



**HAL**  
open science

**Diagnostic de défauts par les Machines à Vecteurs  
Supports : application à différents systèmes  
mutivariabiles nonlinéaires**

Nassim Laouti

► **To cite this version:**

Nassim Laouti. Diagnostic de défauts par les Machines à Vecteurs Supports : application à différents systèmes mutivariabiles nonlinéaires. Autre. Université Claude Bernard - Lyon I, 2012. Français. NNT : 2012LYO10161 . tel-00985437

**HAL Id: tel-00985437**

**<https://theses.hal.science/tel-00985437v1>**

Submitted on 29 Apr 2014

**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.

N° d'ordre: 161-2012

Année 2012

THÈSE DE L'UNIVERSITÉ DE LYON  
Délivrée par  
L'UNIVERSITÉ CLAUDE BERNARD LYON 1

Présentée pour obtenir le grade de docteur en automatique

ÉCOLE DOCTORALE EEA DE LYON  
(arrêté du 7 août 2006)

Par

M. Nassim LAOUTI

# Diagnostic de défauts par les Machines à Vecteurs Supports : application à différents systèmes multivariables non linéaires

Soutenue le 21 septembre 2012, devant le jury composé de

Dr. Mazen ALAMIR  
Professeur Christophe AUBRUN  
Professeur Michel KINNAERT  
Dr. Sami OTHMAN  
Dr. Isabelle QUEINNEC  
Dr. Nida SHEIBAT-OTHMAN

GIPSA-Lab, Grenoble  
CRAN, Université de Lorraine  
Université Libre de Bruxelles  
LAGEP, Université de Lyon  
LAAS, Université de Toulouse  
LAGEP, Université de Lyon

Rapporteur  
Rapporteur  
Examineur  
Co-encadrant  
Président  
Directeur de thèse



# Dédicaces

*Je dédie cette thèse à mes très chers parents à qui je dois tout ce que je suis, et qui m'ont donné un véritable modèle de travail, et qui m'ont appris le respect de l'autre, la gentillesse et la persévérance. J'espère qu'ils trouveront dans ce travail toute ma reconnaissance et tout mon amour.*

*A mes chers frères et sœurs et qui ont toujours été là pour moi: Leila, Mehdi, Sihem,  
Mohamed, Hafid, Rayane Yasmine*

*Je dédie cette thèse au directeur de l'école doctorale EEA de Lyon monsieur Gérard Scorletti.*

*A tous mes amis : Kamel, Kamel, Reda, Bassem, Sidali, Belkacem, Amine, Aziz, Sofiane et  
tous les autres qui m'ont encouragé à finir ce travail*

*Je dédie cette thèse*

***Nassim***

# Remerciements

*Je remercie Dieu de m'avoir donné la volonté et le courage qui m'ont permis de réaliser ce travail, veille-t-il me guider dans le droit chemin.*

*Je remercie Madame Nida Othman et Monsieur Sami Othman, pour leurs gentilleses ainsi que leurs encouragements réguliers face à la difficulté.*

*Je remercie ma sœur Leila pour son aide depuis toujours*

*Je remercie aussi les membres du jury pour avoir accepté de juger ce travail. Je suis très honoré de présenter mes travaux de thèse devant un tel jury.*

*Ce travail a été réalisé au sein du Laboratoire d'Automatique et de Génie des Procédés de l'université Claude Bernard Lyon1 (LAGEP-UMR 5007). Mes remerciements s'adressent à tous les membres du laboratoire LAGEP que ce soit les doctorants, les enseignants, ingénieurs ou les chercheurs pour m'avoir accueilli parmi eux.*

# Table des matières

<b>Introduction générale .....</b>	<b>1</b>
<b>CHAPITRE I : Diagnostic de défauts et théorie des Machines à Vecteurs de Supports .....</b>	<b>5</b>
I.1. Généralités .....	7
I.1.1. Terminologie en matière de diagnostic.....	7
I.1.2. Le principe de détection de défaut.....	9
I.1.3. FDI en boucle ouverte et en boucle fermée .....	10
I.1.4. Domaines concernés par la FDI.....	11
I.1.5. Performances des techniques FDI .....	12
I.2. Les méthodes de diagnostic .....	12
I.2.1. Méthodes à base de modèle explicite .....	13
I.2.2. Méthodes basées sur les données.....	18
I.2.3. Etapes de mise en œuvre des méthodes basées sur les données .....	28
I.3. Les Machines à Vecteurs de Supports .....	39
I.3.1 Théorie de l'apprentissage statistique .....	40
I.3.2. Les machines à Vecteurs de Support pour la classification (SVM) .....	41
I.3.3. Les machines à Vecteurs de supports pour la régression (SVR).....	56
<b>CHAPITRE II : Système de traitement des eaux usées.....</b>	<b>67</b>
II.1. Description du système .....	69
II.1.1. Le procédé de traitement des eaux usées par boues activées .....	70
II.1.2. La modélisation .....	71
II.1.3. Le modèle BSM1.....	73

II.1.4. Variables composites/mesures .....	84
II.2. Commande des systèmes de traitement des eaux usées .....	86
II.3. Revue des méthodes de diagnostic du procédé WWTP .....	87
II.3.1. ACP ( <i>PCA en anglais</i> ) et d'autres méthodes statistiques.....	88
II.3.2. La Projection aux Structures Latentes ( <i>PLS</i> ) .....	90
II.3.3. SVM .....	91
II.4. Stratégie SVR/SVM proposée pour le diagnostic du procédé WWTP .....	92
II.4.1. Schéma SVR/SVM proposé .....	92
II.4.2. Récapitulatif des variables d'état/ d'entrée/ et de sortie.....	94
II.4.3. Les variables à superviser.....	95
II.5. Implémentation de SVR pour la modélisation .....	97
II.5.1. Génération de données pour l'apprentissage et la validation .....	97
II.5.2. Prétraitement des données (normalisation, filtrage).....	99
II.5.3. Analyse de corrélation/Information mutuelle.....	100
II.5.4. Résultats de la modélisation SVR .....	103
II.6. Implémentation de SVM pour le diagnostic de défauts .....	116
II.7. Conclusion.....	129
<b>CHAPITRE III : Diagnostic de défauts dans les éoliennes .....</b>	<b>131</b>
III.1. Introduction.....	133
III.1.1. Revue des méthodes de diagnostic sur les éoliennes .....	135
III.1.2. Revue sur la commande des éoliennes.....	138
III.2. Description du système éolien .....	139
III.2.1. Modèle des pales et pitch.....	141

III.2.2. Modèle convertisseur et génératrice .....	141
III.2.3. Modèle du système d'entraînement .....	142
III.2.4. Le système de contrôle.....	143
III.2.5. Défauts considérés dans ce modèle éolien .....	144
III.3. Diagnostic de défauts dans une éolienne .....	146
III.3.1. Les vecteurs des caractéristiques pour chaque défaut.....	148
III.3.2. Résultats de l'approche de diagnostic .....	151
III.4. Comparaison entre différentes méthodes de diagnostic.....	157
III.5. Combinaison entre l'approche basée sur SVM et un observateur .....	159
III.5.1. Conception de l'observateur de type Kalman .....	161
III.5.2. L'approche basée SVM pour la détection/isolation de défaut actionneur sur le système pitch.....	162
III.6. Conclusion .....	166
<b>CHAPITRE IV : Réacteur chimique parfaitement agité .....</b>	<b>169</b>
IV.1. Revue bibliographique.....	171
IV.2. Présentation du système.....	174
IV.2.1. Le procédé.....	174
IV.2.2. Le modèle .....	175
IV.3. Implémentation du diagnostic de défauts .....	176
IV.3.1 Implémentation des SVMs.....	176
IV.3.2 Implémentation de l'observateur de grand gain .....	178
IV.3.3. Résultats et discussion .....	180
IV.4. Implémentation de SVR.....	189
IV.4.1. Identification d'une dérive de température par régression SVM .....	190

IV.4.2. Identification des U par régression SVM .....	193
IV.5. Conclusions.....	196
<b>Conclusion générale .....</b>	<b>197</b>
<b>Notations .....</b>	<b>201</b>
<b>Annexes .....</b>	<b>209</b>
Annexe A .....	211
Annexe B.....	219
Annexe C.....	223
<b>Bibliographie .....</b>	<b>227</b>

# Table des Figures

Figure I.1	Architecture du fonctionnement d'un système avec une commande tolérante aux défauts .....	9
Figure I.2	Schéma classique de détection des défauts .....	10
Figure I.3	Schéma de classification des méthodes de diagnostic de défauts basées sur le modèle .....	13
Figure I.4	Méthodes de diagnostic dirigées par les données.....	19
Figure I.5	Taux de représentativité des composantes principales dans le modèle.....	21
Figure I.6	L'hyperplan séparateur pour des données bidimensionnelles.....	41
Figure I.7	Séparation linéaire.....	42
Figure I.8	Séparation linéaire avec des erreurs .....	45
Figure I.9	Marge souple .....	46
Figure I.10	Exemple de projection.....	48
Figure I.11	Vecteurs support et surface de décision calculée par la méthode à noyau.....	49
Figure I.12	La fonction de perte insensible à $\epsilon$ dans la régression SVR linéaire.....	58
Figure I.13	Structure de type Hammerstein .....	61
Figure I.14	Structure de type Wiener.....	62
Figure II.1	Présentation générale du procédé BSM1 .....	74
Figure II.2	La matière organique dans le modèle BSM1 .....	79
Figure II.3	Les composés azotés dans le modèle BSM1.....	79
Figure II.4	Stratégie SVR suivi par SVM pour le diagnostic de défauts .....	93
Figure II.5	Concentration de nitrites/nitrates dans la 2 <sup>ème</sup> cuve utilisée pour l'apprentissage et le test du modèle SVR.....	98
Figure II.6	Concentration d'oxygène dissous dans la 5 <sup>ème</sup> cuve utilisée pour l'apprentissage et le test du modèle SVR.....	98
Figure II.7	Boîte à moustache des variables du procédé WWTP non normées.....	99
Figure II.8	Boîte à moustache des variables du procédé WWTP normalisées avec la méthode min/max .....	100
Figure II.9	Analyse des relations entre variables avec l'information mutuelle .....	102
Figure II.10	Analyse des relations entre variables avec l'information mutuelle (pour COD <sub>e</sub> et TSS <sub>e</sub> ) .....	102

Figure II.11	Evolution temporelle des variables d'entrée/sortie des matières en suspension...	104
Figure II.12	Densité de probabilité $Q_{in}$ .....	105
Figure II.13	Concentration $S_{NH,in}$ et sa densité de probabilité .....	106
Figure II.14	L'évolution de la variable $S_{o5}$ par rapport à $K_{la5}$ .....	106
Figure II.15	L'évolution de la variable $S_{no2}$ par rapport à $Q_a$ .....	107
Figure II.16	Tracé de $S_{no2}$ et son estimée.....	109
Figure II.17	Zoom sur le tracé de $S_{no2}$ et son estimé.....	110
Figure II.18	Tracé de $TSS_e$ et son estimé.....	110
Figure II.19	Zoom sur le tracé de $TSS_e$ et son estimé.....	111
Figure II.20	Tracé de $COD_e$ et son estimé .....	111
Figure II.21	Tracé de $S_{o5}$ et son estimé.....	112
Figure II.22	Résidu $S_{o5}$ .....	112
Figure II.23	Résidu $S_{no2}$ .....	113
Figure II.24	Résidu filtrée de $S_{no2}$ .....	113
Figure II.25	Densité de probabilité $S_{no2}$ .....	114
Figure II.26	Autocorrélation du résidu $TSS_e$ .....	114
Figure II.27	Résidu sur $TSS_e$ .....	115
Figure II.28	Densité de probabilité $TSS_e$ .....	115
Figure II.29	Résidu $COD_e$ .....	115
Figure II.30	Évolution du résidu $K_{la5}$ .....	119
Figure II.31	Zoom sur le résidu $K_{la5}$ .....	119
Figure II.32	Évolution du résidu $Q_a$ .....	120
Figure II.33	Évolution du résidu $S_{o5}$ .....	120
Figure II.34	Tracé de la mesure de l'oxygène dissous ( $S_{o5}$ ) et son résidu.....	121
Figure II.35	Zoom sur de la mesure de $S_{o5}$ et son résidu .....	121
Figure II.36	Évolution du résidu $S_{no2}$ .....	122
Figure II.37	La mesure de $S_{no2}$ et son résidu.....	122
Figure II.38	Mesure $TSS_e$ et son résidu .....	123
Figure II.39	Résultat de décision sur le défaut $Q_a$ .....	124
Figure II.40	Zoom sur le résultat du défaut $Q_a$ .....	125
Figure II.41	Résultat de décision sur le défaut $K_{la5}$ .....	126
Figure II.42	Zoom de décision sur le défaut $K_{la5}$ .....	126
Figure II.43	Résultat de décision sur le défaut $S_{no2}$ .....	127
Figure II.44	Zoom sur la décision du défaut $S_{no2}$ .....	127

Figure II.45	Résultat de la décision sur défaut TSS <sub>e</sub> .....	128
Figure II.46	Résultat de la décision sur défaut S <sub>o5</sub> .....	129
Figure III.1	Eoliennes offshore.....	133
Figure III.2	Destruction d'une éolienne par prise de feu .....	134
Figure III.3	Coupe transversale d'une nacelle d'éolienne .....	140
Figure III.4	Une vue d'ensemble du modèle de référence de l'éolienne.....	140
Figure III.5	Puissance de l'éolienne en fonction de la vitesse du vent .....	143
Figure III.6	Séquence vitesse du vent .....	144
Figure III.7	Schéma de diagnostic de défaut pour les éoliennes.....	146
Figure III.8	Stratégie de diagnostic de défaut classique avec SVM .....	147
Figure III.9	Schéma de diagnostic de défauts proposé .....	147
Figure III.10	Stratégie proposée de diagnostic de défaut basée sur SVM .....	147
Figure III.11	Vitesse du vent mesurée .....	152
Figure III.12	Puissance électrique générée .....	152
Figure III.13	Détection et isolation du défaut 2 de type gain (1b) sur le capteur de la position pitch $\beta_{2,m_2}$ .....	153
Figure III.14	Détection et isolation du défaut 4 valeur fixe (type 2a) sur la vitesse du rotor ...	154
Figure III.15	Détection et isolation du défaut 5 de type gain (2b) sur le capteur de la vitesse rotationnelle du rotor $w_{r,m_2}$ .....	155
Figure III.16	Détection et isolation du défaut 5 de type gain (3b) sur le capteur de la vitesse rotationnelle de la génératrice $w_{g,m_1}$ .....	156
Figure III.17	Détection et isolation du défaut 8 de type offset (4a) sur le convertisseur.....	156
Figure III.18	Détection et isolation du défaut 9 (6) sur le système d'entraînement .....	157
Figure III.19	Schéma de combinaison SVM et observateur .....	161
Figure III.20	Détection et isolation du défaut 6 (5a) sur l'actionneur2 et estimation du coefficient d'amortissement $\zeta_{2e}$ et de la pulsation naturelle $w_{n2}$ avec l'observateur .....	164
Figure III.21	Détection et isolation du défaut 7 (5b) sur l'actionneur 3 et estimation du coefficient d'amortissement $\zeta_{3e}$ et de la pulsation naturelle $w_{n3}$ avec l'observateur .....	164
Figure IV.1	Schéma simplifié du réacteur parfaitement agité du laboratoire (Tr1, Tr2, Tj1, Tj2 et THC sont des sondes de température P1, P2, P3 et P4 sont des pompes).....	174
Figure IV.2	Schéma de diagnostic de défauts par SVM en ligne .....	180
Figure IV.3	Dérive dans T <sub>j,in</sub> à 55min.....	181

Figure IV.4 Classification des défauts de $T_{HC}$ . Effet de $\sigma$ (variance du noyau gaussien) sur les hyperplans optimaux (avec $C=100$ ) .....	182
Figure IV.5 Cas d'étude par simulation (avec $C=100$ ) et différentes valeurs de $\sigma$ : Occurrence de défaut dans $T_{HC}$ (objective de commande $T_r^d=70^\circ C$ ) .....	183
Figure IV.6 Classification des défauts de $T_{HC}$ . Effet de $C$ (paramètre de régularisation) sur les hyperplans optimaux (avec $\sigma=10$ ) .....	184
Figure IV.7 Défaut dans le système de chauffe/refroidissement (HC : un actionneur) en changeant les paramètres du PID afin d'écarter $T_{HC}$ de $T_j^d$ ( $T_j^d=70^\circ C$ ) .....	185
Figure IV.8 Dérive dans $T_{r1}$ à 33 min dans le réacteur du laboratoire.....	186
Figure IV.9 Erreur de calibration de $-5^\circ C$ est considérée dans la sonde $T_{r2}$ dans le réacteur pilote de 250 L .....	187
Figure IV.10 Une évaluation expérimentale de la méthode SVM et de l'observateur dans un réacteur pilote de 250 L .....	188
Figure IV.11 Evolution des températures dans le réacteur pilote sans défaut .....	188
Figure IV.12 Schéma de SVR pour l'estimation de $U$ et l'identification de niveaux de défauts pendant la période la chauffe du réacteur .....	189
Figure IV.13 Validation expérimentale de l'identification d'une dérive dans la sonde de température $T_{r1}$ avec différentes valeurs de $\sigma$ ( $\epsilon=0.01$ , $C=100$ , objectif de commande $T_j^d=70^\circ C$ ) .....	191
Figure IV.14 Identification de la dérive dans la sonde $T_{r1}$ avec différentes valeurs de $\epsilon$ ( $\sigma=60$ , $C=100$ , objectif de commande $T_r^d=60^\circ C$ ).....	192
Figure IV.15 Identification de la dérive du capteur $T_{r1}$ avec différentes valeurs de $C$ ( $\epsilon=0.01$ , $\sigma=60$ ).....	193
Figure IV.16 Estimation du coefficient d'échange thermique $U$ pendant la chauffe du réacteur (simulation avec différentes valeurs de $U$ , paramètre de l'observateur $\theta=0.04$ ) .....	195
Figure IV.17 Estimation expérimentale du coefficient d'échange thermique $U$ pendant la chauffe du réacteur.....	195

# Liste des tableaux

Tableau II.1	Les concentrations intervenant dans le modèle BSM1.....	75
Tableau II.2	Paramètres du modèle BSM1 .....	77
Tableau II.3	Paramètres de décantation .....	83
Tableau II.4	Corrélation des variables avec COD <sub>eff</sub> .....	101
Tableau II.5	Performances des modèles SVR.....	109
Tableau II.6	Les différents défauts considérés.....	109
Tableau II.7	Résultats de détection de défauts.....	117
Tableau III.1	Paramètres du modèle de référence de l'éolienne.....	143
Tableau III.2	Défauts considérés dans le modèle éolien de référence .....	145
Tableau III.3	Résultats de diagnostic du défaut 1 .....	158
Tableau III.4	Résultats de diagnostic du défaut 2 .....	159
Tableau III.5	Résultats de détection des défauts.....	166
Tableau IV.1	Vecteurs d'entrés SVM ( $x_i$ ), ( $i$ est l'instant de temps).....	177
Tableau IV.2	Correspondance entre résidu de la température de la double enveloppe et du bain thermostaté (R=résidu).....	178



# Introduction générale

« *Tout ce que je sais, :c'est que je ne sais rien* » Socrate

La société actuelle est très orientée vers la technologie et dépend ainsi de la disponibilité, la sécurité et la fiabilité de technologies de plus en plus complexes. Cependant, la maîtrise parfaite de ces systèmes reste un sujet important et ouvert. Afin de satisfaire le bon fonctionnement, le diagnostic de défaut « *Fault detection and isolation FDI* » apparaît ainsi comme une nécessité qui consiste à identifier les causes du défaut à partir des symptômes ou signaux perçus afin d'éviter toute propagation de ce défaut. Ce dernier peut occasionner notamment :

1. Un arrêt de fonctionnement du système,
2. Des dégâts matériels,
3. La pollution de l'environnement,
4. Des dangers pour l'utilisateur.

Le diagnostic de défauts est, en fait, l'ensemble des actions mises en œuvre pour détecter, localiser et identifier tout dysfonctionnement sur un système. Ceci peut se faire à partir des informations pouvant être récupérées du système, telles que les mesures fournies par les capteurs, les sorties calculées par les contrôleurs, et les connaissances sur la structure ou le modèle du système. Parmi les intérêts du diagnostic de défauts figure l'optimisation des opérations de maintenance, la minimisation de l'arrêt du service, l'assurance d'une bonne qualité du produit, ainsi que la reconfiguration de la commande.

La conception d'un système de diagnostic performant n'est, cependant, pas une tâche facile à réaliser, puisque la majorité des systèmes réels sont de nature non-linéaire faisant intervenir un grand nombre de variables et travaillant dans un environnement entaché de bruit, et affecté par différents types de perturbations. Depuis des années, un grand nombre de chercheurs se sont intéressés à la résolution du problème de diagnostic des défauts. On trouve ainsi plusieurs études dans la littérature qui proposent différentes techniques de diagnostic de défauts. D'une façon générale, on peut décomposer les méthodes de diagnostic en deux grandes familles : des méthodes basées sur le modèle, et des méthodes basées sur les données récupérées à partir du système. Cependant, la complexité des systèmes industriels rend difficile, voire impossible, l'obtention d'un modèle fondamental exact et précis représentant le système dans toutes les conditions de fonctionnement afin de l'utiliser pour réaliser un diagnostic de défaut par modèle. D'un autre côté, la technologie moderne nous offre un grand bénéfice puisqu'elle fournit un grand nombre d'informations (données) mais dont l'analyse n'est pas évidente, ce qui explique

le recours à des méthodes de diagnostic de défauts basées sur les données. En plus la nature des signatures(ou symptômes) de défauts qui sont des grandeurs aléatoires rend la détection d'une différence éventuelle entre la signature nominale et celle mesurée ne peut pas se faire facilement avec les méthodes déterministes.

Pour ces raisons, l'intérêt des industriels pour les méthodes de diagnostic de défauts dirigées par des données notamment celle de l'intelligence artificielle (réseaux de neurones, systèmes experts,..) prend de plus en plus d'ampleur. L'intelligence artificielle a pour but d'élaborer des systèmes ayant des capacités intellectuelles comparables à l'homme capable de prendre des décisions d'une manière automatique dans l'environnement perçu. Dans ce vaste domaine, on trouve un axe de recherche assez important qui est l'apprentissage automatique « *Machine Learning* » (Ao et al., 2010) qui permet de bien comprendre un phénomène ou un système illustré par un nombre limité d'exemples de données afin de construire des modèles pouvant être appliqués à des nouvelles observations (exemples). Dans cette thèse, la méthode d'intelligence artificielle stochastique Machines à Vecteurs Supports (ou Séparateur à Vaste Marge, *Support Vector Machines*, **SVM**) est utilisée pour la conception des méthodes de diagnostic avancées.

La méthode SVM est issue de la théorie de l'apprentissage statistique (Vapnik, 1998) introduite par Vapnik dans les années 90 (Boser et al., 1992). Elle a été conçue au départ pour trouver un discriminateur optimal entre deux classes de données, et a été ensuite étendue pour le cas de la régression. La capacité de généralisation, et l'aspect stochastique ont favorisé l'utilisation de cette méthode d'apprentissage statistique pour le diagnostic de défauts. Dans ce cas, le diagnostic de défaut est assimilé à un problème de classification (en deux classes : normale et anormale). Dans cette méthode, le choix et l'extraction des caractéristiques représentatives de chaque défaut représentent l'étape la plus importante afin d'être sensible aux symptômes du défaut et d'obtenir ainsi un système de diagnostic réactif et précis.

Ce travail contribue à développer une nouvelle approche originale de diagnostic de défauts assez performante en termes de précision, temps de détection et permettant un pourcentage quasi nul des fausses alarmes. La méthode SVM est utilisée en premier choix pour la détection et l'isolation de défaut. Dans certains cas, la modélisation par SVM pour la régression « *SVR* » est réalisée en parallèle afin d'ajouter des estimations au vecteur de mesure. Dans d'autres cas, une combinaison avec des méthodes basées sur le modèle (principalement l'observateur à grand gain) a été nécessaire afin de garantir une bonne isolation de défaut. Cette stratégie est appliquée aux systèmes réels à l'échelle de laboratoire/pilote dans le cas d'un **réacteur chimique**, et en simulation dans le cas de **procédé de traitement des eaux usées** et du système **éolien**.

## **Ce mémoire est organisé comme suit :**

Le **premier chapitre** est dédié à la revue des méthodes de diagnostic et leurs applications suivie d'une présentation de la théorie des méthodes utilisées dans cette thèse, à savoir la méthode SVM, SVR, et quelques observateurs non linéaires.

Le **deuxième chapitre** concerne la première application de cette thèse qui est le système de traitement des eaux usées « *Waste Water Treatment Process WWTP* ». Ce chapitre présente tout d'abord le procédé et son modèle utilisé pour la simulation. Ensuite, la méthode SVR est utilisée pour la modélisation (entrée/sortie) de certaines variables clés. La méthode SVM est également utilisée pour le diagnostic de défaut dans ce système en utilisant les variables estimées par SVR en plus des variables mesurés en ligne. Enfin, quelques résultats de simulation sont discutés.

Le **troisième chapitre** est consacré au système éolien « *Wind Turbine* » en appliquant la méthode SVM et un observateur de type Kalman pour le diagnostic de défaut. On utilisera un benchmark disponible sur internet pour cette application. La stratégie proposée dans ce chapitre par l'auteur de cette thèse (Laouti et al., 2011a) a démontré son efficacité en se classant première par rapport à 12 autres méthodes lors d'une compétition internationale, organisée par la société kk-electronic, sur le diagnostic de défaut dans les éoliennes .

Le **quatrième chapitre** présente l'application de la méthode SVM (classification et régression) aux réacteurs chimiques parfaitement agités « *Continuous Stirred Tank Reactor, CSTR* ».

Dans chaque application, une revue des méthodes de diagnostic employées dans le domaine en question est présentée ainsi que le système et son modèle. Par la suite la manière choisie pour appliquer les SVM est définie et les résultats de simulation sont discutés.

Le mémoire se termine par une conclusion générale résumant les contributions de la thèse et présentant les perspectives de recherche que je souhaite poursuivre dans l'avenir ainsi que quelques réflexions sur ces prochaines voies de recherches.

## **Parmi les contributions de cette thèse on cite :**

1. Une revue de la méthode SVM et ses applications pour le diagnostic de défauts.
2. La combinaison entre une méthode de diagnostic basée sur les données « SVM » avec une méthode basée sur le modèle « Observateur » afin de tirer profit des avantages de chaque méthode.

3. L'application de la méthode SVM dans trois procédés différents : un réacteur parfaitement agité, un système éolien et un système de traitement des eaux usées.

4. Un capteur logiciel basé sur SVM régression « SVR » est développé et sa capacité de prédiction et de généralisation face à un environnement bruité et assez perturbé est étudiée.

# **CHAPITRE I : Diagnostic de défauts et théorie des Machines à Vecteurs de Supports**



Dans ce chapitre, nous présentons d'une manière simplifiée les concepts basiques de diagnostic de défauts, et les fondements théoriques des méthodes utilisées dans cette thèse telles que les Machines à Vecteurs Supports SVM, Observateurs, Analyse en Composantes Principales, ... etc.

## **I.1. Généralités**

La recherche constante d'une meilleure compétitivité pousse les entreprises à produire des machines ayant une grande fiabilité et disponibilité. Cependant, un système quelle que soit sa modernité et sa robustesse, est affecté par des dysfonctionnements qui peuvent compromettre son bon fonctionnement. Il apparaît donc nécessaire d'identifier rapidement les changements inattendus (défauts).

Dans cette première partie, nous allons d'abord présenter les principes de diagnostic de défauts et quelques définitions de la terminologie dans ce domaine.

### **I.1.1. Terminologie en matière de diagnostic**

Partant du fait que le diagnostic de défaut est réparti sur plusieurs domaines technologiques, un comité technique *SAFEPROCESS* de la Fédération internationale de contrôle automatique IFAC a essayé de proposer une terminologie unifiée (Isermann and Ballé, 1997). La terminologie définie par ce comité sera utilisée dans ce mémoire :

**Défaut** : Écart non autorisé d'au moins une propriété caractéristique ou variable du système à partir du comportement acceptable / habituelle /standard du système.

**Panne** : L'interruption permanente d'une capacité du système pour effectuer une fonction requise dans des conditions de fonctionnement spécifiées.

**Dysfonctionnement** : L'irrégularité intermittente dans l'accomplissement de la fonction souhaitée du système.

**Détection de défauts** : Détermination de défauts présents dans un système et l'heure de détection.

**Isolation des défauts** : Détermination du type, de l'emplacement et du temps de détection d'un défaut. Cette étape suit l'étape de détection de défauts.

**Identification de défauts** : Détermination de la taille et variation dans le temps du comportement d'un défaut. Cette étape suit l'étape de l'isolation de défaut.

**Diagnostic des défauts** : Détermination du type, de la taille, de l'emplacement et du temps de détection d'un défaut. Cette étape suit la détection de défaut et inclut l'isolation et l'identification de défaut.

**Résidu** : L'indicateur de défaut, basé sur les écarts entre les mesures et un modèle basé sur des calculs d'équations.

**Symptôme** : Le changement d'une grandeur observable du comportement normal.

**Surveillance** : C'est une tâche continue en temps réel déterminant les conditions possibles d'un système physique, tout en reconnaissant et en indiquant des anomalies du comportement.

**Supervision** : La surveillance d'un système physique tout en prenant des actions appropriées pour maintenir le fonctionnement dans le cas de défauts.

**Sécurité** : Aptitude d'un système à ne pas causer de danger pour les personnes, le matériel ou pour l'environnement.

**Modèle quantitatif** : Utilise des relations statiques et dynamiques entre les variables du système et des paramètres afin de décrire le comportement d'un système en terme quantitatif mathématique.

**Modèle qualitatif** : Utilise des relations statiques et dynamiques entre les variables du système et des paramètres pour décrire le comportement du système en termes qualitatifs tels que les causalités ou les règles.

Pour plus de détails voir (Isermann, 2011, Isermann, 2006).

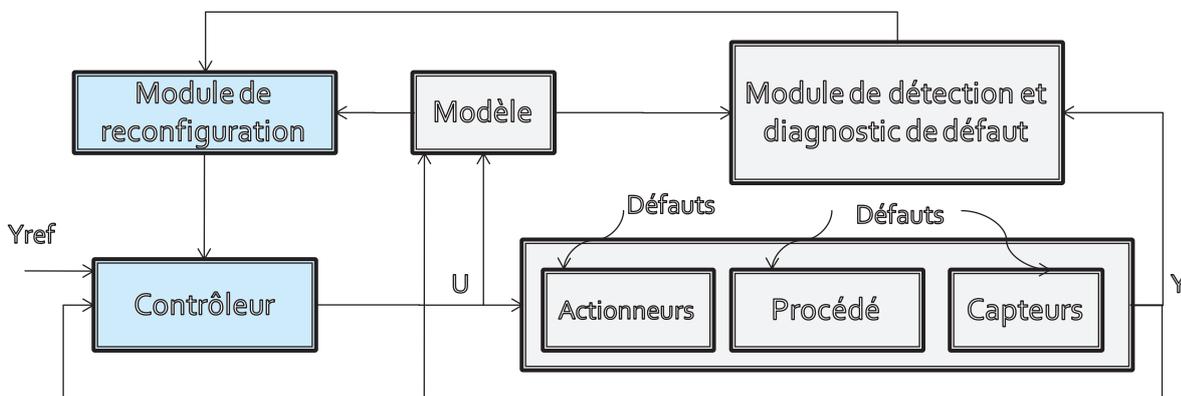
En résumé, le **diagnostic** de défaut consiste à détecter, isoler et estimer le défaut. La **détection** du défaut détermine l'instant d'apparition d'un mauvais fonctionnement dans le système, tandis que **l'isolation** du défaut consiste à cerner (identifier) les causes ou les sources de défaut afin de dire s'il s'agit d'un défaut de type capteur, actionneur ou procédé, et **l'identification** du défaut donne une estimation de certaines caractéristiques du défaut telle que son amplitude.

Le rôle de diagnostic de défaut est d'assurer la sécurité et le fonctionnement optimal du système par l'indication de tout état non désiré (ou statu hors contrôle) permettant ainsi d'éviter tout dommage ou accident. Généralement, le diagnostic de défauts se fait en ligne. Les tâches de détection et de localisation sont accomplies parallèlement ou bien une tâche déclenche l'autre.

Nous pouvons souligner que le terme « diagnostic » a une origine médicale, qui signifie trouver les causes de la maladie (défaut) à partir des symptômes perçus (les signes observables).

### I.1.2. Le principe de détection de défaut

La Figure I.1 montre un schéma global du fonctionnement d'un système pouvant être affecté par différents types de défauts. Comme c'est le cas dans la plupart des systèmes industriels, ce procédé est régulé en boucle fermée avec un contrôleur dont l'objectif est de maintenir la sortie  $Y$  le plus proche possible de la consigne  $Y_{ref}$ . Des instruments de mesure transmettent les données acquises englobant les entrées et les sorties mesurées ainsi que la sortie du contrôleur vers l'organe de détection et de diagnostic de défauts. En plus de ces données, le module de diagnostic de défaut utilise un modèle (fondamental ou statistique) pour réaliser la surveillance du système et informer le module de reconfiguration de la présence de défauts.

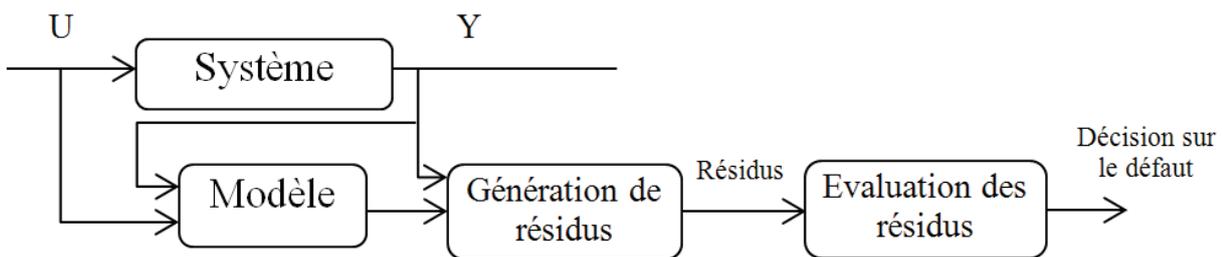


**Figure I.1 Architecture du fonctionnement d'un système avec une commande tolérante aux défauts**

Le bloc de reconfiguration permet de garder le système dans l'état désiré par le réglage des paramètres du contrôleur ou par une adaptation d'une nouvelle loi de commande (Blanke et al., 2006, Blanke et al., 2003, Prakash et al., 2002). On note qu'il existe plusieurs méthodes de reconfiguration, toutefois elles ne seront pas discutées dans ce mémoire. Le but de la reconfiguration est de continuer à faire fonctionner le système tout en maintenant les performances/conditions de sécurité requises malgré la présence de dysfonctionnement dans l'attente d'une opération de maintenance (Patton, 1997).

Afin de réaliser le diagnostic des défauts, une première approche consiste à dupliquer les composants d'instruments de mesure pour augmenter la fiabilité. Cela nous permet de choisir la mesure saine à partir des mesures disponibles. Cette approche s'appelle la **redondance**

**matérielle.** Elle est utilisée dans le diagnostic, surtout dans le domaine aéronautique, mais elle entraîne une augmentation des coûts et une réduction de l'autonomie du système. La deuxième approche, qui est la **redondance analytique**, est très intéressante à la fois sur le plan financier et technique puisqu'elle se base uniquement sur l'exploitation de relations existantes entre les différentes variables mesurées ou estimées. Après la collection des données, on essaie d'extraire les informations utiles à partir de ces données, en d'autres termes, de convertir ces données brutes en quelque chose de significatif pour des objectifs bien précis. Le processus de détection de défauts proprement dit est donc composé essentiellement de deux phases (Chow et al., 1982) : génération de résidus suivie par l'analyse de résidus, comme le montre la Figure I.2



**Figure I.2 Schéma classique de détection des défauts**

**Génération des résidus :** Cette étape consiste à créer un signal calculant la différence entre la sortie réelle et celle estimée. Le résidu est nul (ou négatif dans certaines méthodes) dans le cas du fonctionnement normal, et en cas de dysfonctionnement, il sera proportionnel à la gravité du défaut.

**Évaluation des résidus :** Cette phase permet d'analyser les indicateurs de défaut générés et de définir un seuil (fixe, adaptatif, ou statistique) qui va décider de l'instant de changement d'état de fonctionnement du système. En effet, dans le cas réel, les bruits de mesure, les imprécisions de modélisation et les perturbations ne sont pas des grandeurs déterministes mais de nature aléatoire car ils vont introduire des variations imprévisibles sur le système. Cette étape est donc capitale afin de diminuer les taux de fausse alarme ou de non détection.

### **I.1.3. FDI en boucle ouverte et en boucle fermée**

Un défaut agit d'une manière différente sur le système dépendamment si le fonctionnement (de la commande) est en boucle ouverte ou en boucle fermée. Il influence les sorties et/ou les entrées mesurables et/ou les paramètres internes du système ou les variables d'états non mesurables.

Lorsqu'un système ou procédé fonctionne **en boucle ouverte** et qu'un défaut apparaît, ce dernier va influencer la sortie  $Y$  par un écart (offset)  $\Delta Y$ . Par contre, dans le cas de procédé fonctionnant **en boucle fermée**, la sortie risque de ne montrer qu'une faible déviation qui disparaît rapidement, surtout si un contrôleur de type intégral est utilisé pour asservir cette sortie (contrôleur PI). En revanche, la variable d'entrée manipulée  $U$  va montrer un offset permanent  $\Delta U$ . Il est à souligner que la sortie est également affectée par des bruits, perturbations et des variations dans  $U$ . Cela rend la tâche de diagnostic très difficile.

On note également que les défauts peuvent apparaître de différentes façons, d'une manière brusque (c'est le cas d'une coupure de phase dans un moteur électrique), d'une façon progressive (on parle d'une dérive lente, c'est le cas d'encrassement) ou d'une façon intermittente qui apparaît selon les conditions de fonctionnement (température, pression, concentration...).

#### **I.1.4. Domaines concernés par la FDI**

La communauté de l'automatique a établi un cadre théorique bien fondé pour le diagnostic de défaut pour les systèmes linéaires (Nyberg, 1999, Chow et al., 1982) et non linéaires (Ding, 2008). Une revue détaillée des techniques développées pour le diagnostic de défauts est donnée dans (Venkatasubramanian V. et al., 2003) et un grand nombre d'applications en ingénierie jusqu'en 1996 est fourni dans (Isermann and Ballé, 1997) .

Dans cette littérature, nous pouvons citer les applications suivantes :

- Procédés chimiques : (Caccavale et al., 2011, Verron, 2007, Pierri et al., 2008a, Chetouani, 2006, Ram Maurya et al., 2004)
- Moteurs, machines et roulements : (Verucchi et al., 2008, Ding, 2008, Isermann, 2005)
- Eoliennes : (Wei, 2010, Dolan, 2010), Donders, 2002)
- Procédé de traitement des eaux usées : (Baggiani and Marsili-Libelli, 2009, Wimberger and Verde, 2008, Tharrault, 2008)
- Turbine à gaz industrielle : (Simani et al., 2003)
- Système hydraulique : (Ding, 2008, Hammouri et al., 2002)
- Industrie automobile : (Arnanz et al., 2011, Saludes Rodil et al., 2009, Svard and Nyberg, 2010)
- Réseau de distribution électrique : (Mohamed and Rao, 1995)
- Robot manipulateurs : (Paviglianiti et al., 2010)

- Circuit électronique analogique : (Mohsen and El-Yazeed, 2004)
- Système de contrôle du vol d'avion : (Hajiyev, 2012)
- On peut trouver aussi une revue détaillée sur le diagnostic de défauts appliqué sur les systèmes de chauffage, ventilation, et climatisation : (Katipamula and Brambley, 2005).

### **I.1.5. Performances des techniques FDI**

Afin d'évaluer la performance d'une technique de diagnostic de défauts ou la comparer avec d'autres méthodes, il faut définir certains critères de performances à respecter:

1. La rapidité de détection du défaut.
2. La minimisation de fausses alarmes
3. La robustesse de la méthode face à la présence de bruit, perturbation et erreur de modélisation.
4. L'adaptation face aux variations du procédé.
5. La capacité à identifier des défauts multiples.
6. La capacité à manipuler différents types de défauts (actionneur, capteur, procédé).
7. La diagnosticabilité ou la capacité de distinction entre les différents défauts.
8. La sensibilité face à des petits défauts.
9. La précision de la localisation.
10. La convenance par rapport aux systèmes réels généralement non linéaires.
11. L'aptitude à inclure la commande tolérante aux défauts.
12. La simplicité d'implémentation et d'utilisation.

## **I.2. Les méthodes de diagnostic**

Des chercheurs de domaines variés ont essayé de réaliser la tâche de diagnostic pour différents systèmes en apportant leur connaissance et outils de certains domaines de recherche. Cela conduit à l'enrichissement de méthodes de diagnostic.

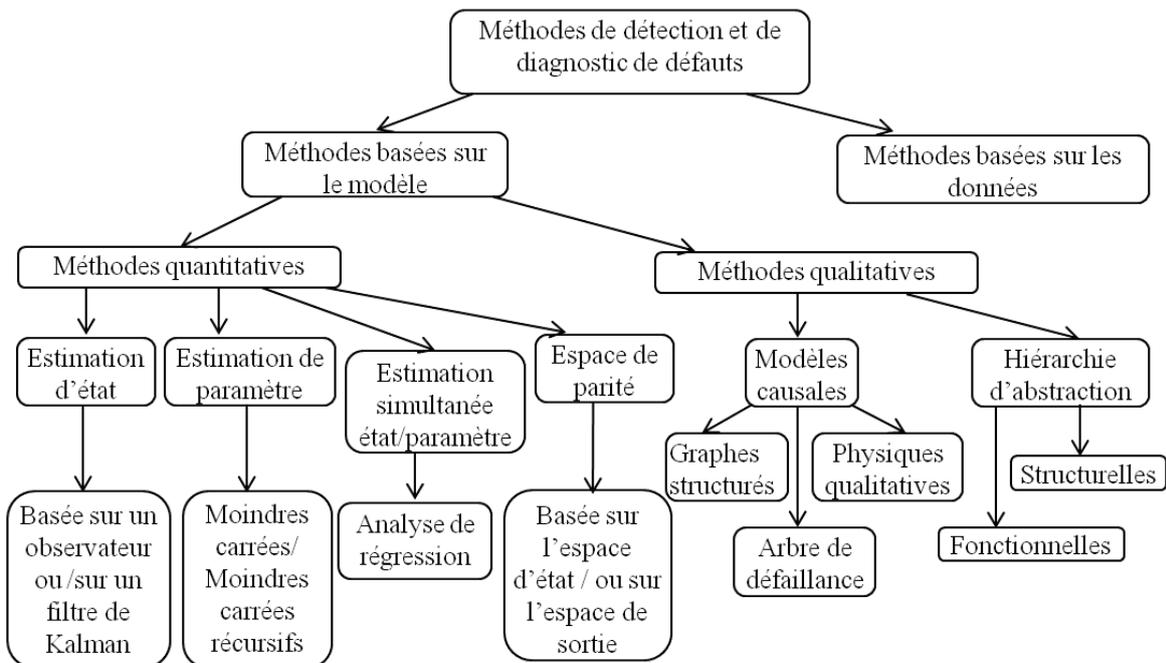
Certains chercheurs décomposent les méthodes de diagnostic en 3 catégories (Isermann, 2011, Ping, 2011, Verron, 2007, Frank et al., 2000). Dans cette thèse, les approches de diagnostic de défauts sont décomposées en deux grandes catégories, comme proposé par (Zhang and Jiang, 2008) et (Venkatasubramanian V. et al., 2003a) :

1. Schémas **basés sur le modèle** (Venkatasubramanian V. et al., 2003a, Venkatasubramanian et al., 2003b)
2. Schémas **basés sur les données** (sans modèle a priori) (Venkatasubramanian, 2003c)

Ces approches sont brièvement discutées dans les paragraphes suivants.

### I.2.1. Méthodes à base de modèle explicite

Les méthodes basées sur un modèle explicite utilisent un modèle décrivant les relations fondamentales gouvernant le système afin de générer des symptômes de défauts (Simani et al., 2003). Ces méthodes ont de nombreux avantages car elles permettent d'identifier le type du défaut et de remonter à la source du défaut pour différentes classes de systèmes. La plupart des méthodes à base de modèle reposent sur le concept de redondance analytique qui consiste à vérifier la cohérence entre le comportement simulé du modèle et celui réellement observé, d'où la nécessité d'avoir un modèle détaillé et suffisamment précis du système afin d'avoir des résultats satisfaisants. Ceci est réalisé en générant des résidus qui permettent à leur tour de détecter un écart entre le comportement défectueux et le comportement normal du système. Plusieurs méthodes sont envisageables pour l'obtention de résidus à partir de modèles analytiques (Figure I.3). On peut citer par exemple les méthodes par estimation de paramètres, par estimation d'état, ou par espace de parité.



**Figure I.3 Schéma de classification des méthodes de diagnostic de défauts basées sur le modèle**

- Approche par estimation de paramètres :

Partant de l'idée qu'un défaut se reflète dans les paramètres physiques du système, des méthodes d'estimation paramétriques sont utilisées pour estimer la valeur actuelle de ces paramètres. Le résidu est ainsi calculé à partir de l'écart entre la valeur des paramètres estimés et celle de références obtenue dans le cas sans défaut (Simani et al., 2003, Gertler, 1998, Fuente et al., 1996).

- Approche par espace de parité :

Cette technique se base sur la réécriture des équations d'état et de mesure de telle sorte que seules les variables connues (entrées et sorties) sont autorisées à apparaître. Ces variables connues sont ensuite utilisées afin de construire un résidu indicateur de défaut.

Soit le modèle  $G_M = \frac{\hat{A}(z)}{\hat{B}(z)}$  décrivant le système  $G_p = \frac{A(z)}{B(z)}$ . Ainsi le résidu peut être défini

comme suit :

$$r(z) = \left( \frac{A(z)}{B(z)} - \frac{\hat{A}(z)}{\hat{B}(z)} \right) u(z) \quad (I.1)$$

Plusieurs travaux ont utilisé cette méthode pour le diagnostic de défauts (Gustafsson, 2007, Gertler, 1998) ou ont combiné cette méthode avec la méthode d'estimation de paramètre (Isermann, 2005).

- Approche par estimation d'état (observateur)

La construction des observateurs d'état représente une des techniques les plus employées dans l'approche basée sur le modèle (Ding, 2008, Pierri et al., 2008b, Li et al., 2008, Isermann, 2006, Simani et al., 2003, Venkatasubramanian V. et al., 2003a, Lootsma, 2001, Lootsma, 1999). La redondance analytique est obtenue grâce aux états estimés à partir d'un modèle adéquat et d'un ensemble de mesures appropriées. En comparant les variables estimées et mesurées, un ensemble de signaux (résidus) sensibles aux défauts est généré. Ces résidus sont ensuite analysés pour détecter et/ou isoler le défaut.

Différents types d'observateurs ont été proposés dans la littérature pour les systèmes linéaires et non linéaires. Parmi les techniques les plus utilisées, on peut citer les observateurs à entrée inconnue (Sotomayor and Odloak, 2005); les observateurs basés sur le placement de structure propre « *The eigen structure assignment approach* » (Patton and Chen, 2000), les observateurs à

mode glissant (Edwards et al., 2000) , le filtre de Kalman étendu qui a été initialement appliqué pour la détection et l'isolation des défauts des systèmes non linéaires par (Huang et al., 2003).

Le problème fondamental de détection et d'isolation des défauts de **systèmes non linéaires** a été introduit initialement par Frank (Frank, 1990). Depuis lors, plusieurs techniques ont été utilisées pour les systèmes non linéaires. Une revue sur les principales approches basées sur les observateurs pour les systèmes non linéaires a été donnée par (Alcorta García and Frank, 1997) . Les auteurs dans (Hammouri et al., 2002), (De Persis and Isidori, 2001), (Kabore et al., 2000) et (Hammouri et al., 1999) ont proposé une solution pour la détection et l'isolation des défauts des systèmes non linéaires en combinant une approche géométrique de découplage avec les techniques des observateurs non linéaires, notamment le grand gain.

Tout l'art de ces approches réside dans le choix du terme correctif et plus particulièrement la synthèse de gain qui a pour objectif non seulement de s'assurer de la convergence de l'estimation vers la sortie réelle du procédé mais aussi le découplage par rapport aux autres défauts.

Dans cette thèse on s'intéresse à deux types d'observateurs ; un observateur à **grand gain** et un autre **de type Kalman** à cause de leur convergence exponentielle. L'observateur à grand gain sera appliqué dans l'estimation d'état d'un réacteur chimique alors que l'observateur de type Kalman trouvera son application dans le système éolien. Un bref rappel de ces types d'observateurs est donné ci-dessous.

### ***1.2.1.1. Observateur à grand gain***

Ce type d'observateur peut s'appliquer à une large classe de systèmes non linéaires qui s'écrivent sous la forme suivante :

$$\begin{cases} \dot{x}_1(t) = A_1(u(t), s(t))x_2 + \psi_1(u(t), s(t), x_1(t)) + \varepsilon_1(t) \\ \vdots \\ \dot{x}_{q-1}(t) = A_{q-1}(u(t), s(t))x_q + \psi_{q-1}(u(t), s(t), x_1(t), \dots, x_{q-1}(t)) + \varepsilon_{q-1}(t) \\ \dot{x}_q(t) = \psi_q(u(t), s(t), x(t)) + \varepsilon_q(t) \\ y(t) = x_1(t) = Cx \end{cases} \quad (I.2)$$

Où  $x = [x_1, \dots, x_q]^T \in R^n$  est le vecteur d'état,  $x_i = [x_{i1}, \dots, x_{ik}]^T$ ,  $u(t)$  est l'entrée,  $A_i(u, s)$  est une matrice carrée de dimension  $k \times k$  ( $q.k = n$ ),  $s(t)$  est un signal connu,  $\psi = [\psi_1, \dots, \psi_q]^T$  sont des

fonctions non linéaires et supposés différentiables,  $\varepsilon = [\varepsilon_1, \dots, \varepsilon_q]^T$  est un vecteur non connu borné,  $C = (I_k, 0, \dots, 0)$ .

Un observateur grand gain peut être construit pour le système précédent en se basant sur les hypothèses suivantes:

**H1)**  $\psi$  est globalement Lipschitz par rapport à  $x$ .

**H2)** Les matrices  $A_i(u(t), s(t))$  et leurs dérivées sont inversibles et bornées

**H3)** Les matrices  $A_i(u(t), s(t))$  sont différentiables par rapport au temps et leurs dérivées sont bornées.

Dans ce cas, l'observateur grand gain du système (I.2) est donné par:

$$\dot{\hat{x}} = A(u(t), s(t))\hat{x} + \psi(u(t), s(t), \hat{x}(t)) + \Lambda^{-1}\Delta_\theta K(C\hat{x} - y) \quad (\text{I.3})$$

Où  $K$  est une matrice de taille  $n \times k$  tel que  $(\bar{A} + KC)$  est une matrice Hurwitz,

$$A(u(t), s(t)) = \begin{pmatrix} 0 & A_1(u(t), s(t)) & 0 & \dots & 0 \\ 0 & 0 & A_2(u(t), s(t)) & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & & & 0 & A_{q-1}(u(t), s(t)) \\ 0 & \dots & & 0 & 0 \end{pmatrix} \quad (\text{I.4})$$

$$\psi(u, t) = \begin{pmatrix} \psi_1(u, s, \hat{x}_1) \\ \vdots \\ \psi_{q-1}(u, s, \hat{x}_1, \dots, \hat{x}_{q-1}) \\ \psi_q(u, s, \hat{x}) \end{pmatrix}, \Delta_\theta = \begin{pmatrix} \theta & 0 & \dots & 0 \\ 0 & \theta^2 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \theta^q \end{pmatrix} \quad (\text{I.5})$$

$$\Lambda(t) = \begin{pmatrix} I_k & 0 & 0 & \dots & 0 \\ 0 & A_1(u(t), s(t)) & & & \vdots \\ \vdots & & \ddots & & \\ 0 & & & & 0 \\ 0 & & & & A_1(u(t), s(t)) \times \dots \times A_{q-1}(u(t), s(t)) \end{pmatrix} \quad (\text{I.6})$$

$$\bar{A} = \begin{pmatrix} 0 & I_k & 0 & \dots & 0 \\ 0 & 0 & I_k & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I_k \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \quad (\text{I.7})$$

$\theta$  est un paramètre de réglage positif,  $I_k$  est la matrice identité de taille  $k \times k$ .

***Théorème: (basé sur les résultats donnés dans Farza et al., 1997, Gauthier et al., 1992)***

Supposons que le système (I.2) satisfait les hypothèses H1), H2) and H3), alors:

$$\|\hat{x}(t) - x(t)\| \leq \eta e^{-\mu t} \|\hat{x}(0) - x(0)\| + \gamma \varepsilon_\infty$$

Où  $\varepsilon_\infty$  est la norme infinie ( $L^\infty$ ) de  $\varepsilon(t)$ ;  $\eta > 0$ ,  $\mu > 0$  et  $\gamma > 0$  sont des constantes. Cela implique que, si  $\varepsilon = 0$ , alors, le système (I.3) est un observateur exponentiel du système (I.2).

