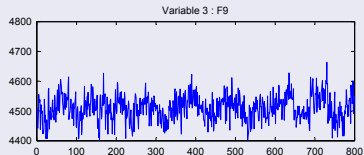
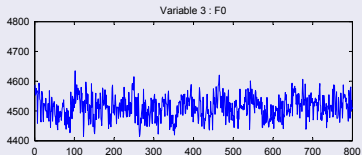
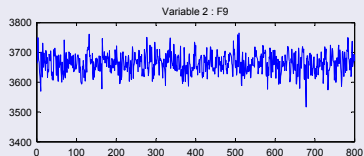
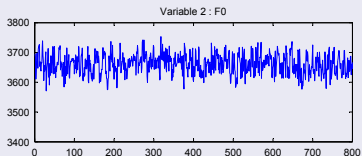
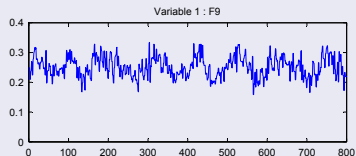
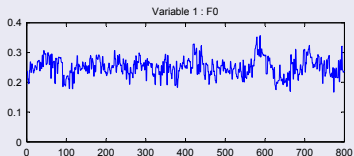
# Performances en détection



# Fautes difficilement détectables

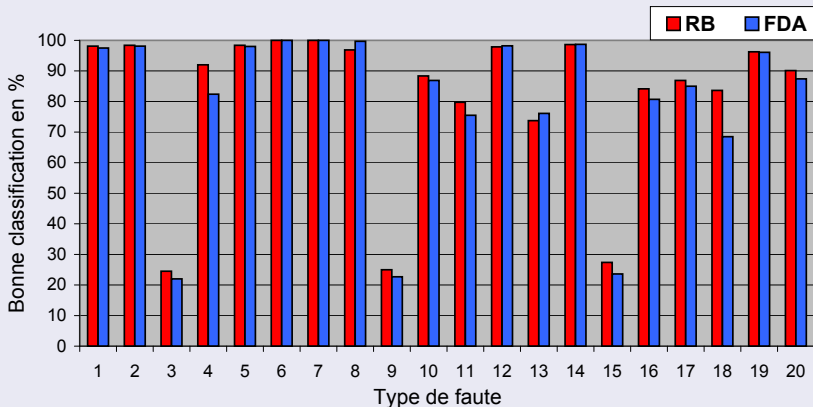
## Exemple de la faute F9

- Pour les trois premières variables d'entrée, on a :





# Performances en diagnostic



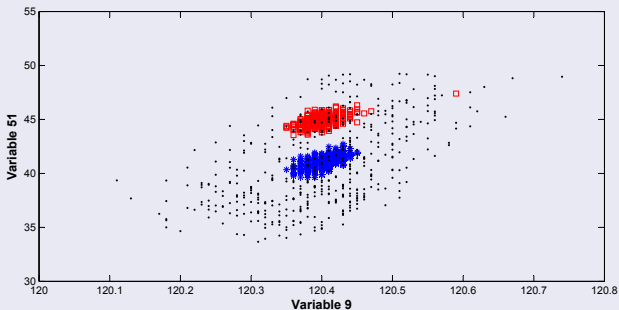
	RB	FDA
Pourcentage moyen de bonne classification	82.00	79.86

# Diagnostic : exemple (1)

## 3 types de faute

Faute	Description	Type
F4	Temp. d'entrée du ref. liq. au réacteur	Saut
F9	Temp. d'alimentation en D	Variation aléatoire
F11	Temp. d'entrée du ref. liq. au réacteur	Variation aléatoire

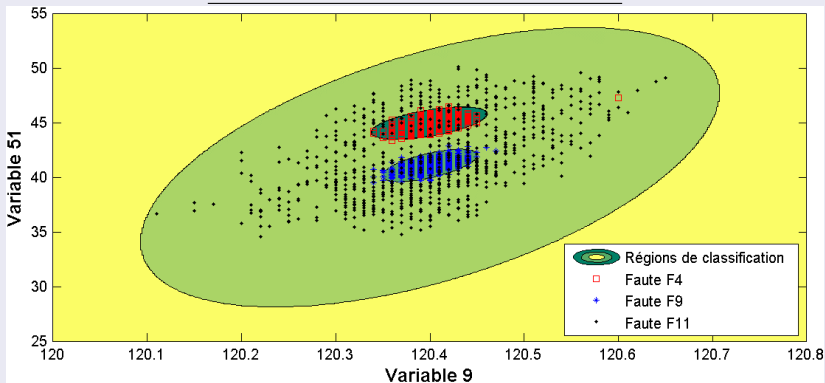
## Données d'apprentissage



# Diagnostic : exemple (2)

## Résultats : Taux d'erreur

Méthodes	Espace total	{51,9}
SVM	44%	6.5%
PSVM	35%	6.0%
ISVM	29.86%	6.0%
RB	18.83%	5.87%



# Conclusions

Dans un réseau bayésien, il est désormais possible de :

- Détecter des situations hors-contrôle pour de faibles et forts sauts d'amplitude.
- Diagnostiquer ces situations dans un espace réduit, tout en prenant en compte l'apparition de nouvelles fautes.
- Identifier les variables responsables d'une situation hors-contrôle.

Grâce aux contributions originales suivantes :

- Classification monoclasse gaussienne (carte de contrôle) dans un réseau bayésien.
- Formule analytique de l'information mutuelle entre une variable multinomiale et une variable gaussienne multivariée.
- Structure déployée sur les trois aspects : détection, diagnostic et caractérisation NF.

# Perspectives

## Les principales perspectives engendrées

- Combinaison des informations liées au diagnostic supervisé et à la caractérisation NF,
- Travaux sur les mélanges de gaussiennes pour passer outre la normalité,
- Extensions vers les diagrammes d'influence pour intégrer la notion de coûts dans les décisions prises.

# Remerciements

Merci de votre attention !

Questions ?