### ***1.2.1.2. Observateur du type Kalman***

L'observateur de type Kalman a été initialement proposé par (Bornard et al., 1989) et (Hammouri and de Leon Morales, 1990) pour une classe de systèmes non linéaires où la matrice d'état peut dépendre du signal d'entrée, du signal de sortie ou du temps et toutes les entrées sont régulièrement persistantes. Ce système peut se mettre sous la forme :

$$\begin{cases} \dot{x} = A(u)x + \varphi(u) \\ y = Cx \end{cases} \quad (\text{I.8})$$

Où  $y \in \mathfrak{R}$ ;  $x \in \mathfrak{R}^n$ ;  $u \in \mathfrak{R}^m$  et  $\varphi(u)$  est un vecteur qui dépend non linéairement de  $u$ .

Notons par  $\phi_u(s, t_0)$ , la matrice de transition, la solution unique de l'équation :

$$\frac{d\phi_u(s, t_0)}{ds} = A(u(s))\phi_u(s, t_0) \quad (\text{I.9})$$

Notons par  $G(u, t_0, t_0 + t_1)$  le Grammien d'observabilité relatif à l'entrée  $u$  sur l'intervalle

$[t_0, t_0 + t_1]$  :

$$G(u, t_0, t_0 + t_1) = \int_{t_0}^{t_0+t_1} \phi_u^*(t, t_0) C^* C \phi_u(t, t_0) dt \quad (\text{I.10})$$

Où  $C^*$  représente la transposée de  $C$ .

#### **Définition :**

L'entrée  $u \in \mathfrak{R}^m$  est dite régulièrement persistante pour le système (I.8) si  $\exists t_1 > 0$ ,  $\exists \alpha_1 > 0$ ,  $\exists \alpha_2 > 0$  and  $\exists t_0 \geq 0$  tels que  $\forall t \geq t_0$  :

$$\begin{aligned} \lambda_{\min}(G(u, t_0, t_0 + t_1)) &\geq \alpha_1 \\ \lambda_{\max}(G(u, t_0, t_0 + t_1)) &\leq \alpha_2 \end{aligned} \quad (\text{I.11})$$

Où  $\lambda_{\min}$  et  $\lambda_{\max}$  représentent la plus petite et la plus grande valeur propre respectivement.

**Théorème:** (Bornard and Hammouri, 1991)

Si  $\mathbf{u}$  est régulièrement persistante, alors pour tout  $\theta > 0$ , le système suivant :

$$\begin{cases} \dot{\hat{x}} = A(u)\hat{x} + G(u) - R^{-1}C^T(Cx - y) \\ \dot{R} = -\theta.R - A^T(u)R - R.A(u) + C^T C \end{cases} \quad (\text{I.12})$$

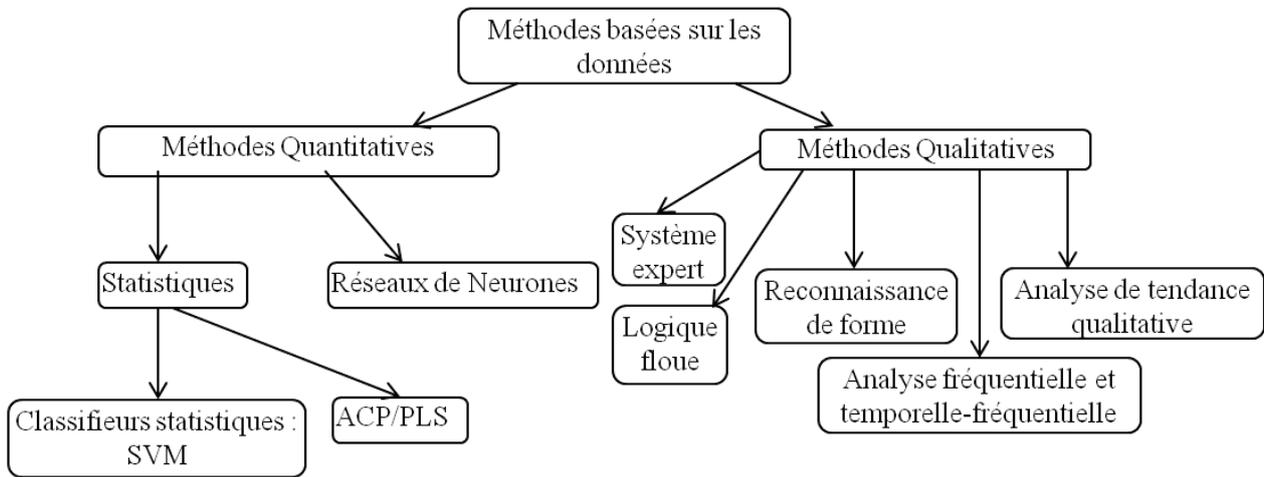
est un observateur exponentiel pour (I.8). Le paramètre de réglage  $\theta$  détermine la vitesse de convergence de l'observateur qui est garantie si la matrice  $\mathbf{R}$  est symétrique définie positive.

La particularité de l'observateur de type Kalman est son gain variable donné par la résolution de l'équation algébrique de Riccati.

### I.2.2. Méthodes basées sur les données

La disponibilité d'une grande quantité des données ainsi que les inconvénients des méthodes basées sur le modèle (dus à la nécessité d'avoir un modèle précis par exemple), ont encouragé la croissance des méthodes de diagnostic basées sur les données. Ces méthodes peuvent permettre des développements rigoureux des données disponibles (mesures historiques ou en temps réel) même dans le cas de systèmes complexes, d'où leur succès dans les applications industrielles. Elles font appel à des procédures d'apprentissage et de reconnaissance de forme, à la logique floue ou à l'intelligence artificielle (Frank and Köppen-Seliger, 1997) afin de générer les symptômes de défaut qui seront analysés pour fournir la décision de diagnostic de défaut (Reiter, 1987).

Les méthodes basées sur les données sont appropriées pratiquement pour n'importe quel type de problème pour lequel d'importantes quantités de données mesurées représentatives du fonctionnement du système sont disponibles. Elles sont d'autant plus intéressantes lorsque le modèle mathématique du système n'est pas disponible, ce qui rend inutilisables les méthodes de diagnostic basées sur le modèle. L'application de ces méthodes est réalisée en différentes étapes d'acquisition et prétraitement des données, de modélisation/apprentissage et de validation. La Figure I.4 répertorie les différentes méthodes de diagnostic de défauts basées sur les données. Un aperçu des méthodes basées sur des données est donné dans (Venkatasubramanian, 2003).



**Figure I.4 Méthodes de diagnostic dirigées par les données**

- **Vérification de limites et tendances**

La méthode la plus simple de supervision dans le cas d'absence de modèle est mono variable et se base sur la vérification de limites (seuils) de la valeur de certaines variables mesurables telles que : la pression, la température, la force, la vitesse et la position. Si cette valeur dépasse la zone de tolérance ( $Y_{\min} < Y < Y_{\max}$  et  $\frac{dY_{\min}}{dt} < Y < \frac{dY_{\max}}{dt}$ ), une alarme se déclenche. L'opérateur doit ensuite cerner le défaut (le localiser) afin de prendre les mesures nécessaires pour le retour à l'état normal.

Certes, cette méthode est simple et facile à mettre en œuvre, mais elle a des limites. En effet, si une variable mesurée dépasse la zone de tolérance, le procédé est probablement déjà dans un état critique. En plus, malgré que ces méthodes fonctionnent quand le procédé est à l'état d'équilibre, la situation devient plus compliquée si la variable contrôlée change de dynamique avec les points de fonctionnement. Aussi, un défaut important dans un procédé complexe peut déclencher plusieurs alarmes en même temps. Cela rend la tâche de l'opérateur qui supervise le procédé difficile pour identifier le défaut en question afin de réagir immédiatement par les actions requises.

- **Méthodes statistiques**

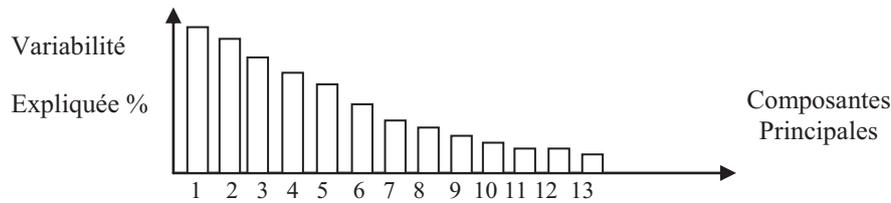
Les méthodes statistiques font partie des méthodes basées sur les données. En étudiant les caractéristiques d'un ensemble d'objets ou d'observations, elles donnent un moyen essentiel pour la connaissance. A partir de ces données, elles fournissent un outil graphique facilement

manipulable et interprétable par l'opérateur. Malheureusement, cela demande la connaissance complète de la population, ce qui rend ces méthodes coûteuses, voire impossibles dans certains cas (comme dans le cas des systèmes régis par des perturbations inconnues). La solution est donc d'utiliser les statistiques inférentielles qui consistent à utiliser la théorie de probabilités afin d'induire les caractéristiques inconnues d'une population à partir de certains échantillons tirés de cette population avec une certaine marge d'erreur, autrement dit la détermination d'une loi de probabilité à partir d'une série d'expériences.

Parmi les méthodes statistiques, on cite l'Analyse en Composantes Principale (ACP), *Partial Least Square* (PLS) et *Support Vector Machines* (SVM). Les méthodes ACP et PLS ont été utilisées dans cette thèse à titre comparatif avec la méthode de référence qui est la méthode SVM. Dans les deux paragraphes suivants, on propose donc de rappeler la théorie et les applications de ces deux méthodes.

### ***1.2.2.1. L'Analyse en Composantes Principales (ACP)***

L'analyse en composantes principales (ACP) « *Principal Component Analysis PCA* » est une méthode géométrique et statistique qui permet de transformer l'espace multivariable d'entrées corrélées en un espace plus petit non corrélé tout en conservant le maximum de variance des données d'entrées. Elle permet de voir les données d'une manière différente (dans un autre repère) par la rotation des axes des variables originales vers des nouveaux axes, qui sont orthogonaux et complètement décorrélés pour éviter d'avoir des informations redondantes. C'est donc une méthode d'analyse et de réduction de dimension de données. Elle génère un nouveau jeu de données appelé **Composantes Principales** (CP) (variables latentes, ou facteurs) à partir desquelles le modèle va être construit, et chaque composante a un taux de représentation du modèle, comme on peut le voir sur la Figure I.5 . Chaque composante est une combinaison linéaire des variables originales. Donc chaque variable peut contribuer à la création du modèle contrairement à la sélection des caractéristiques (voir section I.2.3.3) qui utilisent juste un sous-ensemble des variables originales pour la création du modèle (Ramaswami and Bhaskaran, 2009). Plus le nombre de composantes est important, plus la sensibilité au bruit va augmenter. On se limite généralement à 2 ou 3 composantes, où le maximum d'information (variance) est compressé.



**Figure I.5 Taux de représentativité des composantes principales dans le modèle**

Le champ d'application de cette méthode est très varié allant de la réduction de dimension , réduction et suppression de bruit, visualisation, à la compression de données, détection des défauts (Sun et al., 2005, Dunia et al., 1996), et aussi au prétraitement des modèles empiriques.

### Fondement théorique de la méthode ACP

Soit  $\mathbf{X}$  une matrice de données de taille  $n \times m$  constituée de  $n$  observations ou échantillons et  $m$

variables : 
$$\mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{bmatrix}$$

L'ACP cherche l'axe (utilisé dans la projection des échantillons) qui va maximiser la capture de variabilité ou variance (information) des nuages de points. Ceci est fait en utilisant une matrice de transformation  $\mathbf{P} \in \mathbb{R}^{m \times n}$ , qui projette les données dans un espace de dimension réduit :

$$\hat{\mathbf{X}} = \mathbf{TP}^T \text{ et } \mathbf{E} = \mathbf{X} - \hat{\mathbf{X}} \quad (\text{I.13})$$

$\mathbf{P}_i$  est nommé le vecteur poids « *loading* », qui contient l'information qui concerne la relation entre les variables.  $\hat{\mathbf{X}}$  est la nouvelle matrice de données qui estime  $\mathbf{X}$ .  $\mathbf{E}$  est la matrice des résidus et représente la source de bruit. Les éléments de  $\mathbf{T}$  sont appelés scores, et ils contiennent la relation entre les échantillons. Les scores représentent les valeurs des variables originales mesurées et transformées dans l'espace de dimension réduit. L'espace des données d'origine peut être calculé comme :

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \Rightarrow \mathbf{X} = \sum_{i=1}^a t_i \mathbf{p}_i^T + \mathbf{E} \quad (\text{I.14})$$

Où  $\mathbf{t}_i$  est le vecteur score. Ceci donne à la méthode ACP la possibilité d'enlever le bruit et de laisser seulement l'information pertinente. La matrice de variance-covariance  $\mathbf{R}$  est obtenue par la décomposition en valeurs singulières :

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^T = \frac{\mathbf{X}^T \mathbf{X}}{(n-1)} = \frac{1}{n-1} [\mathbf{X}^T \mathbf{X} + \mathbf{E}^T \mathbf{E}] \quad (\text{I.15})$$

Où  $\Lambda$  est la matrice diagonale contenant les valeurs propres dans un ordre décroissant de la matrice de corrélation  $\mathbf{R}$ . Les colonnes de la matrice  $\mathbf{U}$  sont les vecteurs propres de  $\mathbf{R}$ . La matrice  $\mathbf{P}$  est générée par le choix des vecteurs propres de  $\mathbf{U}$ .

Le choix du nombre  $a$  ( $a < m$ ) des variables latentes (composantes principales, vecteurs colonnes de  $\mathbf{P}$ ) est important puisqu'il permet de construire la matrice  $\mathbf{P}$ , et par la suite  $\hat{X}$  qui représente la source de variabilité dans le procédé. Il détermine donc la qualité du modèle ACP afin de garder la plupart de l'information d'origine (Valle et al., 1999). Ces variables doivent respecter trois propriétés : elles sont non corrélées, elles ont des variances séquentiellement maximales et l'erreur d'approximation quadratique entre les vecteurs d'origines et les premières CP est minimale.

Une des méthodes qui permet de déterminer le nombre de composantes  $a$ , est la **Variance Cumulative en Pourcentage** « *Cumulative Percent Variance* » : Dans cette méthode, le choix se fait de telle sorte que les premières  $a$  composantes doivent avoir une VCP supérieure à 90% :

$$VCP(a) = \frac{\sum_{i=1}^a \lambda_i}{\text{trace}(\mathbf{R})} \times 100 \quad (\text{I.16})$$

Où  $\mathbf{R}$  est la matrice de variance-covariance, et  $\lambda_i$  sont les valeurs propres.

### **Diagnostic de défaut par ACP**

Afin d'utiliser la méthode ACP pour **la détection de défaut**, on analyse les deux distances complémentaires définies ci-dessous qui sont utilisées comme indicateurs de défauts. Cependant, ces deux indices détectent très souvent des situations anormales, ce qui provoque une confusion pour les opérateurs de procédés qui effectuent l'analyse des défauts. En effet, si l'occurrence d'un événement produit un changement dans la covariance du modèle, il sera détecté par une grande valeur de  $T^2$  et  $Q$  qui dépasse les limites de confiance.

**L'isolation de défaut** par ACP peut se faire par la méthode de calcul des contributions approximatives des différentes variables à la valeur de l'indicateur de détection. Les variables avec une forte valeur de contribution sont considérées comme responsables du défaut.

## 1. La carte de contrôle $T^2$ de Hotelling statistique

La carte de contrôle  $T^2$  est définie comme la majeure partie de variation des données à l'intérieur du modèle (l'espace principal). Le processus est considéré étant en fonctionnement normal pour un seuil de confiance  $\alpha$  donné, et une distribution du  $\chi^2$  :

$$T^2(k) = t(k)\Lambda_a^{-1}t^T(k) = x(k)P\Lambda^{-1}P^T x^T(k) \leq \delta_T^2$$

$$T_{a,n,\alpha}^2 = \frac{a(n^2-1)}{n(n-a)} F_\alpha(a, n-a) \quad (\text{I.17})$$

Où  $n$  est le nombre d'échantillons,  $a$  est le nombre de composantes principales,  $F_\alpha(a, n-a)$  est la valeur de la distribution de Fisher-Snedecor.  $\Lambda_a$  est la matrice formée par les premières lignes et colonnes de  $\Lambda$ ,  $\alpha$  représente le niveau approprié de l'importance pour la réalisation du test qui prend généralement la valeur de 0,05 ou 0,01 pour les limites de confiance.

## 2. Indice « Squared Prediction Error SPE »

L'indice SPE est défini comme la somme des carrés de chaque ligne de  $\mathbf{E}$ . Il est perpendiculaire à l'espace construit par le modèle ACP :

$$Q = SPE(k) = \sum_{j=1}^n (x_j(k) - \hat{x}_j(k))^2 = e(k)e(k)^T = X(I - P_a P_a^T)X^T \leq \delta_T^2 \quad (\text{I.18})$$

$$Q = E^T E = \sum_{j=1}^n (x_j(k) - t(k)p_j^T)^2 \leq \delta_T^2 \quad (\text{I.19})$$

La limite supérieure, ou les seuils de confiance de  $\mathbf{Q}$  statistique, peuvent être calculés par sa distribution approximative comme suit :

$$Q_\alpha = \theta_1 \left[ \frac{h_0 \cdot c_\alpha \cdot \sqrt{2\theta_1}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (\text{I.20})$$

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

Où  $c_\alpha$  est l'écart type normal correspondant à la partie supérieure  $(1-\alpha)$  centile,  $\lambda_j$  est la valeur propre associée au  $\mathbf{j}^{ème}$  vecteur poids.  $Q$  statistique représente l'erreur de prédiction quadratique décrivant le bruit dans les données et la variation à l'extérieur du modèle.  $Q$  est utilisé pour surveiller et suivre de près la corrélation entre les variables du procédé. Une augmentation de sa valeur signifie un changement dans les corrélations qui gouvernent les variables ce qui

indiquerait un défaut. Cependant, la distance SPE est sensible aux incertitudes de modélisation provoquant de nombreuses fausses alarmes si les systèmes sont non-linéaires.

Des travaux ont été faits utilisant l'ACP avec les cartes de contrôle Q et  $T^2$  statistiques afin de différencier entre l'occurrence d'un défaut et une condition de fonctionnement normale élargie (Humberstone et al., 2012, Humberstone, 2010). Néanmoins, ces cartes de contrôle présentent certaines restrictions puisqu'elles font des suppositions sur la distribution des données. (Camci et al., 2008) proposent une carte de contrôle basée sur la **distance noyau** utilisant les principes de SVM et démontre qu'il est sensible à tous les types de données en hors contrôle quels que soient leurs distributions et permet aussi de bénéficier et d'apprendre de ces données.

### ***1.2.2.2. Moindres carrés partiels (PLS)***

La méthode de Moindres Carrés Partiels (MCP) ou Projection aux Structures Latentes (PSL) « *Partial Least Squares PLS* » est un algorithme d'analyse statistique multivariable proposé par Wold à la fin des années 60 qui combine les fonctionnalités de l'ACP et la régression multiple (Wold, 1966). Il était destiné au début pour l'économétrie. Il a trouvé un grand succès et pris de l'ampleur en chimie où il est devenu un outil standard d'analyse et de calibrage du spectre pour choisir les longueurs d'onde appropriées. Maintenant, cette technique est appliquée dans plusieurs disciplines telles que la médecine, la modélisation des procédés, les sciences pharmaceutiques...etc. La notation PLS est utilisée par la suite dans cette thèse pour désigner cette méthode.

La PLS cherche les composantes latentes qui maximisent la covariance entre **X** et **Y**, afin de trouver une relation entre les variables indépendantes (entrée **X**) et les variables dépendantes (sortie **Y**) en les décomposant en variables latentes non corrélées avec un processus itératif. La structure latente correspond à l'extraction du maximum de variation de **Y** qui explique au mieux la structure latente de **X**. Cette méthode est similaire à l'**analyse de corrélation canonique**, et est liée aux autres méthodes de régression telles que la **régression en composantes principales**.

La PLS a surmonté les problèmes de colinéarité (qui mènent à une mauvaise interprétation des coefficients de régression  $\beta$ ), et les problèmes de limitation du nombre de variables prédites par rapport aux méthodes de régression standards des moindres carrés. Elle permet d'analyser des données avec plus de variables que d'observations et produit des modèles avec une grande stabilité de prédiction car le risque de sur-apprentissage est minimisé. Elle est très efficace dans la gestion des valeurs manquantes et donc produit des modèles très robustes face aux défauts capteurs par exemple. Elle est aussi capable de détecter les valeurs aberrantes avec une analyse

plus soignée des données améliorant ainsi la qualité des modèles d'ajustement. Cette méthode permet de manipuler plus facilement le bruit et les données redondantes et corrélées et de réaliser une marginalisation des valeurs anormales pour améliorer la robustesse du système. Elle représente aussi une façon très propre d'inverser des matrices, à la différence de l'analyse en composantes principales qui cherche à trouver les composantes principales qui expliquent le maximum de variance d'un seul bloc  $X$  ou  $Y$  sans tenir compte des autres blocs. Enfin, elle a l'avantage de pouvoir prédire soit plusieurs variables réponse à la fois (PLS2) soit juste une seule variable réponse (PLS1). En général, plus les variables réponses sont corrélées meilleure sera la performance de PLS2 par rapport à PLS1 en termes d'interprétation et de précision de prédiction.

Les algorithmes PLS sont numériquement stables, leur simplicité d'implémentation et de calcul leur donne un aspect fort. Une des difficultés de réglage du PLS est le choix du nombre de valeurs latentes qui permet une bonne généralisation. Cela peut se faire avec une validation croisée.

La PLS est un outil très polyvalent qui peut réaliser plusieurs tâches : la régression (Melssen et al., 2007, Kim et al., 2005), la réduction de dimension (Tenenhaus et al., 2007), la classification (Barker and Rayens, 2003, Nguyen and Rocke, 2002), la supervision et le contrôle des procédés (Kourti, 2005), la surveillance des procédés (Zhou et al., 2010), et la modélisation (Shinzawa et al., 2011). Elle a aussi été utilisée avec succès dans la surveillance et la détection des défauts pour les données chimiques (Wise and Gallagher, 1996). La **PLS** est mieux adaptée au diagnostic de défauts que l'ACP sur le plan théorique et pratique (Chiang et al., 2000) car l'ACP s'intéresse seulement à la capture du maximum de variance sans tenir compte de la relation qui existe entre les données et les défauts. La performance et l'efficacité de la méthode PLS a été montrée dans plusieurs travaux et articles (Ferrer et al., 2008, Yin et al., 2011), et elle a été combinée aussi avec d'autres méthodes telles que les SVMs (Rosipal and Trejo, 2003, Lu and Wang, 2008).

Différentes méthodes ont été proposées pour étendre la méthode PLS au cas non linéaire. En utilisant la force des **méthodes noyau**, la PLS peut être étendue d'une manière simple et efficace au cas des systèmes non linéaires. Il y a aussi plusieurs variantes de PLS qui diffèrent selon la façon avec laquelle elles extraient les vecteurs latents. Un des algorithmes les plus élégants qui représentent la façon standard du PLS est le **PLS itératif non linéaire** (Geladi and Kowalski, 1986) qui est le plus approprié pour la supervision des procédés par rapport à la méthode SIMPLS (Li et al., 2010). L'algorithme SIMPLS a cependant l'avantage de calculer les facteurs PLS directement comme une combinaison linéaire des variables d'origine en maximisant un

critère de covariance et donc évite de calculer des matrices dégonflées comme c'est le cas dans l'algorithme NIPALS (de Jong, 1993).

### Fondement théorique de la méthode PLS

Soit un ensemble de données composé des deux matrices  $\mathbf{X}$  des mesures (entrées) et  $\mathbf{Y}$  des réponses (sorties). La PLS détermine la relation entre ces deux blocs, en essayant de garder le maximum de covariance entre eux, par les équations suivantes :

$$\begin{cases} X = TP^T + E \\ Y = UQ^T + F \end{cases} \quad (\text{I.21})$$

$\mathbf{X}$  est la matrice prédicateurs de données de taille  $n \times m$  où chaque ligne représente une observation ou les mesures à un échantillon donné, et chaque colonne représente les mesures d'une variable.  $\mathbf{T}$  est la matrice de scores de  $\mathbf{X}$ ,  $\mathbf{P}$  la matrice de chargement de  $\mathbf{X}$ ,  $\mathbf{Y}$  la matrice de réponse de taille  $n \times m_y$ ,  $\mathbf{U}$  la matrice des scores de  $\mathbf{Y}$ ,  $\mathbf{Q}$  la matrice de chargement de  $\mathbf{Y}$ ,  $\mathbf{E}$  et  $\mathbf{F}$  sont les matrices résiduelles.

La modélisation de  $\mathbf{X}$  et  $\mathbf{Y}$  de cette façon suppose que les deux sont observés avec une erreur (résidu). Le chargement se fait à travers les vecteurs latents  $\mathbf{T}$  qui peuvent être choisis de différentes façons, à condition qu'ils soient orthogonaux (décorrélés) couvrant l'espace des colonnes de  $\mathbf{X}$ . Des conditions additionnelles sont requises afin de spécifier  $\mathbf{T}$  pour que les vecteurs de  $\mathbf{T}$  soient orthogonaux, telle la création d'une combinaison linéaire des colonnes de  $\mathbf{X}$  et  $\mathbf{Y}$  en multipliant ces dernières par les poids  $\mathbf{w}$  et  $\mathbf{c}$  respectivement de telle sorte que leur covariance soit maximale (Geladi and Kowalski, 1986, Rosipal and Krämer, 2006) :

$$\begin{aligned} t &= Xw, u = Yc, w^T w = 1, t^T t = 1 \\ \text{tel que } t^T u &\text{ soit maximale} \end{aligned} \quad (\text{I.22})$$

$\mathbf{t}$ ,  $\mathbf{u}$  sont les vecteurs scores de  $\mathbf{X}$  et  $\mathbf{Y}$  respectivement.  $\mathbf{w}$  et  $\mathbf{c}$  sont les coefficients de régression de  $\mathbf{X}$  et  $\mathbf{Y}$  sur  $\mathbf{u}$ ,  $\mathbf{t}$  respectivement (vecteurs poids de  $\mathbf{X}$  et  $\mathbf{Y}$ ). Quand les premiers vecteurs latents ( $\mathbf{t}$  et  $\mathbf{u}$ ) sont trouvés, ils sont soustraits de  $\mathbf{X}$  et  $\mathbf{Y}$  et la procédure est réitérée jusqu'à ce que  $\mathbf{X}$  devienne une matrice nulle.

$$w = (u^T u)^{-1} X^T u \quad \text{et} \quad c = (t^T t)^{-1} Y^T t \quad (\text{I.23})$$

Avant de commencer le processus itératif, le vecteur  $\mathbf{u}$  est initialisé d'une façon aléatoire par le choix d'une colonne de  $\mathbf{Y}$  (celle avec la plus grande variance).

$$w = \frac{X^T u}{u^T u}, \|w\| \rightarrow 1, t = Xw, c = \frac{Y^T t}{t^T t}, \|c\| \rightarrow 1, u = Yc \quad (\text{I.24})$$

On répète ces étapes jusqu'à la convergence. La PLS est une méthode itérative, qui détermine à chaque itération les vecteurs poids  $w$ ,  $c$  ou les vecteurs scores  $t$ ,  $u$  et les chargements  $p$  sont calculés, ainsi  $X$  et  $Y$  seront ajustés par ces résultats :

$$\begin{cases} X = X - t.p^T = X - tt^T X \\ Y = Y - u.q^T = Y - tt^T Y \end{cases} \quad (\text{I.25})$$

Où  $p$  et  $q$  sont les coefficients de régression de  $X$  et  $Y$  sur  $t$ ,  $u$  respectivement

$$p = (t^T t)^{-1} X^T t \quad \text{et} \quad q = (u^T u)^{-1} Y^T u \quad (\text{I.26})$$

Après chaque itération les vecteurs  $w$ ,  $t$ ,  $p$  et  $q$  sont enregistrés dans les matrices  $W$ ,  $T$ ,  $P$ , et  $Q$  respectivement. On peut voir  $w$  comme le vecteur propre à droite de la matrice  $S = X^T Y$ , et le problème devient :

$$\begin{aligned} X^T Y Y^T X w &= \lambda w \\ t &= X w \end{aligned} \quad (\text{I.27})$$

Dans PLS régression le but est de prédire  $Y$  en utilisant  $X$  et leur structure commune :

$$Y = X\beta + \varepsilon \quad (\text{I.28})$$

D'après (Rosipal and Krämer, 2006), la structure commune est donnée par la relation suivante :

$$\beta = W(P^T W)^{-1} C^T \quad (\text{I.29})$$

ce qui donne :

$$\beta = X^T U (T^T X X^T U)^{-1} T^T Y \quad (\text{I.30})$$

**PLS à noyau** : (KPLS) est une extension non linéaire du PLS linéaire. L'utilisation de fonction noyau (*Kernel*) permet de transformer (mapper) les données complexes non linéaires de l'entrée dans un espace de caractéristiques de plus grande dimension dans lequel la modélisation linéaire peut se faire facilement (Rosipal and Trejo, 2002). Cette notion de noyau va être détaillée par la suite dans la théorie des SVMs. La fonction noyau est donnée comme suit :

$$K_x(x_i, x_j) = \langle \psi(x_i), \psi(x_j) \rangle \quad (\text{I.31})$$

Avant de commencer le processus itératif, le vecteur  $u$  est initialisé d'une façon aléatoire :

$$1) t = \psi \psi^T u = K_x u \quad 2) \|t\| \rightarrow 1 \quad 3) c = Y^T t \quad 4) u = Yc \quad 5) \|u\| \rightarrow 1 \quad (\text{I.32})$$

En utilisant l'équation (I.27) en multipliant les deux termes par  $X$  on obtient :

$$\begin{aligned} XX^T YY^T t &= \lambda t \\ u &= YY^T t \end{aligned} \quad (I.33)$$

Considérons  $K_y = \varphi\varphi$  la matrice noyau du Gram (*Kernel Gram matrix*) de la sortie  $\mathbf{Y}$  et  $\varphi$  la transformation de l'espace de sortie de  $\mathbf{Y}$  vers l'espace de caractéristique, d'où :

$$\begin{aligned} K_x K_y t &= \lambda t \\ u &= K_y t \end{aligned} \quad (I.34)$$

Pour centrer les données transformées dans l'espace des caractéristiques, cela revient à :

$$\begin{aligned} K_x &\leftarrow (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) K_x (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \\ K_y &\leftarrow (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) K_y (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \end{aligned} \quad (I.35)$$

Où  $I_n$  est la matrice identité de taille  $\mathbf{n}$ , et  $\mathbf{1}_n$  est le vecteur unité de taille  $\mathbf{n}$ . Après le calcul des nouveaux scores  $\mathbf{t}$  et  $\mathbf{u}$ , on soustrait des matrices  $K_x$  et  $K_y$  leurs approximations calculées à partir de  $\mathbf{t}$  et  $\mathbf{u}$  :

$$\begin{aligned} K_x &\leftarrow (I_n - t t^T) K_x (I_n - t t^T) \\ Y &= Y - t t^T Y \end{aligned} \quad (I.36)$$

Et donc on obtient :

$$\begin{aligned} \beta &= \psi^T U (T^T K_x U)^{-1} T^T Y \\ \hat{Y} &= \psi \beta = K_x U (T^T K_x U)^{-1} T^T Y = T T^T Y \end{aligned} \quad (I.37)$$

### I.2.3. Etapes de mise en œuvre des méthodes basées sur les données

Malgré le fait qu'on s'intéresse principalement à l'utilisation des méthodes basées sur les données pour le diagnostic de défaut, il est important de noter que ces méthodes peuvent également être utilisées pour la modélisation, donc avec un objectif différent

- Dans l'objectif de **diagnostic**, les méthodes statistiques permettent de relier les entrées et sorties du système au type de situation (défectueuse ou normale pour différents défauts).
- Dans l'objectif de **modélisation (régression)**, on cherche par ces méthodes à trouver une relation empirique entre les entrées et les sorties.

L'application des méthodes basées sur les données pour le diagnostic de défaut est réalisée en plusieurs étapes : l'acquisition et prétraitement des données, l'extraction des caractéristiques

importantes, la modélisation (soit entrée-sortie soit données-type de défaut) et la validation du modèle. Selon l'objectif visé (modélisation ou diagnostic), ces étapes sont à adapter.

### ***1.2.3.1. Acquisition des données***

Cette étape concerne l'accueil (collection) et la préparation (prétraitement) des données. Ces données peuvent être échantillonnées dans le domaine temporel ou fréquentiel avec un intervalle d'échantillonnage adapté à la dynamique du procédé. Des données riches et représentatives des différentes situations sont nécessaires pour l'apprentissage de la dynamique du procédé.

On distinguera deux approches de collections des données :

- pour la **modélisation** (entrées-sorties), la réponse transitoire est nécessaire. Dans ce cas, il est intéressant d'exciter le procédé par une entrée de spectre convenable (par exemple signal pseudo aléatoire) afin de capter le maximum d'information. La réponse transitoire d'un système (la sortie face à une entrée transitoire, tel un échelon ou une impulsion) est souvent utilisée pour la modélisation. Elle permet de comprendre le comportement dynamique du système, tel que le retard et les échelles de temps.
- pour le **diagnostic** (relation mesures-type de défaut), des données dans les situations normale et défectueuse sont nécessaires.

Supposons  $\mathbf{X}$  une matrice de données (contenant les entrées  $\mathbf{U}$  et les sorties mesurées  $\mathbf{Y}$ ) de taille  $n \times m$  constituée de  $\mathbf{n}$  récurrences de mesures (observations, ou encore échantillons) et  $\mathbf{m}$  variables **caractéristiques** ( $X_j$ ) (facteurs ou encore propriétés) dans les colonnes de la matrice:

$$X = [U, Y] = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} = [X_1, \dots, X_j, \dots, X_m]$$

Les données acquises sont généralement séparées en deux groupes : **données d'apprentissage** et **données de test**. Les données d'apprentissage sont utilisées pour le développement du modèle (sélection des caractéristiques et développement du modèle). Les données de test sont utilisées pour la validation (évaluation ou jugement) du modèle et des caractéristiques choisies.

### ***1.2.3.2. Prétraitement des données***

Un prétraitement des données est nécessaire pour réduire le bruit, la suppression des valeurs aberrantes, le traitement des valeurs manquantes et la normalisation des données.

- **Normalisation des données**

La normalisation est très utile lorsque le système a plusieurs variables dont les amplitudes et variances varient dans des plages différentes. Elle consiste à mettre à l'échelle la matrice de données  $X$  en rendant les variables de la nouvelle matrice  $X_N$  comparables (la même plage de variation). Parmi les méthodes de **normalisation** de données, on peut citer :

1. La méthode min/max : met les valeurs de chaque colonne de  $X$  selon le maximum et le minimum de cette colonne, de telle sorte à avoir les valeurs de la matrice normalisé  $X_N$  dans l'intervalle [bas-haut] :

$$X_N = (\text{haut} - \text{bas}) \times \frac{X - \text{minimum}(X)}{\text{maximum}(X) - \text{minimum}(X)} + \text{bas}$$

Où  $\text{minimum}(X)$  et  $\text{maximum}(X)$  remplace respectivement chaque colonne de  $X$  par son minimum et maximum. Ce type de normalisation avec  $\text{haut}=1$  et  $\text{bas}=0$  est utilisé dans cette thèse au chapitre II.

2. La méthode Z-scores : permet de ramener la moyenne de chaque variable de  $X$  à zéro et sa variance à l'unité par l'application de la formule suivante :  $X_N = \frac{X - \text{moyenne}(X)}{\text{écart type}(X)}$
3. La méthode de mise à l'échelle décimale « *decimal scaling* »
4. La méthode de fenêtre coulissante ou glissante.

- **Manipulation des données manquantes**

Dans le monde réel, les données ne sont pas toujours propres et complètes. Souvent elles sont censurées et corrompues. Plusieurs techniques existent pour traiter les données manquantes, mais la méthode la plus simple et la plus utilisée consiste à déterminer la ligne de la matrice où des données sont corrompues puis la supprimer. Malgré sa simplicité, cette méthode réduit la taille de la matrice de données. Cela conduit à une perte de puissance (informations) du résultat. En plus, il y a des cas où la ligne corrompue contient des données importantes qu'on ne peut pas supprimer, ou des cas où la matrice de données a beaucoup de valeurs manquantes et donc la suppression de toutes ces lignes mène à une matrice avec peu de données. Dans ce cas, on a recours à la méthode de substitution qui remplace les données manquantes par une

approximation raisonnable en supposant que ces données sont faciles à modéliser. Parmi les méthodes de substitution se trouve l'interpolation ou la substitution par régression, avec une moyenne ou avec la médiane. La précision de cette méthode dépend de la précision du modèle de données.

- **Valeurs aberrantes**

Les bases de données peuvent également contenir des données erronées ou fausses. La manipulation et la détection des valeurs aberrantes est nécessaire à toute modélisation car leur présence détériore l'élaboration d'un bon modèle. Parmi les techniques de détection des valeurs aberrantes dans les bases de données on cite la distance de Mahalanobis et l'analyse de résidu de la régression PLS avec les réseaux de neurones (Hugo and Sbarbaro, 2011).

### ***1.2.3.3. Sélection et extraction des variables pertinentes***

Un problème sérieux se pose lorsqu'on utilise toutes les variables (caractéristiques ou encore facteurs) pour réaliser le diagnostic ou la modélisation. L'utilisation de toutes les caractéristiques peut rendre le modèle de classification moins précis et non robuste au bruit et donc diminuer ses performances. Les variables qui sont incluses dans un modèle et qui ne lui sont pas bénéfiques ajoutent de la variance aux prédictions. De l'autre côté, les variables qui sont utiles et qui ne sont pas incluses dans le modèle vont causer un biais dans les estimations. La sélection et l'extraction des caractéristiques est donc une étape très importante à la fois pour améliorer la qualité et les performances de la modélisation (ou la classification, ex. éviter le surapprentissage du bruit) et pour surmonter la difficulté de la dimensionnalité (gain du temps de calcul et de mémoire utilisée). En effet, certains algorithmes peuvent même échouer si le nombre de caractéristiques est plus grand que le nombre d'observations.

La première approche pour l'extraction des variables pertinentes consiste à utiliser **l'analyse et la fouille de données** (Witten et al., 2011) pour l'exploration des données afin de trouver et interpréter les relations entre les variables pour une meilleure compréhension et utilisation. Cette discipline est le fruit d'interactions de plusieurs domaines : statistique, intelligence artificielle, et informatique. Elle est utilisée dans de nombreux secteurs d'activités comme par exemple la détection des comportements anormaux (diagnostic de défauts), l'analyse de comportement des consommateurs (finance), ou la prédiction de temps de réponse, etc. L'analyse ou la fouille des données essaie d'analyser et d'extraire des informations (ou des connaissances, ou des variables) les plus importantes, utiles et pertinentes et leur relation les unes avec les autres à partir d'une grande quantité de données pour des objectifs précis. La réduction du nombre de caractéristiques

se fait par une des deux façons suivantes : La **sélection des caractéristiques** ou la **transformation des caractéristiques**.

### 1. La sélection des caractéristiques :

La sélection de caractéristiques joue un rôle important dans la classification. Elle permet de simplifier le modèle en réduisant le coût de calcul et en facilitant son implémentation. On s'intéresse seulement aux mesures qui sont pertinentes en écartant les caractéristiques insignifiantes de l'ensemble de données. Cela rend le modèle plus lisible et compréhensible et plus robuste au bruit et aux incertitudes, ce qui améliorera les performances (précision, erreur quadratique, etc.) de classification/régression des différents modèles. La sélection de caractéristiques permet de contourner la malédiction de la dimensionnalité et d'éviter le sur-apprentissage.

Les algorithmes de sélection choisissent un sous-ensemble de caractéristiques (appelés prédictifs) à partir des caractéristiques originales afin de construire un modèle optimisant les erreurs prédictives. Ces algorithmes se décomposent en deux catégories :

- **Méthodes de filtre** : Elles reposent sur les caractéristiques générales des données afin d'évaluer et de sélectionner les sous-ensembles de caractéristiques sans impliquer l'algorithme d'apprentissage.
- **Méthodes d'enveloppe** ou d'emballage (Wrapper) : Elles utilisent les performances de l'algorithme d'apprentissage choisi pour évaluer chaque sous-ensemble de caractéristique candidat.

Dans (Mahadevan and Shah, 2009) une méthode de sélection de caractéristiques appelée **élimination récursive de caractéristiques par SVM** a été utilisée pour la sélection des caractéristiques pour la discrimination dans le cas de la classification et dans (He et al., 2008) elle a été utilisée pour la prédiction dans le cas de régression.

La sélection et l'extraction des caractéristiques peuvent être réalisées aussi par un **algorithme génétique** (AG). C'est une approche d'optimisation stochastique utilisant la théorie de Darwin de la sélection naturelle et les travaux de Mendel dans la génétique pour simuler l'évolution biologique (Chakraborty, 2008). Elle cherche à trouver une solution optimale sans le besoin d'avoir des connaissances théoriques sur le problème à résoudre. Contrairement aux modèles gradients qui peuvent se coincer sur un optimum local, l'algorithme génétique tente via plusieurs points de recherche de trouver d'autres parties de l'espace de recherche et de surpasser

l'optimum local pour atteindre un optimum global. L'AG cherche à maximiser une fonction Fitness  $J(\theta)$  qui mesure l'aptitude d'un individu  $\theta$  à survivre, à s'accoupler et à produire une descendance (progéniture) dans une population d'individus pour un environnement donné. L'efficacité de l'application de l'algorithme génétique dépend essentiellement de la façon de codage du problème et le choix de la fonction objective (fitness). Plusieurs travaux ont utilisés des algorithmes génétiques pour l'optimisation et la sélection des caractéristiques en général (Khushaba et al., 2011) ou les caractéristiques dans l'apprentissage des SVR (He et al., 2008).

La sélection des caractéristiques peut se faire aussi dans le domaine fréquentiel (voir annexe A.3)

On note qu'on peut utiliser les poids  $w_i$  dans les SVM comme des poids de caractéristiques pour obtenir leur classement respectif. Quelques indices peuvent aider la sélection de caractéristiques dans les problèmes de classification en maximisant la dépendance entre deux quantités ou variables tous en minimisant la redondance :

#### A. L'intercorrélation pour l'estimation du temps de retard :

Une valeur d'inter-corrélation proche de zéro indique une absence de relation linéaire entre ces deux variables, alors qu'une valeur proche de 1 signifie une forte relation entre les deux variables.

$$v(t) = \alpha u(t - D) + \omega(t) \quad (I.38)$$

Ou  $w$  est un bruit,  $D$  le retard,  $u$  signal d'entrée et  $\alpha$  le facteur d'atténuation. Donc, l'intercorrélation entre deux signaux ou variables est donnée par :

$$\begin{aligned} R_{vu}(\tau) &= E\{v(t)u(t + \tau)\} = E\{\alpha u(t - D) + w(t)u(t + \tau)\} \\ &= E\{\alpha u(t - D)u(t + \tau) + w(t)u(t + \tau)\} = \alpha R_{uu}(\tau - D) + R_{wu}(\tau) \end{aligned} \quad (I.39)$$

Par la supposition que le bruit et l'entrée sont non corrélés, c.à.d.  $R_{wu}(\tau) = 0$ , on obtient

$$R_{vu}(\tau - D) = \alpha R_{uu}(\tau - D) \quad (I.40)$$

Donc il y aura un seul pic dans  $R_{vu}$  qui correspond à  $\tau = D$ .

#### B. Information mutuelle :

L'information mutuelle rentre dans le cadre de la théorie d'information et la théorie des probabilités. Dans cette théorie, on calcule  $I(X, Y)$ , une quantité qui mesure la dépendance au sens probabilistique (l'information apportée par une réalisation de  $\mathbf{X}$  sur  $\mathbf{Y}$ ) de deux variables

aléatoires  $\mathbf{X}$  et  $\mathbf{Y}$ . Cette mesure permet de réduire au maximum l'information incertaine qui relie  $\mathbf{X}$  et  $\mathbf{Y}$ .

$$I(X, Y) = - \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (\text{I.41})$$

Où  $\mathbf{p(x)}$ ,  $\mathbf{p(y)}$  et  $\mathbf{p(x,y)}$  sont les densités de probabilité de  $x$ ,  $y$  et de  $(x,y)$  respectivement. La connaissance disponible sur le procédé, ou sur un défaut donné, ou sur tout phénomène aléatoire peut être représentée par la distribution de probabilité :

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X/Y) = H(Y) - H(Y/X) \end{aligned} \quad (\text{I.42})$$

Où  $\mathbf{H(X)}$  et  $\mathbf{H(Y)}$  sont des entropies,  $\mathbf{H(X,Y)}$  l'entropie conjointe, et  $\mathbf{H(X/Y)}$  et  $\mathbf{H(Y/X)}$  sont les entropies conditionnelles (voir l'annexe A.2).

L'information mutuelle peut être appliquée à plusieurs domaines, comme par exemple le diagnostic de défaut dans un procédé Tennessee Eastman (Verron et al., 2007). Néanmoins, l'application principale reste la sélection de caractéristiques dans les problèmes de classification (Peng et al., 2005), où le but est de maximiser la dépendance entre deux quantités ou variables.

Dans cette thèse, l'information mutuelle et l'intercorrélacion sont proposés dans le chapitre II pour aider dans le choix des variables caractéristiques dans un problème de régression.

## 2. La transformation des caractéristiques, fonctionnalités, attributs ou prédicteurs

La famille des transformations des caractéristiques concerne la création de nouvelles variables prédictives à partir des variables initiales, ce qui réduit la dimension de l'espace d'entrée par transformation. Ces méthodes comprennent la technique d'analyse en composantes principales (ACP), Partial least square (PLS), analyse factorielle, etc.

Après plusieurs essais de ces méthodes de transformation et sélection de caractéristiques pour le problème de diagnostic de défauts sur trois procédés différents (éolien, procédé de traitement des eaux, réacteur chimique CSTR), j'ai constaté que parfois ce n'est pas suffisant d'utiliser seulement les caractéristiques d'origine (ou leurs transformés) car elles peuvent être non informatives, redondantes et ayant des interactions entre elles. La solution est de créer de nouvelles caractéristiques qui décrivent au mieux un défaut donné.

## **La génération des caractéristiques « *Feature Generation* » : (GC)**

C'est un nouveau domaine de recherche établi dans l'apprentissage automatique. Cette étape conduit à des améliorations significatives des performances et à une meilleure compréhension du problème (Pachet and Roy, 2009) avec un moins de coût dans plusieurs tâches de l'apprentissage statistique dans la modélisation (régression) ou la classification.

Dans cette thèse, le schéma de diagnostic proposé est basé sur la génération de caractéristiques descriptives des défauts parmi lesquelles se trouve le résidu. Cette génération a amené à un résultat plus performant et précis et a contribué à une meilleure compréhension de la notion d'apprentissage.

La **construction manuelle** des caractéristiques peut donner des résultats intéressants (réalisée dans les applications de l'éolienne et du réacteur parfaitement agité dans ce travail). Mais elle demeure une procédure laborieuse, coûteuse et bien spécifique. Par conséquent, les caractéristiques trouvées sont dépendantes du domaine.

### ***1.2.3.4. Modélisation***

Les méthodes basées sur les données sont utilisées dans cette thèse pour deux finalités :

1. Le **diagnostic de défaut**, où on attend de ces méthodes de relier les entrées + sorties  $\mathbf{X}=[\mathbf{U},\mathbf{Y}]$  au type de situation ( $\mathbf{Z}$ ) (défectueuse ou normale pour différents défauts). Un modèle comportemental (ou empirique) est attendu, contrairement aux méthodes basées sur les modèles qui utilisent le modèle fondamental du système.
2. La **modélisation**, où on attend de ces méthodes de trouver une relation empirique entre les entrées + sorties  $\mathbf{X}=[\mathbf{U},\mathbf{Y}]$  et quelques états du système (mesurés ou non mesurés). Là encore, un modèle comportemental (empirique) est obtenu. Cette modélisation donne une estimée qui servira à enrichir les données dans l'objectif de diagnostic de défauts.

Ce paragraphe discute la modélisation entrées-sorties du système par des modèles empiriques « *data driven modeling* » (Solomatine and Ostfeld, 2008, Niska, 2011).

De manière générale, les modèles (fondamentaux ou empiriques) améliorent la compréhension intrinsèque du système et son comportement. Ils peuvent être utilisés pour l'analyse de la sécurité, la conception, la simulation et le contrôle du système, ainsi que pour l'apprentissage des opérateurs et le diagnostic de défauts. Un modèle doit être précis, simple et général, et il est

préférable qu'il soit facile à construire, à comprendre et à tester. On note que tous les modèles (même fondamentaux) sont faux puisqu'ils donnent juste une image de la réalité (système).

Le choix de la structure du modèle dépend des connaissances a priori du système. Sans connaissance a priori, le modèle de type **boîte noire** (*black box model*) est choisi. Il a une structure flexible et une bonne puissance d'approximation. Avec des connaissances sur la physique décrivant la dynamique du système, on choisit la structure boîte grise (*grey box model*) qui incorpore les connaissances a priori avec une structure d'équations différentielles en plus de certains paramètres ou constantes inconnues. La structure du modèle dépend également du type de la relation entre entrées et sorties, linéaire ou non linéaire. On peut également choisir d'introduire l'influence des perturbations sur les entrées/sorties. En effet, la connaissance de la nature des perturbations est très utile pour prédire les sorties futures du système par la compréhension des corrélations qui existent entre ces perturbations.

La technique la plus courante dans l'ajustement et la modélisation est probablement **la régression**. Elle cherche à trouver une fonction **f** qui explique la relation entre deux variables **y** et **x** de telle sorte à avoir :

$$y = f(x) + \varepsilon, \quad x \in \mathfrak{R}^p \quad (\text{I.43})$$

Où **p** est le nombre des variables explicatives et **ε** est le bruit. Minimiser l'erreur au sens des moindres carrés veut dire minimiser l'erreur entre la donnée réelle **y** et son estimée  $\hat{y}$  :

$$\text{résidu} = y - \hat{y} \quad (\text{I.44})$$

La régression requiert des hypothèses que l'utilisateur doit spécifier pour construire le modèle décrit par la relation entre les entrées et les sorties. Le choix du modèle est basé sur les connaissances sur les dynamiques du procédé. Cela est raisonnable lorsque ces hypothèses sont basées sur les lois physiques ou des détails connus dans le domaine concerné.

La **régression linéaire** consiste à trouver la droite qui passe au mieux (au sens des moindres carrés) à travers le nuage de points représentant les données. Par l'utilisation des fonctions noyaux (Kernel), la régression peut appréhender une **fonction non linéaire**. Les méthodes de régression conventionnelles, comme la régression linéaire multiple, souffrent de plusieurs défauts, comme l'incapacité de déterminer un modèle si le nombre d'observations est inférieur aux variables explicatives, l'incapacité de prendre en compte les données manquantes et l'incapacité de manipuler les variables colinéaires. Ces problèmes peuvent être réglés par des méthodes telles que l'orthogonalisation des prédicteurs (**Régression sur les composantes**

principales), les méthodes de régularisation (*Ridge regression*), la régression PLS (partial least square) ou la régression par machines à vecteurs supports.

### I.2.3.5. Validation du modèle

Le critère de choix du modèle dépend de sa capacité d'ajustement sur les données d'apprentissage (estimation) et de test (validation). Un modèle avec un grand aspect de généralisation est capable de représenter le mieux la réalité. Et si deux modèles ont presque les mêmes performances d'ajustement, on choisit celui dont la structure est la plus simple. Construire un modèle intégrant trop de paramètres (sur-paramétrisation) conduit à une mauvaise généralisation. L'évaluation de la capacité de généralisation d'un nouveau modèle peut être réalisée en utilisant la méthode de *validation croisée*. Une fois le modèle développé, sa robustesse (incertitude, ajustement) peut être évalué en déterminant les intervalles (bandes) de confiance. Cela peut se faire par la méthode *Bootstarp* (Efron, 1988) par exemple. Aussi, une analyse des résidus comparant la sortie du modèle avec la sortie réelle peut confirmer la représentativité du modèle.

L'annexe A (A.4 et A.5) décrit plus en détail les méthodes validation croisé et Bootstrap. Ces deux méthodes sélectionnent le modèle en se basant sur certains critères qui évaluent la performance du modèle comme le compromis biais-variance, l'indice de sensibilité standard d'une variable sur la sortie, le critère d'Information d'Akaike ou le critère de la longueur de description minimale. Dans le chapitre suivant les indices de performances de (I.45) jusqu'à (I.53) sont utilisés pour le choix du meilleur modèle :

1. **Erreur quadratique moyenne (MSE)** « *Mean Squared Error* » :

$$MSE = \frac{1}{N} \sum_{i=1}^N (z_i - \hat{z}_i)^2 \quad (I.45)$$

$z_i$ , et  $\hat{z}_i$  sont respectivement la sortie et son estimée de l'échantillon  $i$

2. **Erreur quadratique moyenne (MAE)** « *Mean Absolute Error* » :

$$MAE = \frac{1}{N} \sum_{i=1}^N |z_i - \hat{z}_i| \quad (I.46)$$

3. **Erreur quadratique moyenne normalisée (NMSE)** « *Normalized Mean Squared Error* » :

$$NMSE = \frac{1}{\sigma^2} \frac{1}{N} \sum_{i=1}^N (z_i - \hat{z}_i)^2 \quad (I.47)$$

Il faut que  $NMSE < 1$

4. **Racine de l'erreur quadratique moyenne (RMSE)** « *Root Mean Squared Error* » :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_i - \hat{z}_i)^2} \quad (I.48)$$

5. **Racine normalisée des erreurs quadratiques moyennes (NRMSE)** « *Normalized Root Mean Squared Error* » :

$$NRMSE = \sqrt{\frac{\sum_{i=1}^N (z_i - \hat{z}_i)^2}{\sum_{i=1}^N z_i^2}} \quad (I.49)$$

6. **La somme relative des erreurs quadratiques (RSSE)** « *Relative Sum of Squared Error* » :

$$RSSE = \frac{\sum_{i=1}^N (z_i - \hat{z}_i)^2}{\sum_{i=1}^N z_i^2} \quad (I.50)$$

7. **Le coefficient de détermination  $R^2$**  :

$$R^2 = 1 - \frac{\sum_{i=1}^N (z_i - \hat{z}_i)^2}{\sum_{i=1}^N (z_i - \bar{z})^2} \quad (I.51)$$

8. **Le critère d'Akaike (AIC)** : (Akaike, 1974), (Bozdogan, 2000) Le degré optimal est défini comme le degré qui minimise AIC :

$$AIC = 2 \times k - 2 \log L(\hat{\theta}) \quad (I.52)$$

Où  $k$  est le nombre de paramètres libres dans le modèle,  $L(\hat{\theta})$  est la fonction de maximum de vraisemblance

9. **Rapport ou ratio des vecteurs supports (SV) dans les SVM** : Le rapport entre le nombre des points supports divisé par le nombre total des données d'apprentissage. Une valeur importante de ce ratio indique que le problème est difficile à résoudre :

$$Ratio\ SV = \frac{NSV}{N} \quad (I.53)$$

10. **Compromis biais-variance** : Un biais désigne un écart systématique entre une grandeur ( $P$ ) et son estimation (prédiction) ( $\hat{P}$ ) :

$$Biais = E[\hat{P}] - P \quad (I.54)$$

Les performances d'un modèle sont mesurées par la valeur de l'erreur quadratique moyenne :

$$EQM = Variance(\hat{P}) + Biais(\hat{P})^2 \quad (I.55)$$

L'incertitude peut être définie comme la racine carrée de la somme des carrés du biais et la variance ( $\sigma$ ) :

$$U = \sqrt{\sigma^2 + \text{biais}^2} \quad (\text{I.56})$$

Pour avoir une erreur minimale et donc un bon modèle, il faut diminuer le biais. Cela peut conduire à l'augmentation du nombre de paramètres (ou degré) du modèle, ce qui entraîne l'augmentation de la variance, puisque le modèle va suivre trop de détails des échantillons (observations). Une grande variance augmente l'erreur quadratique. Il est nécessaire donc de faire un compromis biais-variance de telle sorte à augmenter la précision du modèle sans trop augmenter son degré de façon à minimiser le biais et la variance en même temps.

### I.3. Les Machines à Vecteurs de Supports

La méthode des machines à vecteurs supports (SVM) « *Supports Vector Machines* » fait partie des méthodes basées sur les données. Elle découle de la théorie d'apprentissage statistique. Elle est utilisée dans cette thèse pour la détection de défaut, et également pour la régression (modélisation).

Depuis plus d'un demi-siècle, les chercheurs essaient de résoudre le problème de l'apprentissage automatique supervisé et plus particulièrement celui de la reconnaissance de formes (motifs) avec la meilleure façon possible. En 1956, Frank Rosenblatt's a construit la première machine d'apprentissage appelée *Perceptron* où il a décrit le modèle comme un algorithme capable d'être utilisé par des ordinateurs. Ce modèle reçoit en entrée une somme pondérée de signaux d'entrée plus un seuil appelé biais, et fournit une sortie qui dépend de l'entrée. Parmi les caractéristiques du *Perceptron* est sa capacité de généralisation sur d'autres données non utilisées dans l'apprentissage.

Malgré que les années 60 à 70 n'aient pas donné un résultat révolutionnaire, elles constituent néanmoins les années de la construction des fondements et bases de la *théorie d'apprentissage statistique*. Dans les années 80, il y a eu l'émergence des *réseaux de neurones* ou *multicouches perceptron* inspirés par le système nerveux biologique. Ce type de méthodes est basé sur la minimisation du risque empirique (voir section I.3.1).

La méthode des machines à vecteurs supports (SVM) est un nouvel estimateur non linéaire des fonctions, dont les fondements reposent sur la Théorie Statistique de l'Apprentissage (Vapnik, 2000, Vapnik, 1999b, Vapnik, 1998, Gunn, 1998). Le SVM est le classifieur le plus populaire,

efficace et sophistiqué des algorithmes d'apprentissage automatique introduit par Vapnik dans le début des années 1990 (Boser et al., 1992), conçu pour trouver un séparateur optimal (représenté par les paramètres d'une fonction de décision) qui maximise la marge entre les frontières de chaque classe, par l'utilisation d'un ensemble limité d'entraînement. L'algorithme a été construit au début pour résoudre les problèmes de discrimination (classification) et de reconnaissance de formes. Il a été ensuite étendu pour le cas de régression.

Dans cette thèse, le terme **SVMs** est utilisé pour décrire à la fois la classification et la régression en utilisant les méthodes à vecteurs de support. Le terme **SVM** sera utilisé pour décrire la classification par les machines à vecteurs de Supports tandis que le terme **SVR** sera utilisé pour décrire les méthodes de régression à vecteurs de supports.

### I.3.1 Théorie de l'apprentissage statistique

La théorie « *Statistical Learning Theory* » a été introduite par Vladimir Vapnik (Vapnik, 2000, Vapnik, 1999b, Vapnik, 1998).

Soit l'ensemble d'apprentissage **D** tiré suivant une loi de probabilité de distribution conjointe inconnue  $P(x, y)$  :  $D = \{(x_i, z_i)\}, \quad i = 1, \dots, N$

Le but de l'apprentissage est de trouver une fonction **f** qui estime au mieux les sorties  $z_i$  à partir d'un nombre limité d'entrées  $x_i$ . Les approches traditionnelles cherchent à minimiser la mesure de coûts sur les **N** données d'apprentissage, en d'autres termes minimiser les erreurs sorties  $z_i$  et leurs estimées  $\hat{z}_i = f(x_i)$ . Cela s'appelle la minimisation du risque empirique (**MRE**) :

$$R_{emp}[f] = \frac{1}{N} \sum_{i=1}^N (z_i - \hat{z}_i) = \frac{1}{N} \sum_{i=1}^N L(z_i, f(x_i)) \quad (I.57)$$

Le risque réel (ou fonctionnel) **R** qui représente l'espérance des coûts sur des nouvelles observations (tous les observations possibles) est donné comme suit :

$$R[f] = \int_{x,z} (z - \hat{z}) P(x, z) dx dz \quad (I.58)$$

Vapnik a cherché à trouver une relation qui lie le risque empirique mesuré (**R<sub>emp</sub>**) et le risque réel **R**, et a démontré que :

$$R[f] \leq R_{emp}[f] + \varphi(dVC) \quad (I.59)$$

Où **dVC** est la dimension de Vapnik-Chervonenkis correspondant à la capacité d'un algorithme de classification, qui est lié à sa complexité. Le terme à droite de l'inégalité appelé le risque garanti est composé de deux termes : le risque empirique et une quantité qui dépend du rapport **dVC** appelée **intervalle de confiance**  $\phi(\mathbf{dVC})$  puisqu'elle représente la différence entre le risque empirique et le risque réel. La **MRE** n'est pas suffisante pour minimiser le risque réel. Vapnik et Chervonenkis proposent alors une nouvelle mesure appelée la Minimisation du Risque Structurel (**MRS**) basée sur la minimisation conjointe des deux causes d'erreur, le risque empirique et l'intervalle de confiance. Le but de Vapnik est de trouver la valeur optimale de la dVC afin de minimiser le risque réel.

### I.3.2. Les machines à Vecteurs de Support pour la classification (SVM)

Considérons le problème de séparer l'ensemble **D** composé de **N** paires de données/classes d'apprentissage appartenant à deux classes ( $\Omega_1, \Omega_2$ ) :

$$D = \{(x_1, z_1), \dots, (x_i, z_i), \dots, (x_N, z_N)\}, \quad i = 1 \dots N, \quad x_i \in \mathbb{R}^p, \quad z_i \in \{-1, +1\}$$

$$\Omega_1 = \{x_i | z_i = +1\}, \quad \Omega_2 = \{x_i | z_i = -1\}$$

Chaque exemple d'entrée  $x_i$  (nommé aussi échantillon, observation, forme ou encore motif) est caractérisé par un ensemble de **p** variables descriptives. Ainsi, le vecteur **x** est donné comme suit :  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$

La sortie  $z_i$  définit la classe (ou l'étiquette, la cible ou encore la marque) du vecteur  $x_i$ .

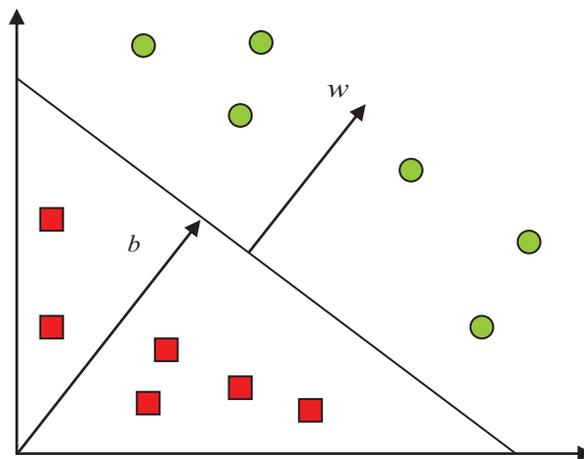


Figure I.6 L'hyperplan séparateur pour des données bidimensionnelles

L'objectif des SVMs est de trouver un hyperplan optimal  $f(x)$  permettant de séparer les données d'apprentissage de sorte que tous les points d'une même classe soient du même côté de l'hyperplan. Donc, cet hyperplan d'un point de vue géométrique divise l'espace en deux comme illustré dans la Figure I.6. Deux cas se présentent : cas des données linéairement séparables, et cas des données non linéairement séparables, que nous détaillons dans la section suivante.

### I.3.2.1 Cas de données linéairement séparables

Supposons la situation la plus simple lorsque les données sont linéairement séparables comme indiqué par la Figure I.7.

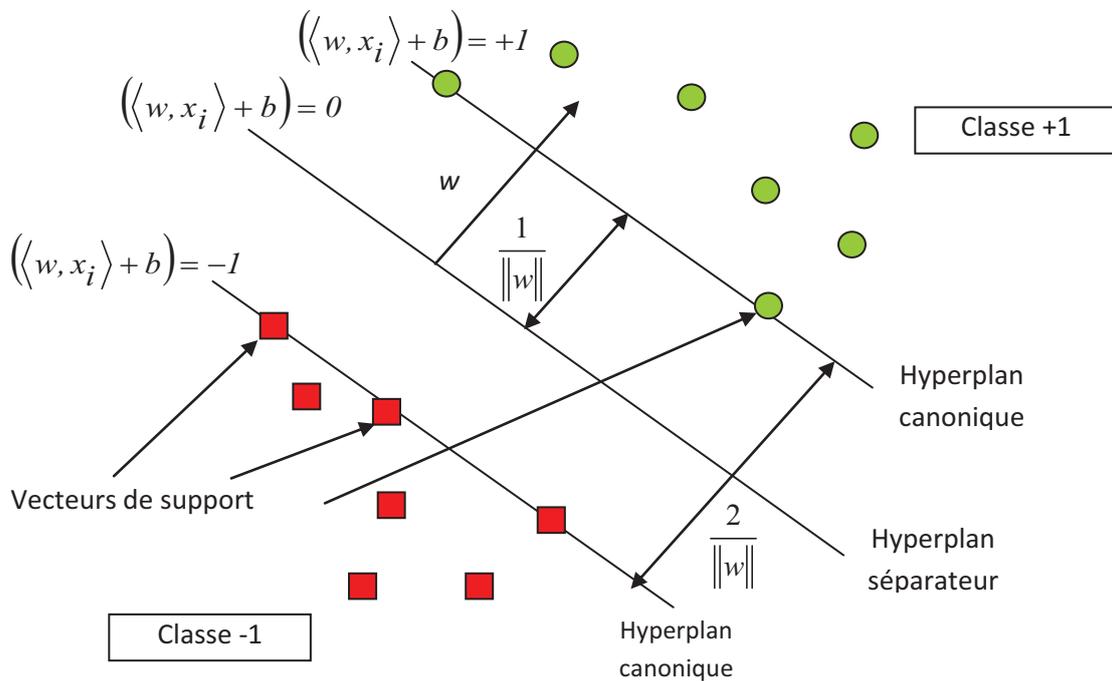


Figure I.7 Séparation linéaire

Il est clair qu'il existe une infinité d'hyperplans séparateurs permettant de séparer l'ensemble  $\mathbf{D}$  en deux classes sans erreurs, mais la question qui se pose est comment trouver le meilleur séparateur. Dans le cas de séparation linéaire, l'hyperplan séparateur est défini par une fonction discriminante linéaire des exemples  $\mathbf{x}$ , qui s'écrit sous la forme suivante :

$$f(x) = \langle w, x \rangle + b \quad (\text{I.60})$$

$w$  et  $b$  sont les paramètres de l'hyperplan séparateur, où  $w$  est le vecteur poids qui définit la direction perpendiculaire à l'hyperplan séparateur et  $b$  est le terme de biais (un scalaire). Ces paramètres sont calculés à partir des données formant l'ensemble  $\mathbf{D}$ .

Cet hyperplan séparateur permet de mettre les points d'une même classe dans un même côté :

$$z_i(\langle w, x_i \rangle + b) \geq 1 \quad \text{pour } i=1, 2, \dots, N. \quad w \in \mathfrak{R}^p \text{ et } b \in \mathfrak{R} \quad (\text{I.61})$$

Vapnik a montré qu'il existe une capacité optimale du classifieur qui est obtenue lorsque la distance entre l'hyperplan et les points les plus proches de chaque classe sera maximisée. Cette distance définie comme  $\frac{1}{\|w\|}$  est appelée **Marge**. L'hyperplan optimal sépare donc bien les deux classes de l'ensemble d'apprentissage, mais tout en minimisant la norme  $\|w\|$ . Par le choix de l'hyperplan qui maximise la marge, les SVM minimisent le risque de surapprentissage. La minimisation de ce critère permet ainsi l'estimation des paramètres de l'hyperplan  $f_w(x)$ .

Note : L'algorithme SVM est aussi appelé classifieur à Marge optimale (maximum) (Boser et al., 1992).

Le problème de minimisation est formulé comme un problème quadratique convexe sous **sa forme primale** par la minimisation de  $\|w\|^2$  tout en respectant des contraintes linéaires :

$$\begin{cases} \min \frac{1}{2} \|w\|^2 \\ \text{sous les contraintes} \\ z_i f(x_i) \geq 1 \quad i = 1, \dots, N \end{cases} \quad (\text{I.62})$$

Où  $f(x_i) = \langle w, x_i \rangle + b$  est la sortie prédite. Ce problème peut être résolu en utilisant les multiplicateurs de Lagrange  $\alpha_i$  et le dual lagrangien de ce problème devient :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i [\langle w, x_i \rangle + b] z_i - 1 \quad (\text{I.63})$$

Le lagrangien doit être minimisé par rapport à  $w$  et  $b$  et maximisé par rapport à  $\alpha$  :

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i x_i z_i \quad (\text{I.64})$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i z_i = 0 \quad (\text{I.65})$$

On obtient ainsi le problème dual suivant :

$$\left\{ \begin{array}{l} \min_{\alpha} \left( \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j z_i z_j \langle x_i, x_j \rangle - \sum_{i=1}^N \alpha_i \right) \\ \text{sous les contraintes} \\ \sum_{i=1}^N \alpha_i z_i = 0 \text{ et } \alpha_i \geq 0 \quad i=1, \dots, N \end{array} \right. \quad (\text{I.66})$$

Cette formulation du problème peut recevoir une solution efficace qui laisse peu de paramètres libres à choisir. La fonction de décision est alors :

$$\begin{aligned} f(x) &= \text{sgn} (\langle w, x \rangle + b) \\ \text{avec } w &= \sum_{i=1}^N \alpha_i z_i x_i \end{aligned} \quad (\text{I.67})$$

La classe d'un nouveau point  $x$  est ensuite donnée par le signe de  $f(x)$ . La fonction de décision est donc influencée uniquement par un petit sous ensemble de  $\mathbf{D}$  correspondant aux points (exemples) ayant  $\alpha_i$  non nuls. Ces points sont appelés les « **vecteurs de support ou Support Vectors (SVs)** ». La solution  $f(x)$  dans ce cas peut se traduire comme une combinaison linéaire des motifs supports.

Note : *Les Supports Vectors (SVs) sont les points qui se trouvent exactement à une distance égale à la marge dont les  $\alpha_i$  sont non nuls.*

Il s'agit là d'une propriété très intéressante des SVMs : seuls les SVs sont nécessaires pour décrire la décision. Le nombre des SVs pour le modèle optimal est généralement petit devant le nombre des données d'apprentissage  $N$ . Les poids  $w$  définissant la fonction de décision discriminante sont calculés à partir de tous les vecteurs de support.

La résolution Lagrangienne qui permet de trouver les différents paramètres, est réalisée de manière à satisfaire les conditions de Karush-Kuhn-Tucker (**KKT**) qui décrivent qu'au point de la solution, le produit entre les variables duales et les contraintes est nul.

**Conditions de KKT** : jouent un rôle essentiel dans la théorie et la pratique d'optimisation sous contraintes. Si on considère le problème (I.63), et en posant :

$$R_i = z_i (w \cdot x_i + b) - 1$$

si les conditions de KKT sont vérifiées alors :

$$1. \alpha_i = 0 \rightarrow R_i \geq 0$$

$$2. 0 < \alpha_i < C \rightarrow R_i = 0$$

$$3. \alpha_i = C \rightarrow R_i \leq 0$$

Les conditions de KKT sont violées dans les deux cas suivants :

$$1. \alpha_i < C \text{ et } R_i < 0$$

$$2. \alpha_i > 0 \text{ et } R_i > 0$$

Où  $C$  est un paramètre de réglage.

### I.3.2.2. Cas de données non linéairement séparables

Dans la plupart des problèmes réels, les données ne sont pas linéairement séparables à cause de la nature des systèmes généralement non linéaires et la présence de bruit dans ces données mesurées, comme indiqué par la Figure I.8 . Pour surmonter ces difficultés, les SVMs ont été généralisés grâce à deux outils : la marge souple et les fonctions noyaux.

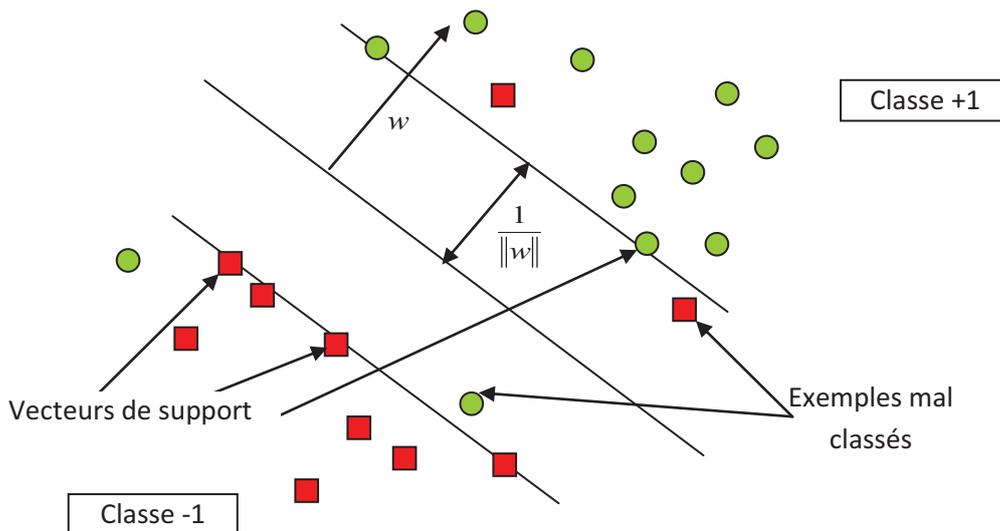
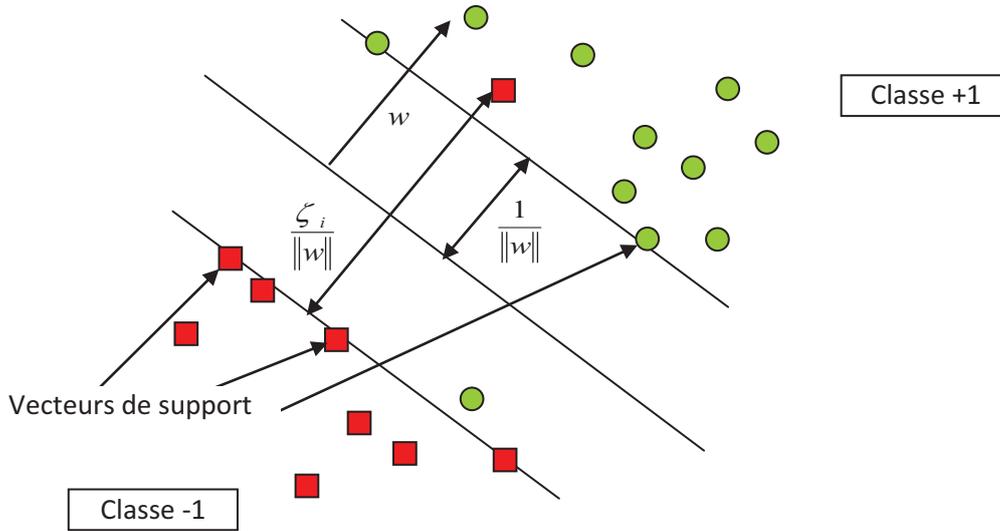


Figure I.8 Séparation linéaire avec des erreurs

#### 1. Marge souple (ou marge douce, *Soft Margin*)

Pour permettre de résoudre certaines limitations pratiques, il y a eu l'idée d'ajouter des variables ressorts dans la maximisation de la marge (Cortes and Vapnik, 1995). Ce principe de la marge souple consiste à autoriser des erreurs de classification, ce qui donne aux SVMs la capacité de manipuler des valeurs aberrantes. Le nouveau problème de séparation optimale est reformulé comme suit : l'hyperplan optimal séparant les deux classes est celui qui sépare les données avec le minimum d'erreurs, et qui satisfait les deux conditions suivantes :

- La distance entre les vecteurs bien classés et l'hyperplan doit être maximale.
- La distance entre les vecteurs mal classés et l'hyperplan doit être minimale.



**Figure I.9 Marge souple**

Des variables de pénalité non-négatives sont introduites,  $\zeta_i$  pour  $i = 1 \dots N$ , appelées variables d'écart (voir Figure I.9) afin de détendre un peu les contraintes sur la marge :

$$z_i (\langle w, x_i \rangle + b) \geq 1 - \zeta_i \quad \text{pour } i = 1, 2, \dots, N \quad (\text{I.68})$$

Avec l'introduction d'un terme de pénalité, on obtient le problème dual suivant :

$$\begin{cases} \min \left( \frac{1}{2} \|w^2\| + C \sum_{i=1}^N \zeta_i \right), \quad C \geq 0 \\ \text{sous les contraintes} \\ z_i [\langle w, x_i \rangle + b] \geq 1 - \zeta_i \quad \text{et } \zeta_i \geq 0 \quad \text{pour } i = 1, \dots, N \end{cases} \quad (\text{I.69})$$

Le paramètre  $C$  ( $C \geq 0$ ) est interprété comme un paramètre de régularisation qui régit la tolérance de mauvaise classification. L'augmentation de la valeur de  $C$  augmente le coût porté sur les points ou les vecteurs mal classés (on autorise plus de points mal classés sur les données d'apprentissage) et en même temps on réduit l'importance donnée à la minimisation de la complexité du modèle (minimiser  $\|w^2\|$ ). Donc par conséquent ce paramètre doit être choisi d'une manière judicieuse afin de garder toujours l'aspect de généralisation des SVMs tout en rejetant les données mal classées.

En introduisant les multiplicateurs de Lagrange ( $\alpha$  et  $\mu$ ), le Lagrangien associé au nouveau problème d'optimisation devient :

$$L(w, b, \zeta, \alpha, \mu) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \zeta_i - \sum_{i=1}^N \alpha_i (z_i \langle w, x_i \rangle + b) + \sum_{i=1}^N \zeta_i - 1 - \sum_{i=1}^N \mu_i \zeta_i \quad (\text{I.70})$$

Le Lagrangien doit être minimisé par rapport à  $w$ ,  $b$ ,  $\zeta$  et maximisé par rapport à  $\alpha$  et  $\mu$ .

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i z_i x_i \quad (\text{I.71})$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i z_i = 0 \quad (\text{I.72})$$

$$\frac{\partial L}{\partial \zeta_i} = 0 \rightarrow \forall i \quad C - \alpha_i - \mu_i = 0 \quad (\text{I.73})$$

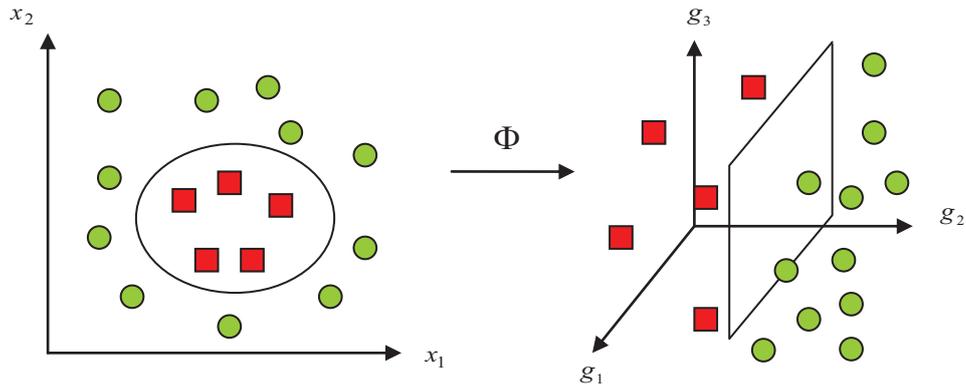
Ceci conduit à un problème dual légèrement différent de celui du cas séparable :

$$\left\{ \begin{array}{l} \max_{\alpha} \left( \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j z_i z_j \langle x_i, x_j \rangle \right) \\ \text{sous les contraintes} \\ 0 \leq \alpha_i \leq C \quad i=1, \dots, N \\ \text{et } \sum_{i=1}^N \alpha_i z_i = 0 \end{array} \right. \quad (\text{I.74})$$

La seule différence avec le cas linéairement séparable est l'introduction d'une borne supérieure pour les paramètres duaux  $\alpha_i$  ( $\alpha_i < C$ ). Ces multiplicateurs de Lagrange représentent une sorte de forces virtuelles découlant des contraintes posées sur le problème à résoudre.

## 2. Fonctions noyaux

Dans le cas de données non linéairement séparables, plus la dimension de l'espace de description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples et les contre-exemples est élevée. En transformant l'espace d'entrée en un espace de redescription (appelé aussi espace de caractéristique) d'une plus grande dimension, éventuellement infinie, il devient possible d'utiliser la méthode des SVMs (Figure I.10).



**Figure I.10 Exemple de projection**

Notons  $\Phi$  une transformation non linéaire ou mappage « *mapping* » de l'espace d'entrée  $\mathbf{x}$  en un espace de redescription  $\Phi(\mathbf{x})$ . Voici un exemple de projection :

$$\mathbf{x} = (x_1, \dots, x_p)^T \rightarrow \Phi(\mathbf{x}) = \langle \Phi_1(\mathbf{x}), \dots, \Phi_n(\mathbf{x}) \rangle \quad (\text{I.75})$$

Où  $\langle, \rangle$  dénote le produit scalaire dans le nouvel espace. Le problème d'optimisation se transcrit dans ce cas par :

$$\left\{ \begin{array}{l} \max_{\alpha} \left( \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j z_i z_j \langle \Phi(x_i), \Phi(x_j) \rangle \right) \\ \text{sous les contraintes} \\ 0 \leq \alpha_i \leq C \\ \text{et } \sum_{i=1}^N \alpha_i z_i = 0 \text{ avec } i=1, \dots, N \end{array} \right. \quad (\text{I.76})$$

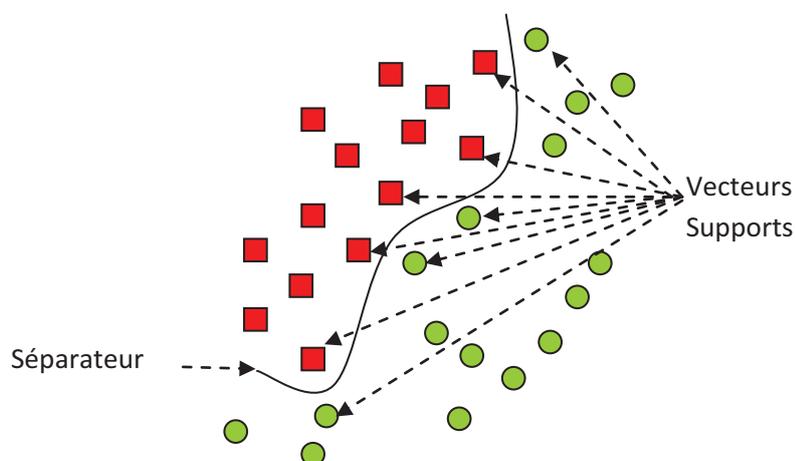
Et la fonction de décision est donnée par :

$$f(\mathbf{x}) = \sum_{i=1}^N z_i \alpha_i \langle \Phi(x_i), \Phi(\mathbf{x}) \rangle + b \quad (\text{I.77})$$

$\langle \Phi(x_i), \Phi(x_j) \rangle$  devient rapidement impossible à calculer quand la dimension de  $\Phi(x_i)$  augmente (sans parler du cas de la dimension infinie). Ceci est d'autant plus difficile si on utilise des transformations non linéaires des descripteurs d'entrée. On peut dans certains cas s'arranger pour court-circuiter le passage par les calculs dans l'espace de redescription. Il existe des fonctions bilinéaires symétriques positives  $K(x_i, x_j)$ , appelées fonctions noyaux (*Kernel functions*), faciles à calculer et dont on peut montrer qu'elles correspondent à un produit scalaire  $\langle \Phi(x_i), \Phi(x_j) \rangle$  dans un espace de grande dimension :

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \quad (I.78)$$

La Figure I.11 ci-dessous, illustre le calcul d'une surface de décision par la méthode à noyau dans le cas des données non linéairement séparables.



**Figure I.11 Vecteurs support et surface de décision calculée par la méthode à noyau**

On peut utiliser n'importe quelle fonction noyau  $K(x_i, x_j)$  valide satisfaisant la condition de Mercer sans avoir besoin de connaître des informations sur la transformation non linéaire qu'il y a derrière. La condition de Mercer nous indique si une fonction est un noyau mais ne fournit aucune information sur la fonction  $\Phi$  (et donc sur l'espace des caractéristiques) induite par ce noyau.

**Condition de Mercer :** Sachant que la matrice contenant les similarités entre tous les exemples de l'entraînement est appelée la matrice de Gram ( $\mathbf{G}$ ), la condition ou le théorème de Mercer s'énonce comme suit :

La fonction  $K(x_i, x_j) : X \times X \rightarrow \mathfrak{R}$  est un *noyau* si et seulement si la matrice de Gram  $G = (K(x_i, x_j))$ , avec  $i, j = 1, \dots, N$  ( $N$  est le nombre d'exemples d'entraînement) est une matrice définie positive.

Notons qu'une fonction  $\mathbf{K} : X \times X \rightarrow \mathfrak{R}$  générant une matrice définie positive  $G$  possède les trois propriétés fondamentales du produit scalaire suivantes :  $\forall (x_i, x_j) \in X$

1. Positivité :  $K(x_i, x_i) \geq 0$
2. Symétrie :  $K(x_i, x_j) = K(x_j, x_i)$
3. Inégalité de Cauchy-Schwartz :  $|K(x_i, x_j)| \leq \|x_i\| \cdot \|x_j\|$

En pratique, la fonction noyau va déterminer la structure du modèle et donne un aspect de flexibilité à la méthode SVM.

Le problème de classification supervisée donnée par le système (I.76) peut s'écrire d'une autre façon :

$$\begin{cases} \max_{\alpha \in \mathbb{R}^N} \left( \alpha^T \mathbf{I} - \frac{1}{2} \alpha^T \cdot G \cdot \alpha \right) \\ \text{sous les contraintes} \\ \alpha^T z = 0 \text{ et } 0 \leq \alpha_i \leq C \quad i = 1, \dots, N \end{cases} \quad (\text{I.79})$$

Avec  $\alpha^T = [\alpha_1, \dots, \alpha_N]$ ,  $z^T = [z_1, \dots, z_N]$ ,  $\mathbf{I}$  est un vecteur unitaire de taille  $\mathbf{N}$ , et  $G = z_i z_j K(x_i, x_j)$  est la matrice d'influence par paires du terme général. Cette matrice est carrée symétrique définie positive de taille  $N \times N$ . Pour décider à quelle classe appartient une donnée  $\mathbf{x}$ , il suffit de prendre le signe de la fonction de décision donné par (I.77) :

$$f(x) = \text{sgn} \left( \sum_{i=1}^N z_i \alpha_i K(x_i, x) + b \right) \quad (\text{I.80})$$

Si le signe de  $\mathbf{f}(\mathbf{x})$  est positif alors la donnée  $\mathbf{x}$  appartient à la classe  $+1$ , sinon elle appartient à la classe  $-1$ . Cette fonction est une somme pondérée de fonction noyaux. Elle dépend juste des  $N_{\text{sup}}$ , car  $\alpha_i = 0$  ailleurs. Donc, on peut remplacer  $\mathbf{N}$  par  $\mathbf{N}_{\text{sup}}$  :

$$f(x) = \text{sgn} \left( \sum_{i=1}^{N_{\text{sup}}} z_i \alpha_i K(x_i, x) + b \right) \quad (\text{I.81})$$

La valeur du biais  $\mathbf{b}$  peut être calculée par l'utilisation de deux supports opposés  $x_A$  et  $x_B$  tels que  $f(x_A) = 1$  et  $f(x_B) = -1$  :

$$b = -\frac{1}{2} \sum_{i=1}^N z_i \alpha_i [K(x_A, x_i) - K(x_B, x_i)] \quad (\text{I.82})$$

### 3. Exemples de noyaux

#### 1. Noyau linéaire

$$K(x_i, x_j) = \langle x_i, x_j \rangle \quad \text{avec } \Phi(x_i) = x_i \quad (\text{la fonction identité})$$

## 2. Noyau polynomial

$$K(x_i, x_j) = \langle x_i, x_j \rangle^d$$

Où  $d$  est le degré du polynôme. Ce noyau réalise implicitement un produit scalaire dans l'espace des descripteurs correspondant à tous les produits à  $d$  dimensions.

## 3. Noyau gaussien (ou noyau RBF)

Cette fonction noyau est typique et très largement utilisée puisqu'elle permet avec un réglage approprié de représenter n'importe quelle fonction. Le noyau gaussien représente une mesure de similarité entre les échantillons  $(x_i, x_j)$  et s'écrit sous la forme suivante

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

Où  $\sigma$  est la variance, qui devient un paramètre de régulation.

## 4. Noyau sigmoïde

$$K(x_i, x_j) = \tanh(\langle x_i, x_j \rangle + 1)$$

D'autres variantes de SVM existent. On cite par exemple SVM à une classe,  $\nu$ -SVM (voir l'annexe A.6 pour plus de détails)

### ***1.3.2.3. Extension des SVM binaires au cas multi classe***

L'expression (I.81) de SVM peut résoudre des problèmes de classification binaire. Or, les problèmes les plus rencontrés en pratique sont ceux où l'on a plusieurs classes. L'extension de SVM aux problèmes de plus de deux classes est donc nécessaire. Un travail de synthèse intéressant concernant la théorie et les applications des SVM multi classe est donné dans (Guermeur, 2007).

Considérons un ensemble  $D$  de  $N$  données d'apprentissage séparables en  $k$  classes  $(\Omega_1, \Omega_2, \dots, \Omega_k)$ , tel que  $D = \{(x_1, z_1), (x_2, z_2), \dots, (x_N, z_N)\}$ ,  $x_i \in \mathcal{R}^p$ ,  $z_i \in Z = \{1, \dots, k\}$

Dans ce cas, il ne s'agit plus de trouver une seule séparatrice entre deux classes mais il faudra classer les exemples en plusieurs classes, ce qui revient à construire  $k$  hyperplans linéaires d'équations données comme suit :

$$w_k^T x + b_k = 0 \text{ avec } k \in z \quad (\text{I.83})$$

Généraliser les SVMs dans le cas multi classes revient à résoudre le problème de programmation quadratique suivant :

$$\left\{ \begin{array}{l} \min_{w_k, b_k, \zeta} \left( \frac{1}{2} \sum_{k \in z} \|w_k\|^2 + \sum_{i=1}^N \zeta_i^k \right) \\ \text{sous les contraintes} \\ w_{z_i}^T x_i + b_{z_i} - (w_k^T x_i + b_k) \geq 1 - \zeta_i^k \text{ avec } \zeta_i^k \geq 0, i = 1, \dots, N \text{ et } k \in z \end{array} \right. \quad (\text{I.84})$$

La fonction de décision est alors :

$$f(x) = \operatorname{argmax}_{k \in z} f_k(x) \text{ où } f_k(x) = w_k^T x + b_k \quad k \in z \quad (\text{I.85})$$

Donc, un nouvel individu aura comme classe, la classe  $z_i$  tel que :

$$f_{z_i}(x) \geq f_k(x) \quad k \in z \quad (\text{I.86})$$

Plusieurs méthodes ont été proposées pour résoudre ce problème, les plus utilisées sont les suivantes :

### 1. Un Contre tous (OAA pour *One Against All*)

Pour chaque classe, on détermine un hyperplan séparant celle-ci de toutes les autres, en considérant cette dernière comme la classe **+1** et les autres classes comme étant la classe **-1**, ce qui résulte en  $k$  SVMs binaires (pour un problème à  $k$  classes) (Liu and Zheng, 2005) .

### 2. Un contre un (OAO pour *One Against One*)

Dans ce cas, on construit un SVM pour chaque paire de classes. Ainsi, on sera amené à calculer  $\frac{k(k-1)}{2}$  SVMs binaires. Cette approche est associée en général à une opération de vote majoritaire pour la fusion des décisions des  $\frac{k(k-1)}{2}$  sous-classifieurs SVMs.

### 3. Arbre ou graphe de décision

Dans un arbre de décision, chaque nœud fait la décision si une donnée appartient à un certain ensemble de classes (Platt et al., 2000). La succession de décisions par d'autres nœuds fait réduire le nombre de classes jusqu'à ce qu'on arrive à une seule classe. Chaque nœud de l'arbre

contient un classifieur SVM destiné à distinguer entre deux classes seulement parmi les autres. A chaque niveau, on élimine une classe jusqu'à ce qu'une seule classe reste.

#### ***1.3.2.4. Les algorithmes d'optimisation pour SVM***

L'entraînement par SVM consiste à résoudre un problème d'optimisation quadratique convexe sous contraintes. En raison de son immense taille, le problème résultant de l'approche SVM ne peut être résolu facilement par l'intermédiaire des techniques standards de programmation quadratique. La forme quadratique du problème implique une matrice qui a un nombre d'éléments égale au carré du nombre d'exemples d'entraînement. A cause de sa taille, cette matrice ne peut pas être traitée correctement.

Les algorithmes d'optimisation sont nécessaires afin de réduire le temps de calcul lors de l'apprentissage, mais d'autres peuvent également servir pour l'optimisation paramétrique tel que « *Differential evolution, Artificial immune algorithm, Genetic algorithms* ». Parmi les méthodes les plus efficaces utilisées pour l'entraînement des SVMs, on trouve :

- Les multiplicateurs de Lagrange.
- La méthode d'optimisation séquentielle minimale « *Sequential Minimal Optimisation, SMO* », la plus recommandée.
- La méthode d'Optimisation par essaims particulaires (PSO) « *Particle Swarm Optimization based on chaos searching* ».
- La méthode *Chunking*.
- La méthode de décomposition.

De la même manière que la résolution lagrangienne, toutes ces méthodes doivent vérifier les conditions de **KKT**. La méthode SMO est considérée comme une des plus efficaces méthodes décrites ci-dessus, elle est la base de plusieurs algorithmes tels que la librairie LIBSVM (Chang and Lin, 2001) qui va être utilisée dans cette thèse pour la construction des modèles de prédiction par SVR dans le chapitre II.

- **Méthode d'optimisation séquentielle minimale (SMO)**

En 1999, John Platt (Platt, 1999) proposa un nouvel algorithme pour l'entraînement des SVMs qu'il appela algorithme d'Optimisation Séquentielle Minimale (SMO). Cet algorithme

décompose le problème de (I.62) mais choisit de résoudre le plus petit sous-problème possible à chaque étape d'optimisation de la fonction *objectif*. Cela est réalisable en choisissant d'optimiser, non pas un ensemble de multiplicateurs de Lagrange à la fois, mais deux seulement. De ce fait, l'algorithme SMO ne nécessite aucun stockage supplémentaire de matrice. Du côté des performances en temps de réponse, l'algorithme SMO est plus efficace. En effet, il nécessite plus d'itérations pour converger, mais chaque itération effectue moins d'opérations.

L'algorithme SMO est basé sur trois éléments :

1. Une méthode analytique pour résoudre les deux multiplicateurs de Lagrange.
2. Deux heuristiques pour choisir quels sont les multiplicateurs à optimiser.
3. Une méthode pour le calcul du paramètre **b** utilisé dans la fonction de décision.

### ***1.3.2.5. Avantages / applications des SVMs***

Les SVMs ont un nombre d'avantages sur les autres méthodes statistiques ou les réseaux de neurones :

- Capacité de généralisation induite par la théorie statistique d'apprentissage et le concept de minimisation du risque structurel.
- Pas de problème avec une grande dimension des données. Faire face à des situations complexes lorsque le nombre de variables est plus grand que celui d'observations.
- La solution est parcimonieuse, unique et l'optimum est global.
- Peu de paramètres de réglage.
- Implémentation facile, pas de problème d'instabilité numérique et la rapidité de calcul.
- Capacité d'obtenir un résultat précis même dans le cas de taille réduite de l'ensemble d'apprentissage.
- Manipulation des valeurs aberrantes, et du bruit.
- Élimination des données non significatives par la concentration de l'information utile sur les vecteurs supports.
- Robustesse face aux variations des paramètres du système.
- Construction de modèle non linéaire grâce aux fonctions noyaux.

Grâce à ces avantages, les SVMs ont montré leur efficacité dans plusieurs domaines et applications, surtout dans les problèmes de grande taille grâce à leur solide fondement théorique qui cherche à minimiser la borne supérieure de l'erreur de généralisation permettant de récupérer

juste l'information utile à partir des données d'apprentissage. Parmi ces applications, on peut citer :

- Reconnaissance de chiffres manuscrits (Vapnik, 1998, Cortes and Vapnik, 1995, Boser et al., 1992)
- Reconnaissance et classification d'image
- Catégorisation du texte (Dumais et al., 1998, Basu et al., 2003)
- Classification de la couverture des terres (Huang et al., 2002)
- Classification des données audio (Aggarwal and Aggarwal, 2011)
- Applications en chimie (Li et al., 2009, Ivanciuc, 2007),
- Estimation de densité (Mukherjee and Vapnik, 1999)
- Prédiction: séries temporelles financières (Gestel et al., 2001), capacité thermique (Xue et al., 2004)
- Surveillance des variations des procédés (Cheng and Cheng, 2008, Camci et al., 2008)
- Classification des défauts : ex. analyse du transformateur de gaz dissous (Tang and Wu, 2011), où l'algorithme génétique extrait les caractéristiques des défauts et la méthode Bootstrap aide à améliorer le prétraitement par l'égalisation des nombres d'échantillons pour différents types de défauts. Ensuite, un classifieur comme le SVM est utilisé pour le diagnostic de défauts.

(Widodo and Yang, 2007) présente une revue de la littérature utilisant les SVM dans la surveillance et le diagnostic des machines roulements, moteur à induction, HVAC, machine, boîte à vitesse, pompe, vanne, compresseurs. Dans la littérature, plusieurs travaux se sont intéressés à l'utilisation de SVM pour résoudre le problème de diagnostic de défauts :

- Boîte à vitesse (Saravanan et al., 2010, Saravanan and Ramachandran, 2009, Saravanan et al., 2008, Davy et al., 2006, Samanta, 2004, Samanta et al., 2003)
- Roulement (Sugumaran et al., 2008, Abbasian et al., 2007)
- Turbo compresseur et pompe à rotor (Yuan and Chu, 2007)
- Les machines (Yang and Widodo, 2008)
- Moteur à induction (Baccarini et al., 2011, Pöyhönen et al., 2005)
- Turbine à vapeur (Salahshoor et al., 2010)
- Transmission de puissance (Samantaray et al., 2007, Ravikumar et al., 2008, Ganyun et al., 2005)
- Procédé de traitement des eaux usées (Laouti et al., 2011b, Zeng et al., 2006)
- Procédé Tennessee Eastman (Monroy et al., 2010, Zhang, 2009, Yélamos et al., 2007)

- Traitement de minéraux (Jemwa and Aldrich, 2006)
- Moteur à induction (Baccarini et al., 2011)
- Turbine à vapeur (Salahshoor et al., 2010)
- Ligne de transmission de puissance (Samantaray et al., 2007)
- Les systèmes de chauffage, ventilation et de conditionnement (Liang and Du, 2007)
- Système réfrigérateur (Ren et al., 2008)
- Les accidents de la circulation (Chen et al., 2009)

Il existe plusieurs programmes et progiciels qui implémentent les SVMs avec plusieurs langages de programmation comprenant des routines d'optimisation numériques très efficaces. Parmi ces logiciels on trouve :

- **SVM<sup>light</sup>** : l'implémentation des SVMs la plus connue réalisée par Thorsten Joachims (Joachims, 1999) (<http://svmlight.joachims.org/>).
- **La boîte à outils SV** : Écrite complètement dans Matlab® par Stéphane Canu et al. (2005), elle est une implémentation facile à manipuler pour les gens familiarisés avec Matlab. [<http://asi.insa-rouen.fr/enseignants/~arakotom/toolbox/index.html>].
- **LIBSVM** : représente une bibliothèque complète des SVMs présentée par Chang et Lin (Chang and Lin, 2011 (Chang and Lin, 2001) écrite dans le langage C/C++ et Java utilisant la même méthode de décomposition du problème quadratique que celle de l'algorithme SMO. Ce package effectue à la fois les tâches de classification et de régression, et propose plusieurs interfaces pour les autres langages tel que : Python®, Matlab®. Cette bibliothèque peut être téléchargée sur :

<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

Pour plus de détails sur d'autres packages, voir le lien ci-dessous :

[http://www.support-vector-machines.org/SVM\\_soft.html](http://www.support-vector-machines.org/SVM_soft.html).

### **I.3.3. Les machines à Vecteurs de supports pour la régression (SVR)**

L'obtention des modèles de fonctionnement des procédés complexes s'avère une tâche difficile qui nécessite des compétences et du temps. L'utilisation des Séparateurs à vastes marges étendue au cas de la régression constitue une alternative prometteuse pour la modélisation de ces

systèmes. L'information de la modélisation sera utilisée dans cette thèse afin d'enrichir les données dans l'objectif de détection de défauts.

### ***1.3.3.1. Théorie***

Soit l'ensemble  $\mathbf{D}_2$  de  $N$  paires données, ayant un vecteur de données  $\mathbf{x}_i$  comme entrée et l'étiquette  $\mathbf{z}_i$  de ce vecteur comme sortie, qui peut maintenant prendre n'importe quelle valeur réelle.

$$D_2 = \{(x_1, z_1), \dots, (x_i, z_i), \dots, (x_N, z_N)\}, \text{ pour } i = 1, \dots, N \text{ avec: } x_i \in R^p, z_i \in R$$

Comme précédemment, habituellement dans une régression le but est trouver une fonction pouvant mettre en relation les entrées  $\mathbf{x}_i$  et les sorties  $\mathbf{z}_i$  :

$$z_i = f(x_i)$$

Les étapes nécessaires pour réaliser une régression classique sont :

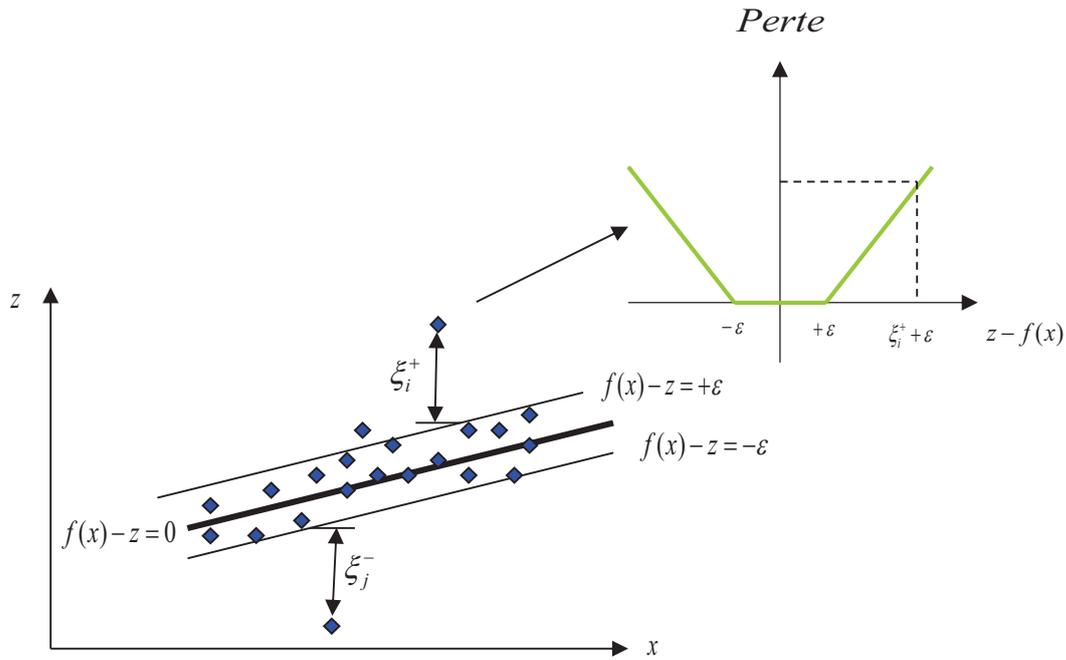
- Choisir un modèle qui représente au mieux ces données, par exemple une droite:

$$f(x_i) = a.x_i + b$$

- Optimiser un critère pour trouver les paramètres  $\mathbf{a}$  et  $\mathbf{b}$  de ce modèle, par exemple

moindres carrés :  $\min \sum_{i=1}^N (z_i - f(x_i))^2$ , où  $z_i$  est la sortie réelle.

Avec une fonction de coût de type quadratique (moyenne des erreurs quadratique **MSE**), on obtient la solution de *maximum de vraisemblance* quand le bruit est gaussien ou normalement distribué. Mais, la présence d'une valeur aberrante risque de biaiser le modèle obtenu représenté par la fonction  $f$ . L'idée qui permet de généraliser l'utilisation des SVMs de la classification à la régression est l'adoption d'une nouvelle fonction de perte (coût) «perte de Vapnik » qui inclut la notion de distance. Cette fonction de perte ignore les erreurs qui se trouvent à une certaine distance  $\epsilon$  de la vraie valeur. Ce type de fonction est considéré comme une fonction de perte  *$\epsilon$ -insensible*. Elle a l'avantage de diminuer ainsi le surapprentissage.



**Figure I.12 La fonction de perte insensible à  $\varepsilon$  dans la régression SVR linéaire**

Donc l'objectif de SVR est de trouver une fonction plate  $z = f(x) = \langle w, x \rangle + b$  qui rapproche toutes les paires  $(x_i, z_i)$  avec une précision de  $\varepsilon$  (voir Figure I.12), ou en d'autres termes une fonction qui estime les cibles bruités  $z_i$  avec au maximum un écart  $\varepsilon$ . Une façon pour assurer la platitude de la fonction  $f(x)$  est de minimiser  $\|w\|$ .

Afin d'autoriser certaines erreurs (plus grande de  $\varepsilon$ ), des variables d'écart  $(\zeta^+, \zeta^-)$  peuvent être introduites pour adoucir la marge avec le paramètre de régulation  $C$  qui fait le compromis entre les erreurs et le terme de régularisation  $\frac{1}{2}\|w\|^2$ .

$$\begin{aligned}
 & \text{Minimiser } \frac{1}{2}\|w\|^2 + C \sum_{i=1}^l (\zeta_i^+, \zeta_i^-) \\
 & \text{sujet à } \begin{cases} z_i - \langle w, \phi(x_i) \rangle - b \leq \varepsilon + \zeta_i^+ \\ \langle w, \phi(x_i) \rangle + b - z_i \leq \varepsilon + \zeta_i^- \\ \zeta_i^+, \zeta_i^- \geq 0 \end{cases} \quad (\text{I.87})
 \end{aligned}$$

$C$  est un paramètre d'équilibrage qui permet de contrôler implicitement le niveau de « bruit » du problème et donc la régularité de la solution. L'introduction d'un tube de largeur  $\varepsilon$  mène à la fonction de perte appelée Vapnik insensible à  $\varepsilon$  donné par :

$$|\zeta|_\varepsilon = \begin{cases} 0 & \text{si } |\zeta| \leq \varepsilon \\ |\zeta| - \varepsilon & \text{sinon} \end{cases} \quad (\text{I.88})$$

En introduisant les multiplicateurs de Lagrange ( $\alpha_i, \alpha_i^*, \mu_i, \mu_i^* \geq 0$ ), le Lagrangien associé au nouveau problème d'optimisation devient :

$$\begin{aligned} L = & \frac{1}{2} \|w\|^2 + \sum_{i=1}^N C(\zeta_i + \zeta_i^*) - \sum_{i=1}^N (\mu_i \zeta_i + \mu_i^* \zeta_i^*) \\ & - \sum_{i=1}^N \alpha_i (\varepsilon + \zeta_i - z_i + \langle w, x_i \rangle + b) \\ & - \sum_{i=1}^N \alpha_i^* (\varepsilon + \zeta_i^* - z_i - \langle w, x_i \rangle - b) \end{aligned} \quad (\text{I.89})$$

Comme pour le cas de classification, le problème peut se résoudre par l'utilisation du dual lagrangien qui conduit à :

$$f(x) = \sum_{i=1}^{N_{\text{sup}}} (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (\text{I.90})$$

$$\text{Avec, } w = \sum_{i=1}^{N_{\text{sup}}} (\alpha_i - \alpha_i^*) \phi(x_i) \quad (\text{I.91})$$

Où  $\alpha_i, \alpha_i^*$  sont les multiplicateurs de Lagrange. Seules les valeurs non nulles des multiplicateurs de Lagrange sont utiles pour la détermination de la courbe de régression. Les points de données associés à ces coefficients (qui se trouvent sur ou à l'extérieur du tube  $\pm\varepsilon$ ) sont appelés les vecteurs de support.

Le fait d'ajouter des variables d'écart et de manipuler le paramètre de régularisation  $C$  donne aux SVR un aspect flexible et rigoureux à la fois. Pour voir l'influence des différents paramètres d'une manière visuelle, voir le tutorial sur les SVM et SVR donné dans (Brereton and Lloyd, 2010).

La méthode classique de moindres carrés bien qu'elle soit considérée comme le meilleur algorithme d'estimation polynomiale, reste inférieure en termes de performances ( $R^2$ , erreur quadratique) par rapport à l'approche SVR, Ceci a été démontré dans (Wang et al., 2007), ou dans (Maalouf et al., 2008) pour la prédiction du module d'élasticité des matériaux de chaussée dans le domaine de génie civil.

La comparaison de SVR avec d'autres méthodes en termes de précision et robustesse a balancé en faveur du SVR dans la quasi-totalité des cas :

- SVR par rapport à la méthode processus gaussien (Deswal, 2011)
- SVR par rapport à filtre de Kalman, moindres carrés et H infinie (Zhang et al., 2007).
- SVR par rapport à la méthode de calcul de la dynamique de fluide (Varol et al., 2008)
- SVR par rapport à la méthode de moyenne mobile et la méthode des modèles autorégressifs (Bermolen and Rossi, 2009)
- SVR par rapport à la méthode régression linéaire multiple, et aussi face au réseau neuronal à base d'une fonction radiale (Xue et al., 2004)
- LSVR par rapport à régression linéaire et régression non linéaire (Baylar et al., 2009)

### ***I.3.3.2. Implémentation de SVR***

La régression par machines à vecteurs de supports est utilisée pour l'estimation (Lahiri and Ghanta, 2008, Varol et al., 2008, Smola et al., 1996), la prédiction des séries temporelles (Crone and Kourentzes, 2010, He et al., 2008, Gestel et al., 2001), l'identification des systèmes (Yu et al., 2005) dans différents types d'applications, comme par exemple :

- Prédiction de charge de liaison d'un réseau (Priyadarshin et al., 2011, Bermolen and Rossi, 2009).
- Modélisation de type boîte noire du comportement d'un bras de robot hydraulique (Gretton et al., 2001).
- Prédiction du coefficient global de transfert d'oxygène (Deswal, 2011).
- Prédiction de la force de compression du béton à haute résistance en génie civil (Gupta, 2007).
- Estimation du décalage vers le rouge photométrique (Wang et al., 2007).
- Estimation des champs thermiques et d'écoulement dus à la convection naturelle (Varol et al., 2008)
- Estimation de la vitesse critique de l'écoulement de suspension solide-liquide (Lahiri and Ghanta, 2008)
- Estimation en ligne de la concentration de la biomasse dans un procédé de fermentation (Wang, Yu et al., 2006)
- Prédiction de la capacité thermique (Xue et al., 2004) .
- Prédiction des propriétés mécaniques des matériaux en acier (Wang, Mu et al., 2006).
- Prédiction de l'hybridation croisée avec l'utilisation de la teneur en nucléotide des sondes ADN comme variables d'entrée (Sun et al., 2006).

- Capteur logiciel d'indices de fusion « *melt indices* » pour deux applications : processus de neutralisation de pH en simulation et un procédé d'extrusion de polymère à échelle de laboratoire (Chitralkha and Shah, 2010).
- Identification des défauts (Cheng et al., 2011) et a montré sa supériorité par rapport aux réseaux de neurones et à la méthode de carte de contrôle des sommes cumulées CUSUM (*Cumulative Sum Control chart*)

Parmi les algorithmes d'entraînement les plus populaires et efficaces des SVR, on trouve l'algorithme classique SMO. Néanmoins, ce dernier présente une source d'inefficacité due au fonctionnement avec une valeur de seuil unique en plus de sa vitesse de convergence lente. Ainsi, plusieurs travaux proposent une amélioration de l'algorithme SMO (Shevade et al., 2000), (Flake and Lawrence, 2002).

(Wang et al., 2009) ont proposé un algorithme SVR en ligne écrit sur Matlab pour modéliser la fonctionnalité dynamique du processus de fermentation. Cet algorithme présente un inconvénient puisqu'il est inadapté aux grandes dimensions. (Ma et al., 2003) ont développé et implémenté un algorithme SVR en ligne capable d'être appliqué pour la prédiction de séries temporelles et la validation croisée sans le besoin de refaire l'apprentissage quand un nouvel échantillon est ajouté ou supprimé. Un autre algorithme SVR en ligne a été développé pour prédire les tensions nécessaires à appliquer pour compenser la charge de gravité du bras d'un robot (Parrella, 2007).

Dans l'identification des systèmes non linéaires, le système peut être représenté comme une série de blocs liés entre eux. On cite par exemple :

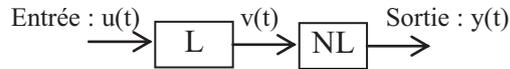
- **Le modèle Hammerstein** : qui est composé d'un bloc non linéaire statique suivi d'un bloc ou élément linéaire dynamique comme le montre la Figure I.13. L'obtention du modèle Hammerstein peut se faire par l'algorithme d'identification sous-espace N4SID qui utilise SVR à moindres carrés *LS-SVR* (Goethals et al., 2005), et dans (Mujahed and Westwick, 2011) un modèle SVR pour capter la non linéarité.



**Figure I.13 Structure de type Hammerstein**

- **Le modèle de type Wiener** : constitué d'un élément dynamique linéaire suivi d'un

élément non-linéaire statique comme le montre la Figure I.14. Dans (Tötterman and Toivonen, 2009), le composant linéaire du modèle Wiener est représenté par un filtre étendu « *basis filter expansion* » et le bloc non linéaire est déterminé par un modèle SVR à noyau.



**Figure I.14 Structure de type Wiener**

- **La variante LS-SVM** : a été également utilisée pour l'identification où un algorithme basé sur les SVM à moindres carrés pondérés identifie les paramètres structurels et leur variations afin de suivre les changements soudains ou lents et détecter la gravité du dommage s'il y a lieu (Tang et al., 2006).

#### **Modèles de prédiction (séries temporelles) :**

- **Filtre à réponse impulsionnelle finie (FRIF)** : ce filtre est stable et est caractérisé par une réponse uniquement basée sur les valeurs du signal d'entrée.
- **Processus autorégressif (AR)** : Modèle de régression pour séries temporelles dans lequel la série est expliquée par ses valeurs passées plutôt que par d'autres variables (Zhang et al., 2007).
- **Modèle ARMA** : est un outil pour comprendre et prédire les valeurs futures d'une série temporelle. Il est composé de deux parties : une part autorégressive (AR) et une part moyenne-mobile (MA). Le modèle est généralement noté ARMA(p,q), où **p** est l'ordre de la partie AR et **q** l'ordre de la partie MA (Zhang et al., 2007, Mohsen and El-Yazeed, 2004).
- **L'astuce noyau dans SVR** : transforme les données d'entrées non linéairement séparables en un espace où elles pourraient être séparables : espace de Hilbert à noyau reproduisant (EHNR), ce qui permet au SVR de capter les non linéarité du système dans un modèle de type ARMA par exemple (Martínez-Ramón et al., 2006) .
- **Modèle ARX** : Autorégressif avec entrées exogènes. Une méthode de réseaux de neurones ARX adaptatives utilisant une fenêtre mobile qui a été proposée pour prédire des variables. Il a montré une capacité pour capter les changements dynamiques lents du procédé de filtration anaérobie à pleine échelle de traitement des eaux usées (Lee et al., 2005). (Adachi et al., 2001) ont proposé un modèle de régression linéaire de type ARX et FIR qui a montré sa robustesse contre les valeurs aberrantes.

- **Modèle NARX**: non linéaire autorégressif avec entrées exogènes

### ***I.3.3.3. Réglage des paramètres SVM et SVR***

La qualité du modèle et la performance de la prédiction du SVR dépend de plusieurs paramètres :

- 1- **Paramètres liés au procédé** : tels que la dimension des variables du procédé, les échelles du temps, l'incertitude et la variabilité.
- 2- **Paramètres d'apprentissage** :  $C$ , la tolérance  $\epsilon$ , la variance du noyau gaussien  $\sigma$ , le nombre d'échantillons  $N$ .

L'analyse de sensibilité consiste à faire varier ces paramètres et à évaluer la réponse du modèle. Il y a deux façons d'aborder la question : par intervalles (valeur basse et valeur haute du paramètre en cause) et par la méthode de Monte Carlo (voir annexe A.1), en supposant que les valeurs prises par le paramètre en cause peuvent être représentées par une distribution de probabilité (normale, log-normale, uniforme...) :

- Construction d'une grille d'optimisation : on change les paramètres de  $C$ ,  $\sigma$  et  $\epsilon$  1000 fois afin de voir l'effet sur l'erreur.
- Mesurer également l'impact du changement de la taille de données  $N$  utilisées dans l'apprentissage, de  $\frac{1}{32}N$  jusqu'à  $\frac{1}{2}N$ .
- Calculer l'intervalle de confiance à 95%. Si de nombreuses prédictions sortent de l'intervalle de confiance, cela veut dire qu'il faut corriger le modèle (précision insuffisante).

$C$  et  $\sigma$  sont des paramètres continus et donc c'est difficile, voire impossible, de tester toutes leurs valeurs. Une solution pratique sur le choix de ces paramètres consiste en une recherche de grille avec des séquences en croissance exponentielle.

Jusqu'à aujourd'hui, n'y a pas une méthode définitive qui permet de calculer les paramètres optimaux de SVMs. Néanmoins, plusieurs travaux ont été effectués dans ce sens afin de proposer des valeurs pratiques (Schittkowski, 2005, Shitong et al., 2004, Cherkassky and Ma, 2002, Schölkopf et al., 1999, Smola et al., 1998)

(Mjalli et al., 2007) et (Gunter and Zhu, 2007) ont proposé un algorithme qui calcul le chemin de la solution complète de la régression par SVR. Une estimation non biaisée pour le degré de

liberté du modèle SVR permettant de sélectionner le paramètre  $C$  a été proposé. Une version améliorée de l'algorithme génétique appelée algorithme à évolution différentielle est utilisée pour le choix des meilleurs paramètres d'apprentissage de SVR (Lahiri and Ghanta, 2008).

**Paramètre  $C$  :**

Le paramètre  $C$  borne les multiplicateurs de Lagrange. Il permet de créer un compromis entre la pénalisation d'une mauvaise classification et la minimisation de la complexité de fonction. Donc il influence le degré de tolérance face aux valeurs aberrantes « *outliers* ». Par contre, il a un effet négligeable sur la performance de généralisation s'il dépasse un certain seuil.

L'obtention de la valeur optimale de  $C$  par l'utilisation de la méthode validation croisée est présentée dans (Shawe-Taylor and Cristianini, 2004). L'algorithme génétique connu comme un optimiseur optimal et plus particulièrement l'algorithme à évolution génétique a été utilisé pour choisir la meilleur valeur du paramètre  $C$  (Lahiri and Ghanta, 2008). (Smola et al., 1996) présente une méthode efficace pour estimer des fonctions réelles qui demandent moins d'efforts de calcul en utilisant les données. Cette méthode peut être vue comme une façon de compression des données. (Cherkassky and Ma, 2004) ont proposé une méthodologie pour choisir ces méta-paramètres basés sur l'ensemble d'apprentissage au lieu des méthodes classiques basées sur les approches de rééchantillonnage telles que la validation croisée. Une des faiblesses des approches de rééchantillonnage est le coût de calcul assez conséquent.

(Cherkassky and Ma, 2004) ont proposé une relation pour calculer analytiquement  $C$  à partir des données et le niveau de bruit ( $v$ ) qui est la suivante:  $C = \max(|\bar{z} + 3v_z|, |\bar{z} - 3v_z|)$

Où  $\bar{z}$  est la moyenne des réponses de l'ensemble d'apprentissage.  $v_z$  est l'écart type des réponses de l'ensemble d'apprentissage. Cette prescription permet aussi de manipuler les valeurs aberrantes.

**Paramètre  $\sigma$  (variance de la distribution gaussienne) :**

C'est l'unique paramètre du noyau gaussien. Le paramètre  $\sigma$  permet de régler la largeur de la gaussienne, et donc va contrôler la capacité de généralisation des SVMs. (Cherkassky and Ma, 2004) ont proposé cette relation pour calculer  $\sigma$  :

$$\hat{\sigma}^2 = 1.5 \frac{1}{N} \sum_{i=1}^n (z_i - \hat{z}_i)^2$$

### **Paramètre $\epsilon$** (la largeur du tube en SVR)

Ce paramètre n'intervient pas dans le cas de la classification SVM, mais seulement dans le cas de la régression SVR. Il influence la complexité et la régularité de la fonction de régression, le nombre de points supports ( $N_{SV}$ ), la capacité de généralisation et la précision du modèle SVR.

Une valeur trop faible de  $\epsilon$  risque de donner un surapprentissage. (Mattera and Haykin, 1999) propose de choisir une valeur de  $\epsilon$  qui permet d'avoir  $N_{SV}$  égal à la moitié du nombre de données d'apprentissage  $N$ . (Smola et al., 1998) montre qu'il y a une dépendance linéaire entre  $\epsilon$  et la variance de bruit sur les données d'apprentissage, et propose de prendre  $\epsilon$  proportionnel à la variance de bruit ( $\upsilon$ ) comme un choix asymptotiquement optimal. (Kwok and Tsang, 2003, Cherkassky and Ma, 2002, Vapnik, 1999, et Smola et al., 1998) proposent également de choisir  $\epsilon$  proportionnel à la variance de bruit en correspondance avec la formulation d'origine de SVR. Mais cette proposition ne tient pas compte de la taille des données d'apprentissage  $N$ . (Cherkassky and Ma, 2004) ont proposé une équation empirique pour calculer  $\epsilon$  à partir de  $N$  et

$\upsilon$  (la variance du bruit). Pour des grandes valeurs de  $N$ , cette relation est  $\epsilon = \tau \upsilon \sqrt{\frac{\ln N}{N}}$ , avec la

constante  $\tau = 3$



## **CHAPITRE II : Système de traitement des eaux usées**



Dans ce chapitre, la stratégie de diagnostic de défaut proposée dans cette thèse est appliquée au procédé de traitement des eaux usées (station d'épuration) « *Waste Water Treatment Process WWTP* ». Dans la première section, nous présentons le procédé et son modèle. La deuxième section présente une revue bibliographique des méthodes basées sur les données qui ont été utilisées pour la modélisation et la détection de défaut dans le procédé WWTP. Dans la troisième section, nous considérons l'application de la méthode SVR pour la modélisation de certaines variables du procédé (oxygène dissous  $S_{O_5}$ , azote oxydé  $S_{NO_2}$ , Matières En Suspension à la sortie MES ( $TSS_e$  en anglais) .... ) et ensuite la méthode SVM pour le diagnostic de défauts de capteurs ( $S_{O_5}$ ,  $S_{NO_2}$ ,...) et aussi d'actionneurs (débit de recyclage interne  $Q_{intr}$ , coefficient de transfert d'oxygène  $K_{La5}$ ).

## II.1. Description du système

La vie de l'homme, des plantes et des animaux nécessite obligatoirement la présence d'une ressource importante qui est l'eau. Les besoins de cette matière ne cessent d'augmenter avec la croissance de la population mondiale. Cependant, cette ressource est de plus en plus rare à cause des changements climatiques provoquant la sécheresse et la diminution des nappes. La bonne gestion et la revalorisation de l'eau constitue ainsi un enjeu majeur afin de protéger ce patrimoine.

La complexité croissante des pollutions provenant de différentes natures, ménagères (urbaines), agricoles, industrielles et pharmaceutiques nocives à l'environnement et la diffusion de ces pollutions, nous amène à rechercher des méthodes de dépollution plus efficaces avec le moindre coût. En effet, la pollution peut être due à des structures chimiques organiques (carbone, protéines, matières grasses, huiles, pesticides, azote organique, ...) ou inorganiques (métaux lourds, azote ammoniacal, nitrates et nitrites, phosphates, sulfates, ...) présentes dans l'eau. Le traitement des eaux usées permet d'économiser cette ressource et de préserver l'écologie et les milieux aquatiques en éliminant la majeure partie des différents déchets et pollutions carbonées, azotées et phosphorées ainsi que les matières en suspension contenues dans ces eaux (Pons et al., 2008). On note que dans le modèle utilisé par la suite le traitement du phosphore n'est pas considéré.

Les normes européennes ont établi une réglementation stricte et rigoureuse des limites de rejet des eaux usées dans l'environnement. Plusieurs paramètres appelés **variables composites** sont définis pour évaluer la qualité de l'eau. En plus de ces variables, il faut respecter d'autres paramètres tels que la transparence, la couleur, l'odeur et la saveur de l'eau.

Le traitement des eaux usées se fait dans **les stations d'épuration**. Ces stations font intervenir un ensemble d'installations et d'équipements qui permettent d'enlever la pollution et de rendre l'eau conforme selon les normes environnementales avant son rejet dans le milieu naturel, sa réutilisation pour l'irrigation ou son recyclage. Ce sont des systèmes complexes, non linéaires, soumis à des perturbations/incertitudes importantes dans les débits influents (ou affluents) et dans la composition des eaux (la charge polluante) (Belia et al., 2009). En plus de la difficulté liée à la composition de l'entrant, ces réactions ont lieu dans des conditions considérées non optimales (produits toxiques, faible température, oxygène insuffisant, écart de pH, faible concentration en biomasse ...) et il y a une nécessité de fonctionner d'une manière continue.

### **II.1.1. Le procédé de traitement des eaux usées par boues activées**

Le procédé dit à « boues activées » ou à cultures libres est le procédé le plus employé actuellement. Il est basé sur le principe d'une culture bactérienne en suspension et aérée qui permet d'oxyder la matière organique (Hauduc, 2010). Dans ce système, le traitement des eaux résiduaires s'effectue sur plusieurs **étapes** de traitements physico-chimiques et biologiques. Les eaux usées collectées subissent initialement une étape de **prétraitement** dont le but est de faciliter la tâche de traitement et d'éviter tout dommage d'équipement. Le prétraitement se fait dans un premier temps par une **opération de dégrillage/dessablage**. Cette opération consiste en l'exclusion des éléments grossiers difficiles à traiter par les bactéries et l'élimination des sables, pierres et toute matière assez lourde. En deuxième lieu vient **l'opération de dégraissage** qui a pour objectif d'enlever les graisses et huiles qui se trouvent dans l'eau.

Ensuite, lorsque l'eau arrive au bassin d'aération, elle suit une **épuration biologique** par contact avec la masse bactérienne (biomasse) épuratrice par brassage et aération afin d'enlever les pollutions organiques, azotées et le phosphate qui se trouve dans l'eau reçue. Il existe différentes dispositions possibles pour le bassin d'aération suivant les critères à respecter. Un ou plusieurs bassins sont utilisés, en parallèle ou en séries, aérés ou non, ou avec une aération cadencée. En effet, la dénitrification (transformation des nitrates en diazote) nécessite un milieu anoxique tandis que l'élimination de phosphore a besoin d'une zone anaérobie. L'aération a pour but de permettre l'activité biologique des bactéries et de faciliter ainsi la dégradation de la matière organique et des agents polluants (la présence de l'oxygène permet que la matière organique soit avalée par la biomasse qui va ainsi grandir, (Heymann, 2010)). L'azote ammoniacal et les phosphates sont aussi utilisés en tant que nutriments.

Après un séjour suffisant dans le réacteur biologique, l'eau à traiter est ensuite envoyée vers un **clarificateur** ou un décanteur qui va permettre de séparer la biomasse (solide) de l'eau épurée (liquide) par l'effet de gravitation. Une décantabilité convenable des boues signifie une eau de bonne qualité assez claire avec une faible turbidité. L'eau épurée se trouve en haut du décanteur et se déverse à l'extérieur par effet de débordement, tandis que la biomasse (boue) se trouve en bas du clarificateur. Une partie de biomasse est enlevée et jetée comme déchets et la partie qui reste va être recyclée afin de maintenir une concentration assez élevée de boue dans le bassin d'aération qui permet une intensification des processus biologique. Il faut noter que le stockage et le transport des boues produites comme des déchets constituent une charge importante dans les stations d'épuration.

### II.1.2. La modélisation

**La modélisation** des procédés de traitement des eaux est devenue une étape importante utilisée en ingénierie (Jeppsson, 1996 , Gernaey et al., 2004, Rieger et al., 2010). La connaissance des cinétiques biologiques mises en jeu sont essentielles pour assurer une bonne gestion des installations, la supervision et la commande (Spérandio et al., 2007). Par exemple, ceci permet la prédiction de la quantité nécessaire d'oxygène et la production de la biomasse qui sont des informations importantes pour la gestion du procédé. La modélisation a permis de réaliser des simulations réalistes du procédé de traitement à boues activées qui ont été appliquées pour :

- L'optimisation du traitement (Rivas et al., 2008) et de la consommation d'énergie (Jonasson, 2007)
- La conception d'ouvrage d'assainissement (Artan et al., 2004)
- Le développement et le test de nouvelles commandes (David, Vande Wouwer, et al., 2009), (Mulas, 2006)
- Le test des schémas de surveillance et de diagnostic de défauts (Caccavale et al., 2010)

Parmi ces modèles, on peut citer le modèle **ASM1** (*Activated Sludge Modelno1*) qui décrit les processus importants qui se produisent dans les procédés de traitement des eaux usées à boues activées (Henze et al., 1987). Il a été réalisé par le groupe de travail sur la modélisation mathématique pour la conception et le fonctionnement du procédé des boues activées appelé en anglais « *Task Group on Mathematical Modelling for Design and Operation of the Activated Sludge Process* » dans le cadre de l'association IWA en 1986. Ce Modèle est considéré comme la base de plusieurs modèles de référence et il est encore largement utilisé dans les logiciels de

simulation dynamique. Par la suite, différents modèles ont été développés ASM2, ASM2d, ASM3 (Henze et al., 2000).

Par rapport à ASM1 (Nelson and Sidhu, 2009) qui a pour objectif d'éliminer l'azote et le carbone, les modèles ASM2 et ASM2d décrivent en plus les processus liés à l'élimination biologique du phosphore (Henze et al., 1995). Le modèle ASM3 est une version améliorée de ASM1, il a été élaboré par l'introduction de nouveaux processus comme l'incorporation des processus de stockage de la matière organique, et des simplifications comme le changement du processus de mort et régénération par un processus de respiration endogène (Gujer et al., 1999). Des tentatives pour construire un guide pour la modélisation à boues activées ont été faites comme par exemple celle du groupe de travail "Good Modeling Practice" (GMP-TG) de l'International Water Association (IWA) (<https://iwa-gmp-tg.cemagref.fr/>).

Le développement des modèles représentant les procédés de traitement des eaux usées reste une tâche assez délicate car ils intègrent un grand nombre de paramètres et variables. La détermination de la valeur de ces paramètres pose un problème en raison de leur dépendance aux conditions de fonctionnement (température, climat), (Hauduc, 2010). La manipulation de ces modèles, pour l'automatisation du procédé est également difficile, comme ASM1 par exemple, puisqu'ils sont complexes et de nature non linéaire. La **linéarisation** de ces modèles est souvent réalisée afin d'obtenir des modèles linéaires à temps variant plus facilement commandable et interprétable (Benhalla et al., 2010). L'utilisation du **modèle réduit** (plus simple mais assez représentatif du système, (Costa et al., 2008, Queinnec and Gómez-Quintero, 2009) est également courante dans le traitement des eaux usées à boues activées, pour des fins de commande (Chachuat et al., 2003, Ekman, 2003), ou de construction de capteur logiciel (Mulas, 2006), et d'estimation de variables d'états non mesurées (Queinnec and Gómez-Quintero, 2009), (Theilliol et al., 2003). La réduction du modèle peut se faire par la suppression des dynamiques lentes, comme les concentrations des composantes azotées (Chachuat et al., 2003).

Dû aux incertitudes et perturbations liées à ces systèmes, les modèles mécanistiques ont cependant des difficultés à représenter le comportement du système. Pour ces raisons, l'utilisation des méthodes de modélisation dirigées par des données comme les réseaux de neurones (Jeppsson et al. 2002, Lee, Joung, et al., 2005, Mjalli et al., 2007, Ráduly et al., 2007), les moindres carrés partiels PLS (Lee et al., 2007), ou la combinaison des deux (Lee, Lee, et al., 2006a) devient intéressante pour l'automatisation du système.

Le modèle utilisé dans ce travail est le modèle de simulation **BSM1** (Copp, 2002), appelé en anglais « *the Benchmark Simulation Model No 1* ». Il a été développé au sein du groupe de

travail COST « action 682 et 624 » et le groupe de travail sur l'analyse comparative des stratégies de contrôle des installations de traitement des eaux usées de l'association internationale de l'eau (IWA), appelé en anglais « *Task Group on Benchmarking of Control Strategies for Wastewater Treatment Plants* ». Ce modèle décrit le fonctionnement du système et particulièrement le réacteur biologique et le clarificateur intégrant le contrôle et les critères d'évaluation de performances de ces systèmes. Pour plus de détails voir les liens suivants:

<http://www.benchmarkwwtp.org/>

[http://www.primodal.com/iwa\\_taskgroup\\_bsm.htm](http://www.primodal.com/iwa_taskgroup_bsm.htm)

Une extension du modèle BSM1 pour une implémentation complète des systèmes de management d'eaux usées est réalisée par Jeppsson dans (Jeppsson et al., 2007).

Ces différents modèles sont programmés dans plusieurs **logiciels** commerciaux utiles pour la modélisation, l'analyse, l'aide à la conception des procédés ou les tests des stratégies de contrôle, on cite par exemple :

BIOWIN: <http://www.envirosim.com/products/bw32/bw32intro.php>

GPS-X: <http://www.hydromantis.com/GPS-X.html>

SIMBA: <http://www.ifak-system.com/services/dynamic-simulation.html>

Dans cette thèse, nous avons programmé le modèle BSM1 dans Matlab® afin d'y incorporer le bloc FDI, sans l'utilisation des logiciels commerciaux.

### II.1.3. Le modèle BSM1

#### II.1.3.1. Le procédé

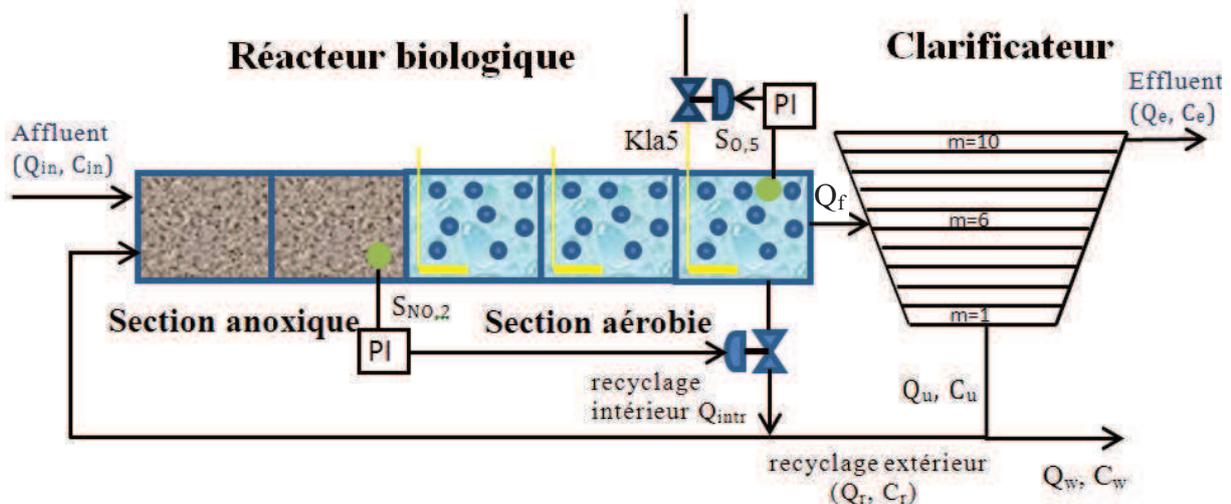
Le procédé de référence BSM1 est constitué d'un réacteur biologique et d'un clarificateur (**Figure II.1**). Le bassin de boues activées (réacteur biologique) est traité comme une succession de cuves en nombre ajustable, égale à 5 par défaut (Copp, 2002). Ces compartiments sont divisés en deux zones comme suit :

- **Une section anoxique** : constituée de deux cuves/compartiments non aérées (absence d'oxygène) mais avec la présence de nitrate. Chaque compartiment a un volume égal à  $V_1 = V_2 = 1000 m^3$ .

- **Une section aérobie** : constituée de trois cuves aérées d'un volume égal  $V_3 = V_4 = V_5 = 1333 \text{ m}^3$ . Les cuves 3 et 4 sont excitées par un coefficient de transfert d'oxygène fixe égal à  $K_{La} = 240 \text{ j}^{-1}$ . En revanche, sa valeur est variable pour la cuve 5, car on va manipuler  $K_{La}$  afin de réguler la concentration de l'oxygène dissous à  $2 \text{ g. (DCO)/m}^3$ .

Où **DCO** représente la Demande Chimique en Oxygène.

Afin de compenser la perte de biomasse nécessaire dans l'opération d'épuration qui est provoquée par le débit de sortie du 5<sup>ème</sup> compartiment du réacteur biologique  $Q_5 = Q_f$ , une partie de la boue du clarificateur est recyclée par pompage vers le réacteur avec un débit  $Q_r$ . Cette opération s'appelle le **recyclage extérieur**. L'autre partie de boue en excès est purgée, avec un débit de purge égal à  $Q_w$ , qui est ensuite déshydratée ou stockée. Les taux de **recyclage** et de purge permettent de maîtriser l'*âge des boues* c'est-à-dire le temps au bout duquel la biomasse sera entièrement renouvelée. Ce temps caractérise la forme physiologique de la biomasse dans le réacteur mais aussi, indirectement, la composition de celle-ci.



**Figure II.1 Présentation générale du procédé BSM1**

Le pilotage de ce procédé est fait par la **régulation** des deux variables suivantes :

1. La concentration  $S_{NO,2}$  de nitrates dans le 2<sup>ème</sup> compartiment à une valeur égale à  $1 \text{ g.(N).m}^{-3}$  par la manipulation d'un débit de recyclage interne  $Q_{intr}$  (ou  $Q_a$ ).
2. La concentration  $S_{O,5}$  de l'oxygène dissous dans le 5<sup>ème</sup> compartiment à une valeur égale à  $2 \text{ g. (DCO).m}^{-3}$  par la manipulation du coefficient du transfert d'oxygène  $K_{la5}$ .

Dans le modèle BSM1, la désignation **X** signifie particulaire et **S** soluble, l'indice **i** signifie inerte et l'indice **s** (exemple **X<sub>s</sub>**) caractérise la biodégradabilité (Dispan, 2003) (Tableau II.1).

**Tableau II.1 Les concentrations intervenant dans le modèle BSM1**

	Définition	Notation
	Alcalinité	$S_{ALK}$
	Matière soluble inerte	$S_I$
	Oxygène dissous	$S_O$
	Azote sous forme de nitrites ( $NO_2^-$ ) et de nitrates ( $NO_3^-$ )	$S_{NO}$
	Azote ammonium ( $NH_4^+$ ) et ammoniac ( $NH_3$ )	$S_{NH}$
	Azote organique soluble et biodégradable	$S_{ND}$
	Substrat soluble facilement biodégradable	$S_S$
Particulaire	Biomasse active hétérotrophe	$X_{B,H}$
	Biomasse active autotrophe	$X_{B,A}$
	Matière organique particulaire inerte	$X_I$
	Azote organique particulaire et biodégradable	$X_{ND}$
	Produits particuliers résultant de décomposition de la biomasse	$X_P$
	Substrat lentement biodégradable	$X_S$

Dans ce modèle, la pollution organique est classée en plusieurs types :

- Substrat soluble facilement biodégradable détérioré et assimilé directement par les microorganismes comme les bactéries,
- Substrat lentement biodégradable nécessitent des processus intermédiaires tels que l'hydrolyse afin d'être assimilé par les bactéries,
- Substrat particulaire.

### **II.1.3.2. Termes de croissances/mortalité**

Les processus biologiques sont modélisés par des équations de type **Monod** (relation empirique) afin d'exprimer les taux de croissance de microorganismes (Monod, 1949):

$$\frac{dX}{dt} = \mu \cdot X \text{ avec } \mu = \mu_m \frac{S}{(K + S)} \quad (\text{II.1})$$

Où  $\mu$  représente le taux de croissance des microorganismes, **K** la constante de demi-saturation en substrat au dessous de laquelle le taux de croissance devient très dépendant de la concentration en substrat et exprime la facilité d'assimilation du substrat par la biomasse,  $\mu_m$  le taux de croissance spécifique maximal des microorganismes et **S** est la concentration du substrat limitant la croissance.

Le processus est modélisé selon l'hypothèse de la reconstitution et du décès. Les organismes décèdent à un certain taux et une partie de la matière organique est considérée comme non biodégradable et s'ajoute au produit particulaire inerte  $X_I$ . Le reste s'ajoute à la quantité du substrat biodégradable. Les réactions biologiques et biochimiques ayant lieu dans le bioréacteur peuvent être résumées en 8 processus basiques décrits par les termes de vitesse de réactions (cinétique), ( $i=1, \dots, 8$ ) :

### 1. Croissance aérobie des hétérotrophes

Les bactéries hétérotrophes nécessitent la présence d'un substrat organique pour se développer. Elles sont incapables de fabriquer leur propre nourriture organique :

$$\rho_1 = \mu_H \left( \frac{S_O}{K_{O,H} + S_O} \right) \left( \frac{S_S}{K_S + S_S} \right) X_{B,H} \quad (\text{II.2})$$

### 2. Croissance anoxie des hétérotrophes

$$\rho_2 = \mu_H \left( \frac{S_{NO}}{K_{NO} + S_{NO}} \right) \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_S}{K_S + S_S} \right) \eta_g X_{B,H} \quad (\text{II.3})$$

Les bactéries hétérotrophes sont en grande partie responsables de la décomposition des matières organiques telles que la matière carbonée en aérobie ou en anoxie. En anoxie elles utilisent les nitrates ( $N_{O_3}^-$ ) ou nitrites ( $N_{O_2}^-$ ) comme source d'oxygène. Les processus anoxies sont inhibés par l'oxygène.

### 3. Croissance aérobie des autotrophes

Les bactéries autotrophes sont capables de se multiplier en utilisant la matière minérale (oxygène, hydrogène, carbone, azote et sels minéraux divers) dans un milieu aéré (aérobie) afin de produire un composant organique complexe qui représente leur nutriment.

$$\rho_3 = \mu_A \left( \frac{S_O}{K_{O,A} + S_O} \right) \left( \frac{S_{NH}}{K_{NH} + S_{NH}} \right) X_{B,A} \quad (\text{II.4})$$

Les équations des processus de croissance de (II.2) à (II.4) sont représentées par un modèle classique de Monod couplé à des fonctions de commutation de type Michaelis-Menten. Ces fonctions vont permettre plus au moins d'activer ou désactiver la croissance (ajout des conditions de limitation) suivant les conditions environnementales concernant la présence de certains éléments nécessaires à la croissance.

#### 4. Mortalité des hétérotrophes

$$\rho_4 = b_H X_{B,H} \quad (\text{II.5})$$

#### 5. Mortalité des autotrophes

$$\rho_5 = b_A X_{B,A} \quad (\text{II.6})$$

#### 6. Ammonification de l'azote organique soluble

$$\rho_6 = k_a S_{ND} X_{B,H} \quad (\text{II.7})$$

#### 7. Hydrolyse des composés organiques absorbés

$$\rho_7 = k_h \left[ \left( \frac{S_O}{K_{O,H} + S_O} \right) + \eta_h \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right] \frac{X_S / X_{B,H}}{K_X + X_S / X_{B,H}} X_{B,H} \quad (\text{II.8})$$

Où le ratio  $X_S / X_{B,H}$  traduit l'abondance de matière enzymatique qui va faciliter la réaction d'hydrolyse.

#### 8. Hydrolyse de l'azote organique absorbé

$$\begin{aligned} \rho_8 &= k_h \left[ \left( \frac{S_O}{K_{O,H} + S_O} \right) + \eta_h \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right] \frac{X_S / X_{B,H}}{K_X + X_S / X_{B,H}} \frac{X_{ND}}{X_S} X_{B,H} \\ &= \rho_7 \frac{X_{ND}}{X_S} \end{aligned} \quad (\text{II.9})$$

Les paramètres du modèle BSM1 dépendent de la température. Par exemple le taux de croissance des bactéries ( $\mu_A$ ,  $\mu_H$ ) et leurs taux de mortalité ( $b_A$ ,  $b_H$ ) varient suivant la température. Les valeurs données dans le Tableau II.2 correspondent à une température aux alentours de 15°C.

**Tableau II.2 Paramètres du modèle BSM1**

	Description	Symbole	Valeur	Unité
Paramètres Stoechiométrique	Rendement de croissance autotrophe	$Y_A$	0.24	$\frac{\text{g DCO}_{X_{B,A}}}{\text{g N}}$
	Rendement de croissance hétérotrophe	$Y_H$	0.67	$\frac{\text{g DCO}_{X_{B,H}}}{\text{g N}}$
	Fraction de biomasse recyclée en produits particuliers	$f_p$	0.08	-
	Fraction d'azote dans la biomasse	$i_{XB}$	0.08	$\frac{\text{g N}}{\text{g DCO}}$
	Fraction d'azote dans les produits de lyse	$i_{XP}$	0.06	$\frac{\text{g N}}{\text{g DCO}}$

Paramètres cinétiques	Taux de croissance maximal de la biomasse hétérotrophe	$\mu_H$	4	$j^{-1}$
	Constante de demi-saturation hétérotrophe	$K_S$	10	$g \text{ DCO} \cdot m^{-3}$
	Constante de demi-saturation de l'oxygène pour la biomasse hétérotrophe	$K_{O,H}$	0.2	$g \text{ O}_2 \cdot m^{-3}$
	Constante de demi-saturation du nitrate pour la biomasse hétérotrophe	$K_{NO}$	0.5	$g \text{ NO}_3\text{-N} \cdot m^{-3}$
	Taux de décès hétérotrophe	$b_H$	0.3	$j^{-1}$
	Facteur de correction pour la croissance anoxique des hétérotrophes	$\eta_g$	0.8	-
	Facteur de correction pour l'hydrolyse sous condition anoxie	$\eta_h$	0.8	-
	Taux d'hydrolyse spécifique maximal	$k_h$	3	$\frac{g \text{ DCO}_x}{g \text{ DCO}_{x,B,H}} \cdot j^{-1}$
	Constante de demi-saturation pour l'hydrolyse du substrat lentement biodégradable $X_S$	$K_X$	0.1	$\frac{g \text{ DCO}_{x,S}}{g \text{ DCO}_{x,B,H}}$
	Taux de croissance maximal de la biomasse autotrophe	$\mu_A$	0.5	$j^{-1}$
	Constante de demi-saturation de l'azote ammoniacal pour la biomasse autotrophe	$K_{NH}$	1	$g (\text{NH}_4\text{-N}) \cdot m^{-3}$
	Taux de décès autotrophe de la biomasse autotrophe	$b_A$	0.05	$j^{-1}$
	Constante de demi-saturation de l'oxygène pour la biomasse autotrophe	$K_{O,A}$	0.4	$g (\text{O}_2) \cdot m^{-3}$
	Taux d'ammonification	$k_a$	0.05	$m^3 (g (\text{DCO}) j)^{-1}$

Le traitement des différents types de pollution (organique, azotée) se fait par des méthodes distinctes, décrites dans les sections suivantes. On note que lorsque les matières organiques (Figure II.2) sont abondantes, cela va causer le dépassement de la capacité d'autoépuration naturelle du milieu récepteur. En effet, la dégradation de ces substances (oxydation par des microorganismes) provoque une consommation excessive d'oxygène au détriment des organismes aquatiques qui vont être asphyxiés.

La reproduction des microorganismes intervient en conditions favorables en favorisant le métabolisme des bactéries. Lorsque leur croissance est importante et que les bactéries se mettent à se diviser, un flocculant est ajouté leur permettant de s'accumuler en floccs décantables produisant des boues. Cette flocculation va permettre d'éliminer les composantes carbonées (graisses, sucres, protéines, etc.) et les éléments polluants qui se trouvent dans l'eau reçue.

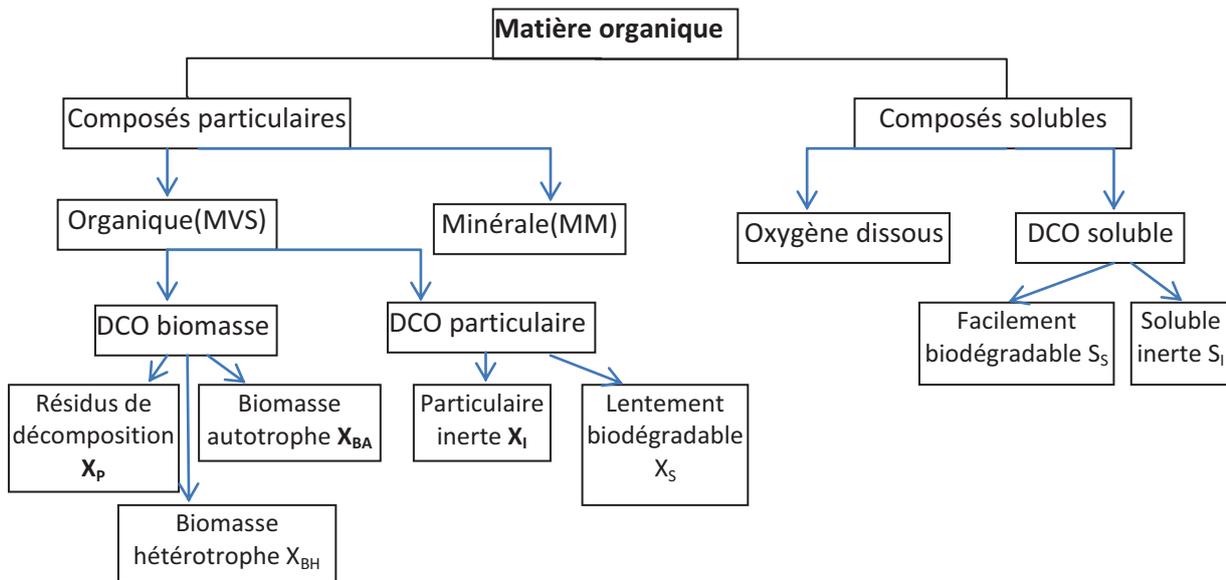


Figure II.2 La matière organique dans le modèle BSM1

### II.1.3.3. Traitement de la pollution azotée

La grande partie de l'azote qui se trouve dans l'eau résiduaire est présente sous forme d'azote ammoniacal (ammoniac  $NH_3$ , ion ammonium  $NH_4^+$ , nitrates  $NO_3^-$  ou nitrites  $NO_2^-$ , Figure II.3).

On note que l'ammoniac et les nitrites sont des produits toxiques.

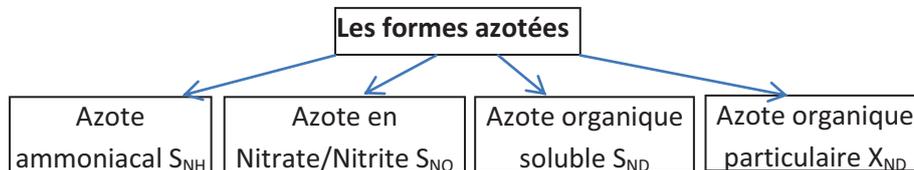
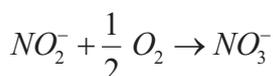
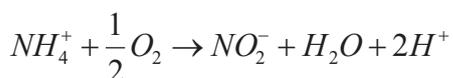


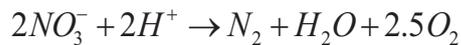
Figure II.3 Les composés azotés dans le modèle BSM1

L'azote est supprimé par les opérations de nitrification et dénitrification (Gujer, 2010) :

1. **Nitrification biologique:** c'est le processus de transformation de l'ammonium par oxydation. Cela s'effectue en deux étapes en nitrite d'abord ( $NO_2^-$ ) puis rapidement en nitrate ( $NO_3^-$ ) :



2. **Dénitrification biologique:** cette opération survient dans un environnement anoxique, et a comme but d'éliminer les nitrates ( $NO_3^-$ ) en produisant le gaz d'azote :



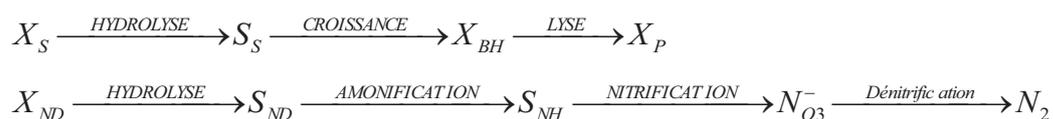
#### II.1.3.4. Traitement de la pollution phosphorique

La présence de phosphates dans les eaux résiduaires provoque l'augmentation de la turbidité et la multiplication des algues et d'autres espèces qui déséquilibre totalement le milieu et gêne beaucoup le passage de la lumière aux couches inférieures, entraînant ainsi le phénomène d'eutrophisation. L'élimination de phosphore dans les stations d'épuration modernes est assez neuve. Cette opération peut se faire de deux façons, soit chimique ou biologique :

- **Traitement chimique :** cette méthode se fait par un dosage d'un produit chimique (sel d'aluminium par exemple) qui permet la précipitation des molécules de phosphates qui se mettent en liaison, ensuite une coagulation se crée et enfin des floes se forment. Ces derniers peuvent être enlevés par sédimentation. Ce type de traitement a plusieurs inconvénients, par exemple la salinisation de l'eau, l'augmentation de la quantité de boues, sans oublier le prix des précipitants qui augmente.
- **Traitement biologique :** ce type de traitement nécessite la présence d'une zone anaérobie appelée digestion anaérobie afin de libérer les phosphates, ensuite les phosphates sont absorbés dans des conditions anoxiques ou aérobies. Comme l'absorption est plus grande que la libération, elle conduit à une absorption nette de phosphore.

#### II.1.3.5. Hydrolyse

Dans certains cas, quand le substrat (par exemple  $X_S$ ) n'est pas assimilable par les microorganismes, une opération d'hydrolyse s'effectue à l'aide des exoenzymes afin de le rendre plus facilement assimilable par les microorganismes et donc dégradable.



### II.1.3.6. Vitesses des réactions

Les vitesses des réactions  $r_j$  ( $j=1\dots 13$ ) pour chaque élément sont une combinaison des 8 processus précédents :

$$\left\{ \begin{array}{l}
 \text{Pour } S_i(j=1): \quad r_1 = 0 \\
 \text{Pour } S_s(j=2): \quad r_2 = -\frac{1}{Y_H} \rho_1 - \frac{1}{Y_H} \rho_2 + \rho_7 \\
 \text{Pour } X_I(j=3): \quad r_3 = 0 \\
 \text{Pour } X_S(j=4): \quad r_4 = (1-f_p)\rho_4 + (1-f_p)\rho_5 - \rho_7 \\
 \text{Pour } X_{B,H}(j=5): \quad r_5 = \rho_1 + \rho_2 - \rho_4 \\
 \text{Pour } X_{B,A}(j=6): \quad r_6 = \rho_3 - \rho_5 \\
 \text{Pour } X_P(j=7): \quad r_7 = f_p\rho_4 + f_p\rho_5 \\
 \text{Pour } S_O(j=8): \quad r_8 = -\frac{1-Y_H}{Y_H} \rho_1 - \frac{4.57-Y_A}{Y_A} \rho_3 \\
 \text{Pour } S_{NO}(j=9): \quad r_9 = -\frac{1-Y_H}{2.86Y_H} \rho_2 + \frac{1}{Y_A} \rho_3 \\
 \text{Pour } S_{NH}(j=10): \quad r_{10} = -i_{XB}\rho_1 - i_{XB}\rho_2 + \left(i_{XB} + \frac{1}{Y_A}\right)\rho_3 + \rho_6 \\
 \text{Pour } S_{ND}(j=11): \quad r_{11} = -\rho_6 + \rho_8 \\
 \text{Pour } X_{ND}(j=12): \quad r_{12} = (i_{XB} - f_P i_{XP})\rho_4 + (i_{XB} - f_P i_{XP})\rho_5 - \rho_8 \\
 \text{Pour } S_{ALK}(j=13): \quad r_{13} = -\frac{i_{XB}}{14} \rho_1 + \left(\frac{1-Y_H}{14.2.86Y_H} - \frac{i_{XB}}{14}\right)\rho_2 + \left(\frac{i_{XB}}{14} + \frac{1}{7Y_A}\right)\rho_3 + \frac{1}{14} \rho_6
 \end{array} \right. \quad (\text{II.10})$$

Avec  $r_j = \sum_i v_{ij} \rho_i$  , avec  $i = 1\dots 8$ , et  $j = 1\dots 13$

$v_{ij}$  sont les coefficients stœchiométriques, impliquant chaque variable dans chaque réaction. Leurs valeurs sont négatives pour les composés consommés et positives pour ceux produits.

Les équations de réactions précédentes montrent que ce modèle fait intervenir des relations de liaison entre les différentes composantes qui créent ainsi une sorte de compétition entre ces

composantes. Certaines vont se développer plus aisément au détriment d'autres qui deviennent une minorité ou vont disparaître. Tout ceci donne un modèle complexe et non linéaire.

### II.1.3.7. Bilans de matière

Le modèle mathématique du procédé de traitement des eaux usées est établi à partir des bilans de matière pour chaque composante sur l'ensemble des bassins :

$$\text{Accumulation} = \text{Flux d'entrée} + \text{Terme de réaction} - \text{Flux de sortie}$$

Les concentrations des différentes composantes sont réunies dans un vecteur d'état  $\mathbf{C}$  de 13 variables :  $C^T = [S_I, S_S, X_I, X_S, X_{B,H}, X_{B,A}, X_P, S_O, S_{NO}, S_{NH}, S_{ND}, X_{ND}, S_{ALK}]$

Les équations dynamiques de ces concentrations font intervenir les vitesses des réactions détaillées plus haut. Pour le premier compartiment  $k=1$  (ou 1<sup>er</sup> unité) :

$$\frac{dC_1}{dt} = \frac{1}{V_1} (Q_{in} C_{in} + Q_r C_r + Q_a C_a + r_1 V_1 - Q_1 C_1) \quad \text{avec} \quad Q_1 = Q_{in} + Q_r + Q_a \quad (\text{II.11})$$

Pour  $k=2$  à  $5$  :

$$\frac{dC_k}{dt} = \frac{1}{V_k} (Q_{k-1} C_{k-1} + r_k V_k - Q_k C_k) \quad \text{avec} \quad Q_k = Q_{k-1} \quad (\text{II.12})$$

Et :

$$\begin{aligned} Q_f &= Q_5 - Q_a = Q_e + Q_w + Q_r = Q_e + Q_u \\ C_a &= C_5, \quad C_f = C_5, \quad C_w = C_r \end{aligned} \quad (\text{II.13})$$

Il faut noter que l'équation différentielle de l'oxygène dissous représente un cas spécial dans le sens où c'est la seule variable d'état dont la dynamique dépend d'un terme de transfert qui caractérise le couple réacteur/système d'aération. Son expression est donnée comme suit :

$$\frac{dS_{O,k}}{dt} = \frac{1}{V_k} (Q_{k-1} S_{O,k-1} + r_k V_k - Q_k S_{O,k} + (K_L a)_k V_k (S_O^* - S_{O,k}) S_{O,k}) \quad (\text{II.14})$$

Avec  $S_O^* = 8 \text{ g.m}^{-3}$  la concentration de saturation pour l'oxygène.

### II.1.3.8. Dynamique du clarificateur

Le clarificateur ou décanteur secondaire a pour rôle de séparer les matières en suspension de l'eau épurée sans réaction, juste par effet de sédimentation ou décantation. Il est modélisé de 10 couches par défaut d'une hauteur totale de 4 m avec un volume de 6000 m<sup>3</sup> et une surface

circulaire de 1500 m<sup>2</sup>. Le modèle de simulation du clarificateur utilisé est un modèle unidimensionnel qui décrit le processus de décantation le long de l'axe vertical avec l'hypothèse d'un bon mixage dans chaque couche (Takács, 2008), (Takács et al., 1991), (David, Saucez, et al., 2009). D'autres façons de modéliser le décanteur existent, par exemple la stratégie basée sur les équations non linéaires à dérivés partielles (Diehl and Jeppsson, 1998).

**Tableau II.3 Paramètres de décantation**

Paramètre	Symbole	Valeur	Unité
Vitesse maximale de décantation	$v'_0$	250	m.j <sup>-1</sup>
Vitesse maximale de décantation (Vitesse de Vesilind)	$v_0$	474	m.j <sup>-1</sup>
La décantation freinée	$r_h$	0.000576	m <sup>3</sup> /g de TSS
La décantation de particules isolées	$r_p$	0.00286	m <sup>3</sup> /g de TSS
Fraction non décantable	$f_{ns}$	0.00228	-
Flux de particules qui décantent de la couche j vers la couche j+1	$J_{s,j}$	-	m <sup>3</sup> .j <sup>-1</sup>
Concentration en solides non décantables	$X_{\min}$	-	g.m <sup>-3</sup>
Vitesse de décantation dans la couche j	$v_{s,j}$	-	m.j <sup>-1</sup>

Le flux solide dû à la pesanteur est donné par :

$$J_s = v_s(X)X \quad (\text{II.15})$$

Où X est la concentration totale des boues. Suivant (Takács et al., 1991) la vitesse du clarificateur peut s'écrire sous la forme suivante :

$$v_s(X) = \max \left[ 0, \min \left\{ v'_0, v_0 \left( e^{-r_h(X-X_{\min})} - e^{-r_p(X-X_{\min})} \right) \right\} \right] \quad \text{avec } X_{\min} = f_{ns} X_f \quad (\text{II.16})$$

La hauteur de chaque couche **m** est égal à  $z_m = 0.4m$ . Sur chaque couche, un bilan de matière est établi afin de déduire la concentration des matières en suspension. Pour la couche du fond (**m=1**) :

$$\frac{dX_1}{dt} = \frac{v_{dn}(X_2 - X_1) + \min(J_{s,2} - J_{s,1})}{z_1} \quad (\text{II.17})$$

Pour les couches intermédiaires inférieures (**m entre 2 et 5**) :

$$\frac{dX_m}{dt} = \frac{v_{dn}(X_{m+1} - X_m) + \min(J_{s,m} - J_{s,m+1}) - \min(J_{s,m} - J_{s,m-1})}{z_m} \quad (\text{II.18})$$

Pour la couche d'alimentation (**m=6**) :

$$\frac{dX_m}{dt} = \frac{\frac{Q_f X_f}{A} + J_{clar,m+1} - (v_{up} - v_{dn})X_m - \min(J_{s,m} - J_{s,m-1})}{z_m} \quad (\text{II.19})$$

Pour les couches intermédiaires supérieures (**m entre 7 et 9**) :

$$\frac{dX_m}{dt} = \frac{v_{up}(X_{m-1} - X_m) + J_{clar,m+1} - J_{clar,m}}{z_m} \quad (II.20)$$

$$J_{clar,j} = \begin{cases} \min(v_{s,j}X_j, v_{s,j-1}X_{j-1}) & \text{si } X_{j-1} > X_{seuil} \\ v_{s,j}X_j & \text{si } X_{j-1} \leq X_{seuil} \end{cases}$$

$X_{seuil}$  est la concentration seuil en solides en suspension qui est égale à  $3000 \text{ g.m}^{-3}$ .

Pour la couche supérieure (**m=10**) :

$$\frac{dX_{10}}{dt} = \frac{v_{up}(X_9 - X_{10}) - J_{clar,10}}{z_{10}} \quad (II.21)$$

Pour plus de détails sur le modèle de référence BSM1, le lecteur est conseillé de voir (Alex et al., 2008).

## II.1.4. Variables composites/mesures

### II.1.4.1. Les normes

Les normes de rejet sont basées sur certaines mesures qui déterminent la qualité de l'eau de l'effluent.

- **DCO** ( $\text{kg.j}^{-1}$ ) : La Demande Chimique en Oxygène (*Chemical Oxygen Demand (COD) en anglais*)

Elle représente la quantité d'oxygène nécessaire pour l'oxydation (classiquement par dichromate de potassium) de la totalité des matières (susceptibles de consommer de l'oxygène) contenues dans l'eau par voie chimique, en milieu acide et chaud. DCO doit être inférieure à  $100 \text{ g.m}^{-3}$  d'oxygène dissous :

$$DCO = X_{B,A} + X_{B,H} + X_P + X_S + X_I + S_S + S_I \quad (II.22)$$

- **DBO5** ( $\text{kg.j}^{-1}$ ) : Demande Biologique en Oxygène sur 5 jours (*Biological Oxygen Demand after five days (BOD<sub>5</sub>) en anglais*)

Elle représente la quantité d'oxygène nécessaire pour oxyder les matières organiques (biodégradables) par voie biologique (par des bactéries). Cette quantité indique la teneur en matières organiques biodégradables de la charge polluante carbonée des eaux usées. **DBO<sub>5</sub>** doit être inférieure à  $10 \text{ g.m}^{-3}$  d'oxygène dissous au bout de 5 jours.

$$DBO_5 = 0.25 \times (X_S + S_S + (1 - f_p)(X_{B,A} + X_{B,H})) \quad (II.23)$$

- **MES** ( $\text{kg.j}^{-1}$ ) : Les Matières En Suspension (*Total Suspended Solids (TSS) en anglais*)  
Il s'agit de l'ensemble des particules solides fines non dissoutes visibles à l'œil nu qui se trouvent en suspension dans l'eau. Ces matières augmentent la turbidité, empêchent la bonne pénétration de la lumière et affectent négativement le milieu aquatique, les animaux et les plantes. MES doivent être inférieurs à **30  $\text{g.m}^{-3}$**  pour satisfaire les normes de rejet.

$$MES = 0.75 \times (X_P + X_I + X_S + X_{B,A} + X_{B,H}) \quad (\text{II.24})$$

- **NTK** ( $\text{kg.j}^{-1}$ ) : Azote total (ammoniacal et organique) (*Total Kjeldahl Nitrogen (TKN) en anglais*)

Pour l'**azote**, 5 formes peuvent être trouvées dans le procédé de traitement d'eau usée : L'azote organique (constitue les cellules vivantes végétales ou animales principalement composées d'urée et d'acides aminés), l'azote ammoniacal ( $N_{H4}^+$ ), les nitrites ( $N_{O2}^-$ ), les nitrates ( $N_{O3}^-$ ) et l'azote gazeux ( $N_2$ ). **NTK** correspond à la somme de l'azote sous forme ammoniacale et organique. Elle doit être inférieure à **18  $\text{g.m}^{-3}$** .

$$\begin{aligned} NTK &= S_{NH} + S_{ND} + X_{ND} + i_{XB} \times (X_{B,A} + X_{B,H}) + i_{XP} \times (X_P + X_I) \\ N_{Total} &= S_{NO} + NTK \end{aligned} \quad (\text{II.25})$$

Dans cette thèse, ces variables sont notées COD, BOD<sub>5</sub>, TSS et TKN par la suite.

#### **II.1.4.2. Les capteurs de mesures**

Les grandeurs généralement mesurées en ligne sont les suivantes : la température, la pression, le pH, les débits, le potentiel d'oxydo-réduction (*Oxidation Reduction Potential (ORP) en anglais*) pour la réaction, les concentrations d'oxygène, d'azote total, d'azote ammoniacal, des nitrites/nitrates, et de phosphore, Les Matières Volatiles en Suspension (TSS), la demande chimique en oxygène (COD), la demande biologique (ou biochimique) en oxygène sur 5 jours (BOD<sub>5</sub>), etc. (Vanrolleghem and Lee, 2003, Ingildsen, 2002). Ces mesures ont une grande importance pour la commande et l'amélioration de la performance de fonctionnement qui peut aller jusqu'à la suppression de l'azote, carbone de 20 à 50% (Olsson, 2007). Elles permettent d'évaluer l'efficacité de l'épuration, de s'assurer du respect des limites de rejet, et d'aider à cerner les défauts ou dysfonctionnements possibles dans les équipements de la station d'épuration.

Les capteurs pour la surveillance du processus d'élimination des nutriments de procédé WWTP se décomposent en trois catégories (Gernaey et al., 1998): sondes directes (oxygène dissous DO, pH, Potentiel d'Oxydo-Réduction qui mesure la réactivité des espèces chimiques, turbidité)

sondes indirectes (exemple analyseurs  $N_{H4}^+$  et  $N_{H3}^-$  en ligne) et biocapteurs (exemple: la demande d'oxygène rapide et le testeur de toxicité (Rapid Oxygen Demand and Toxicity tester RODTOX), *NITRification tOXicity tester (NITROX)*, *DEnitrification CARbon source DOsage System; patent pending (DECADOS)*, ...). Par exemple, afin de sauvegarder les microorganismes responsables de l'épuration contre des influents toxiques, un pH-mètre ou mieux un capteur de toxicité par respirométrie peut être installé sur l'eau à traiter. L'aération peut être régulée par sonde à oxygène dissous ou encore mieux par potentiel redox.

## II.2. Commande des systèmes de traitement des eaux usées

Le contrôle des procédés de traitement des eaux usées est important pour satisfaire les performances désirées, réduire les émissions de gaz à effet de serre, améliorer la qualité de l'effluent, protéger l'environnement et optimiser les coûts d'énergie et des produits chimiques utilisés. On trouve plusieurs travaux dans la littérature tels que (Mulas, 2006), (Ekman, 2003), (Ingildsen, 2002).

- Trois éléments sont nécessaires pour définir la stratégie de contrôle :
  - 1- Les variables manipulées : variables affectées par les actionneurs permettent de manipuler le système.
  - 2- Les variables mesurées : obtenues à partir des capteurs.
  - 3- L'algorithme de contrôle.

Les informations concernant le procédé sont transmises par les capteurs, qui sont comme les yeux et les oreilles chez l'homme. Ces données transmises sont nécessaires pour réaliser un bon contrôle en ligne du procédé de traitement des eaux (Schütze et al., 2004).

La commande des procédés WWTP consiste à commander l'aération (Ekman et al., 2006), pompage de boues et les aliments chimiques par les manipulations d'actionneurs (vanne, pompe, etc..) et le contrôle par le débit de purge permet de régler le temps de séjour des boues dans le procédé.

Parmi les stratégies de contrôle, on cite le correcteur PID qui est la façon la plus simple et en même temps efficace pour réguler des systèmes sans avoir à connaître des informations autres que la mesure de la variable à contrôler sur ces derniers. Cependant, cette commande présente des inconvénients tels que la limite à rejeter des perturbations et rencontre des problèmes face à des procédés de grande constante de temps. Autre stratégie, la commande anticipatrice

(*feedforward control*) qui permet d'entreprendre une action corrective avant même que les perturbations n'affectent les sorties du système. Cette commande est utile pour les procédés lents comme c'est le cas dans le procédé WWTP. Une commande adaptative basée sur architecture en réseau de neurones est appliquée à la régulation de l'oxygène dissous dans un procédé WWTP (Han and Qiao, 2011), une approche combinant un réseau neuronal avec une commande floue (Waewsak et al., 2010). Un bon rapport qui discute les méthodes de contrôle des procédés WWTP municipales peut être revu dans (Haimi et al., 2009)

Le maintien des objectifs de contrôle quand le procédé de traitement des eaux est affecté par un défaut est fondamental. Cette tâche se réalise par une commande tolérante aux défauts (Blanke et al., 2006) qui consiste à soit accommoder le défaut si ce dernier n'est pas très dangereux par l'adaptation des paramètres de contrôleur aux propriétés dynamique du procédé défectueux, ou carrément par la reconfiguration de la loi de commande ainsi que la structure du contrôleur. Une stratégie de commande tolérante au défaut a été proposée par (Zumoffen and Basualdo, 2008a). Celle-ci est basée sur une commande prédictive en ligne et est appliquée à un procédé WWTP.

### **II.3. Revue des méthodes de diagnostic du procédé WWTP**

Afin d'augmenter la robustesse et la sensibilité de la surveillance du procédé WWTP, il faut un système de diagnostic précis, puissant et efficace qui permet la détection rapide des défauts. Le bloc de diagnostic de défaut (*FDI en anglais*) et le bloc de la commande doivent être conçus de façon à réduire la vulnérabilité face aux défauts, éliminer les fausses alarmes, diminuer les coûts, économiser l'énergie et la consommation de produits chimiques, améliorer l'élimination des nutriments et de la pollution en général et améliorer la sécurité et protéger les écosystèmes aquatiques, l'environnement et les eaux.

L'immersion de grandes quantités de données obtenues par la modernisation des systèmes et l'apparition de nouveaux capteurs facilitent la tâche de la surveillance du procédé WWTP. Les données accueillies à partir du procédé sont cependant de nature non linéaires, non stationnaires, souvent corrélées et bruitées, et multi échelles dues aux variations journalières, hebdomadaires et saisonnières. Ceci rend la tâche de diagnostic de défaut plus difficile.

Différentes méthodes ont été appliquées pour le diagnostic de défaut dans le procédé WWTP. Notre revue bibliographique ne considérera que les méthodes basées sur les données qui représentent le sujet de cette thèse, sachant qu'un nombre important de travaux concerne les

méthodes basées sur les modèles (ex. l'estimation de paramètres (Fuentes et al., 1996), les observateurs adaptatifs (Fragkoulis et al., 2011)).

Parmi les outils basés sur les données qui ont été utilisés dans le traitement des eaux usées, nous pouvons citer les **réseaux de neurones** (Lee et al., 2008), (Caccavale et al., 2010)) le **réseau Bayésien** ((Chong and Walley, 1996), les modèles Bonds graphs (Zhang and Hoo, 2011) et les méthodes statistiques (ex. ACP (*PCA en anglais*), PLS et SVM).

La construction de modèles de prédiction de certaines variables du procédé WWTP comme les concentrations de l'effluent ( $COD_{eff}$ ,  $BOD_{eff}$ , TSS,TKN ,...) est intéressant pour surveiller le comportement du procédé WWTP. Cette estimation peut se faire par l'utilisation soit d'un Réseau de Neurones Artificiel (ANN en anglais) ((Ráduly et al., 2007), (Mjalli et al., 2007)) ou par l'application de méthodes de régression telles que l'Analyse Factorielle Parallèle et la Projection en Structures Latentes non linéaire (Singh et al., 2009). Dans cette thèse on s'intéresse aussi à la construction de modèle de prédiction par SVR.

### II.3.1. ACP (*PCA en anglais*) et d'autres méthodes statistiques

Différentes méthodes statistiques ont été appliquées dans le diagnostic de défaut dans le procédé WWTP. Parmi ces approches on peut citer :

- Analyse en composantes principales « *Principal Components Analysis (PCA) en anglais* » appliqué à l'analyse d'un procédé WWTP dans (Yoo et al., 2008), (Yoo, Villez, et al., 2006), (Tomita et al., 2002).
- Analyse en composantes principales dynamique « *Dynamic principal components analysis (DPCA) en anglais* », appliquée au diagnostic de défaut dans un procédé WWTP comme (Lee, Choi, et al., 2004), (Yoo et al., 2002).
- Analyse en composantes principales dynamique à noyau « *Dynamic Kernel Principal Components Analysis (DKPCA) en anglais* » appliquée à la surveillance et à la détection de défauts dans un procédé WWTP (Choi and Lee, 2004).
- Analyse en Composantes Indépendantes (ACI) « *Independent Component Analysis (ICA) en anglais* » appliquée à la surveillance d'un procédé WWTP (Lee, Yoo, et al., 2004).
- Analyse en Composantes Indépendantes Dynamique (ACID) « *Dynamic Independent Component Analysis (DICA) en anglais* » (Stefatos and Ben Hamza, 2010).
- Analyse en Composantes Indépendantes à noyau « *Kernel Independent Component Analysis (KICA) en anglais* » (Wang and Shi, 2010).

- Analyse en Composantes Indépendantes modifiée (**ACIM**) « *Modified Independent Component Analysis (MICA) en anglais* » appliqué pour la détection, l'isolation, et la surveillance du procédé WWTP (Lee, Qin, et al., 2006)
- Analyse canonique (**AC**) « *canonical variate analysis (CVA) en anglais* » (Juricek et al., 2004), (Russell et al., 2000).
- Analyse Discriminante de Fisher à noyau (**ADFN**) « *Kernel Fisher's Discriminant Analysis (KFDA) en anglais* » utilisée pour déterminer les variables qui permettent une discrimination entre deux ou plusieurs groupes naturels (Li and Cui, 2009).
- Analyse Factorielle des Correspondances (**AFC**) « *Correspondence Analysis (CA) en anglais* » (Pusha et al., 2009), (Detroja et al., 2007).

Ces approches multivariées ont été appliquées sur différents systèmes, mais leur principe se ressemble puisqu'il se base sur:

1. Construire un modèle sans défaut dans les conditions normales de fonctionnement
2. Choix d'une métrique et d'un seuil
3. Projeter une donnée de test sur ce modèle, le calcul de la distance métrique et convenablement la classer comme une donnée normale ou défectueuse
4. Dans le cas d'un défaut, l'identification des variables qui sont associées au défaut en utilisant des mesures de contribution

D'autres variétés ont aussi été utilisées pour ce système telles que l'ICA multidimensionnelle "*multi-way en anglais*" (Yoo, Vanrolleghem, et al., 2006), une méthode combinant la déformation temporelle dynamique "*Dynamic Time Warping (DTW) en anglais*" avec l'analyse discriminante (Jun, 2011).

Une évolution intéressante des méthodes statistiques concerne l'utilisation de **noyaux non linéaires**. En effet, les méthodes classiques statistiques multivariées comme l'ACP et l'ACI peuvent rencontrer des difficultés lors de la manipulation des données réelles dynamiques à cause des changements saisonniers comme dans le cas WWTP et des non linéarités. Des études comparatives ont montré que l'utilisation d'un modèle non linéaire avec un noyau gaussien rend de meilleurs résultats par rapport à l'ACI dans le diagnostic de défaut d'un procédé WWTP (Yu, 2012). D'autres limitations de la méthode ACP sont liées au fait qu'elle ne permet pas d'analyser des procédés qui admettent des variations sur plusieurs échelles de temps. Dans ce cas, l'utilisation de l'approche **ACP multi-échelles**, qui utilise la transformée d'ondelettes (par

exemple) peut être utile pour la surveillance du procédé de traitement des eaux (Rosen and Lennox, 2001). La transformée d'ondelettes est souvent utilisée comme étape de prétraitement des signaux afin de déterminer les caractéristiques et les singularités de transitions (Zhang and Hoo, 2011), (Saravanan et al., 2008), (Abbasion et al., 2007), (Ruiz et al., 2000). Cette méthode permet de décomposer le signal en plusieurs échelles de temps, ce qui aide à analyser le signal localement et ainsi faire face aux perturbations. Ceci est réalisé en multipliant le signal original par les ondelettes de différentes tailles et à différents intervalles de temps, une autre difficulté surgie lors de l'utilisation de l'ACP est liée au non stationnarité du procédé. Cette difficulté peut être surmontée en utilisant l'**ACP adaptative** qui met à jour les propriétés statistiques (variance, moyenne, covariance..) du modèle ACP initial. Cette approche est appliquée à la surveillance du procédé WWTP dans (Rosen and Lennox, 2001). On note que la transformée de Fourier est incapable de traiter de façon globale les signaux non-stationnaires et perd aussi le comportement temporel du signal. Enfin, la méthode **ACI** est utilisée pour la transformation d'espace dans l'objectif de projeter les données dans un espace de caractéristique de faible dimension construit par des variables latentes qui sont à la fois non gaussiennes et aussi indépendantes statistiquement entre elles (Zhang, 2008).

### **II.3.2. La Projection aux Structures Latentes ( *PLS* )**

L'approche PLS « *Partial Least Square* » est très utilisée en traitement des eaux pour construire des modèles pour le suivi de plusieurs paramètres de rejet tels que COD, BOD et TSS (ex. procédé de traitement d'un parc de carburant, Lourenço et al., 2008). Elle a été utilisée pour quantifier les incertitudes des capteurs (Rieger et al. 2005). Elle a également été utilisée afin de construire un modèle non paramétrique en utilisant un réseau neuronal combiné avec un modèle mécanistique dans une configuration parallèle qui a été appliquée pour la détection et l'isolation de défauts dans un procédé WWTP (Lee, Vanrolleghem, et al., 2005). Afin de surmonter les difficultés liées à la rigidité des modèles statistiques et au manque de souplesse face aux changements du procédé, une **PLS adaptative robuste** a été développée pour la prédiction et la surveillance du procédé WWTP (Lee et al., 2007). Afin de prendre en compte les nonlinéarités et les dynamiques du procédé, une approche hybride (NDPLS) combinant **PLS nonlinéaire**, réalisée par réseaux de neurone NN avec une méthode de modélisation dynamique (réponse impulsionnelle finie (RIF)), a été proposée pour la prédiction de variables dans un procédé WWTP (Lee, Lee, et al., 2006b). La réduction de la complexité de l'analyse du procédé WWTP et la décentralisation de la surveillance et du diagnostic de défaut d'une manière hiérarchique

peuvent être faites par **PLS multi-blocs** qui a été utilisée pour l'estimation du COD en utilisant les cartes de contrôle Q et T<sup>2</sup> pour détecter les défauts (Choi and Lee, 2005).

### II.3.3. SVM

Les approches de base de SVM sont la classification (SVM) et la régression (SVR). L'approche SVR est appliquée pour la modélisation (prédiction, estimation ou identification) alors que l'approche SVM est appliquée pour le diagnostic de défauts. On cite par exemple l'utilisation de SVR pour la prédiction du coefficient global de transfert d'oxygène (Deswal, 2011) et l'estimation en ligne de la concentration de la biomasse dans un procédé de fermentation (Wang, Yu et al., 2006). Un modèle SVR a également été développé pour estimer la qualité de l'eau à travers la prédiction des concentrations COD, NH<sub>4</sub><sup>+</sup>, TN et TP de l'effluent (Fang et al., 2011).

Une variante de SVM appelée **SVM à moindres carrés** (*Least square SVM en anglais, aussi LSSVR*) a été utilisée afin de prédire le taux d'entraînement d'air dans les bassins (Baylar et al., 2009). Cette approche diffère un peu dans la formulation de l'approche SVM classique dans les considérations faites sur les contraintes choisies (Des contraintes de type égalité sont choisies pour résoudre des équations linéaires au lieu des contraintes d'inégalité qui génèrent un problème de programmation quadratique). Cette même approche LSSVM a été appliquée avec succès dans un procédé de traitement des eaux usées pour l'estimation de la concentration du nitrite-nitrate dans l'effluent (S<sub>NO</sub>) (Yang, 2006, Yang, 2005). Dans ces travaux, des connaissances sur le procédé et plus particulièrement sur le processus d'élimination de l'azote ont été utilisées pour choisir les variables d'entrée pour entraîner l'algorithme LSSVR afin d'estimer la concentration S<sub>NO</sub> de l'effluent. Les variables choisies ont été COD, TSS, TKN, TN de l'influent, K<sub>la5</sub> et S<sub>no2</sub>.

La première approche proposée par l'auteur de cette thèse appliquée au procédé de traitement des eaux est basée uniquement sur SVM classification (Laouti et al., 2011b). Cependant, cette étude n'était pas très poussée puisqu'elle demande un effort assez conséquent pour le réglage des paramètres d'apprentissage  $C$  et  $\sigma$  sans garantir l'annulation des fausses alarmes, en plus elle été appliqué sur une variante simpliste du modèle ASM1. Afin d'améliorer ce résultat une deuxième approche originale est proposée qui va être détaillée par la suite de ce chapitre.

## II.4. Stratégie SVR/SVM proposée pour le diagnostic du procédé WWTP

### II.4.1. Schéma SVR/SVM proposé

Dans cette stratégie, la méthode SVR est utilisée pour la modélisation du procédé WWTP due à sa solide base théorique et ses nombreux avantages tels que sa robustesse face aux incertitudes (aspect de généralisation face à de nouvelles données et nouveaux modes de fonctionnement), sa tolérance au bruit, le traitement des valeurs aberrantes et la capacité à capter la non linéarité du système. En deuxième temps, la méthode SVM est utilisée pour la détection et l'isolation de défauts.

L'implémentation de la méthode de diagnostic peut être divisée en deux parties : utilisation de SVR pour la modélisation de certains variables du système et ensuite utilisation de SVM pour le diagnostic (détection et isolation) de défaut en utilisant les résultats de prédiction obtenus par SVR. Ces deux étapes sont précédées par une étape de prétraitement des données et suivies d'une étape de validation.

- Acquisition et raffinement des données
  1. Génération des séquences de variables d'entrées aléatoires sans défauts pour SVR, et sans et avec défaut pour SVM.
  2. Recueil des données entrées/sorties ( $X=[U \ Y]$ ).
  3. Filtrage et normalisation des données.
  4. Définition des groupes des données d'apprentissage et de validation.
- Construction des modèles de prédiction **SVR** pour quelques variables clés (des variables clés dont l'estimation est supposée nécessaire dans le diagnostic de défauts).
  1. Choix des variables/caractéristiques pertinentes (régresseurs  $x$ ). La sélection de variables d'entrée permet de choisir les meilleures variables explicatives, et d'étudier l'intercorrélation entre une variable à prédire (de sortie) et une variable prédictive ou régresseur permettant ainsi de déterminer s'il y a une dépendance entre ces variables. Dans le cas de liaison positif, elle permet de calculer la valeur du retard temporel.
  2. Choix du noyau, et réglage des paramètres  $C$  et  $\sigma$ .

3. Evaluation du modèle (ajustement + analyse des résidus et incertitudes, par exemple par la matrice de la covariance).

- Réalisation de modèles de décision SVM pour chaque défaut
  1. Détermination des caractéristiques descriptives du défaut (choix du vecteur  $x$  pour SVM classification). Elles peuvent contenir des estimations obtenues par SVR.
  2. Réglage des paramètres du modèle SVM (noyau,  $C$ ,  $\sigma$ )
  3. Validation du modèle SVM.
- Mise en œuvre en ligne de la stratégie.

Soit  $N_c$  le nombre de capteurs, et  $N_a$  le nombre d'actionneurs dans le système en question,  $N_d$  le nombre de défauts étudiés, et  $E_m$  les entrées choisies comme prédicteurs ou régresseurs utilisés par l'algorithme SVR pour construire le capteur logiciel (modèle de prédiction). La Figure II.4 montre la stratégie de SVR suivie par SVM pour le diagnostic de défauts.

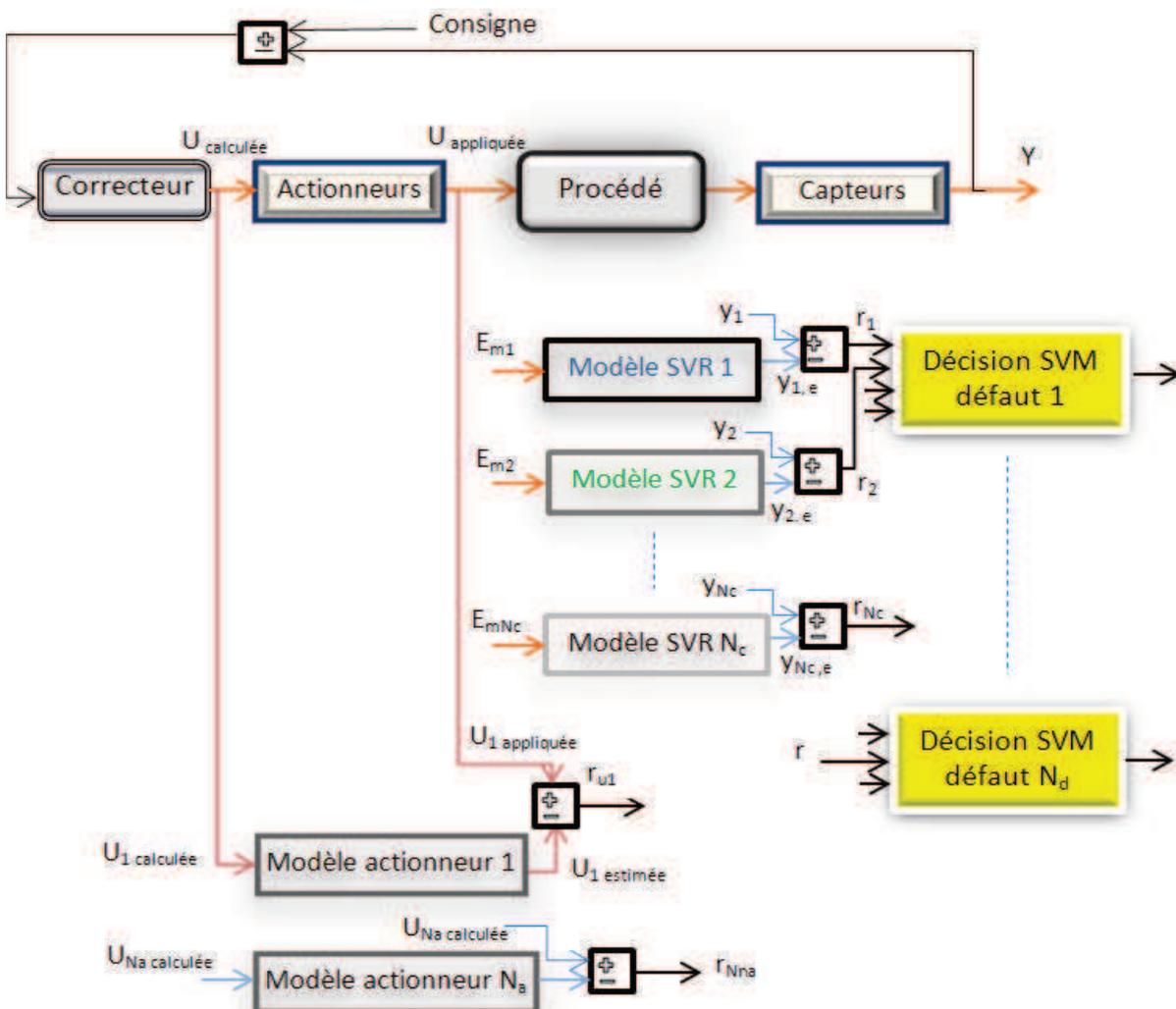


Figure II.4 Stratégie SVR suivie par SVM pour le diagnostic de défauts

L'idée dans cette stratégie est donc de créer des caractéristiques, observations spécifiques ou signaux symptômes du défaut, et ne pas se contenter des entrées/sorties disponibles. Une de ces caractéristiques les plus importantes est le résidu qui est la différence entre un signal mesuré et celui estimé. Par la suite, les caractéristiques sont utilisées comme entrées de SVM classification afin de construire une fonction de décision claire et précise permettant de spécifier ou de prédire la nature de la nouvelle observation. Le fait d'utiliser la même théorie pour la modélisation et le diagnostic est un des points forts de notre stratégie. Le développement d'un estimateur pour prédire les valeurs de différentes variables de sortie Y par SVR est simple et rapide en termes de temps de calcul en ligne puisque il utilise un nombre réduit de données qui sont les supports. Malgré les avantages du SVR, et sa robustesse face aux variations des paramètres et la propriété de flexibilité, l'utilisation des SVR dans la communauté des automaticiens reste en grande partie peu explorée, c'est pour cela, cette thèse apportera à mon avis un plus à la communauté des automaticiens.

#### II.4.2. Récapitulatif des variables d'état/ d'entrée/ et de sortie

Les **variables d'état** sont les différentes concentrations dans les 5 cuves (13 états par cuve) plus la concentration totale des boues dans les 10 couches du clarificateur ( $X_m$ ,  $m=1:10$ ) :

$$C^T = [S_I, S_S, X_I, X_S, X_{B,H}, X_{B,A}, X_P, S_O, S_{NO}, S_{NH}, S_{ND}, X_{ND}, S_{ALK}, X_1, X_2, \dots, X_{10}]$$

Les entrées U ou variables manipulées (ou commande) du système sont :

$K_{la5}$  : le coefficient de transfert d'oxygène pour la 5<sup>ème</sup> cuve

$S_{no2}$  : azote sous forme de nitrates-nitrites dans la 2<sup>ème</sup> cuve.

L'entrée de perturbation est l'influent ( $Q_{in}, C_{in}$ )

Les sorties mesurées (Y) du système considérées dans ce travail sont :  $Q_{in}, S_{NH,in}, COD_{in}, TKN_{in}, TSS_2, COD_2, BOD_2, K_{la5}, S_{no2}, Q_{intr}, S_{o5}, S_{nhe}, S_{noe}, TSS_4, TSS_5, S_{nhe}, TSS_u, COD_u$

Les sorties régulées sont :  $y_1=S_{no2}$  et  $y_2=S_{o5}$

On note que les variables d'état ne sont pas toutes mesurables.

### II.4.3. Les variables à superviser

Plusieurs variables doivent être supervisées dans les procédés de traitement des eaux usées afin de s'assurer du bon fonctionnement et de la sécurité des procédés de traitement des eaux et rendre ainsi une eau propre en sortie (ex. l'activité des microorganismes, la température et l'influent). Toute erreur dans les capteurs de mesures ou les actionneurs doit également être détectée rapidement afin d'éviter la propagation d'erreurs et la détérioration de matériels et la diminution de la qualité du rendement.

On peut dénombrer les défauts qui peuvent avoir lieu dans ce procédé (ou les paramètres à superviser) comme suit :

1. **La concentration d'oxygène** : L'oxygène représente une condition fondamentale de la vie des microorganismes afin de permettre de maintenir une concentration élevée de la biomasse dans le réacteur biologique. Cela n'est pas toujours facile puisque ce système subit constamment des perturbations d'entrée variables qui peuvent empoisonner les microorganismes quand l'eau contient certains métaux. La provision d'oxygène se fait par aération des bassins. Le maintien d'une aération adéquate permet d'améliorer l'activité biologique en encourageant le contact entre les microorganismes et la matière organique, ce qui permet d'optimiser la consommation d'énergie (Jianxu, 2011). Dans ce travail, on a choisi de superviser la concentration d'oxygène dissous dans le 5<sup>ème</sup> compartiment car c'est une variable importante dans le bon fonctionnement du procédé WWTP.
2. **Le coefficient du transfert d'oxygène  $K_{La5}$**  : des défauts peuvent surgir dans cet actionneur qui permet de contrôler la concentration  $S_{O5}$  de l'oxygène dissous dans le 5<sup>ème</sup> compartiment à la valeur désirée de  $2 \text{ g.m}^{-3}$ . Le suivi de ce coefficient est donc aussi important que les concentrations d'oxygène dissous.
3. **La concentration d'azote sous forme de nitrites et de nitrates  $S_{no2}$**  : ce capteur qui nous informe sur le bon déroulement de l'opération de dénitrification (dégradation de l'azote) peut être affecté par un dysfonctionnement.
4. **Le Débit de recyclage interne  $Q_a$**  : important pour la suppression des polluants azotés par la régulation de  $S_{no2}$  à  $1 \text{ g.m}^{-3}$ , cet actionneur aussi peut subir des pannes.
5. **Choc de toxicité** : ce défaut est dû à la présence de matières toxiques dans l'eau comme les pesticides, ce qui va influencer négativement la croissance des organismes hétérotrophes ( $\mu_{h,max}$ ) qui sont responsables de l'élimination de la pollution.

6. **Inhibition** : ce défaut est plus violent que le premier, produit par les déchets hospitaliers ou métallurgiques. Il va non seulement réduire la croissance des organismes hétérotrophes ( $\mu_{h\max}$ ), mais aussi accélérer la mort de ces organismes ( $b_h$ ). Inhibition de la nitrification et toxicité sont causées par les mêmes composés (l'inhibition produit des effets à des concentrations faibles tandis que la toxicité se produit à des concentrations plus élevées).
7. **Gonflement ou prise de masse des boues** (*bulking en anglais*) : ce phénomène est dû à la croissance de microorganismes filamenteux dans les boues activées, ce qui provoque l'impossibilité de décantation dans le décanteur. Il coïncide avec les périodes d'été : haute température et faible débit à 266 j et 567 j. Cet événement est représenté dans le modèle par les paramètres de décantation de la fonction double-exponentielle et plus précisément la vitesse de sédimentation de 250 à 200 m.d<sup>-1</sup>.

Les défauts procédé (choc de toxicité, inhibition, gonflement de la masse des boues) ne sont pas considérés dans ce travail, néanmoins il est envisagé de les traiter dans un futur proche.

D'autres capteurs à savoir TSS<sub>e</sub> et COD<sub>e</sub> sont d'une importance capitale puisqu'ils permettent de caractériser les performances de traitement de l'eau de ce système.

En fonction de cette analyse, les **défauts considérés (à superviser) dans cette thèse** sont les défauts de **capteurs** de mesures de :

1. **TSS<sub>e</sub>** : les matières en suspensions de l'effluent en sortie
2. **S<sub>no2</sub>** : la concentration de l'azote sous forme de nitrites-nitrates dans le 2<sup>ème</sup> compartiment.
3. **S<sub>o5</sub>** : la concentration d'oxygène dissous dans le 2<sup>ème</sup> compartiment

On considérera des défauts de type : **offset, dérive, défaut fixe (blocage de mesure) et arrêt complet** pour ces différents capteurs.

Ainsi que des défauts d'actionneurs

4. **Q<sub>a</sub> ou Q<sub>intr</sub>** : débit de recyclage interne
5. **K<sub>la5</sub>** : le coefficient du transfert d'oxygène K<sub>la5</sub> dans le 5<sup>ème</sup> compartiment

Avec les types de défauts suivants : **offset, retard, blocage de l'actionneur et arrêt ou fermeture complète**.

**Les variables à modéliser :**

En fonction de ces défauts, nous avons choisi d'estimer (**modéliser**) les **variables** suivantes qui sont susceptibles d'apporter plus d'information concernant certains défauts :

$S_{no2}$ ,  $S_{no5}$ ,  $COD_{eff}$  et  $TSS_e$ .

## II.5. Implémentation de SVR pour la modélisation

### II.5.1. Génération de données pour l'apprentissage et la validation

Le modèle de référence BSM1 est programmé et simulé sous Matlab®. L'état quasi stationnaire est atteint dans ce modèle au bout de 100 j (Gernaey et al., 2006). Les données d'apprentissage et de test sont obtenues en injectant un ensemble de données d'influent de 609 j récupéré sur le site du benchmark BSM1 <http://www.benchmarkwwtp.org/>. Les multiples données (entrées/ sorties/ variables de commande) qui caractérisent le procédé sont récupérées à partir du programme avec une période d'échantillonnage de 15 min et décomposées en deux parties. Une première partie de 245 j est utilisée pour l'apprentissage et la construction des différents modèles. La partie restante de 364 j est utilisée pour le test des modèles de prédiction.

Plusieurs paramètres peuvent affecter sérieusement les résultats de l'apprentissage (Lauer and Bloch, 2006). Par exemple, le fonctionnement en commande en boucle ouverte ou en **boucle fermée** ne donne pas les mêmes résultats. On a fait le choix de recueillir des données qui ressembleront aux données en ligne, et donc le fonctionnement en boucle fermée est maintenu. Les tests avec apprentissage du modèle SVR en boucle ouverte n'ont pas donné de bons résultats, ce qui s'explique par le fait qu'une erreur faible dans le modèle en boucle ouverte peut devenir assez conséquente en boucle fermée par l'effet de bouclage.

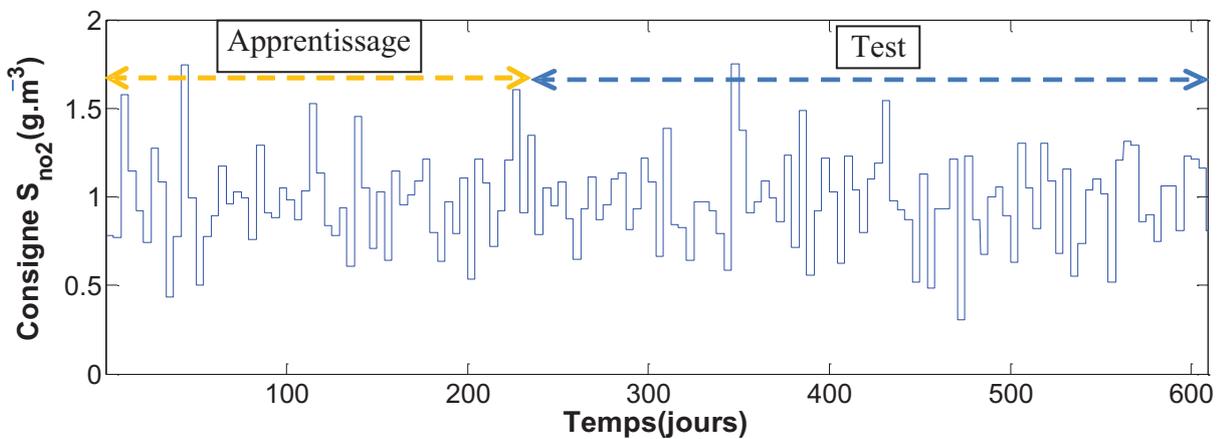
On se pose également la question de savoir s'il faut recueillir des données à l'**état stationnaire** ou transitoire. Le risque en prenant des données stationnaires est que les dérives de cet état stationnaire peuvent être interprétées comme des défauts. En même temps, le fonctionnement transitoire est plus difficile à modéliser (nécessitant plus de données).

Il convient également d'établir des règles de gestion des valeurs manquantes et aberrantes, des perturbations mesurées  $S_{NH,in}$  et  $Q_{in}$ , de normalisation, et le choix des variables pertinentes pour la régression (régresseurs).

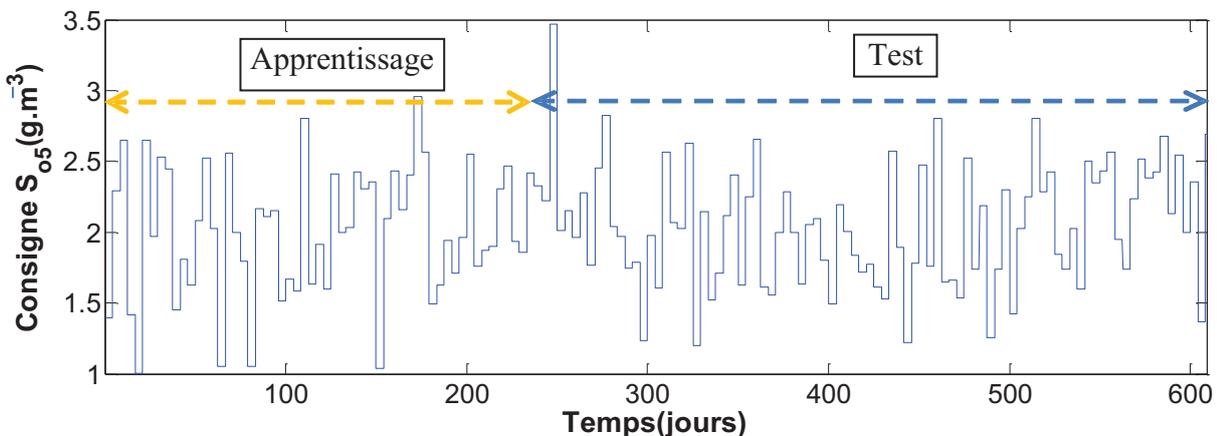
Une dernière condition afin d'obtenir une relation entrée/sortie assez robuste, consiste à choisir **une entrée suffisamment riche en information**. La fréquence et l'amplitude de cette entrée sont à choisir de manière à exciter le système suffisamment pour capter la relation entrée/sortie

(ce qui revient à choisir l'entrée de manière à rendre le système observable). En effet, l'observabilité des systèmes non linéaires dépend généralement de l'entrée du système. L'entrée doit permettre d'avoir un bon suivi de trajectoire et en même temps d'être continuellement excitante (ou excitation persistante) pour obtenir une bonne identification qui permet d'extraire les dynamiques du système.

Les entrées dans ce système sont la concentration d'oxygène dissous dans le 5<sup>ème</sup> compartiment et la concentration de nitrites/nitrates dans le 2<sup>ème</sup> compartiment. Ces concentrations sont considérées comme entrée puisqu'elles sont les variables à réguler. La Figure II.5 montre la concentration de nitrites/nitrates dans la 2<sup>ème</sup> cuve ( $S_{no2}$ ) pour les périodes d'apprentissage et de test (et la Figure II.6 montre celle de  $S_{o5}$ ).



**Figure II.5 Concentration de nitrites/nitrates dans la 2<sup>ème</sup> cuve utilisée pour l'apprentissage et le test du modèle SVR**



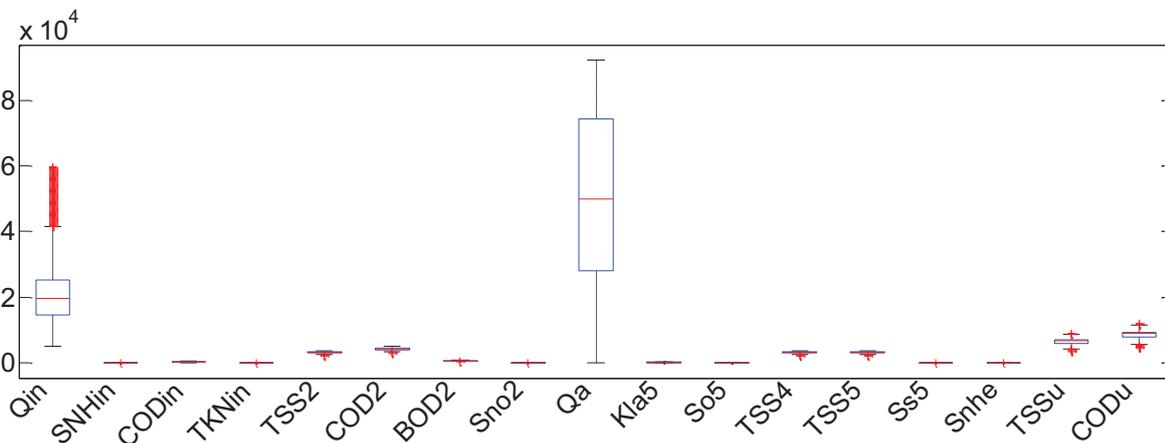
**Figure II.6 Concentration d'oxygène dissous dans la 5<sup>ème</sup> cuve utilisée pour l'apprentissage et le test du modèle SVR**

Ces mêmes données sont également utilisées pour l'implémentation de SVM pour le diagnostic de défauts, à la différence que des situations défectueuses sont également considérées en plus des situations saines. Chaque défaut (de capteur, d'actionneur ou de procédé) a lieu sur un intervalle de temps bien distinct des intervalles des autres défauts.

## II.5.2. Prétraitement des données (normalisation, filtrage)

L'obtention d'un bon modèle dépend fortement de la qualité et la représentabilité de la base de données utilisée pour l'apprentissage. Premièrement, **la normalisation** des données (Ogasawara et al., 2010) joue un rôle important dans la régression (modélisation) et dans le diagnostic de défauts car elle permet de bien conditionner les matrices contenant les signaux du système avec une distribution uniforme.

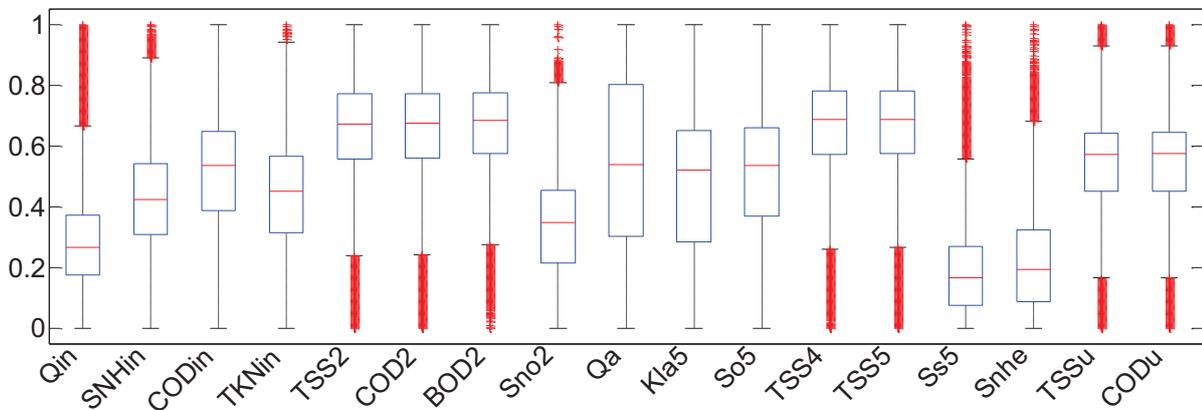
Les données mises à l'échelle (normalisées) peuvent être analysées statistiquement plus facilement. La Figure II.7 montre un exemple de génération d'une boîte à moustache dans le cas de données non normées. Les données sont résumées en 3 composantes : une ligne centrale indiquant la tendance centrale, une case pour indiquer la variabilité autour de la tendance centrale et les moustaches montrant la portée. Après l'essai des différentes méthodes de normalisation mentionnées en chapitre I, le choix a été porté sur la normalisation min/max qui a permis d'obtenir les meilleurs résultats (distribution uniforme) comme le montre la Figure II.8 .



**Figure II.7 Boîte à moustache des variables du procédé WWTP non normées**

Les données ont aussi parfois besoin de **filtrage** afin de diminuer l'impact du bruit et de pouvoir décrypter les tendances. Le lissage peut être fait par exemple en utilisant une moyenne mobile afin d'éliminer les pics au courts du temps. La suppression des valeurs aberrantes peut être faite par la suppression des mesures qui sont en dehors de  $\pm 3\sigma$  par exemple et reconstituer ou

supprimer les valeurs manquantes. SVR peut être utilisée pour estimer les valeurs manquantes comme c'est le cas pour les données d'expression géniques (Wang et al., 2006). Dans ce travail, un filtre passe bas est utilisé pour diminuer le bruit.



**Figure II.8 Boîte à moustache des variables du procédé WWTP normalisées avec la méthode min/max**

### II.5.3. Analyse de corrélation/Information mutuelle

Pour l'apprentissage SVM ou SVR, il faut définir le **vecteur d'entrée x**. Le choix de ce vecteur peut provenir des connaissances a priori sur le procédé. Ainsi, on va chercher la solution dans une direction bien précise ou dans un domaine plus restreint ce qui permet d'obtenir un bon modèle de régression. Ces connaissances peuvent provenir du modèle fondamental (modèle mécanistique, hydraulique, chimique, physico-chimique, aéronautique, biochimique, mécanique, hydrodynamique, etc.) qui gouvernent le procédé.

Dans ce travail, le choix du vecteur  $x$  est fait simplement à partir de cette connaissance a priori pour SVM classification alors que pour SVM régression (SVR), le vecteur des régresseurs  $x$  est déterminé en réalisant des analyses de corrélation et d'information mutuelle (décrites dans le chapitre I).

On cherche à modéliser les variables  $COD_{eff}$ ,  $TSS_e$ ,  $S_{o5}$  et  $S_{no2}$ . Il convient donc de trouver les relations entre ces variables et chaque élément dans le vecteur de mesures et d'entrées du système.

L'étude de **l'inter-corrélation** est une façon qui permet d'avoir une idée sur la dépendance et l'interaction entre deux variables. Par contre l'autocorrélation détermine la dépendance temporelle entre les observations d'un même signal du procédé. Cela peut aider dans le choix des

régresseurs. Dans le cas du procédé WWTP, l'étude de l'inter-corrélation entre les concentrations d'influents en entrée et les concentrations d'effluents de sortie permet de quantifier la relation entre entrées et sorties et elle donne une information sur le décalage temporel (le temps que prend l'entrée pour affecter une sortie donnée). Une forte corrélation à un décalage spécifique peut indiquer un temps de retard dans le procédé. Ce genre d'information peut être utile pour indiquer un dysfonctionnement dans le procédé. Quand le procédé WWTP fonctionne normalement, l'influent subit des transformations qui peuvent s'observer sur l'effluent avec un certain retard (effet tampon) dû aux grands volumes de liquide qui s'entassent dans le réacteur biologique et clarificateur, ce qui donne des concentrations d'effluent un peu lisses par rapport aux dynamiques des influents.

D'après l'analyse de corrélation, il ressort que les variables qui influent le plus sur  $COD_{eff}$  sont celles données dans le tableau II.4. Les coefficients prennent des valeurs entre 0 et 1 dans la méthode inter-corrélation. Le tableau présente également le décalage temporel qui donne la corrélation maximale entre les différentes variables et  $COD_{eff}$ . On note que pour un nombre de variables le retard est égal à zéro par rapport à  $COD_{eff}$ , ce qui signifie que l'effet de l'entrée ne se fait pas significativement sentir sur la sortie ( $COD_{eff}$ ) dans le temps de simulation.

**Tableau II.4 Corrélation des variables avec  $COD_{eff}$**

Variables	$Q_{in}$	$K_{la5}$	$TKN_{in}$	$TN_{in}$	$BOD_{in}$	$COD_{in}$	$S_{NH,in}$	$TSS_{in}$
Corrélation avec $COD_{eff}$	0.5	0.47	0.44	0.44	0.41	0.4	0.39	0.36
Retard (échantillons)	2	0	96	96	0	0	96	0

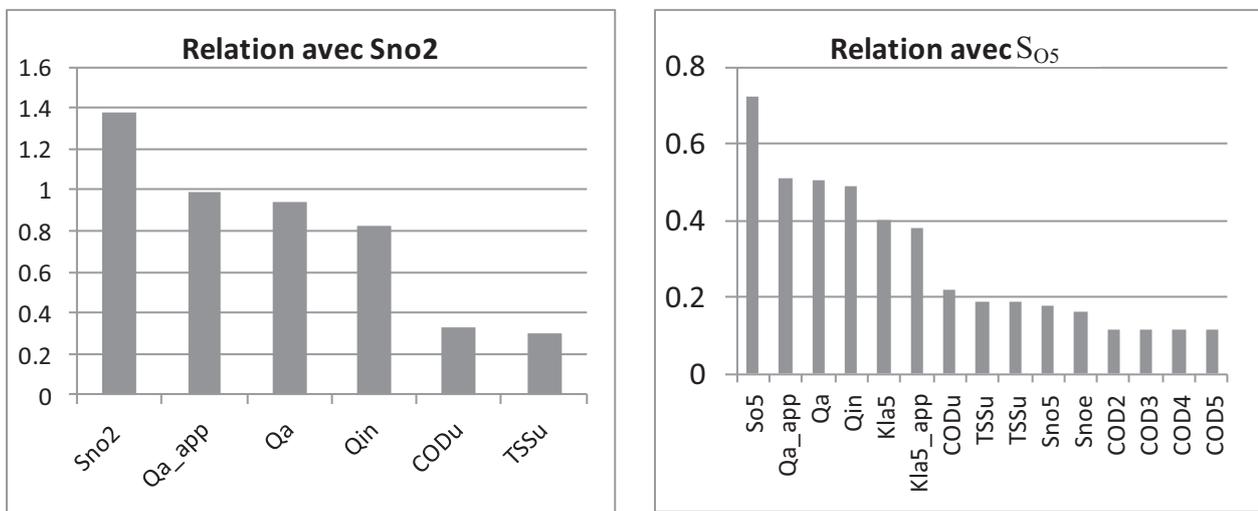
Pour la variable  $S_{no2}$  elle est affectée le plus par :  $TSS_1$ ,  $COD_1$ ,  $BOD_1$ ,  $TN_1$ ,  $TSS_2$ ,  $COD_2$ ,  $BOD_2$ ,  $TSS_3$ ,  $COD_3$ ,  $BOD_3$ . Et les corrélations maximales avec  $TSS_e$ ,  $S_{no2}$  et  $S_{no5}$  sont données pour un retard de :

2 échantillons entre  $TSS_e$  et  $Q_{in}$ . 16 échantillons entre  $S_{no2}$  et  $K_{la5}$ . 22 échantillons entre  $S_{no2}$  et  $S_{NH,in}$ . 12 échantillons entre  $S_{o5}$  et  $K_{la5}$ . 20 échantillons entre  $S_{o5}$  et  $S_{NH,in}$

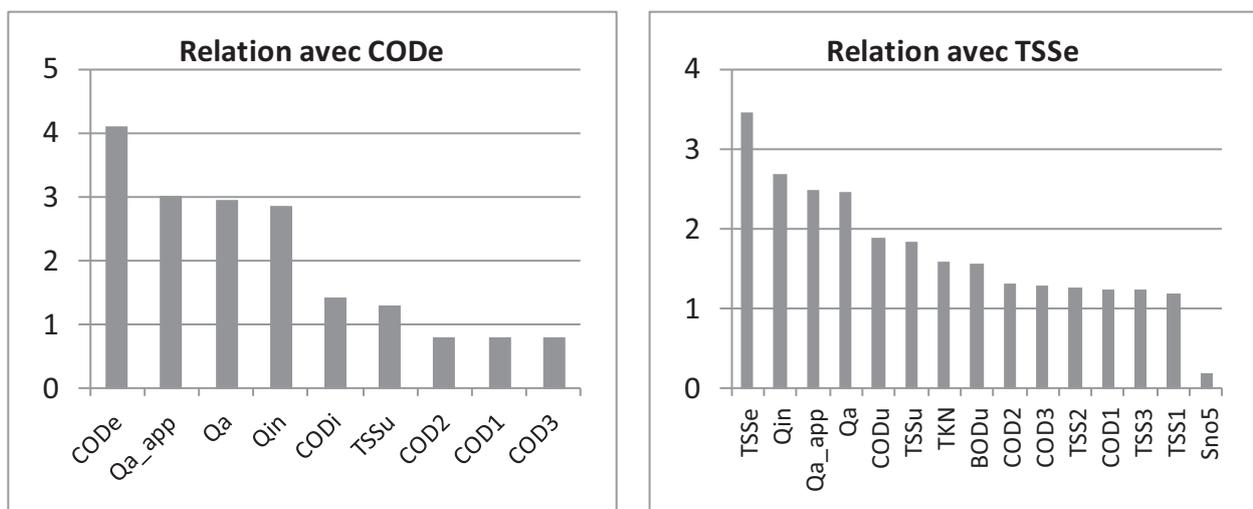
On remarque que l'analyse de l'inter-corrélation n'a pas permis de trouver une dépendance linéaire entre  $S_{no2}$  et  $Q_a$ , alors que d'après la connaissance du procédé on sait qu'il existe une relation entre ces deux variables puisque  $Q_a$  est la variable manipulée permettant la régulation de  $S_{no2}$ . Ceci est dû au fait que l'étude de la corrélation et de l'inter-corrélation informe uniquement sur les relations de dépendance linéaire entre les variables. Cependant, le procédé de traitement des eaux usées est un système complexe et fortement non linéaire. Afin de réaliser une

investigation plus poussée pour l'extraction et la sélection des meilleures caractéristiques (régresseurs, prédicteurs ou entrées susceptibles d'être utilisées pour construire la modèle SVR), on propose de réaliser l'étude de la dépendance à travers **l'information mutuelle**.

La Figure II.9 montre la relation entre  $S_{no2}$  et quelques variables mesurées, obtenue par l'analyse d'information mutuelle. On remarque que contrairement à l'analyse de l'inter-corrélation, l'analyse de l'information mutuelle montre une dépendance entre les variables  $S_{no2}$  et  $Q_a$ . La Figure II.9 et la Figure II.10 montrent la relation entre  $S_{no2}$ ,  $S_{o5}$ ,  $COD_e$  et  $TSS_e$  et les autres variables du procédé. Plus la valeur de relation est grande plus la liaison est forte. Les variables les plus influentes seront utilisées dans le schéma de prédiction SVR de chaque variable considérée.



**Figure II.9** Analyse des relations entre variables avec l'information mutuelle (pour  $S_{no2}$  et  $S_{no5}$ )



**Figure II.10** Analyse des relations entre variables avec l'information mutuelle (pour  $COD_e$  et  $TSS_e$ )

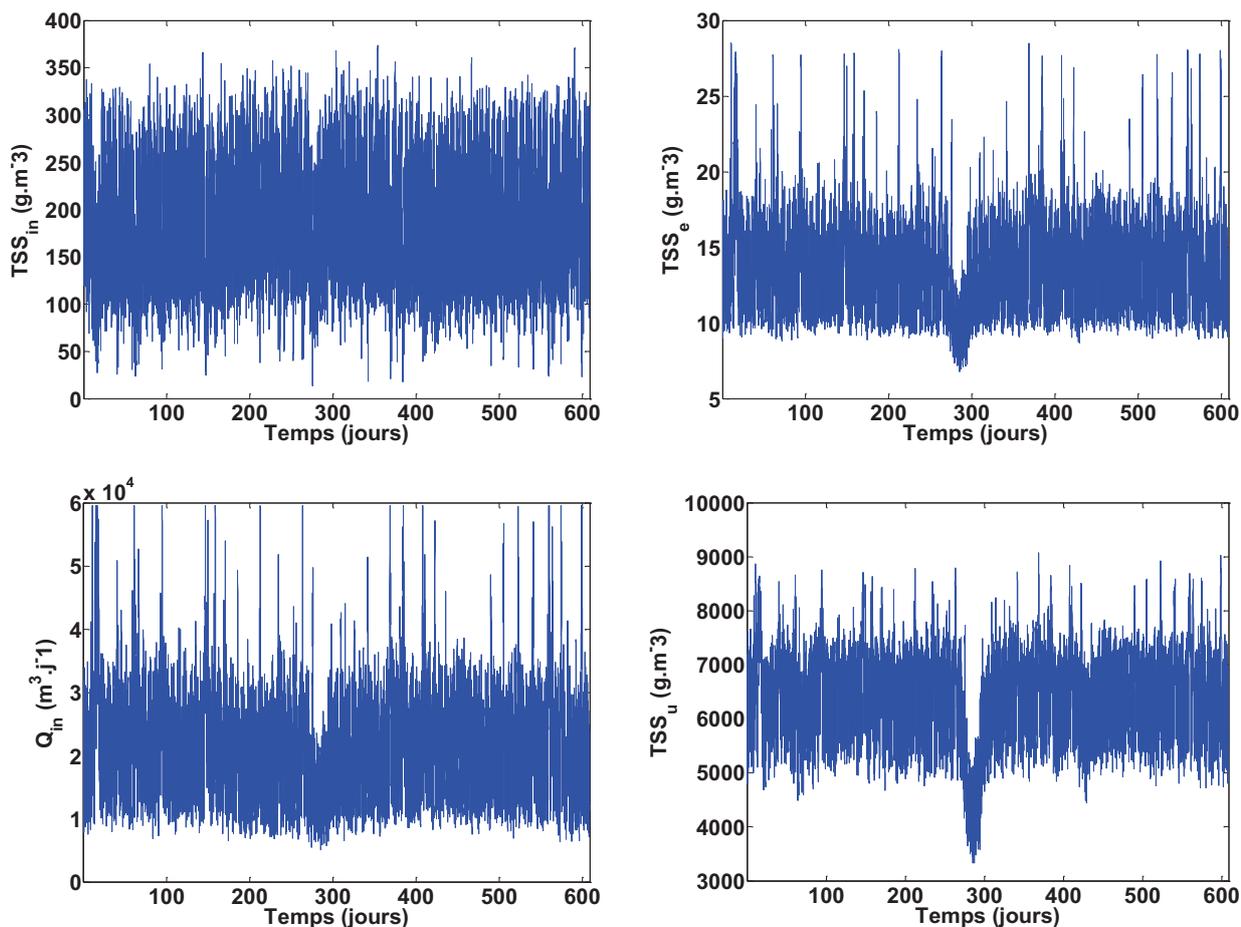
#### II.5.4. Résultats de la modélisation SVR

En fonction de l'analyse de corrélation et d'information mutuelle, le vecteur d'entrée  $E_m$  du modèle SVR a été choisi comme montré par la suite pour les 4 variables à modéliser :

Les séries temporelles des différentes variables du procédé WWTP sont visualisées dans les figures suivantes. Ces figures permettent de visualiser les dynamiques et la variabilité assez conséquente qui se trouvent dans le procédé WWTP, et montrent l'étendue temporelle importante utilisée dans ce travail qui est de 609 jours avec une période entre deux observations égale à 15 minutes ce qui fait un total de 58464 échantillons. Cela permet de tester des situations multiples rencontrées dans les différentes saisons (été, hiver, printemps,...). L'approche basée sur les SVMs utilisée dans ce chapitre permet de mieux évaluer les performances d'estimation et de diagnostic de défauts dans le procédé WWTP par rapport à des méthodes testées juste sur une faible période de temps comme (Wang and Shi, 2010), (Lee, Qin, et al., 2006) avec 1400 échantillons, (Baggiani and Marsili-Libelli, 2009) testé sur 32 jours, (Zumoffen and Basualdo, 2008b) testés sur 14 jours, etc..

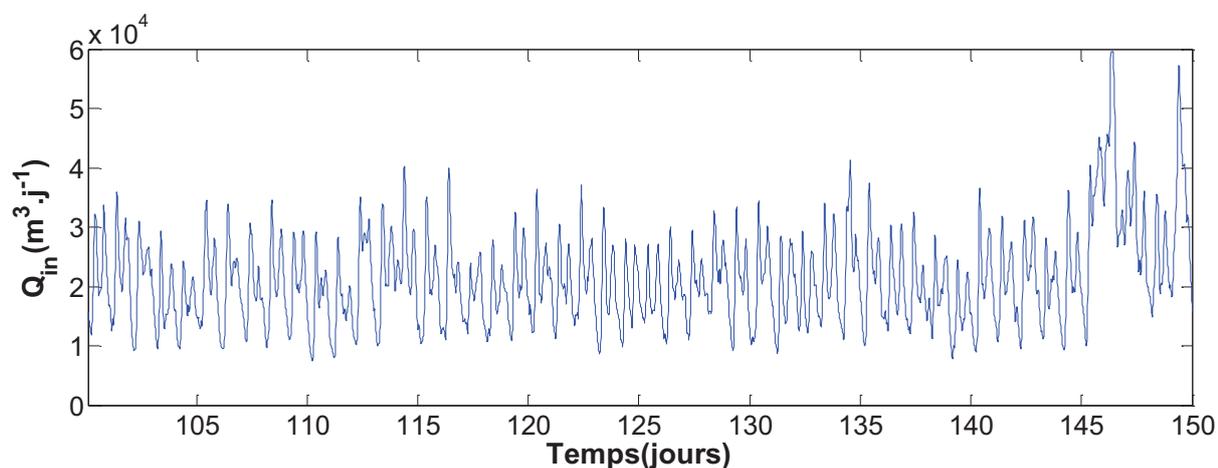
Plusieurs travaux utilisent un modèle simple du procédé WWTP. Par contre, le modèle de référence BSM1 pour les systèmes de traitement des eaux usées utilisé dans cette étude est très proche de la réalité et inclut beaucoup de variables ce qui rend la tâche de surveillance et de diagnostic un vrai défi à relever par rapport à d'autres articles.

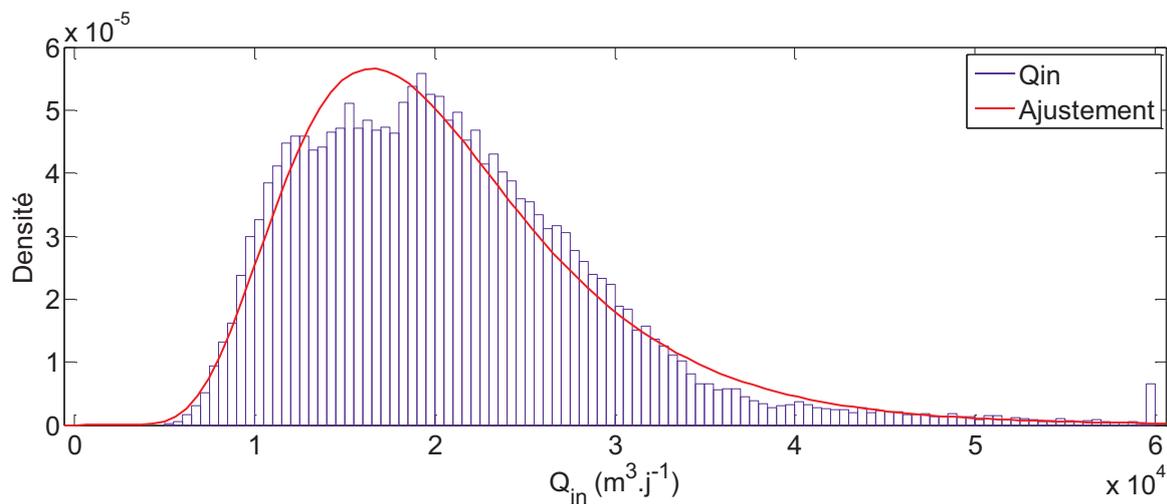
La Figure II.11 illustre le rôle (traitement) du procédé WWTP. Ainsi,  $TSS_{in}$  qui constitue les matières en suspension à l'entrée ayant une grande valeur va être traité pour donner à la sortie du procédé dans la couche supérieure du clarificateur  $TSS_e$  effluent une faible valeur de concentration par rapport à l'entrée, tandis que dans la couche basse du clarificateur l'effet de sédimentation est bien clair avec la valeur de  $TSS_u$ . La courbe de débit à l'entrée illustre bien la variabilité assez importante de débit qui varie suivant les saisons (été, hiver)



**Figure II.11 Evolution temporelle des variables d'entrée/sortie des matières en suspension**

Le zoom ci-dessous sur la figure du débit d'entrée  $Q_{in}$  du procédé WWTP montre que cette perturbation est très oscillatoire.  $Q_{in}$  a une dynamique avec plusieurs échelles de temps. Ceci est dû au fait que le débit des eaux usées diminue la nuit et pendant la période estivale mais augmente en journée et pendant la période hivernale avec l'arrivée d'eau de pluie.

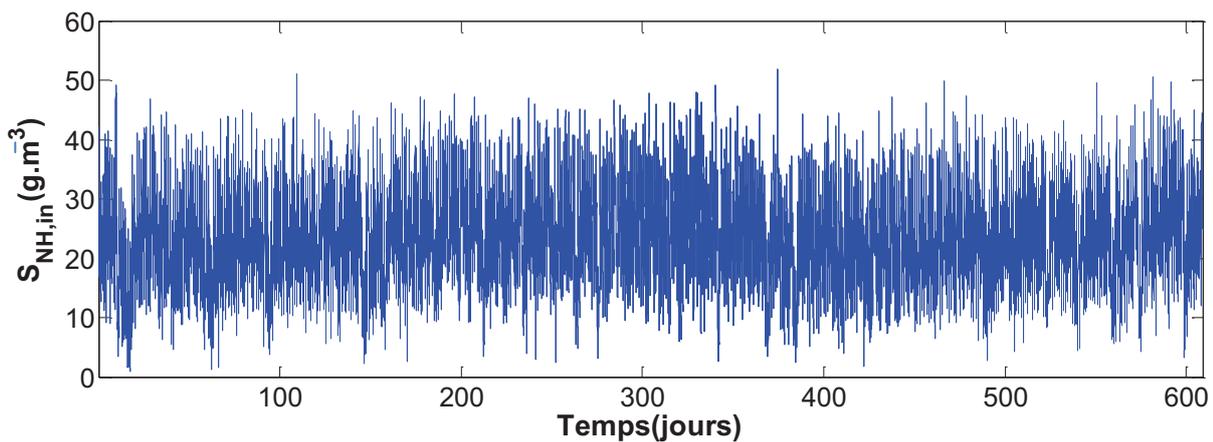


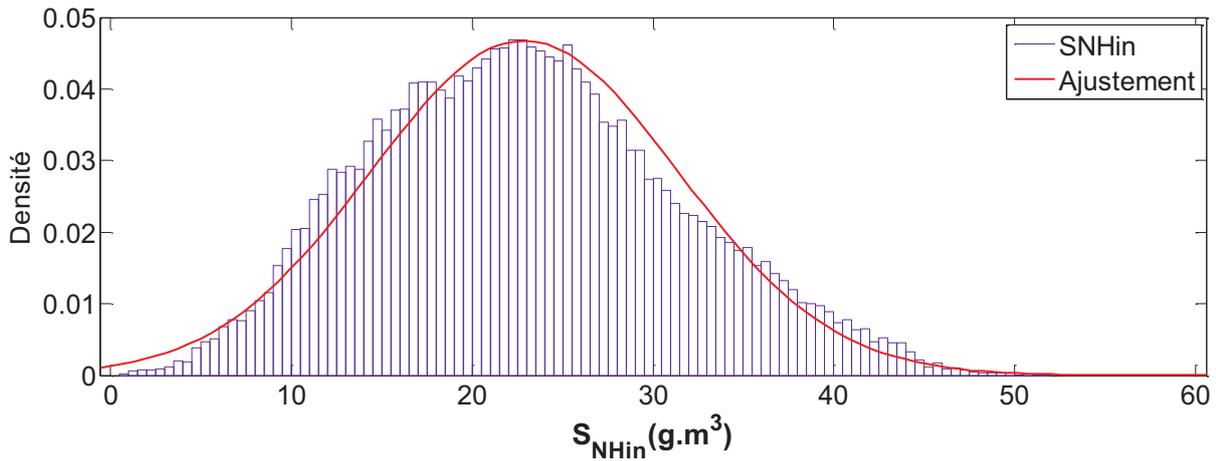


**Figure II.12 Densité de probabilité  $Q_{in}$**

La distribution log normal représente bien les données de  $Q_{in}$  comme illustré dans la Figure II.12. L'étude de cette distribution montre l'étendue de la variation de cette variable. Cette distribution peut être utilisée par la suite pour la génération de séquences aléatoires du  $Q_{in}$  avec la méthode de Monte Carlo.

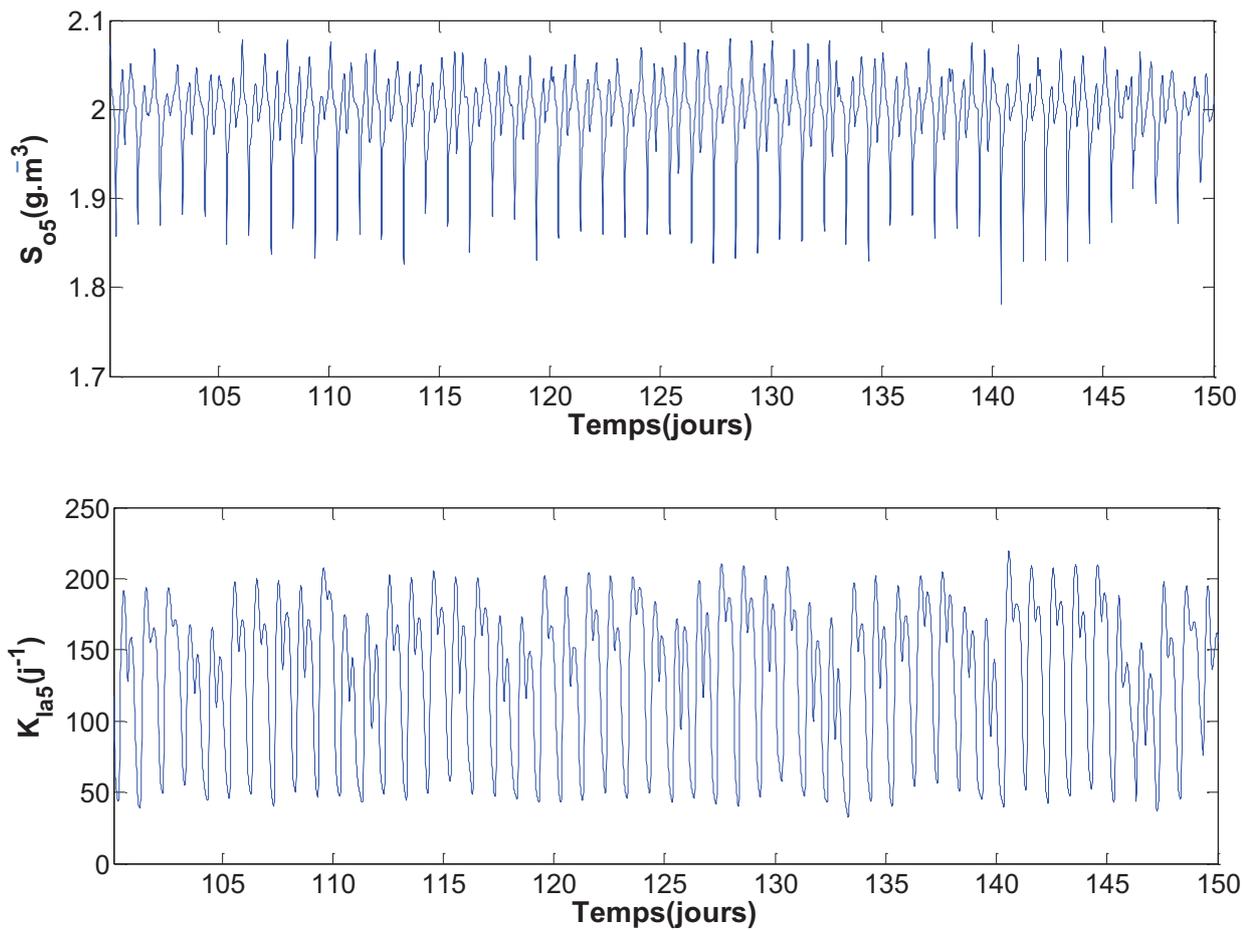
L'interprétation de ces figures de séries temporelles n'est pas tellement facile. Cependant, quelques remarques peuvent être tirées non seulement sur la variabilité, mais aussi sur les pics dans les périodes de (jour, hiver) pour les variables, mais aussi des minimas dans les périodes (été, nuit).





**Figure II.13 Concentration  $S_{NH,in}$  et sa densité de probabilité**

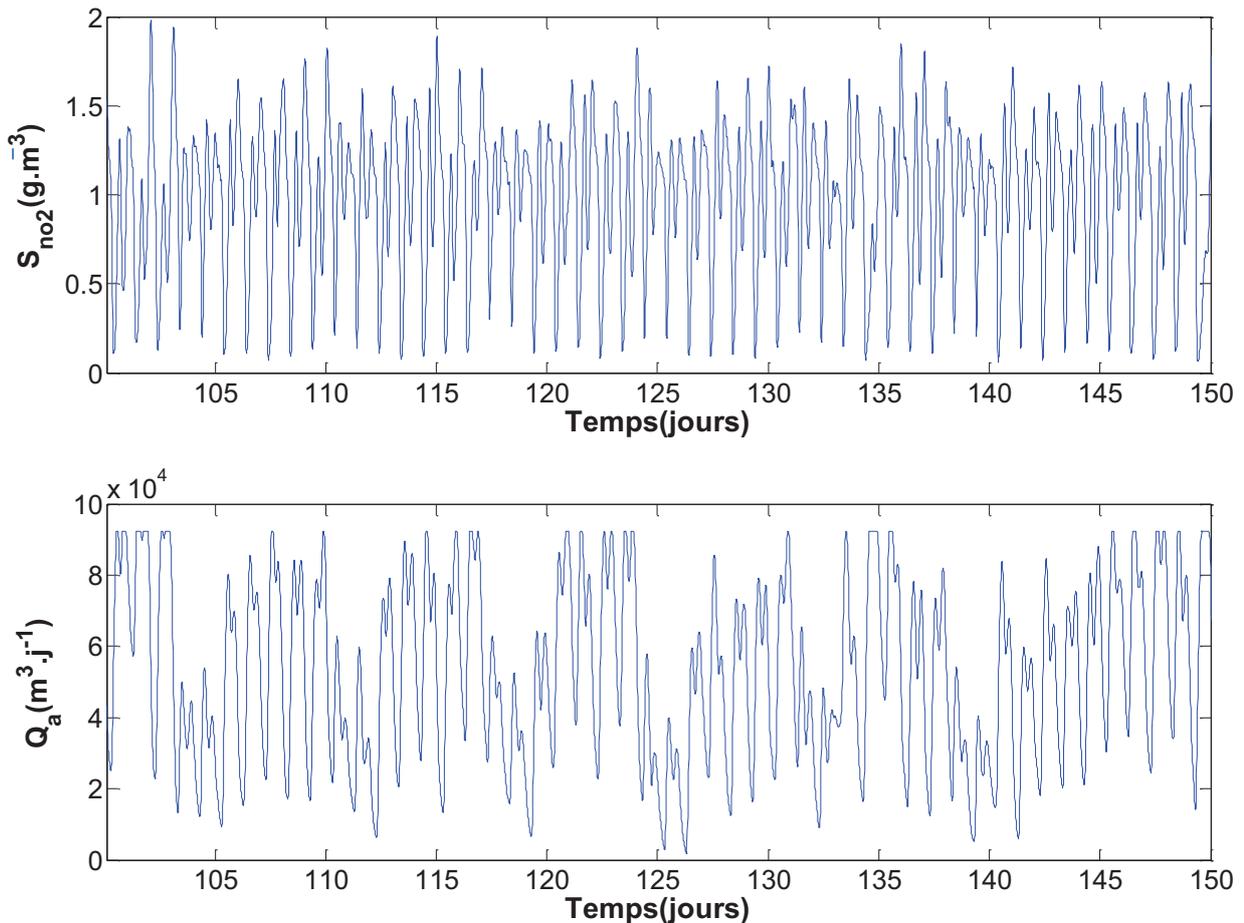
La distribution normale représente bien les données de  $S_{NH,in}$ . Les variables  $S_{no2}$  et  $S_{o5}$  sont des variables à réguler choisies à cause de leur importance dans le bon fonctionnement du procédé WWTP.



**Figure II.14 L'évolution de la variable  $S_{o5}$  par rapport à  $K_{la5}$**

L'interprétation des figures ci-dessus n'est pas facile, dû à la variabilité du procédé et à la grande période d'évaluation. Un zoom (Figure II.14) montre bien que la variable  $K_{La5}$  permet de bien réguler l'oxygène dissous  $S_{O_2}$  à la valeur  $2 \text{ g.m}^{-3}$  (sa valeur varie entre 1.78 et 2.1).

Un zoom sur le graphe de  $S_{NO_2}$  permet de mieux voir les dynamiques sur cette variable (Figure II.15), et la difficulté liée à la régulation de cette variable puisque le but est de stabiliser la valeur de la concentration  $S_{NO_2}$  à  $1 \text{ g.m}^{-3}$ .



**Figure II.15 L'évolution de la variable  $S_{NO_2}$  par rapport à  $Q_a$**

Au départ, j'ai effectué plusieurs essais afin de tester la possibilité de modéliser l'entrée  $S_{NH,in}$  et les entrées de commande  $K_{La5}$  et  $Q_a$  à partir des sorties. Cela n'a pas donné de bons résultats. On peut expliquer ceci par le fait que  $S_{NH,in}$  est une perturbation de nature aléatoire.

D'autres tentatives sur le choix automatique des caractéristiques d'entrée de SVM par l'utilisation de l'ACP n'ont pas donné les résultats attendus. Cela peut se comprendre puisque l'ACP cherche seulement à choisir de nouvelles composantes qui maximisent la récupération d'information sur les données de départ sans se soucier de la relation de ces derniers avec un défaut donné. L'approche PLS peut être vue comme un bon candidat pour le choix des

caractéristiques grâce à sa capacité de mise en relation entre deux quantités (entrée/sortie). J'ai choisi la matrice de données comme entrée et un défaut donné comme sortie. PLS classique linéaire n'a pas permis de modéliser cette relation non linéaire entre le défaut et les données du procédé. Par la suite, PLS à noyau a été utilisée et a permis de capter la relation non linéaire. Cependant, les résultats obtenus étaient loin d'être satisfaisants, dus au nombre élevé de fausses alarmes et détection manquées.

Des études de Monte Carlo peuvent-être envisagées dans un futur proche afin de mieux caractériser les incertitudes avec des générations aléatoires de différentes variables d'entrée comme  $S_{NH,in}$ ,  $Q_{in}$  utilisant leurs distributions respectives, changement des valeurs nominales des paramètres du modèle ainsi que des changements sur le bruit.

Les vecteurs d'entrées des différents modèles SVR sont donnés comme suit:

### **Le vecteur d'entrée pour modéliser $S_{no2}$ par SVR**

L'incorporation de décalages temporels des variables d'entrées et sorties a permis d'améliorer la qualité du modèle de prédiction.

$$E_{m1} = (TSS_2, TSS_4, COD_2, BOD_2, S_{NH,in}, K_{la5}, Q_a, Q_a(t-1), Q_a(t-2), TSS_u, COD_u, S_{no2}(t-12.T_e))^T$$

### **Le vecteur d'entrée pour modéliser $S_{o5}$ par SVR**

$$E_{m2} = (Q_{in}, S_{NH,in}, S_{no2}, S_{no2}(t-1), K_{la5}, K_{la5}(t-1), TSS_5, S_{NH5}, S_{NH_e}, S_{noe}, S_{o5}(t-50.T_e))^T$$

### **Le vecteur d'entrée pour modéliser $TSS_e$ par SVR**

$$E_{m3} = (Q_{in}, Q_{in}(t-1), Q_{in}(t-2), S_{NH,in}, S_{NH,in}(t-1), Q_a)^T$$

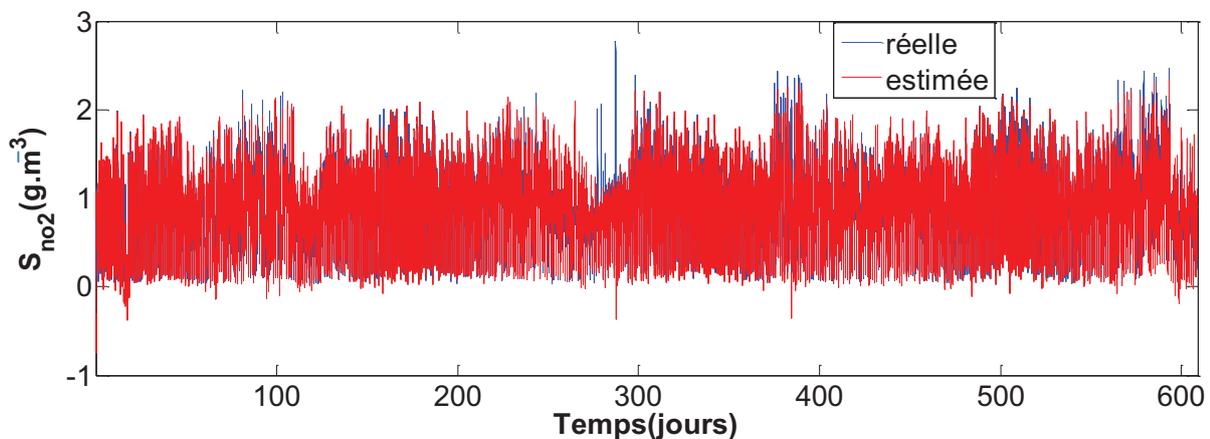
Après une multitude d'essais, le tableau II.5 suivant donne les valeurs des différents indices de performances concernant les meilleurs modèles de prédiction construit par SVR. Ces performances démontrent bien l'efficacité de la méthode de modélisation proposée. Deux modèles sont construits pour estimer  $S_{no2}$ . Le 2<sup>ème</sup> modèle est plus performant que le premier car son coefficient de détermination  $R^2$  qui détermine la qualité de l'ajustement du modèle est plus élevé ( $R^2$  doit être le plus proche de 1), ses erreurs (MSE, MAE, NMSE, RMSE, NRMSE, RSSE) sont plus petites que celles du premier modèle  $S_{no2}$ .

**Tableau II.5 Performances des modèles SVR**

Modèle	Modèle1 Sno2	Modèle2 Sno2	Modèle2 TSS <sub>eff</sub>	Modèle1 So5	Modèle3 So5	Modèle COD <sub>eff</sub>
<b>R<sup>2</sup></b>	0.9044	0.9270	0.9982	0.8492	0.8626	0.8450
<b>MSE</b>	0.0206	0.0157	0.0171	0.0176	0.0160	2.8139
<b>AIC</b>	-10 <sup>5</sup>	-1.1 10 <sup>5</sup>	-1.1 10 <sup>5</sup>	-10 <sup>5</sup>	-1.1 10 <sup>5</sup>	1.85 10 <sup>5</sup>
<b>MAE</b>	0.1039	0.0885	0.0942	0.0437	0.0413	1.1628
<b>NMSE</b>	0.0956	0.0730	0.0018	0.1508	0.1374	0.1550
<b>RMSE</b>	0.1434	0.1253	0.1306	0.1326	0.1266	1.6775
<b>NRMSE</b>	0.1474	0.1288	0.0357	0.0939	0.0897	0.2456
<b>RSSE</b>	0.0217	0.0166	0.0013	0.0088	0.0080	0.0603
<b>RatioSV</b>	0.0378	0.0587	0.0413	0.0329	0.0613	0.0650

#### II.5.4.1. Modèle de prédiction SVR de $S_{no2}$

La Figure II.16 montre que le modèle de régression SVR construit pour estimer la concentration  $S_{no2}$  effectue un bon travail sur une grande période de temps de 609 jours. Cependant, ce modèle SVR donne quelques valeurs inférieurs à zéro, ce qui n'est pas vrai puisque la concentration est une mesure positive ou nulle. L'ajout d'information à la sortie de l'algorithme peut améliorer la qualité du modèle.



**Figure II.16 Tracé de  $S_{no2}$  et son estimée**

La Figure II.17 est un zoom sur la figure précédente qui montre le bon suivi de l'estimation de  $S_{no2}$  (construit par SVR) à la mesure  $S_{no2}$  dans une partie de zone de test allant de 300 jusqu'à 320 jours.

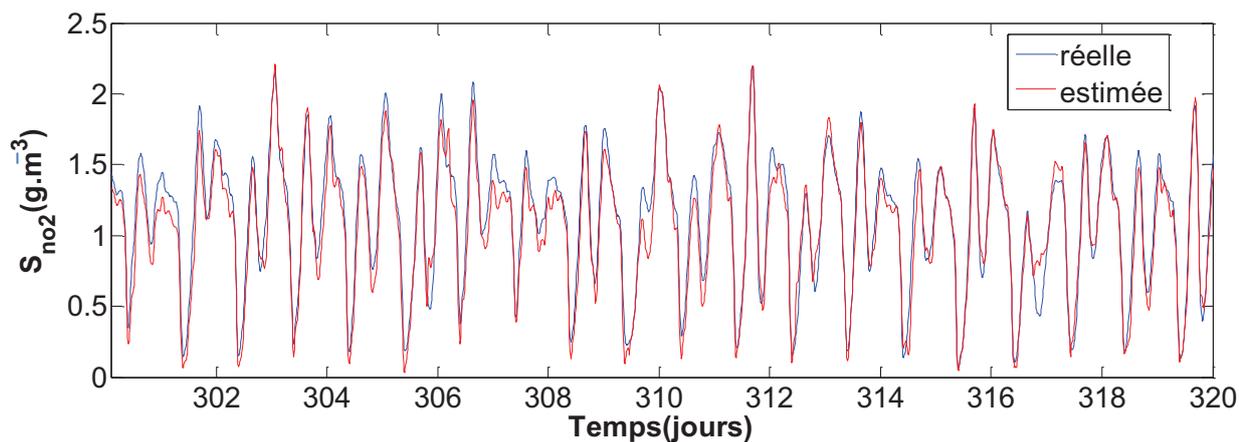


Figure II.17 Zoom sur le tracé de  $S_{no2}$  et son estimé

#### II.5.4.2. Modèle de prédiction SVR de $TSS_e$ de l'effluent

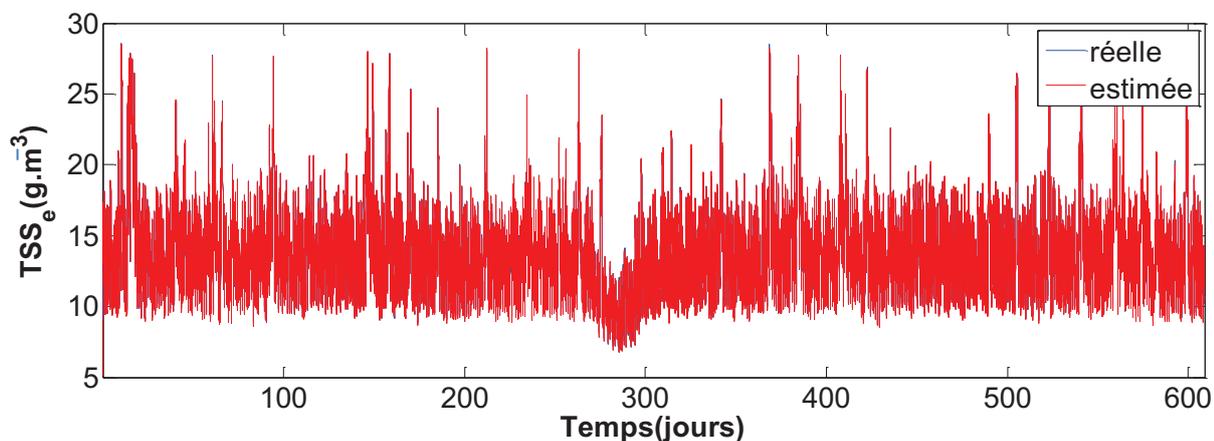


Figure II.18 Tracé de  $TSS_e$  et son estimé

La Figure II.18 illustre bien la parfaite concordance entre  $TSS_e$  et son estimé. Un zoom sur cette figure permet de mieux voir le bon travail d'estimation réalisé par l'approche proposée basée sur SVR, dans une partie de zone de test allant de 300 jusqu'à 350 jours.

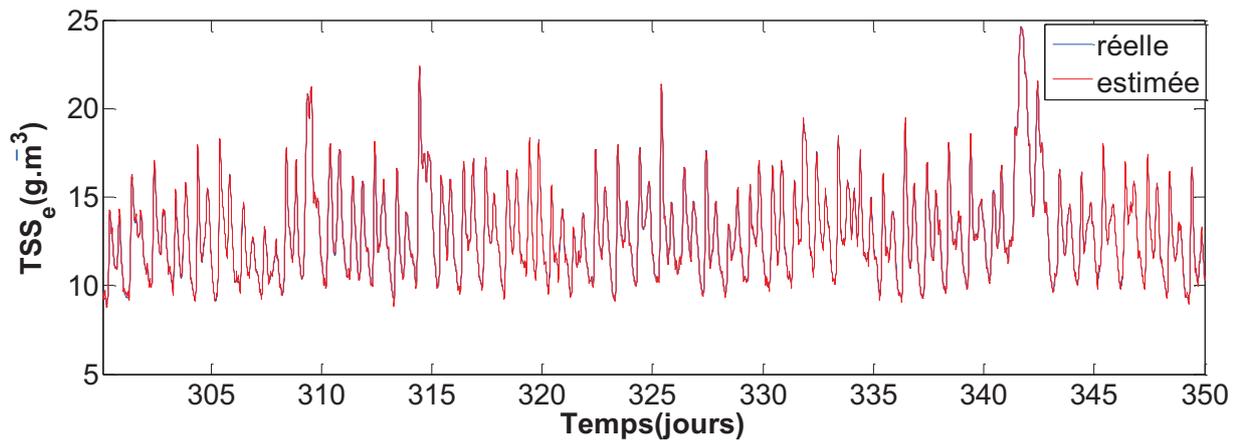


Figure II.19 Zoom sur le tracé de  $TSS_e$  et son estimé

#### II.5.4.3. Modèle de prédiction SVR de COD de l'effluent

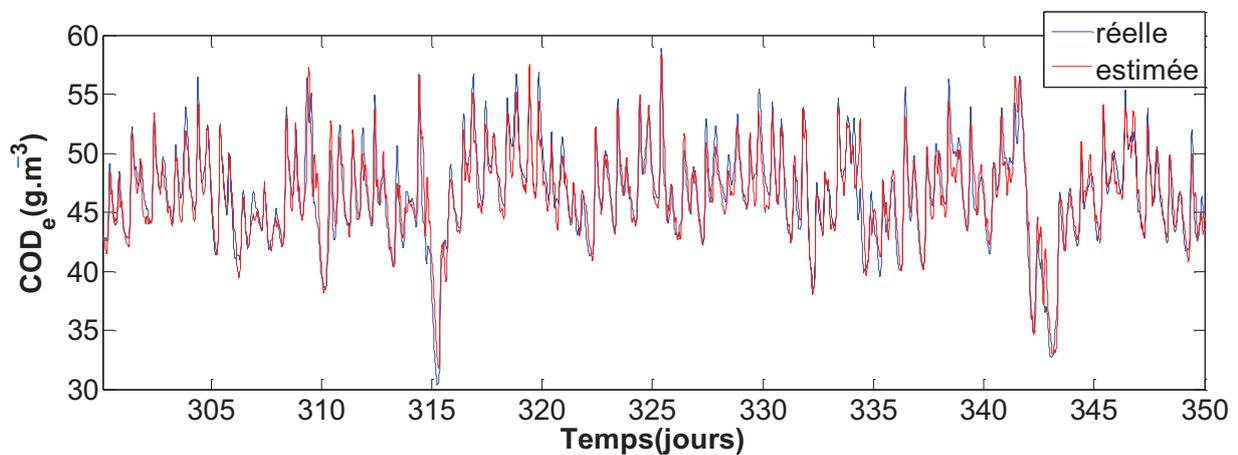


Figure II.20 Tracé de  $COD_e$  et son estimé

La Figure II.20 montre bien l'adéquation entre  $COD_e$  et son estimée sur une période de test allant de 300 jusqu'à 350 jours.

#### II.5.4.4. Modèle de prédiction SVR de $S_{05}$

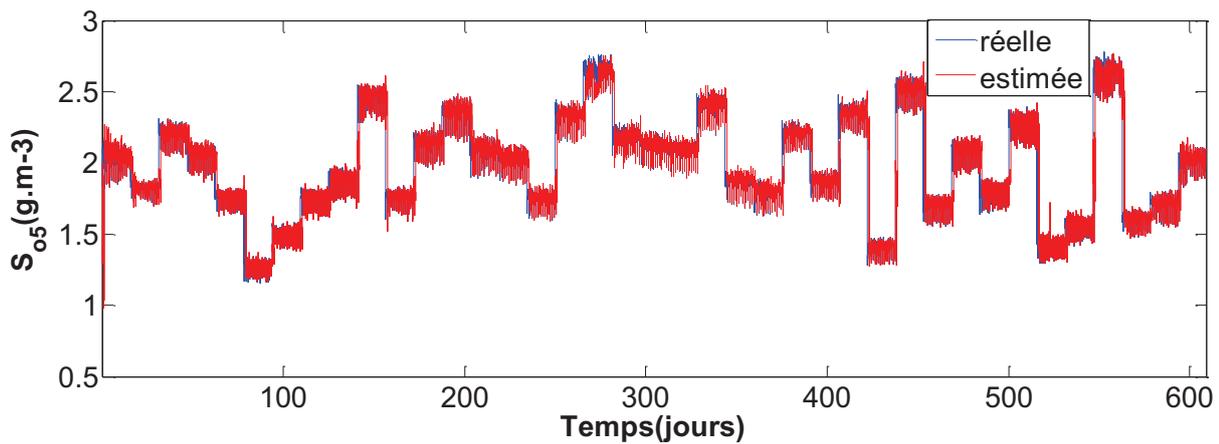


Figure II.21 Tracé de  $S_{05}$  et son estimé

L'estimation de la variable  $S_{05}$  est bien réalisée par l'approche proposée comme le montrent la Figure II.21 ainsi que la Figure II.22 du résidu avec sa faible valeur. Il faut noter que dans cette simulation la consigne  $S_{05,ref}$  n'est pas fixe à  $2 \text{ g.m}^{-3}$  mais variable afin de bien vérifier l'aspect prédictif du modèle SVR de  $S_{05}$ .

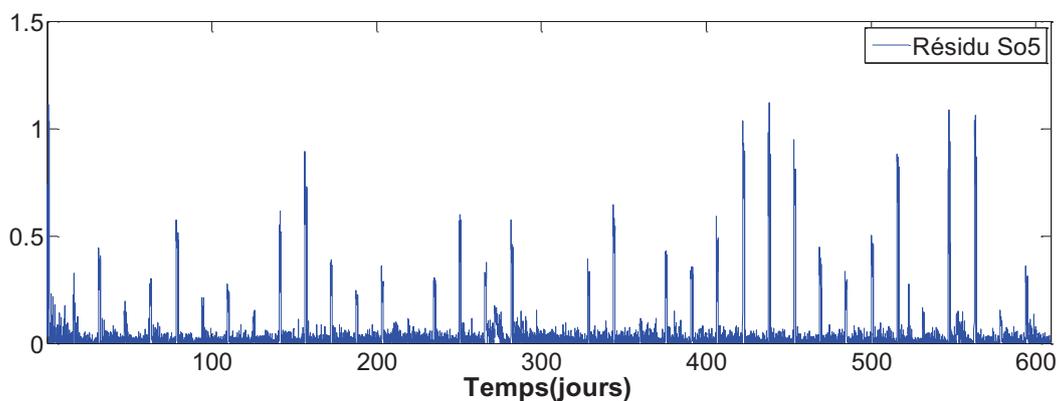


Figure II.22 Résidu  $S_{05}$

#### II.5.4.5. Analyse de résidus

Cette étude aussi permet l'évaluation de la qualité du modèle de prédiction. Si les résidus admettent une certaine tendance ou qu'ils sont négatifs ou positifs d'une façon systématique dans une grande partie de la plage de données. Cela indique que le modèle ajuste mal les données, si par contre les résidus sont petits et aléatoirement distribués autour de zéro, cela indique que le modèle décrit parfaitement les données.

### Résidu Modèle $S_{no2}$

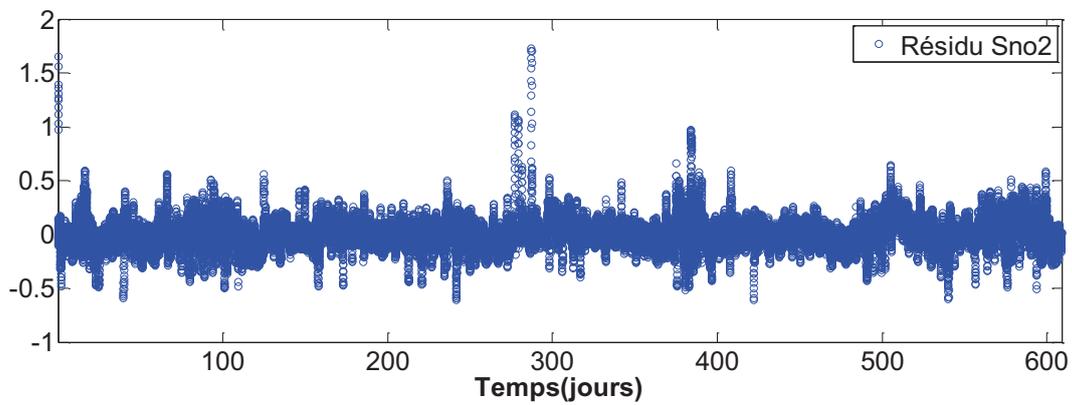


Figure II.23 Résidu  $S_{no2}$

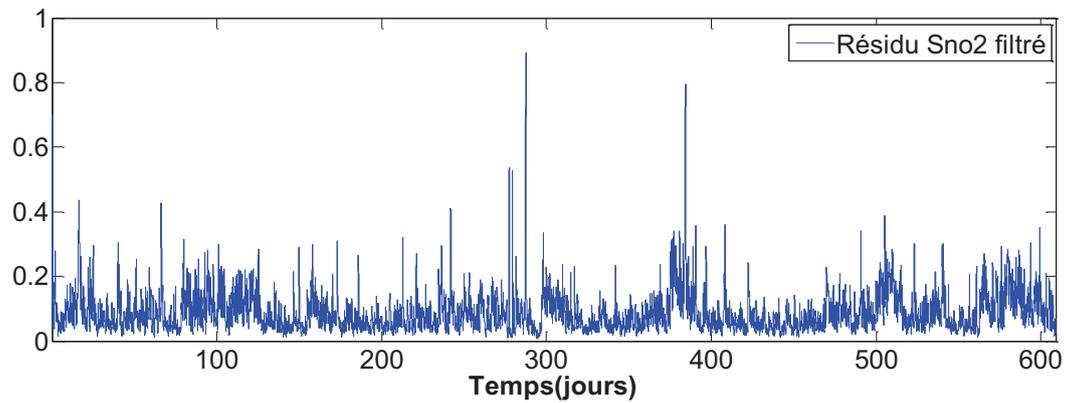
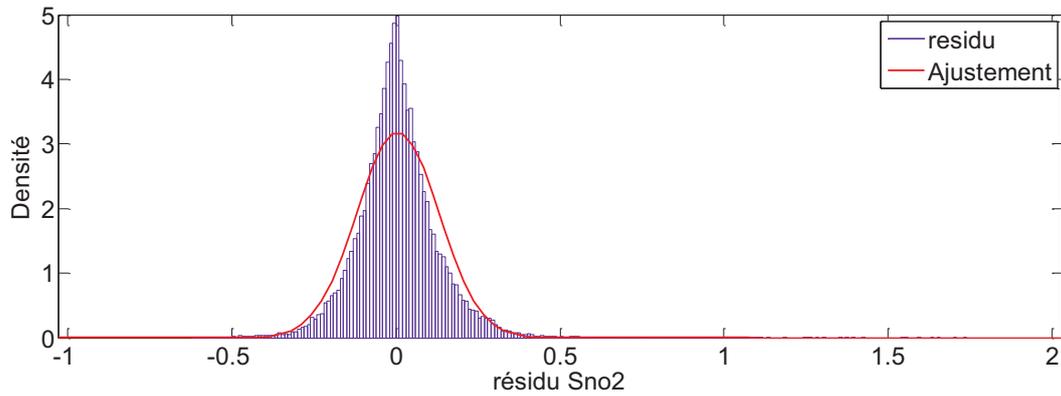


Figure II.24 Résidu filtrée de  $S_{no2}$

L'utilisation de filtre passe-bas permet d'atténuer l'effet des valeurs aberrantes et d'obtenir ainsi un résidu plus faible comme montré dans la Figure II.24.

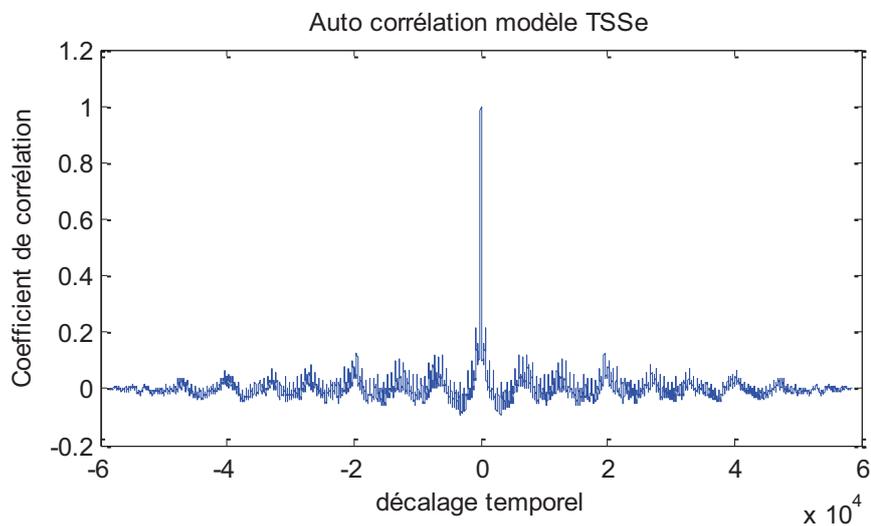


**Figure II.25 Densité de probabilité  $S_{no2}$**

L’histogramme de résidus est une façon de voir la forme de la distribution, et si elle est normalement distribuée ou non.

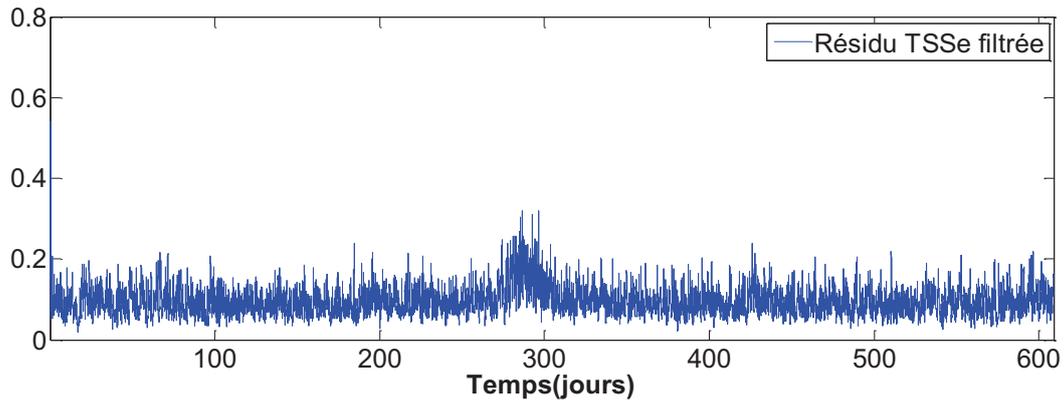
### Résidus modèle $TSS_{eff}$

Le tracé de la structure d'autocorrélation suivant montre bien qu'il existe une faible périodicité de phénomènes à évolution lente dans le procédé WWTP. Aussi cela permet de dire, si on s'éloigne du zéro du point de vue décalage temporel, que la corrélation entre les résidus est assez faible, ce qui montre qu'il n'y a pas trop de dépendance entre les résidus et donc le modèle est bien réalisé.



**Figure II.26 Autocorrélation du résidu  $TSS_e$**

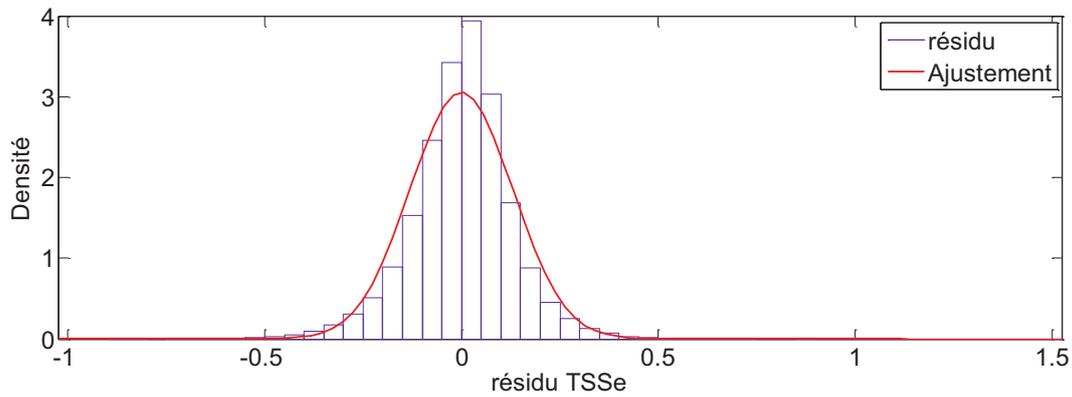
La Figure II.27 trace le résidu filtré en fonction d'une grande période de temps qui a une valeur faible. Cela indique que le modèle SVR dédié à l'estimation de  $TSS_e$  réalise un bon travail.



**Figure II.27 Résidu sur TSS<sub>e</sub>**

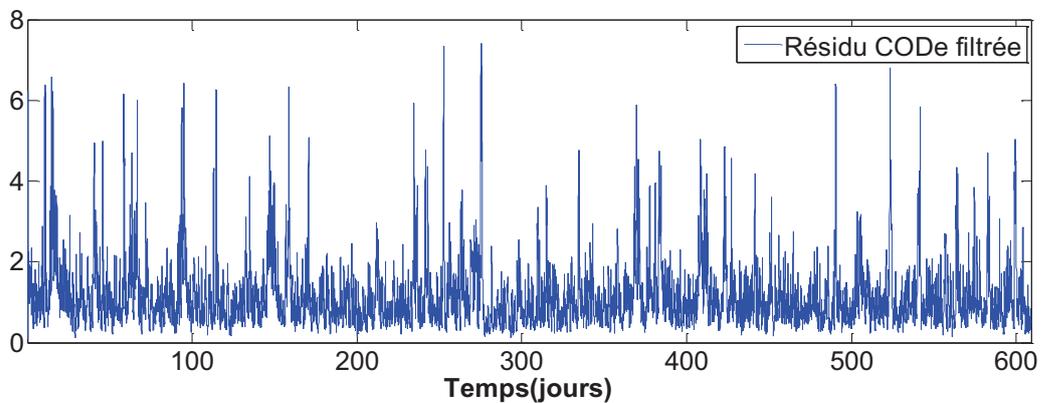
On peut encore vérifier la normalité de la distribution des résidus TSS<sub>e</sub>

Moyenne: -0.000437909 ; Variance: 0.0170659; écart-type: 0.130636



**Figure II.28 Densité de probabilité TSS<sub>e</sub>**

### Résidus modèle COD<sub>e</sub>



**Figure II.29 Résidu COD<sub>e</sub>**

La Figure II.29 montre une valeur inférieure à 8. Cependant, malgré qu'elle reste une valeur convenable par rapport à l'ordre de grandeur de  $COD_e$  ( $50 \text{ g.m}^{-3}$ ), une amélioration peut être envisagée dans un futur proche.

Malgré la difficulté rencontrée avec ce type de procédé, et par rapport à d'autres approches qui dépendent des conditions initiales, les résultats obtenus sont assez satisfaisants du point vue précision mais aussi de simplicité de la mise en forme de ce type d'approche de modélisation.

## II.6. Implémentation de SVM pour le diagnostic de défauts

Dans ce travail, différentes situations de défauts ont été intégrées dans les modèles d'actionneurs et capteurs afin de réaliser une simulation très réaliste. Des travaux ont été réalisés dans ce sens en se basant par exemple sur des chaînes de Markov (Rosen et al., 2008).

**Tableau II.6 : Les différents défauts considérés**

Numéro du défaut	Description du défaut	Type	Occurrence du défaut (j)		Valeur du défaut	
			apprentissage	Test	apprentissage	test
1	Capteur $TSS_e$	Offset	[62:63] [63.5:64.5] [65:66.5]	[260:262] [265:267] [285.5:287]	$-2 \text{ g.m}^{-3}$ $2 \text{ g.m}^{-3}$ $5 \text{ g.m}^{-3}$	$2 \text{ g.m}^{-3}$ $5 \text{ g.m}^{-3}$ $10 \text{ g.m}^{-3}$
2	Capteur $TSS_e$	Dérive	[72:80]	[270:280]	$0.2 \text{ j}^{-1}$	$0.2 \text{ j}^{-1}$
3	Capteur $TSS_e$	Fixe	[68.5 :70]	[310:315]	-	-
4	Capteur $TSS_e$	complet	[67:68]	[300:305]	-	-
5	Capteur $S_{no2}$	Offset	[82:83] [83.5:84.5] [85:86.5]	[320:322] [325:327] [345.5:347]	$-0.5 \text{ g.m}^{-3}$ $0.5 \text{ g.m}^{-3}$ $1 \text{ g.m}^{-3}$	$0.5 \text{ g.m}^{-3}$ $1 \text{ g.m}^{-3}$ $2 \text{ g.m}^{-3}$
6	Capteur $S_{no2}$	Dérive	[92:100]	[330:340]	$0.2 \text{ g.m}^{-3}\text{j}^{-1}$	$0.2 \text{ g.m}^{-3}\text{j}^{-1}$
7	Capteur $S_{no2}$	Fixe	[88.5 :90]	[370:375]	-	-
8	Capteur $S_{no2}$	complet	[87:88]	[360:365]	-	-
9	Capteur $S_{o5}$	Offset	[102:103] [103.5:104.5] [105:106.5]	[370:372] [375:377] [375.5:377]	$-0.5 \text{ g.m}^{-3}$ $0.5 \text{ g.m}^{-3}$ $1 \text{ g.m}^{-3}$	$-1 \text{ g.m}^{-3}$ $1 \text{ g.m}^{-3}$ $2 \text{ g.m}^{-3}$
10	Capteur $S_{o5}$	Dérive	[112:120]	[380:400]	$0.2 \text{ g.m}^{-3}\text{j}^{-1}$	$0.2 \text{ g.m}^{-3}\text{j}^{-1}$
11	Capteur $S_{o5}$	Fixe	[108.5 :110]	[410:415]	-	-
12	Capteur $S_{o5}$	complet	[107:108]	[400:405]	-	-
13	Capteur $Q_a$	Offset	[122:123] [123.5:124.5] [125:126.5]	[420:422] [425:427] [445:447]	$-800 \text{ m}^3\text{j}^{-1}$ $800 \text{ m}^3\text{j}^{-1}$ $1600 \text{ m}^3\text{j}^{-1}$	$1000 \text{ m}^3\text{j}^{-1}$ $-1000 \text{ m}^3\text{j}^{-1}$ $2000 \text{ m}^3\text{j}^{-1}$
14	Actionneur $Q_a$	Retard	[132:140]	[430:440]	$0.2 \text{ j}$	$0.2 \text{ j}$
15	Actionneur $Q_a$	Blocage	[128.5 :130]	[460:465]	-	-
16	Actionneur $Q_a$	Fermeture complète	[127:128]	[450:455]	-	-
17	Actionneur $K_{la5}$	Offset	[142:143] [143.5:144.5] [145:146.5]	[470:472] [475:477] [495:497]	$-5 \text{ j}^{-1}$ $5 \text{ j}^{-1}$ $10 \text{ j}^{-1}$	$5 \text{ j}^{-1}$ $15 \text{ j}^{-1}$ $-10 \text{ j}^{-1}$
18	Actionneur $K_{la5}$	Retard	[152:160]	[480:490]		

19	Actionneur $K_{la5}$	Blocage	[148.5 :150]	[510:515]	-	-
20	Actionneur $K_{la5}$	Fermeture complète	[147:148]	[500:505]	-	-

Les valeurs de défauts montrés dans le Tableau II.6 correspondent à un seul scénario, sachant que d'autres scénarios ont été testés par le changement des valeurs de défauts.

Dans les procédés réels, il ya plusieurs retards normaux dus au transport à longue distance ou associé au retard des mesures. Dans ce travail, en plus des autres défauts supposés sur les actionneurs  $K_{la5}$  et  $Q_a$ , j'ai considéré que c'est intéressant d'ajouter le défaut retard supplémentaire qui se produit souvent en pratique et qui peut provoquer un changement du comportement du système commandé en boucle fermée dégradant ainsi les performances de régulation et peut même causer une instabilité.

Les variables manipulées  $K_{la5}$  et  $Q_a$  pour la régulation du procédé WWTP sont contraints par des valeurs limites:  $K_{la5}=[0-240 \text{ j}^{-1}]$  et  $Q_a=[0-92230 \text{ m}^3\text{j}^{-1}]$

Le problème d'évaluation de performances des méthodes de diagnostic de défaut dans les procédés WWTP a été traité dans (Corominas et al., 2011), où un indice a été proposé pour examiner, déterminer les limites et comparer entre des méthodes de diagnostic dans le modèle de référence d'un procédé de traitement des eaux usés BSM1. D'autres indices de comparaison de performance ont été proposés dans (Yélamos et al., 2006).

Les critères de détection du défaut considérés dans ce travail sont :

Temps de détection du défaut (TD), fausses alertes(FA)

**Tableau II.7 Résultats de détection de défauts**

Défaut	Type	Valeur	Temps d'occurrence	TD (j)	FA
Capteur TSSe	Offset	2 g.m <sup>-3</sup>	260 j	260.1	0
		5 g.m <sup>-3</sup>	265 j	265.04	
		10 g.m <sup>-3</sup>	285.5 j	285.021	
Capteur TSSe	Dérive	0.2 g.m <sup>-3</sup> .j <sup>-1</sup>	270 j	275.5	0
Capteur TSSe	Fixe	-	310 j	310.333	0
Capteur TSSe	Complet	-	300 j	300.021	0
Capteur S <sub>no2</sub>	Offset	0.5 g.m <sup>-3</sup>	320 j	-	0
		1 g.m <sup>-3</sup>	325 j	-	
		2 g.m <sup>-3</sup>	345.5 j	345.8	
Capteur S <sub>no2</sub>	Dérive	0.2 g.m <sup>-3</sup> .j <sup>-1</sup>	330 j	334.6	0
Capteur S <sub>no2</sub>	Fixe	-	370 j	-	0

Capteur S <sub>no2</sub>	Complet	-	360 j	361.1	0
Capteur S <sub>o5</sub>	Offset	-1 g.m <sup>-3</sup>	370 j	-	0
		1 g.m <sup>-3</sup>	375 j	-	
		2 g.m <sup>-3</sup>	375.5 j	-	
Capteur S <sub>o5</sub>	Dérive	0.2 g.m <sup>-3</sup> .j <sup>-1</sup>	380 j	385.2	0
Capteur S <sub>o5</sub>	Fixe	-	410 j	410.6	0
Capteur S <sub>o5</sub>	Complet	-	400 j	400.4	0
Actionneur Q <sub>a</sub>	Offset	1000 m <sup>3</sup> .j <sup>-1</sup>	420 j	-	0
		-1000 m <sup>3</sup> .j <sup>-1</sup>	425 j	-	
		2000 m <sup>3</sup> .j <sup>-1</sup>	445 j	-	
Actionneur Q <sub>a</sub>	Retard	0.2 j	430 j	430.137	0
Actionneur Q <sub>a</sub>	Blocage	-	460 j	460.19	0
Actionneur Q <sub>a</sub>	Fermeture complète	-	450 j	450.073	0
Actionneur K <sub>la5</sub>	Offset	5 j <sup>-1</sup>	470 j	470.8	0
		15 j <sup>-1</sup>	475 j	475.187	
		-10 j <sup>-1</sup>	495 j	495.2	
Actionneur K <sub>la5</sub>	Retard	0.2 j	480 j	480.1	0
Actionneur K <sub>la5</sub>	Blocage	-	510 j	510.2	0
Actionneur K <sub>la5</sub>	Fermeture complète	-	500 j	500.1	0

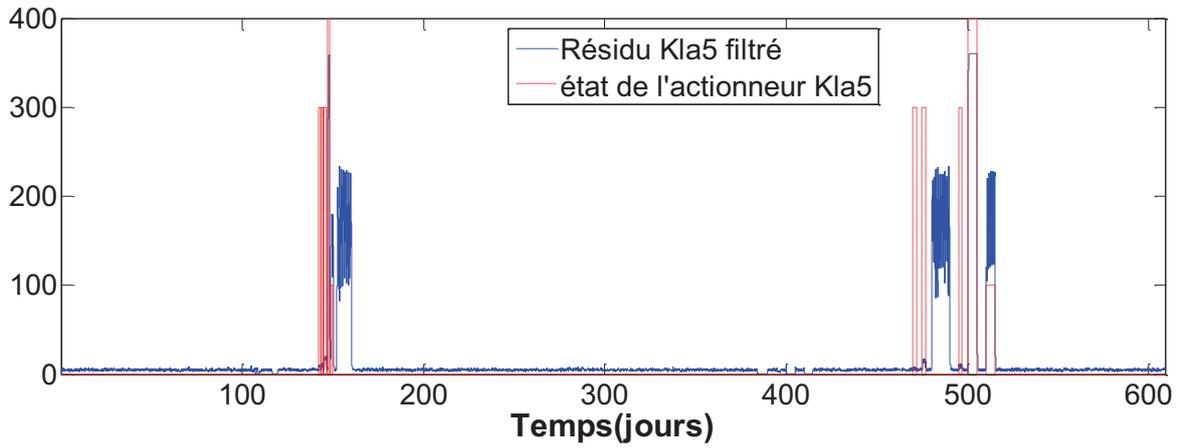
Il est difficile de détecter des défauts offset sur S<sub>no2</sub> et aussi sur S<sub>o5</sub> puisque ces concentrations sont régulées par un PID qui va masquer l'effet de ce défaut.

**Les types de défauts** considérés ou suspectés dans ce travail sont montrés dans le Tableau II.7 ci-dessus.

### III.6.1. Résultats de diagnostic par la stratégie proposée basée sur les SVMs

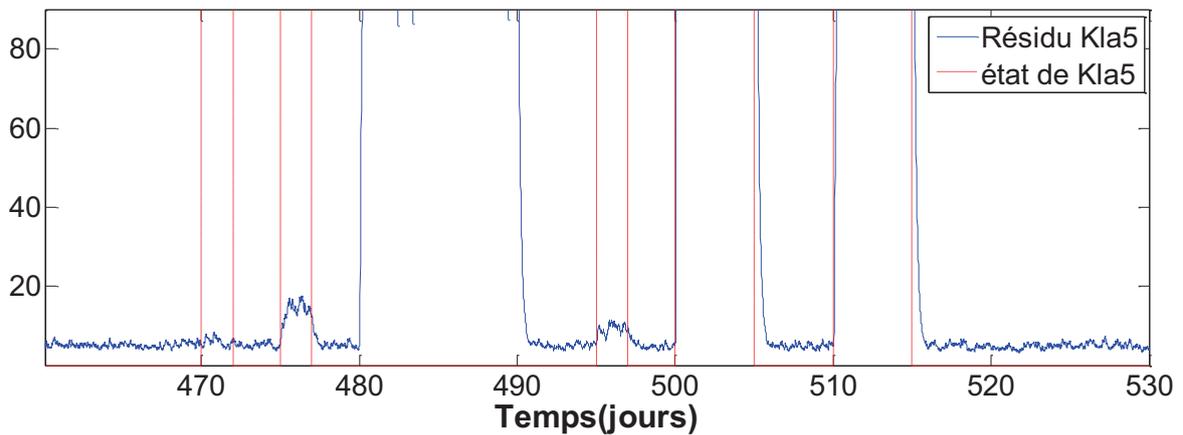
Afin d'évaluer la capacité de l'approche proposée, plusieurs défauts sont introduits à différents moments sur une large période de temps (609 jours).

Le résidu reste parmi les caractéristiques principales utilisées comme entrée par la décision SVM, elle est la plus importante à mon avis. La Figure II.30 trace le résidu filtré de K<sub>la5</sub> en bleu qui est sensible seulement à l'état de défaut sur l'actionneur K<sub>la5</sub> présenté par une valeur non nulle dans la courbe rouge en pointillés.



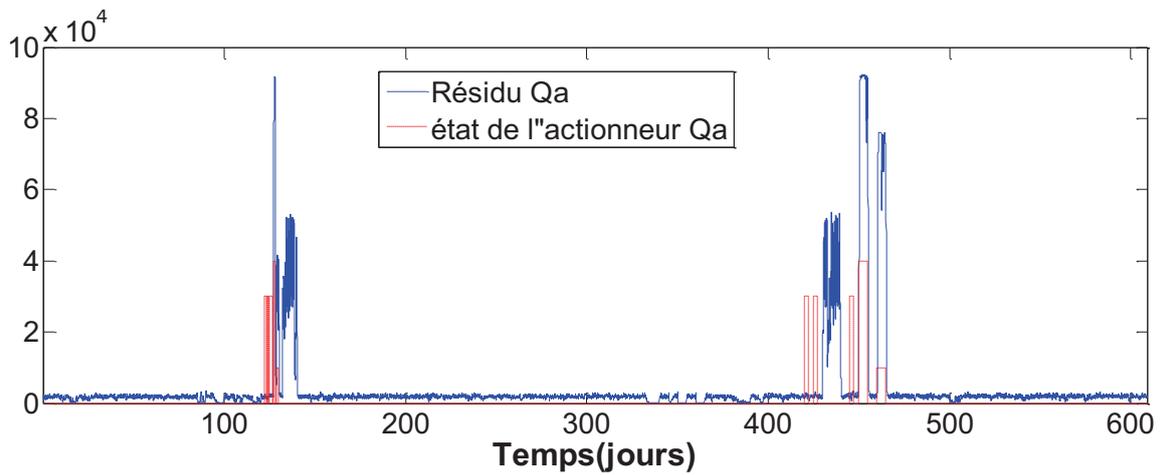
**Figure II.30** Évolution du résidu  $K_{la5}$

Un zoom sur la figure précédente permet de mieux voir la sensibilité face à l'apparition du défaut  $K_{la5}$  (voir Figure II.31)



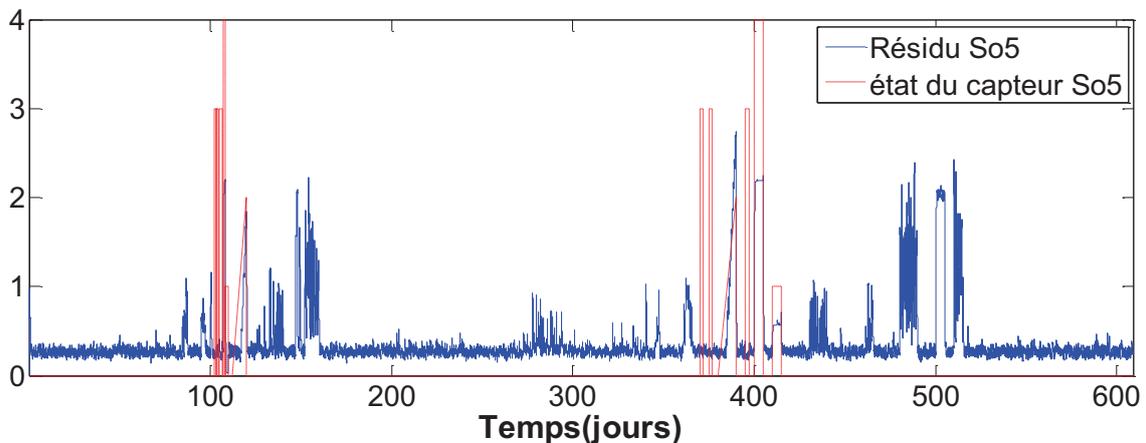
**Figure II.31** Zoom sur le résidu  $K_{la5}$

Comme pour le résidu  $K_{la5}$ , la Figure II.32 ci-dessous montre bien que le résidu  $Q_a$  filtré est sensible uniquement à l'apparition du défaut  $Q_a$ .



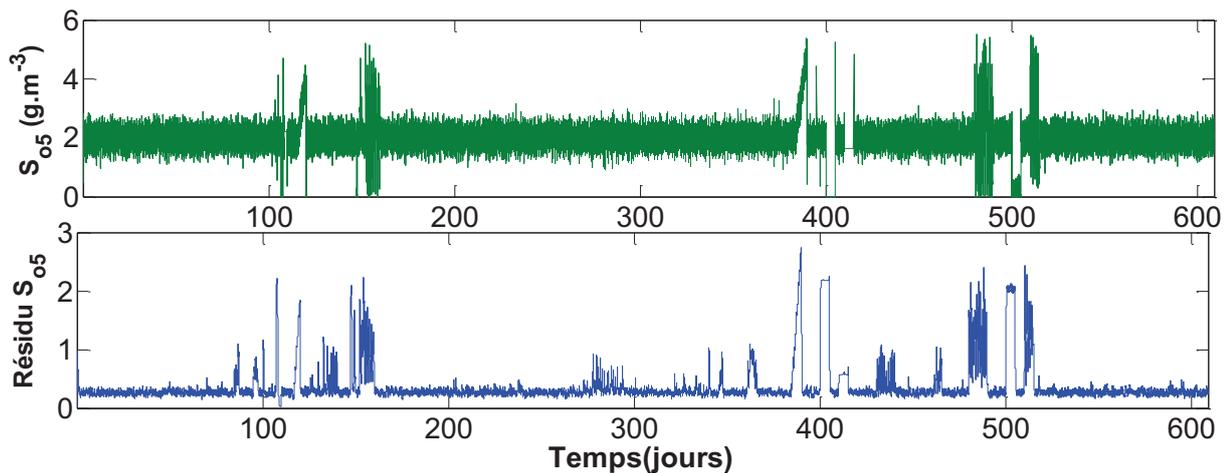
**Figure II.32 Évolution du résidu  $Q_a$**

La Figure II.33 suivante montre bien que le défaut capteur  $S_{05}$  a bien été détecté mais pas isolé dans la zone d'apprentissage mais aussi dans la zone de test. Le résidu filtré  $S_{05}$  est sensible à plusieurs défauts dont le défaut  $S_{05}$ , rappelant que la courbe rouge en pointillés définit l'état de capteur. Elle prend la valeur zéro si le capteur est dans son état normal ou une valeur différente à zéro dans le cas d'un défaut sur  $S_{05}$ .



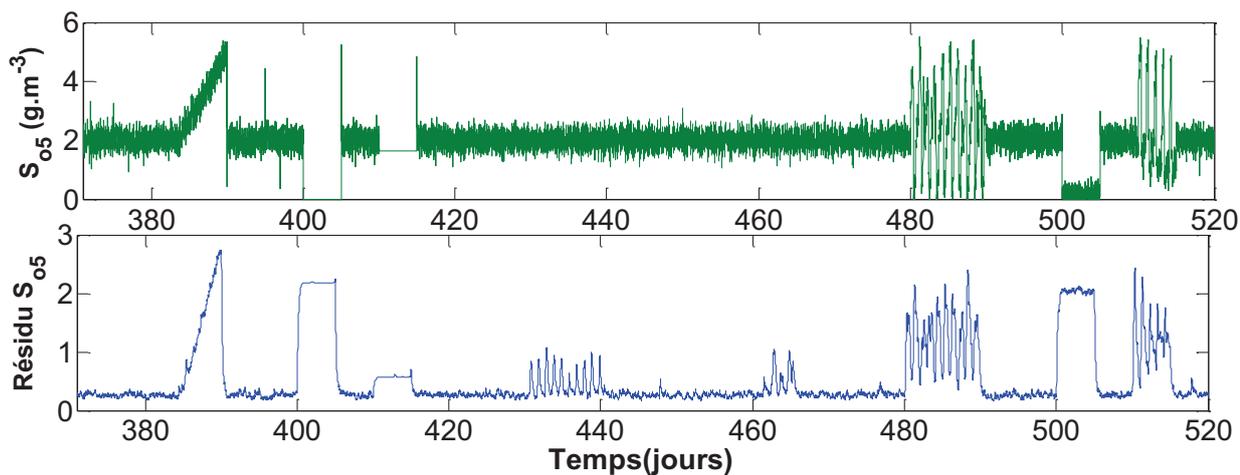
**Figure II.33 Évolution du résidu  $S_{05}$**

La Figure II.34 suivante trace le résidu de  $S_{05}$  ainsi que la mesure  $S_{05}$ . Il apparaît bien que la variable  $S_{05}$  est affectée non seulement par les bruits, perturbations et du défaut capteur  $S_{05}$  ainsi que d'autres défauts. Cela est dû au fait que cette variable à réguler est importante puisqu'elle indique la quantité d'oxygène qui est nécessaire à la survie de l'écosystème dans le procédé WWTP.



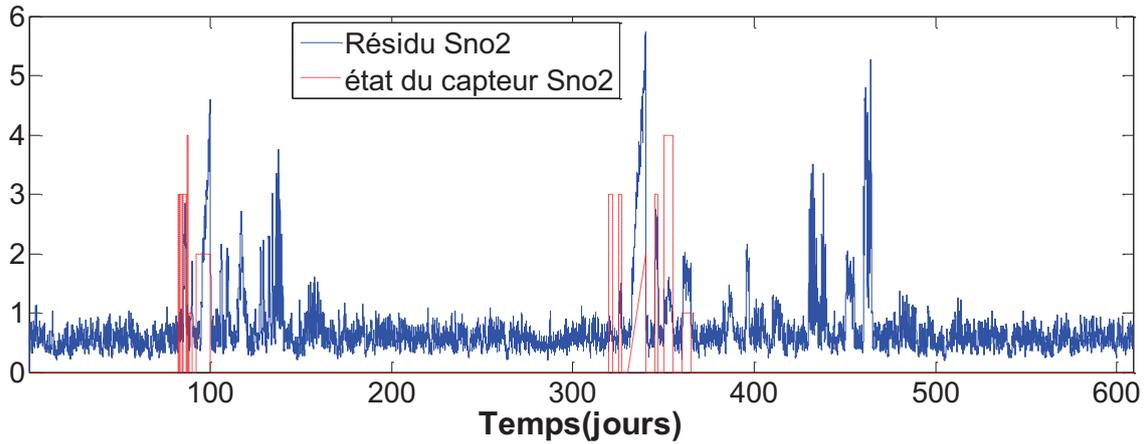
**Figure II.34 Tracé de la mesure de l'oxygène dissous ( $S_{o5}$ ) et son résidu**

La Figure II.35 du zoom permet de bien voir l'effet des différents types de défauts capteur  $S_{o5}$  à savoir valeur fixe, dérive, biais et défaut complet.



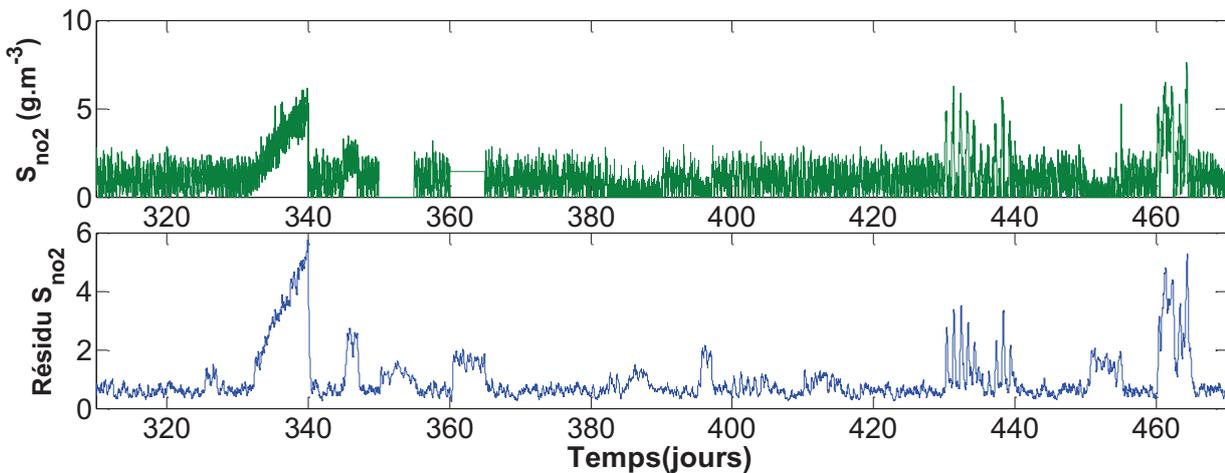
**Figure II.35 Zoom sur de la mesure de  $S_{o5}$  et son résidu**

La Figure II.36 suivante trace le résidu  $S_{no2}$  qui est affecté par d'autres défauts en plus du défaut de capteur sur  $S_{no2}$  illustré par la courbe rouge en pointillés.



**Figure II.36** Évolution du résidu  $S_{no2}$

La Figure II.37 trace le résidu de  $S_{no2}$  par rapport à la mesure  $S_{no2}$  qui doit être commandée à la valeur  $1 \text{ g. (N).m}^{-3}$ . Cependant, il est clair d'après cette figure qu'elle varie beaucoup à cause des bruits, perturbations et aussi des difficultés de la réguler juste par le débit  $Q_a$ , sans oublier bien sur les défauts considérés.



**Figure II.37** La mesure de  $S_{no2}$  et son résidu

La Figure II.38 montre bien que malgré la variabilité de la variable  $TSS_e$  à la sortie. Le résidu arrive non seulement à détecter mais aussi à isoler le défaut sur le capteur  $TSS_e$  puisqu'il arrive à rejeter les autres défauts et cerne seulement le défaut en question montré par la courbe rouge.

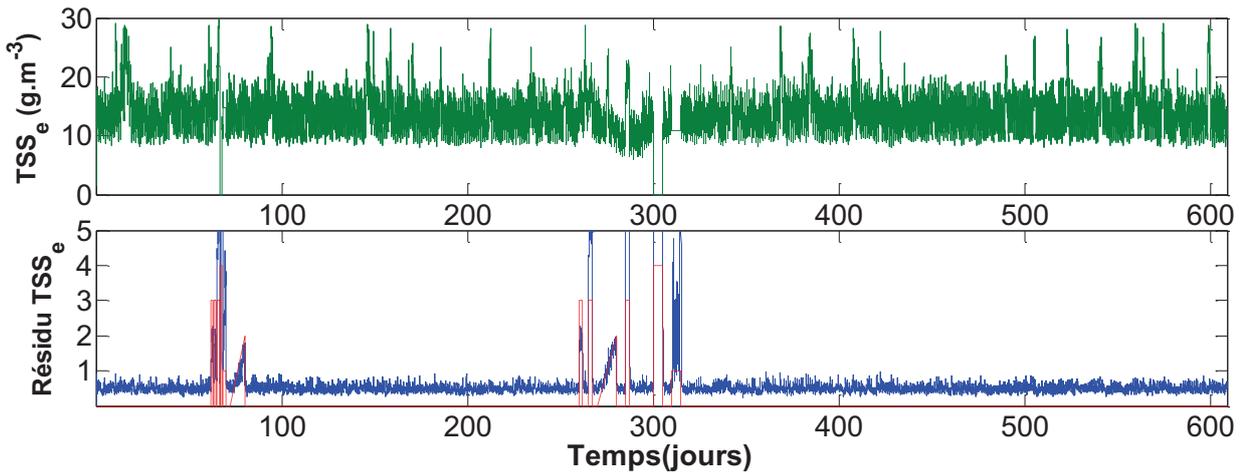


Figure II.38 Mesure  $TSS_e$  et son résidu

## II.6.2. Réglage des paramètres $\sigma$ et $\varepsilon$ en SVM

### Les décisions sur les différents défauts

Parmi les défauts étudiés dans ce travail le défaut d'actionneur de type retard qui va influencer négativement sur les performances de régulation en boucle fermée et provoque de sérieux problèmes dans le comportement du système, à savoir la diminution de temps de réponse, l'apparition des oscillations, voire même carrément l'instabilité du système.

Une grande valeur de  $\sigma$  signifie moins de propagation ou plus de comportement non linéaire du noyau, donc elle permet de gouverner le degré de non linéarité. Dans la conception des modèles de décision SVM,  $\sigma$  est choisi entre 2 et 17 par la méthode traditionnelle d'essai-erreur. La constante de régulation  $C$  qui contrôle le surapprentissage du modèle est choisie entre 7 et 30. Un seuil est fixé pour différencier l'état normal de celui défectueux. Sa valeur est choisie minutieusement afin de détecter le défaut le plus rapidement possible tout en évitant au maximum les fausses alarmes, en tenant compte que SVM classe les données en deux catégories ayant respectivement les labels +1 et -1. Le seuil est choisi d'une valeur plus grande que zéro et plus précisément des valeurs entre 0.7 et 1 ont été utilisés pour la prise de décision. Le noyau utilisé dans cette étude est le noyau gaussien qui est le mieux approprié à manipuler avec les données réelles incluant des bruits et des perturbations.

### II.6.2.1. Défaut actionneur $Q_a$

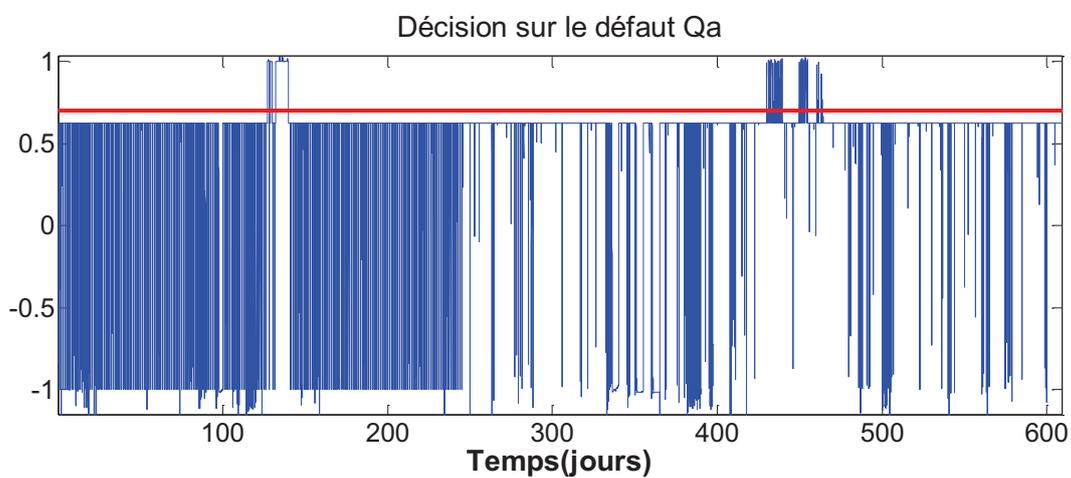
Le vecteur de caractéristiques utilisé comme entrée est le suivant:

$$x = \begin{pmatrix} |\text{résidu } Q_a| \\ |\text{résidu } S_{no2}| \\ |Q_a(t) - Q_a(t-1)| \end{pmatrix}$$

Les paramètres d'apprentissage utilisés pour construire ce modèle de décision sont :

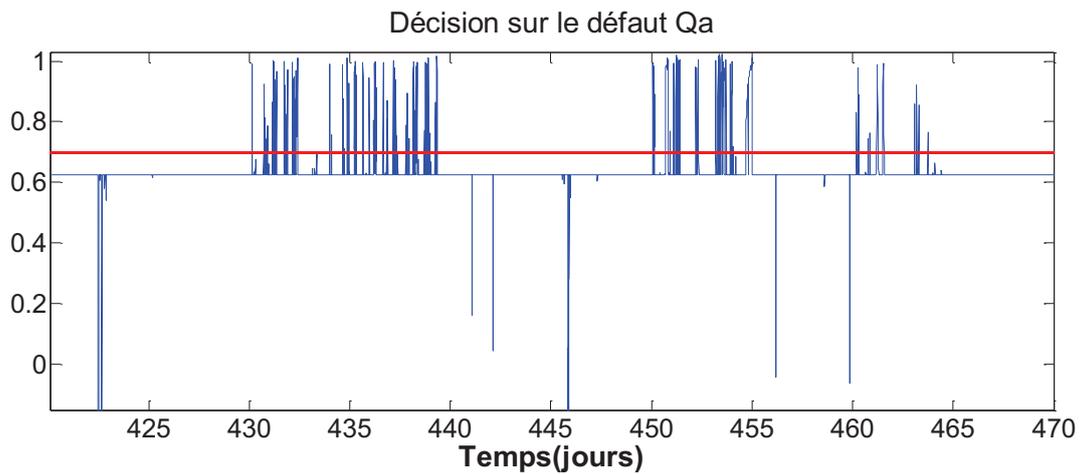
$C = 30, \sigma = 15$  la valeur du seuil est égale à 0.7 pour ce défaut

La Figure II.39 ci-dessous montre bien la capacité décisionnelle du classifieur SVM, qui permet de détecter l'apparition du défaut  $Q_a$  rapidement dans les deux zones d'apprentissage et de test et cela sans fausses alarmes.



**Figure II.39** Résultat de décision sur le défaut  $Q_a$

Un zoom sur le résultat de décision du défaut  $Q_a$  (Figure II.40) dans la zone de test permet de montrer la détection de défaut  $Q_a$  de différents types : retard, fermeture complète, blocage.



**Figure II.40 Zoom sur le résultat du défaut Q<sub>a</sub>**

### ***II.6.2.2. Défaut actionneur K<sub>la5</sub>***

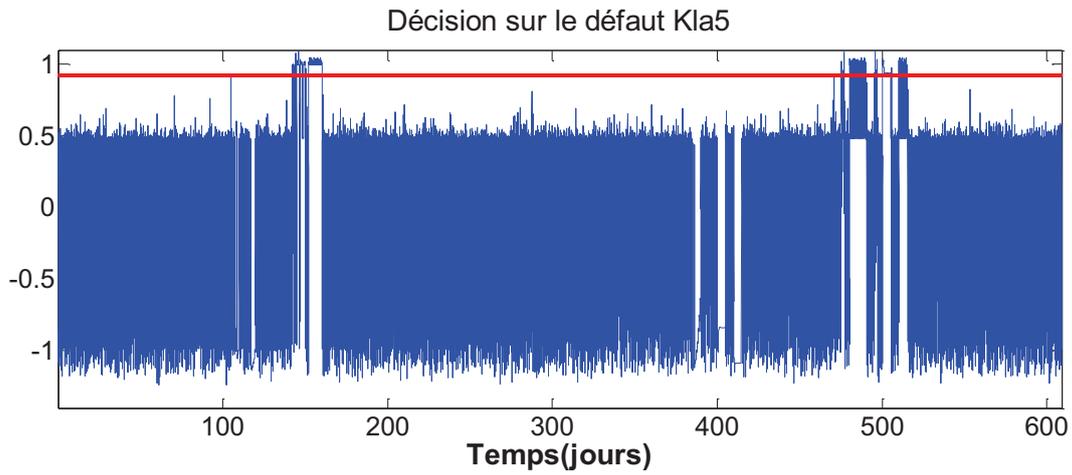
Le vecteur de caractéristiques utilisé comme entrée est donné par:

$$x = \begin{pmatrix} |résidu K_{la5}| \\ |résidu S_{o5}| \\ |K_{la5}(t) - K_{la5}(t-1)| \end{pmatrix}$$

Les paramètres d'apprentissage utilisés pour construire ce modèle de décision sont :

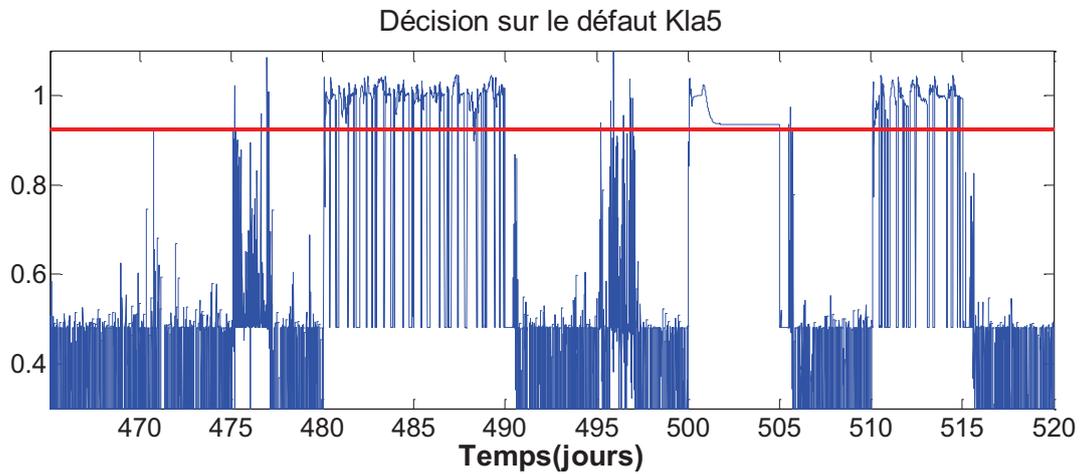
$C = 10, \sigma = 2$  la valeur du seuil est égale à 0.925 pour ce défaut.

La Figure II.41 ci-dessous montre bien la capacité décisionnelle du classifieur SVM, qui permet de détecter l'apparition du défaut K<sub>la5</sub> rapidement dans les deux zones d'apprentissage et de test et cela sans fausses alarmes.



**Figure II.41** Résultat de décision sur le défaut  $K_{1a5}$

Avec un zoom sur la zone de test on peut mieux apprécier le résultat (Figure II.42)



**Figure II.42** Zoom de décision sur le défaut  $K_{1a5}$

### **II.6.2.3. Défaut Capteur $S_{no2}$**

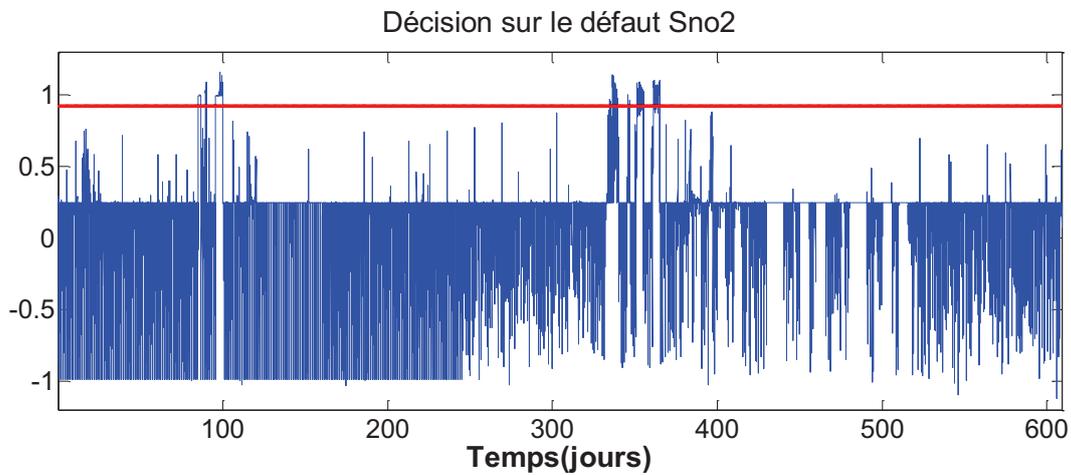
Le vecteur de caractéristiques utilisé dans ce cas est le suivant :

$$x = \begin{pmatrix} |résidu S_{no2}| \\ |résidu S_{o5}| \\ |résidu K_{1a5}| \\ |résidu Q_a| \\ |S_{no2}(t) - S_{no2}(t-1)| \end{pmatrix}$$

Les paramètres d'apprentissage utilisés pour construire ce modèle de décision sont :

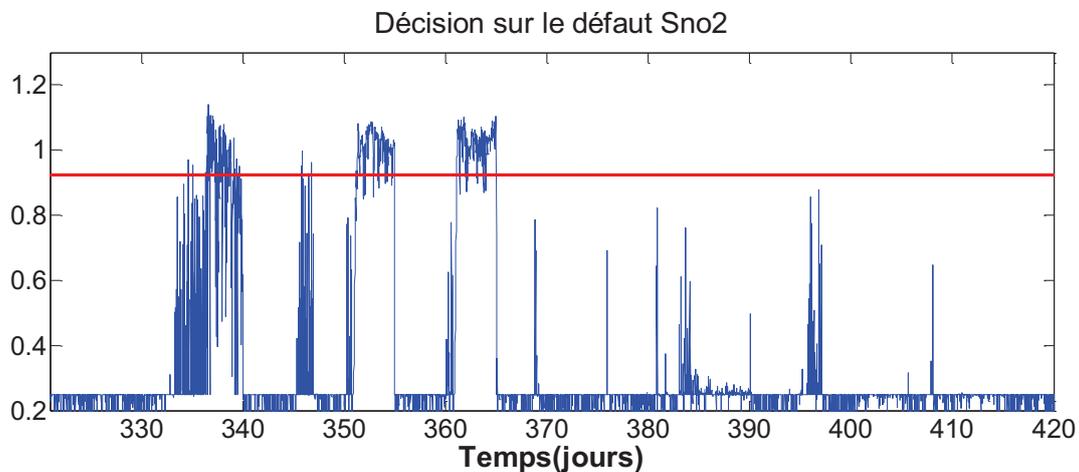
$C = 10, \sigma = 2$ , la valeur du seuil est égale à 0.925 pour ce défaut.

La Figure II.43 ci-dessous montre bien la capacité décisionnelle du classifieur SVM, qui permet de détecter l'apparition du défaut  $Kla_5$  rapidement dans les deux zones d'apprentissage et de test et cela sans fausses alarmes.



**Figure II.43** Résultat de décision sur le défaut  $S_{no2}$

Malgré la difficulté sur ce défaut due à la très grande variabilité de cette mesure et qui peut avoir des valeurs de  $0 \text{ g.m}^{-3}$  mais qui correspond à un fonctionnement normal, ce défaut est bien détecté. Un zoom va montrer plus de détails (voir Figure II.44)



**Figure II.44** Zoom sur la décision du défaut  $S_{no2}$

#### **II.6.2.4. Défaut capteur $TSS_e$**

Le vecteur de caractéristiques utilisé est donné comme suit :  $x = \begin{pmatrix} |résidu TSS_e| \\ |TSS_e(t) - TSS_e(t-1)| \end{pmatrix}$

Les paramètres d'apprentissage utilisés pour construire ce modèle de décision sont :

$C = 10, \sigma = 20$ , la valeur du seuil est égale à 0.925 pour ce défaut. La Figure II.45 illustre le bon résultat obtenu pour le défaut sur TSS<sub>e</sub> caractérisé par une rapidité de détection et l'absence de fausses alarmes.

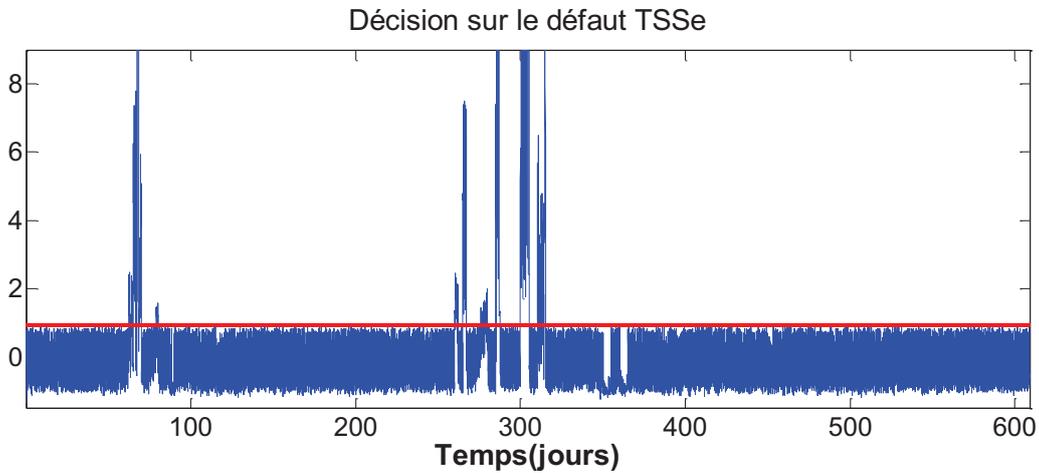


Figure II.45 Résultat de la décision sur défaut TSS<sub>e</sub>

#### II.6.2.5. Défaut capteur $S_{o5}$

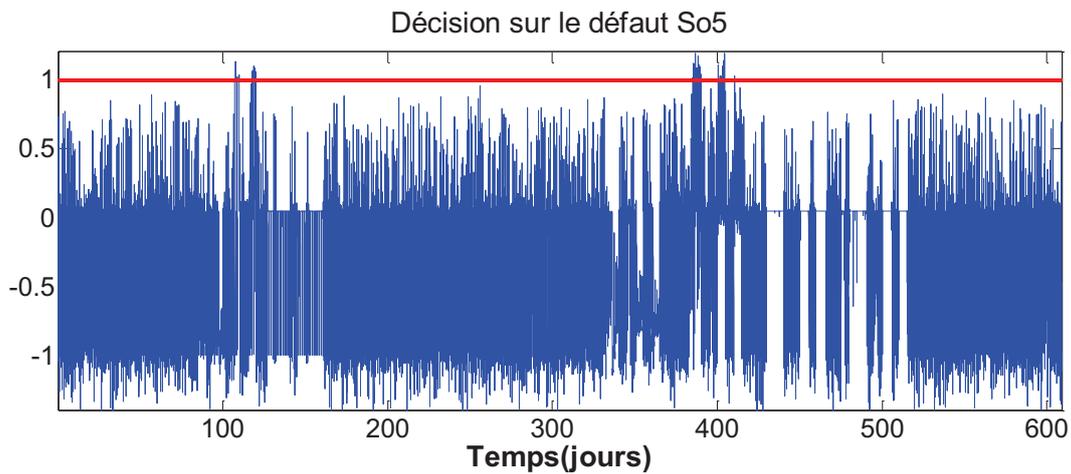
Le vecteur de caractéristiques utilisé comme entrée est le suivant:

$$x = \begin{pmatrix} |résiduS_{o5}| \\ |résiduS_{no2}| \\ |résiduK_{La5}| \\ |résiduQ_a| \end{pmatrix}$$

Les paramètres d'apprentissage utilisés pour construire ce modèle de décision sont :

$C = 7, \sigma = 2$ , la valeur du seuil est égale à 1 pour ce défaut.

La Figure II.46 illustre que les défauts sur le capteur  $S_{o5}$  de type valeur fixe, dérive ou complet ont été bien détectés. Cependant, le défaut de type offset ne l'est pas puisque il est masqué par l'action du PID. On rappelle toujours la justesse de décision sur le défaut avec un pourcentage nul de fausses alarmes.



**Figure II.46** Résultat de la décision sur défaut  $S_{05}$

Il est à noter que malgré les fortes perturbations et bruits dans le procédé de traitement des eaux, cette stratégie proposée dans ce travail a permis de surmonter ces difficultés par la bonne détection et isolation du défaut en question avec un grand avantage par rapport aux autres méthodes statistiques telles que l'ACP qui est le pourcentage quasi nul des fausses alarmes

## II.7. Conclusion

Le procédé de traitement des eaux usées comme la plupart des procédés réels est de nature stochastique variable dû en grande partie à la variation de l'influent. C'est pour cela que les approches statistiques deviennent de plus en plus populaires dans divers secteurs industriels.

La stratégie proposée dans cette thèse délivre une méthodologie simple, utile et fructueuse pour modéliser les systèmes réels. Il est démontré la capacité de SVR à réaliser une modélisation flexible qui capte la non-linéarité et le maximum de variabilité des données et qui décrit le comportement dynamique du système en présence de l'incertitude.

Les performances intéressantes de prédiction de SVR ont été soulignées dans ce chapitre.

Je crois que la stratégie de diagnostic de défaut proposée dans ce chapitre qui utilise uniquement la théorie des SVMs peut être appliquée à d'autres systèmes complexes.

Je pense que SVM et SVR vont être de plus en plus connus et utilisés et peuvent devenir ainsi un outil important et indispensable dans l'analyse de données scientifiques dans les années à venir. C'est pourquoi la compréhension de leur théorie est essentielle, un très bon tutoriel a cherché à expliquer les SVMs et leurs paramètres avec un plan visuel illustré par des graphes (Brereton and Lloyd, 2010).

Malgré les bons résultats obtenus concernant la modélisation SVR, une amélioration peut être faite par la suite par l'incorporation des connaissances du procédé WWTP dans le modèle SVR. Aussi, pour que ce dernier soit très efficace dans les applications industrielles, je pense qu'il faut tenir compte de changements dans le temps des propriétés et caractéristiques propres du procédé WWTP. L'idée dans le futur est d'ajouter un aspect d'adaptation au SVR standard par l'utilisation d'un terme correctif ou d'ajouter en parallèle du modèle SVR standard un algorithme SVR en ligne qui va apprendre le comportement du procédé au fur et à mesure.

# **CHAPITRE III : Diagnostic de défauts dans les éoliennes**



### III.1. Introduction

L'énergie est l'objet de beaucoup de débats mondiaux aujourd'hui. L'objectif des pays est de satisfaire la demande en énergie qui ne cesse d'augmenter tout en respectant l'environnement et la sécurité des gens. Diversifier les sources d'énergie est essentiel à cause du début d'épuisement des ressources d'énergie naturelles comme le pétrole et le gaz, afin d'éviter la dépendance énergétique pour les pays non producteurs des hydrocarbures, et aussi afin de préserver la nature. Ainsi, l'attention s'est portée sur les énergies renouvelables.

Ces dernières années, l'énergie solaire et l'énergie produite par les éoliennes font partie de ces énergies à la mode. La maîtrise de ces énergies nouvelles demande des progrès non seulement technologiques mais aussi scientifiques.

L'éolienne est un système d'énergie renouvelable qui permet d'exploiter la ressource du vent et de la transformer en énergie électrique. L'énergie éolienne est en pleine expansion aujourd'hui. Plusieurs installations de parcs éoliens se trouvent en mer (*offshore*) (Figure III.1). Cette tendance va se confirmer encore plus dans le futur pour éviter de gêner les citoyens au niveau du paysage et du bruit, et aussi pour avoir un vent fort, fréquent et assez régulier.



**Figure III.1 Eoliennes offshore**

Ces implantations flottantes dans le large nécessitent une étude approfondie du vent, de l'aérodynamique, des courants marins, de l'hydrodynamique, de l'élasticité, du contrôle de l'éolienne, ainsi que la plate-forme afin de déterminer leur faisabilité technique et économique. Ceci exige aussi d'augmenter la disponibilité et la fiabilité de l'éolienne, de réduire les arrêts et les visites de maintenance qui sont plus difficiles et plus coûteuses par rapport aux installations de l'éolienne sur la terre. Une façon de satisfaire ces exigences est de concevoir un système de

surveillance et de diagnostic de défauts assez performant. La surveillance et le diagnostic de défauts des éoliennes permettent de prédire et détecter rapidement toute occurrence d'une défaillance avant qu'elle ne se propage et conduise à l'arrêt ou voire même à la destruction de l'appareil éolien comme dans la Figure III.2.



**Figure III.2 Destruction d'une éolienne par prise de feu**

La taille des éoliennes ne cesse d'augmenter. Les plus grandes éoliennes peuvent présenter une puissance de l'ordre 6 MW (par exemple celle développée et réalisée par Alstom) avec des pales allant jusqu'à 73.5 mètres produisant un diamètre total de 150 mètres. Ces grandes structures sont sujettes à de fortes charges.

L'objectif de la stratégie de Diagnostic de Défaut est d'aider à l'obtention d'un meilleur fonctionnement de l'éolienne afin de maximiser la génération de l'électricité avec un coût minimal tout en respectant les conditions de sécurité. Toutefois, plusieurs difficultés et défis sont rencontrés par le concepteur du système FDI dont :

- La structure complexe de l'éolienne.
- La non-linéarité, l'instabilité et la complexité de l'aérodynamique de l'éolienne.
- Le cisaillement du vent à cause des forces de frottement dues à la rencontre du vent avec les différents obstacles possibles.
- Le vent, cette ressource abondante est nécessaire pour le fonctionnement de l'éolienne puisque c'est la force d'entraînement de cette machine. Cependant, sa nature aléatoire et turbulente nous amène à le considérer aussi comme une perturbation.
- Les fluctuations de la puissance de l'éolienne qui sont dues à des processus stochastiques du vent ainsi qu'à des phénomènes d'oscillations périodiques comme le cisaillement du vent et la réorientation du vent en raison de la structure de la tour de l'éolienne « *tower shadow* ».

Inclure la modélisation de ces phénomènes dans le modèle de l'éolienne permet d'aider le concepteur de l'éolienne dans son travail (Das et al., 2011).

- Les bruits de mesures et les perturbations.
- La commutation de la commande.

Le travail présenté dans ce chapitre porte sur la surveillance et le diagnostic des défauts de capteurs, d'actionneurs et du système sur l'éolienne par l'analyse et l'exploration des données. L'approche générale proposée basée sur SVM permet de manipuler les non linéarités du système éolien et d'obtenir des résultats très satisfaisants.

Au début du chapitre, une revue bibliographique est présentée, puis une description du système éolien et son modèle mathématique est donné. Ensuite la stratégie de diagnostic de défaut basée sur SVM est expliquée. Par la suite, les résultats de cette approche sont montrés ainsi qu'une comparaison avec différentes méthodes de diagnostic. A la fin du chapitre, une combinaison entre une approche basée sur les données « SVM » et une autre basée sur le modèle « observateur de type Kalman » est présentée.

### **III.1.1. Revue des méthodes de diagnostic sur les éoliennes**

Même les éoliennes de petite taille demandent un investissement considérable, afin de rendre le rapport production d'énergie-prix de l'éolienne intéressant et d'encourager la croissance de l'énergie produite par ces machines. La détection des défauts au tout début de leur apparition est nécessaire et fondamentale pour l'efficacité et la fiabilité de ces systèmes en évitant la dégradation des matériaux et d'autres effets secondaires. Ainsi, la conception automatisée d'un système FDI pour l'éolienne visant à minimiser au maximum le besoin de décision de l'opérateur est nécessaire (Svård and Nyberg, 2011). Dans ce sens (Echavarria et al., 2007) propose une méthodologie basée sur le concept de reconfiguration par l'utilisation des redondances fonctionnelles qui existent dans le système et les informations sur le système ainsi que les interactions entre ses différentes composantes afin de garder l'éolienne en marche même si elle perd en capacité de production.

Différentes approches basées sur les modèles sont utilisées pour le FDI dans les éoliennes. Un observateur à entrée inconnue est utilisé pour détecter les défauts de capteurs de vitesse angulaire de rotor ( $w_r$ ) et de la génératrice ( $w_g$ ) (Odgaard and Stoustrup, 2010) et un observateur linéaire à Paramètre Variant sert à estimer la sortie et à adapter le seuil de détection. La tâche d'isolation est effectuée par le concept de sensibilité des résidus de défauts à travers une matrice de

sensibilité de défaut (Negre et al., 2011). Une autre approche linéaire est celle qui utilise la transformation de Coleman et l'identification de sous espaces pour obtenir un modèle linéaire à temps invariant (LTI) (Wei et al., 2009), permettant de construire un filtre de Kalman qui génère des résidus avec comme but la détection des défauts. Cependant, ces méthodes linéaires ont des difficultés à cause de la non-linéarité du comportement entrée-sortie du système éolien.

Des approches basées sur les données ont aussi été appliquées pour la surveillance et la détection de défauts, comme par exemple l'utilisation des réseaux de neurones qui imbrique une forme non linéaire (Brandao et al., 2010, Elhor et al., 1999). L'approche réseau de neurones a été également utilisée pour l'estimation de la puissance générée par l'éolienne (Li et al., 2001). Une comparaison entre un réseau de neurones et un modèle de régression polynomiale a montré la supériorité du réseau de neurones dans la détection de défauts dans une éolienne (Schlechtingen and Ferreira Santos, 2011).

Beaucoup de travaux s'intéressent au diagnostic de défauts sur une partie ou un composant d'éolienne:

- Détection de défaut des capteurs d'angle de calage : une méthode a été proposé par (Simani et al., 2011) utilisant un système de modélisation s'appuyant sur les modèles affinés par morceaux. Chacun d'eux est valable dans une région de fonctionnement, dont les paramètres sont identifiés à travers les acquis d'entrées-sorties des mesures touchées par l'incertitude de mesure. Par conséquent, ces modèles sont exploités pour générer des signaux résiduels.
- Détection de défaut du multiplicateur de vitesse : dans (Lu, 2012) des ondelettes extraient des caractéristiques liées au défaut. Par la suite ces caractéristiques avec les données de différents capteurs sont fusionnées par une approche statistique de pondération basée sur l'ACP. Ces données fusionnées sont utilisées comme entrée pour un algorithme de décision sur les défauts.
- Détection de défaut du convertisseur : (Odgaard and Stoustrup, 2009) utilise un observateur à entrée inconnue.
  - Détection de défaut du Frein mécanique (Entezami et al., 2012)
  - Fissure des dents dans un engrenage : (Barszcz and Randall, 2009) utilise le Kurtosis Spectral, un outil statistique qui peut indiquer la présence d'une série de phénomènes transitoires non gaussiens et leurs emplacements dans le domaine fréquentiel.
  - Détection de défaut des Roulements dans une génératrice à induction à double alimentation : (Amirat et al., 2010a), (Amirat et al., 2010b) utilise la transformation de Hilbert comme une méthode de démodulation d'amplitude du courant.

Des revues qui décrivent l'état de l'art en terme de surveillance de l'état de l'éolienne ainsi que le diagnostic de défaut peuvent être trouvées dans (Hameed et al., 2009), (Amirat et al., 2009) et (Lu et al., 2009). Ces travaux mettent aussi l'accent sur les outils d'analyse et prétraitement temps–fréquence, tels que les ondelettes. La transformée d'ondelettes permet non seulement d'éliminer le bruit, mais aussi de bien analyser les signaux non stationnaires comme celui de la puissance. Elle a été utilisée pour la surveillance d'état et le diagnostic de défauts mécaniques et électriques (comme : le déséquilibre du rotor, un court-circuit dans le générateur, l'enroulement de générateur) pouvant se produire sur une éolienne à générateur synchrone (Yang et al., 2008). Aussi, cette transformée est utilisée pour distinguer entre le situation normale et défectueuse afin de détecter les défauts de capteurs de position « *pitch* », d'actionneurs ainsi que des défauts dans le système d'entraînement de l'éolienne (Dinesh Kumar et al., 2011).

Dans le printemps de l'année 2010, une compétition internationale était lancée par la société kk-electronic, MathWorks et l'université d'Aalborg afin de trouver la meilleure approche de détection et isolation des défauts dans une turbine éolienne. Plusieurs solutions ont été développées pour relever le défi posé couvrant différentes méthodes à savoir :

- Une approche basée sur le rapport de vraisemblance généralisée « *the Generalized Likelihood Ratio en anglais (GLR)* » (Kiasi et al., 2011)
- Une solution basée sur un compteur réversible « *Up-down counter* » (Ozdemir et al., 2011)
- Une solution basée sur l'estimation par observateur (Zhang et al., 2011)
- Une solution basée sur des modèles affines (Simani et al., 2011) .
- Une solution basée sur les équations de parité qui génère des résidus primaires, suivie d'une deuxième génération/filtrage réalisée par une optimisation (Pisu and Ayalew, 2011).
- Une méthode basée sur les données (Dong and Verhaegen, 2011)
- Une solution basée sur un observateur où un filtre de Kalman est utilisé pour générer les résidus (Chen et al., 2011).
- Notre stratégie (Laouti, et al., 2011a) a démontré son efficacité et précision et a été classée en première place (voir annexe III.1).

Pour plus de détails concernant cette compétition merci de voir le lien suivant :

<http://www.kk-electronic.com/Default.aspx?ID=9442&Action=1&NewsId=43&PID=34329>

Très peu d'études utilisent SVM dans les éoliennes. Une étude pas très poussée qui traite le diagnostic de défauts uniquement dans le système de transmission de l'éolienne par l'utilisation de SVM et la sélection de caractéristiques de vibration du signal de l'arbre principal est donnée par (An et al., 2011). Des modèles de prédiction de défauts ont été développés par le choix de la vitesse du vent ainsi que de la puissance de sortie comme entrées de SVMs (Kusiak and Li, 2011). Mais, d'après les auteurs, la prédiction de ces modèles est un peu basse pour être acceptable.

### III.1.2. Revue sur la commande des éoliennes

La commande des éoliennes est indispensable non seulement pour satisfaire les performances désirées en production d'électricité et maintenir un bon régime de fonctionnement, mais aussi pour diminuer les contraintes mécaniques et la charge quand le vent est fort.

L'objectif de commande est d'étendre la durée de vie d'une éolienne et de réduire les coûts de production d'électricité, ainsi que d'assurer l'optimisation et la régulation de puissance des éoliennes (Boukhezzar, 2006). Plusieurs travaux utilisant différentes méthodes de commande ont été appliqués à la commande de l'éolienne :

- Approche basée sur la construction de contrôleur Linéaire à Paramètre Variant (Sloth et al., 2011, Østergaard et al., 2009)
- Une commande linéaire quadratique gaussienne optimale avec une commande anticipative « *feedforward* » utilisant des signaux d'estimation du vent permettant de rejeter les perturbations et de réduire la charge sur les pales (Selvam et al., 2009).
- Une commande à mode glissant pour réguler la puissance de production dans une éolienne à vitesse variable (Beltran et al., 2009).
- Une commande anticipative utilisant une base orthogonale qui permet d'extraire les tendances générales dans le signal de vent (Odgaard and Stoustrup, 2011).
- Une commande adaptative et une commande floue (Musyafa et al., 2011).
- Une commande prédictive non linéaire (Henriksen et al., 2011).

Le système de contrôle cherche à trouver le meilleur angle de calage des pales (optimisation) à l'aide de vérins hydrauliques autour d'un axe longitudinal pour profiter au maximum du vent instantané et maximiser ainsi la production de puissance, tout en limitant la puissance pour des vitesses de vent supérieures à la vitesse nominale. La variation de l'angle de calage entraîne une

diminution ou une augmentation de la portance de la pale, donc du couple moteur (Biegel et al., 2011).

Un bon tutoriel sur les dynamiques et le contrôle des éoliennes qui décrit l'évolution récente dans les contrôleurs avancés ainsi qu'un aperçu de nombreux problèmes ouverts dans les domaines de la modélisation et du contrôle des éoliennes peut être consulté dans (Pao and Johnson, 2009).

Le système de contrôle doit avoir la capacité de travailler même si l'éolienne est exposée à des défauts. Une commande tolérante aux défauts doit être mise en œuvre indépendamment de la structure de la commande nominale (Esbensen and Sloth, 2009).

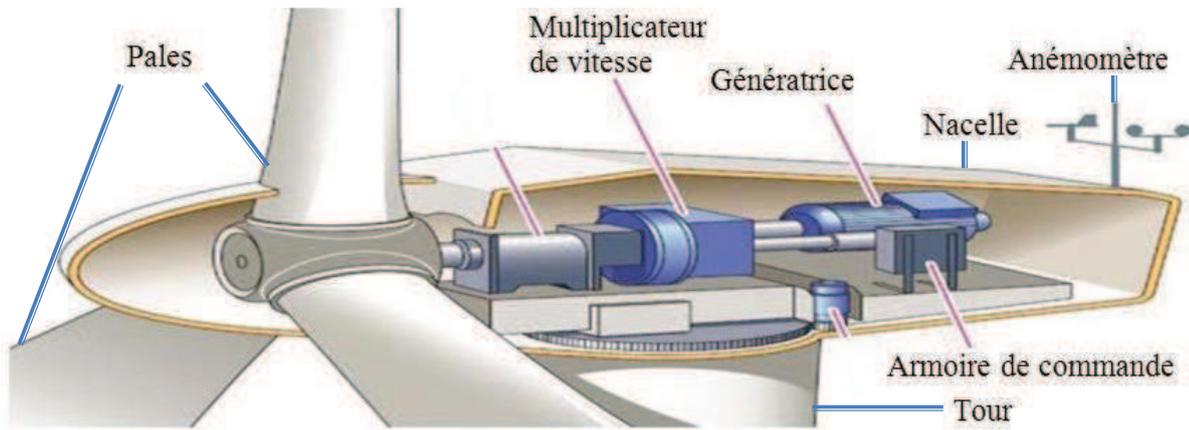
La détection de défauts est également intéressante pour la reconfiguration de commande pour garantir une puissance optimale en cas de défaut partiel. Pour plus d'informations sur les progrès récents dans le diagnostic de défauts et la commande tolérante aux défauts dans les éoliennes, le lecteur est conseillé de voir (Pourmohammad and Fekih, 2011).

## **III.2. Description du système éolien**

Afin de fournir quelques connaissances basiques sur l'éolienne et ces composants, la liste qui suit montre les différentes composantes d'une éolienne (Figure III.3) :

- Le rotor qui comprend généralement 3 pales est fixé à la nacelle.
- La Tour ou le mât permet de placer le rotor à une hauteur suffisante pour permettre son mouvement et qu'il soit entraîné par le vent.
- La nacelle montée au-dessous de la Tour, peut tourner pour orienter la machine éolienne dans la bonne direction (perpendiculaire à la direction du vent). Elle accueille l'installation de génération de l'énergie électrique, qui se compose de : multiplicateur de vitesse, la génératrice (permet de transformer l'énergie mécanique en énergie électrique), armoire de commande,... etc.

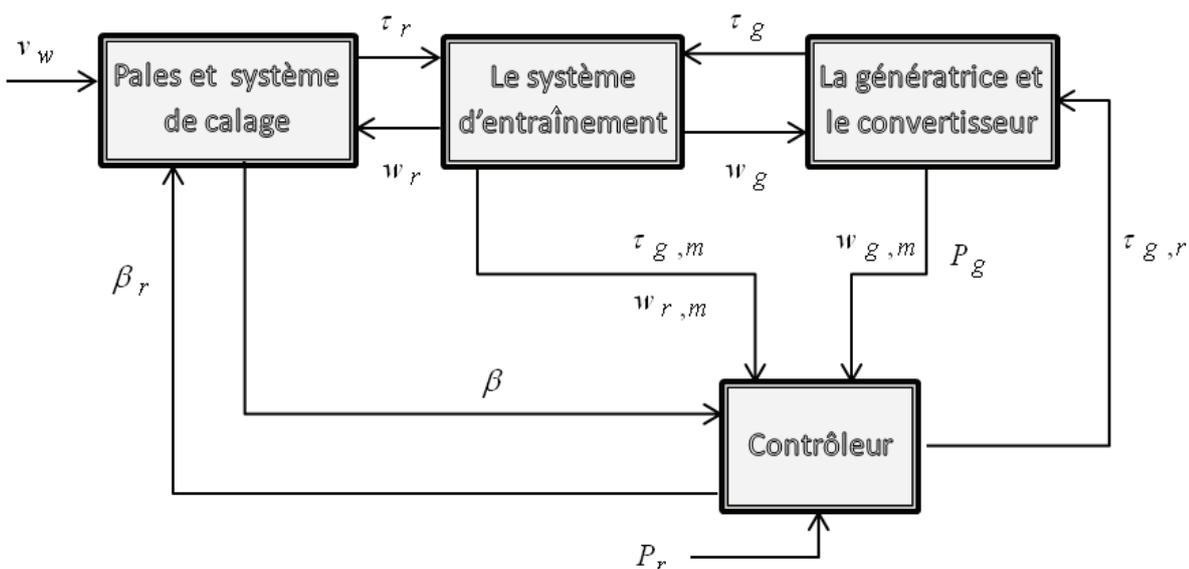
Un anémomètre est placé au niveau de la nacelle pour mesurer la vitesse du vent qui a un impact important sur les dynamiques de l'éolien. Il faut noter que cette mesure est très bruitée, une estimation efficace de la valeur de vitesse du vent pouvant être réalisée par l'utilisation de l'angle du pitch et un observateur (Østergaard et al., 2007).



**Figure III.3 Coupe transversale d'une nacelle d'éolienne**

Il faut noter que la modélisation des éoliennes tient compte de l'endroit de son implantation, une modélisation de l'éolienne placée en terre ferme étant différente de celle placée dans la mer offshore (Jonkman, 2009, Matha, 2009). Dans ce travail, le modèle de référence utilisé est supposé pour un éolien en terre introduit par (Odgaard et al., 2009) incluant des défauts de capteurs, d'actionneurs et du système afin d'évaluer les stratégies de diagnostic de défauts. Ce modèle décrit un type spécifique d'éoliennes à vitesse variable à axe horizontal d'une puissance nominale de 4.8 Mw qui est une hélice perpendiculaire au vent, composée de trois pales et d'un convertisseur complet.

La modélisation de l'éolienne représente principalement ses caractéristiques aérodynamiques, mécaniques et électrotechniques. Un schéma-bloc (Figure III.4) est donné pour montrer la structure du modèle éolien utilisé dans ce chapitre.



**Figure III.4 Une vue d'ensemble du modèle de référence de l'éolienne**

Ce modèle contient un rotor avec les pales qui permettent de capter l'énergie cinétique du vent. Un train d'entraînement est utilisé pour augmenter la vitesse de rotation du rotor au générateur. Un convertisseur intégral couplé à un générateur permet de convertir l'énergie mécanique en énergie électrique. Un contrôleur cherche à optimiser le fonctionnement de l'éolienne.

### III.2.1. Modèle des pales et pitch

Ce système est une combinaison du modèle aérodynamique et du modèle de calage des pales.

**Modèle aérodynamique de l'éolienne** (Johnson et al., 2006) : le couple aérodynamique qui agit sur les pales est donné comme suit

$$\tau_r(t) = \frac{\rho \cdot \pi \cdot R^3 C_q(\lambda(t), \beta(t)) v_w(t)^2}{2} \quad (\text{III.1})$$

Où  $C_q(\lambda, \beta)$  est le coefficient de couple aérodynamique,  $\beta$  est l'angle de calage des pales « pitch » et  $R$  le rayon de pales,  $v_w$  est la vitesse du vent,  $\rho$  est la densité de l'air et  $\lambda = \frac{w_r \cdot R}{v_w}$  correspond au rapport de vitesse à la pointe (vitesse spécifique).

**Modèle du système de calage des pales « pitch »** : ce système actionneur hydraulique peut être modélisé par une fonction de transfert de 2<sup>ème</sup> ordre avec des contraintes sur chaque pale :

$$\frac{\beta_{k,m}(s)}{\beta_k^d(s)} = \frac{w_n^2}{s^2 + 2 \cdot \zeta \cdot w_n s + w_n^2} \quad (\text{III.2})$$

Où  $\beta_{k,m}(s)$  et  $\beta_k^d(s)$  sont les positions mesurées et désirées d'angle de calage des pales.

$k=1, 2, 3$ ,  $[w_n, \zeta] = [11.11 \text{ rad/s}, 0.6]$  sont les paramètres nominaux du modèle sachant que  $w_n$  est la pulsation et  $\zeta$  est le coefficient d'amortissement. La baisse de pression est réalisée par le changement des paramètres du modèle vers les valeurs  $w_{n2}, \zeta_2$ .

### III.2.2. Modèle convertisseur et génératrice

Les dynamiques du convertisseur sont modélisées par une fonction de transfert du 1<sup>ème</sup> ordre:

$$\frac{\tau_{g,m}(s)}{\tau_{g,r}(s)} = \frac{\alpha_{gc}}{s + \alpha_{gc}} \quad (\text{III.3})$$

Où  $\alpha_{gc} = 50 \text{ s}^{-1}$  est le paramètre du modèle. On peut l'écrire sous cette forme

$$\frac{\tau_{g,m}(s)}{\tau_{g,r}(s)} = \frac{1}{\tau \cdot s + 1} \quad (\text{III.4})$$

Où  $\tau_{g,m}$  et  $\tau_{g,r}$  sont respectivement le couple réel et celui désiré, et la constante du temps est  $\tau=0.02$  s. Le couple réel étant non mesuré, il est calculé à partir de la vitesse du générateur mesuré  $w_{g,mi}$ . La puissance produite par la génératrice est :

$$P_g(t) = \eta_g w_g(t) \tau_g(t) \quad (\text{III.5})$$

Où  $\eta_g$  est le rendement de la génératrice.

### III.2.3. Modèle du système d'entraînement

Le système d'entraînement « *drive train* » est modélisé par un modèle à 3 variables d'états en tenant compte des inerties du générateur et du rotor. Les dynamiques de ce système sont données comme suit :

$$\begin{bmatrix} \dot{w}_r(t) \\ w_g(t) \\ \theta_\Delta(t) \end{bmatrix} = A \begin{bmatrix} w_r(t) \\ w_g(t) \\ \theta_\Delta(t) \end{bmatrix} + B \begin{bmatrix} \tau_r(t) \\ \tau_g(t) \end{bmatrix} \text{ avec } A = \begin{bmatrix} \frac{-B_{dt} - B}{J_r} & \frac{B_{dt}}{N_g J_r} & \frac{-K_{dt}}{J_r} \\ \frac{\eta_{dt} B_{dt}}{N_g J_g} & \frac{-\eta_{dt} B_{dt}}{N_g^2} - B_g & \frac{\eta_{dt} K_{dt}}{N_g J_g} \\ 1 & \frac{-1}{N_g} & 0 \end{bmatrix}, B = \begin{bmatrix} \frac{1}{J_r} & 0 \\ 0 & \frac{-1}{J_g} \\ 0 & 0 \end{bmatrix} \quad (\text{III.6})$$

Où  $J_r, J_g$  sont respectivement le moment d'inertie de l'arbre à faible vitesse et à grande vitesse.  $K_{dt}$  est la rigidité de torsion du train d'entraînement,  $B_{dt}$  le coefficient de torsion amorti du train d'entraînement,  $B_g$  le frottement visqueux de l'arbre à grande vitesse,  $N_g$  le rapport de transmission ou d'engrenage,  $\theta_\Delta(t)$  est l'angle de torsion du train d'entraînement, et  $\eta_{dt}$  le rendement du train d'entraînement.

Le défaut système de faible transmission est modélisé par un autre paramètre  $\eta_{dt_2}$ .

### Modèle des capteurs

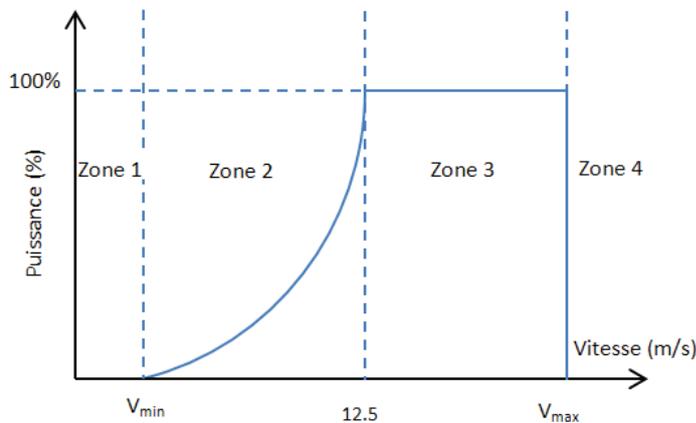
Les capteurs sont modélisés par la valeur de la variable réelle ajoutée avec un bruit stochastique réalisé par le bloc de bruit aléatoire à bande limitée dans Matlab. Les valeurs des paramètres du modèle de l'éolienne de référence sont données dans le Tableau III.1.

**Tableau III.1 Paramètres du modèle de référence de l'éolienne**

Paramètres du modèle pales et pitch								
R	$\rho$	$\zeta$	$w_n$	$\zeta_2$	$w_{n2}$	$\zeta_3$	$w_{n3}$	
57.5 m	1.225 Kg/m <sup>3</sup>	0.6	11.11 rad/s	0.45	5.73 rad/s	0.9	3.42 rad/s	
Paramètres du modèle de transmission								
B <sub>dt</sub>	B <sub>r</sub>	B <sub>g</sub>	N <sub>g</sub>	K <sub>dt</sub>	$\eta_{dt}$	$\eta_{dt_2}$	J <sub>g</sub>	J <sub>r</sub>
775.49 $\frac{N.m.s}{rad}$	7.11 $\frac{N.m.s}{rad}$	45.6 $\frac{N.m.s}{rad}$	95	2.7 10 <sup>9</sup> $\frac{N.m.s}{rad}$	0.97	0.92	390 Kg m <sup>2</sup>	55.10 <sup>6</sup> Kg m <sup>2</sup>
Paramètres du modèle convertisseur et génératrice								
$\alpha_{gc}$				$\eta_{gc}$				
50 s <sup>-1</sup>				0.98				

### III.2.4. Le système de contrôle

La vitesse du vent (précisément sa moyenne) détermine la zone de fonctionnement du contrôleur et par conséquent de la machine éolienne. La Figure III.5 illustre bien différentes zones. Lorsque la vitesse est inférieure à une certaine valeur  $v_{min}$ , l'éolienne est à l'arrêt (zone1). Si  $v_{min}$  est dépassée mais inférieure à 12.5 m/s, le contrôleur fonctionne dans la zone 2 qui correspond à l'optimisation de la puissance. La zone 3 correspond à la saturation (valeur constante) dans la production de puissance quand la vitesse du vent est entre 12.5m/s et  $v_{max}$ . Le dépassement de la valeur  $v_{max}$  par la vitesse du vent peut avoir des conséquences néfastes sur l'éolienne. C'est pour cette raison que dans cette zone 4 l'éolienne est arrêtée.



**Figure III.5 Puissance de l'éolienne en fonction de la vitesse du vent**

Par la suite, la détection de défauts sera étudiée à l'aide de la simulation en boucle fermée dans la zone 2 (optimisation de la puissance) et la zone 3 (production constante d'énergie) avec une séquence réaliste de vent de 4400s montrée dans la Figure III.6.

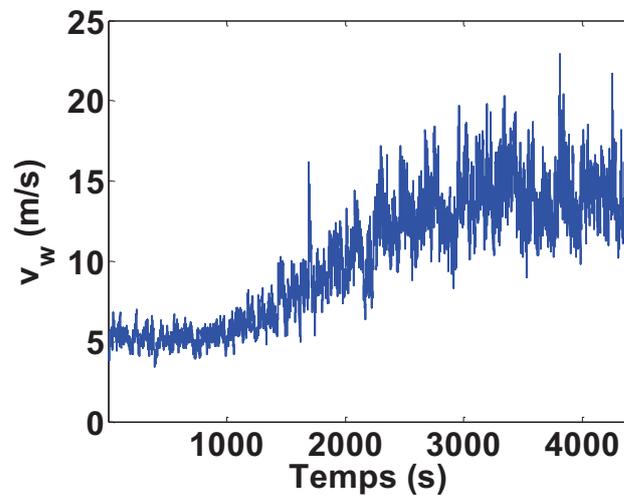


Figure III.6 Séquence vitesse du vent

### III.2.5. Défauts considérés dans ce modèle éolien

#### III.2.5.1. Défauts capteurs

Le système est équipé de capteurs dupliqués permettant de mesurer les trois positions angulaires de calage des pales ( $\beta_{k,m_i}$ ,  $k=1,2,3$ ,  $i=1,2$ ) et les vitesses de la génératrice et du rotor ( $w_{g,m_i}$ ,  $w_{r,m_i}$ ,  $i=1,2$ ). Cela donne un total de dix capteurs tous soumis à deux types de défauts: valeur fixe et gain. Par conséquent, vingt défauts doivent être détectés avec un temps de détection ( $T_D$ ) qui doit être inférieur à 10 fois le temps d'échantillonnage ( $T_e=0,01$  s).

#### III.2.5.2. Défauts actionneurs

Les actionneurs sont les trois systèmes hydrauliques « pitch » et le convertisseur. Ils permettent respectivement le calage des pales et le réglage du couple générateur pour commander la vitesse de rotation du générateur et le rotor. Ces actionneurs sont également soumis à un dysfonctionnement. Le système de convertisseur de couple peut avoir un décalage qui doit être détecté rapidement ( $T_D < 5.T_e$ ). Les trois systèmes de calage des pales peuvent avoir un changement dans la dynamique qui peut être dû à une chute de pression dans la pompe (5a) ou à haute teneur en air dans l'huile à un rythme plus lent (5b) (voir le Tableau III.2). Dans ce cas, le nombre total de défauts d'actionneurs est de sept.

**III.2.5.3. Défaut système** une panne de système peut se produire dans la conduite d'un train en raison du changement de frottement avec le temps. Le nombre total de défauts devant être contrôlés est donc de 28.

Cependant, certains défauts sont similaires. Par exemple, un défaut sur le capteur de mesure de position de la pale 1 « *blade 1* » ( $\Delta\beta_{1,m_i}$ ) est semblable à ceux dans les capteurs de mesure des positions des pales 2 et 3. De cette manière, on peut voir que nous avons dix différents types de défauts à prendre en considération distinctement qui sont classés dans le Tableau III.2. Le processus a d'autres capteurs, mesurant par exemple la vitesse du vent, qui ne sont pas supervisés dans ce travail.

**Tableau III.2 Défauts considérés dans le modèle éolien de référence**

N.	Défaut	Symbole	Type	Exigences
1a	Capteur	$\Delta\beta_{k,m_i}, k = 1,2,3, i = 1,2$	Valeur fixe	$T_D < 10.T_e$
1b	Capteur	$\Delta\beta_{k,m_i}, k = 1,2,3, i = 1,2$	Gain	$T_D < 10.T_e$
2a	Capteur	$\Delta w_{r,m_i}, i = 1,2$	Valeur fixe	$T_D < 10.T_e$
2b	capteur	$\Delta w_{r,m_i}, i = 1,2$	Gain	$T_D < 10.T_e$
3a	Capteur	$\Delta w_{g,m_i}, i = 1,2$	Valeur fixe	$T_D < 10.T_e$
3b	Capteur	$\Delta w_{g,m_i}, i = 1,2$	Gain	$T_D < 10.T_e$
4a	Actionneur	$\Delta\tau_g$	Offset	$T_D < 5.T_e$
5a	Actionneur	$\Delta\beta_k, k = 1,2,3$	Changement brusque de dynamique dû à une chute de pression dans la pompe	$T_D < 8.T_e$
5b	Actionneur	$\Delta\beta_k, k = 1,2,3$	Changement lent de dynamique dû à la présence de l'air dans l'huile	$T_D < 100.T_e$
6	Système	$\Delta w_r, \Delta w_g$	Changement de dynamique dû à l'augmentation de vibration dans le système de transmission	-

Il est à noter que l'exigence en termes de rapidité de temps de détection du défaut dépend de la gravité de ce dernier. Et le nombre de fausses alarmes doit être le plus faible possible.

### III.3. Diagnostic de défauts dans une éolienne

Dans ce chapitre, une méthode de diagnostic de défauts par SVM est proposée. SVM classification est utilisé pour évaluer les résidus générés ainsi que d'autres caractéristiques afin de conclure sur l'état de fonctionnement du système. Il est essentiel de bien comprendre le système éolien et comment il fonctionne avant d'envisager de réaliser un schéma de diagnostic de défaut (Figure III.7). L'idée derrière ce schéma de diagnostic de défaut dédié aux éoliennes est de voir le problème de diagnostic comme un problème de classification.

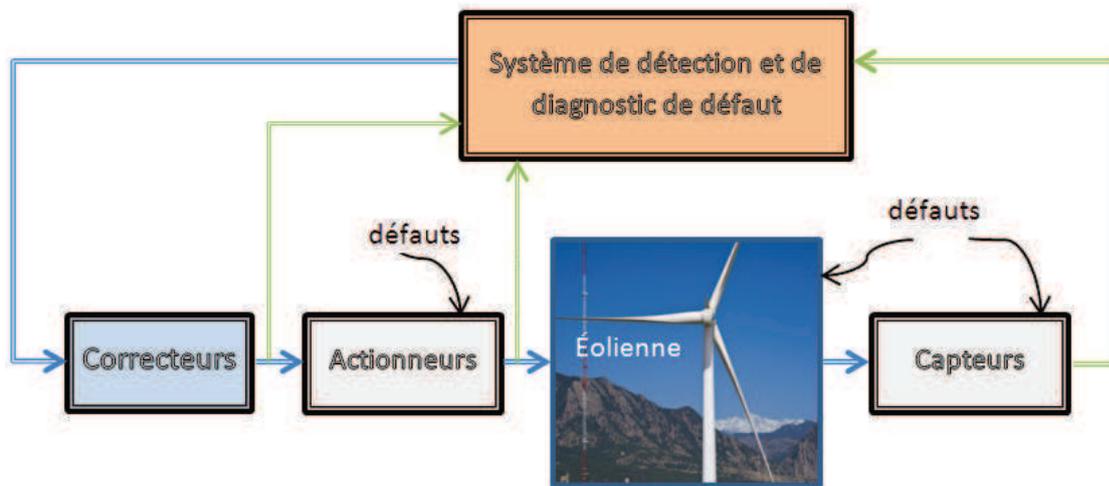
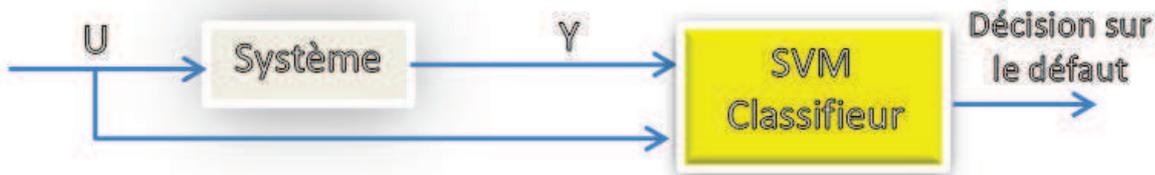


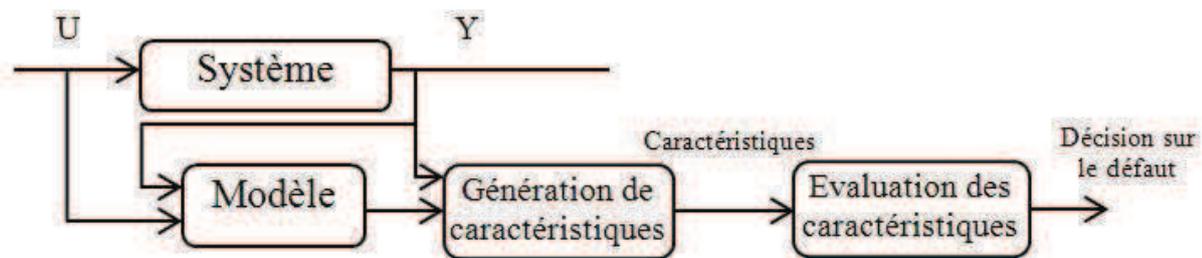
Figure III.7 Schéma de diagnostic de défaut pour les éoliennes

La méthode SVM a été utilisée pour le diagnostic de défaut dans la majorité des cas par l'utilisation directe des signaux d'entrées/sorties (Yélamos et al., 2009, Ge et al., 2004) comme montré dans la Figure III.8 ou par l'ajout d'une phase de prétraitement avant l'application des SVMs : transformée d'ondelettes (Saravanan et al., 2008), transformé de Walsh (Xiang et al., 2008). Le deuxième choix apparaît le plus approprié car le premier schéma risque d'embrouiller le travail de l'algorithme SVM qui va chercher la solution dans toutes les directions. Dans la littérature, des chercheurs ont essayé de combiner les SVMs avec d'autres techniques de prétraitement telles que l'analyse en composantes principales (Widodo and Yang, 2008, Jemwa and Aldrich, 2006). Bien que cette combinaison améliore un peu les résultats des SVMs appliqués sur les données brutes, des fausses alertes restent toujours présentes. Cela peut se comprendre car ACP cherche à compresser les données brutes vers une dimension plus réduite par la capture du maximum de variance des données brutes sans s'intéresser aux symptômes des défauts.



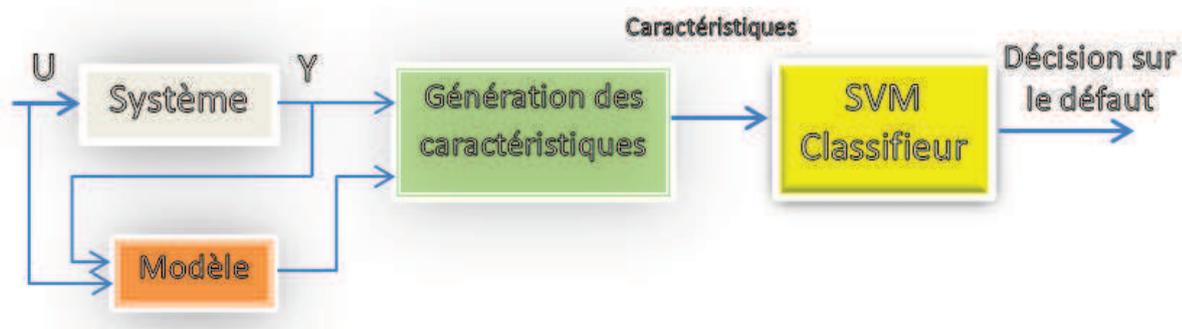
**Figure III.8 Stratégie de diagnostic de défaut classique avec SVM**

Dans cette thèse, je propose de voir le processus de détection de défauts d'une façon plus générale que celle de génération de résidu suivie par l'évaluation de résidu montrée dans le chapitre I. Il s'agit de créer des caractéristiques spécifiques à chaque défaut. Parmi ces caractéristiques se trouve le résidu. On peut avoir des mesures pertinentes informatives de défaut, combinaison des mesures, mesures filtrées, etc., la Figure III.9 suivante illustre la vision proposée pour résoudre le problème de diagnostic des différents systèmes.



**Figure III.9 Schéma de diagnostic de défauts proposé**

La construction des caractéristiques dans ce travail appliqué au cas des éoliennes est réalisée d'une façon manuelle en se basant sur les connaissances du système éolien. Par la suite, SVM est utilisé comme une étape d'évaluation des caractéristiques d'un défaut donné comme le montre la Figure III.10.



**Figure III.10 Stratégie proposée de diagnostic de défaut basée sur SVM**

La détection de défauts par SVM est développée en deux parties. En premier lieu, un ensemble de données avec et sans défauts est utilisé pour apprendre les modèles de détection de chaque défaut par l'utilisation d'une séquence de vent donnée comme entrée. Les modèles obtenus sont validés sur un nouveau scénario de défauts comme dans les autres applications étudiées dans cette thèse.

### III.3.1. Les vecteurs des caractéristiques pour chaque défaut

L'idée clé dans l'apprentissage d'un nouveau modèle pour la détection de défaut par SVM est la définition du vecteur  $x$  à utiliser pour la classification. Ce vecteur doit inclure l'information la plus pertinente du comportement du système relative à un défaut donné.

Il peut contenir les sorties mesurées, des mesures décalées, les entrées, les consignes, une combinaison de celles-ci ou la variation des sorties avec le temps. Afin de construire un vecteur utile, il faut observer attentivement les résultats des sorties du processus pour chaque anomalie et proposer une combinaison qui assure un fort impact de défaut considéré dans le vecteur  $x$ .

Plusieurs vecteurs sont proposés pour différents types de défauts. La redondance matérielle (capteurs dupliqués) est utilisée comme une caractéristique dans la plupart de ces vecteurs. Le noyau utilisé pour l'apprentissage de tous les défauts est le noyau gaussien qui est le plus adapté aux données réelles bruitées et perturbées grâce à sa grande capacité d'apprentissage. La plupart des données sont filtrées par un filtre de premier ordre avec une constante de temps  $\tau$  afin de réduire la sensibilité aux perturbations du procédé ou aux bruits de mesures.

#### III.3.1.1. Diagnostic des défauts de capteurs de pitch

Pour les 6 capteurs mesurant les positions pitch ( $\Delta\beta_{k,m_i}$ ,  $k=1,2,3$ ,  $i=1,2$ ), le paramètre  $\sigma$  de la fonction noyau gaussienne est régulé à 10.

**Type valeur fixe (1a) :** pour ces défauts, le vecteur suivant est utilisé pour la détection et l'isolation:

$$x = \begin{bmatrix} \left| \hat{\beta}_{k,m1}(t_j) - \hat{\beta}_{k,m2}(t_j) \right| \\ \left| \beta_{k,m1}(t_j) - \beta_{k,m1}(t_{j-1}) \right| \\ \left| \beta_{k,m2}(t_j) - \beta_{k,m2}(t_{j-1}) \right| \end{bmatrix} \quad (\text{III.7})$$

Où  $t_j$  et  $t_{j-1}$  sont les instants du temps  $j$  et  $j-1$  respectivement et  $\hat{\beta}$  est la valeur filtrée de  $\beta$  utilisant un filtre passe-bas avec une constante de temps  $\tau=0.06$  s. Notez que les valeurs absolues

sont utilisées dans  $x$ . Lorsque  $|\beta_{k,m_i}(t_j) - \beta_{k,m_i}(t_{j-1})| = 0, i = 1,2$ , ce terme est remplacé par une large valeur constante (5000) afin d'améliorer la distinction entre le défaut de type valeur fixe et le cas normal (sans défaut).

**Type gain (1b) :** ce défaut est détecté et isolé en deux étapes. Tout d'abord, le défaut est détecté utilisant le vecteur suivant:

$$x = \begin{bmatrix} |\hat{\beta}_{k,m1}(t_j) - \hat{\beta}_{k,m2}(t_j)| \\ |\hat{\beta}_{k,m1}(t_j) - \hat{\beta}_{k,m1}(t_{j-1})| \\ |\hat{\beta}_{k,m2}(t_j) - \hat{\beta}_{k,m2}(t_{j-1})| \end{bmatrix} \quad (III.8)$$

Les deuxième et troisième lignes dans l'équation ci-dessus sont importantes afin d'exclure les défauts de type 1a. Par la suite, si le défaut de type b est détecté, le vecteur suivant est utilisé pour l'isolation entre les capteurs 1 et 2:

$$x = \begin{bmatrix} |\hat{\beta}_{k,r}(t_j) - \hat{\beta}_{k,m1}(t_j)| \\ |\hat{\beta}_{k,r}(t_j) - \hat{\beta}_{k,m2}(t_j)| \end{bmatrix} \quad (III.9)$$

Où  $\beta_{k,r}$  est la valeur désirée de l'angle pitch  $\beta_k$  et  $\hat{\beta}$  est la valeur filtrée par un filtre passe-bas de constante de temps  $\tau=0.08$  s.

### III.3.1.2. Diagnostic des défauts capteurs des vitesses de génératrice et du rotor

Pour des défauts capteurs dans les vitesses de la génératrice et du rotor ( $w_{g,mi}, w_{r,mi}, i = 1,2$ ), la variance gaussienne est réglée à  $\sigma = 15$  pour améliorer la croissance de la capacité de détection. Il est à noter qu'une grande valeur de  $\sigma$  peut conduire à des fausses alarmes.

**Types valeurs fixes 2a et 3a:** pour la détection des défauts de type a, le vecteur suivant est utilisé pour la détection et l'isolation:

$$x = \begin{bmatrix} |\hat{w}_{p,m1}(t_j) - \hat{w}_{p,m2}(t_j)| \\ |w_{p,m1}(t_j) - w_{p,m1}(t_{j-1})| \\ |w_{p,m2}(t_j) - w_{p,m2}(t_{j-1})| \end{bmatrix}, p = g, r \quad (III.10)$$

$\hat{w}_g$  est obtenu en utilisant un filtre avec  $\tau=0.02$ s et  $\hat{w}_r$  avec  $\tau=0.6$ s. Lorsque le terme  $w_{p,mi}(t_j) - w_{p,mi}(t_{j-1}) = 0, i = 1,2$ , il est remplacé par une grande valeur constante dans le vecteur de caractéristique  $x$  afin d'améliorer la distinction entre le défaut de valeur fixe et le cas normal (sans défaut) et faciliter la tâche du classifieur SVM.

**Types gains 2b et 3b:** pour la détection des défauts de type gain b (sans les défauts de type a), le vecteur suivant est utilisé en première étape :

$$x = \begin{bmatrix} |\hat{w}_{p,m1}(t_j) - \hat{w}_{p,m2}(t_j)| \\ |\hat{w}_{p,m1}(t_j) - \hat{w}_{p,m1}(t_{j-1})| \\ |\hat{w}_{p,m2}(t_j) - \hat{w}_{p,m2}(t_{j-1})| \end{bmatrix}, p = g, r \quad (\text{III.11})$$

Les mesures sont filtrées par un filtre passe-bas avec une constante de temps égale à  $\tau=0.6$  s pour l'estimation des défauts  $\hat{w}_r$  et  $\tau=0.02$ s pour  $\hat{w}_g$ .

Dans une seconde étape, l'isolation entre les capteurs 1 et 2 dans le cas de défaut de type b est faite en utilisant le vecteur suivant:

$$x = \begin{bmatrix} \frac{\hat{\tau}_{g,m} \times \hat{w}_{p,m1}(t_j) \times c}{\hat{P}_g^m} \\ \frac{\hat{\tau}_{g,m} \times \hat{w}_{p,m2}(t_j) \times c}{\hat{P}_{g,m}} \end{bmatrix}, p = g, r \quad \text{avec } c \text{ une constante } c = \begin{cases} c_3 = 10 \rightarrow p = g \\ c_4 = 1000 \rightarrow p = r \end{cases} \quad (\text{III.12})$$

Où,  $P_{g,m}$  est la mesure de la puissance générée de la génératrice. Les mesures sont filtrées par un passe bas avec une constante de temps égal à  $\tau=0.6$  s pour l'estimation des défauts  $\hat{w}_r$  et  $\tau=0.06$  s pour  $\hat{w}_g$ .

### III.3.1.3. Diagnostic des défauts du convertisseur et du système

Pour les défauts (4a and 6), le vecteur suivant est utilisé:

$$x = \begin{bmatrix} |w_{g,m1}(t_j) - w_{g,m2}(t_j)| \\ |\tau_{g,r}(t_j) - \tau_{g,m}(t_j)| \\ \lambda_2 \times |w_g^d(t_j) - (w_{g,m1}(t_j) + w_{g,m2}(t_j))/2| \end{bmatrix} \quad (\text{III.13})$$

Où  $w_g^d = \frac{P_r}{\tau_{g,r}}$  est la vitesse de la génératrice désirée calculée à partir du couple générateur désiré

$\tau_{g,r}$  obtenu par le contrôleur avec  $P_r$  la puissance désirée. Le facteur  $\lambda_2 = 10^{-10} \times v_w^6$  dans la 3<sup>ème</sup> composante de  $x$  est utilisé pour tenir compte de la vitesse du vent et pour la normalisation (avoir le même ordre de grandeur). Notez également que  $\tau_{g,r}$  est filtré en utilisant un filtre du premier ordre avec une constante de temps  $\tau = 0,02$  s. L'objectif de ce filtre est de prendre en

compte la dynamique du système de contrôle (temps nécessaire pour  $\tau_{g,m}$  pour atteindre  $\tau_{g,r}$ ) et non de rejeter le bruit de mesure ou les perturbations. Le paramètre  $\sigma=10$  pour défaut de type 4a.

### III.3.1.4. Détection / isolation de défaut actionneur sur le système pitch

Pour la détection de défauts 5a et 5b, le vecteur suivant est utilisé avec  $\sigma=10$ :

$$x = \begin{bmatrix} |w_{g,m1}(t_j) - w_{g,m2}(t_j)| \\ |\beta_{k,m1}(t_j) - \beta_{k,m2}(t_j)| \\ |\beta_{k,m1}(t_j) - \beta_{k,m1}(t_{j-1})| \\ |\beta_{k,m2}(t_j) - \beta_{k,m2}(t_{j-1})| \end{bmatrix} \quad (\text{III.14})$$

Une fois les vecteurs d'apprentissage définis pour chaque défaut, les scénarios de défauts sont simulés et à chaque échantillon est attribué  $z = \pm 1$  (avec ou sans faute) pour chaque type de défauts.

## III.3.2. Résultats de l'approche de diagnostic

Environ sept scénarios ont été envisagés pour chaque défaut avec des amplitudes de défauts différents. L'algorithme d'apprentissage SVM utilise les vecteurs de caractéristiques ( $x$ ) et les valeurs  $z$  correspondantes pour identifier  $\alpha_i$  et les vecteurs de support ( $x_{\text{sup}}$ ) nécessaires pour la prise de décision. L'identification dépend du type du noyau et des paramètres de réglage:  $\sigma$  et  $C$ . Un bon réglage des paramètres d'apprentissage de SVM évite le surapprentissage. L'approche essai-test a été adoptée dans cette application. Elle consiste à réaliser plusieurs essais avec différentes valeurs des paramètres  $C$  et  $\sigma$  et de les tester sur un ensemble de test. Notez que le même «modèle» ( $x_{\text{sup}}$  et  $\alpha_i$ ) est utilisé pour tous les défauts de type 1a, un autre «modèle» pour tous les défauts de type 1b et ainsi de suite. Dix «modèles» ont donc été développés.

La solution de diagnostic de défaut proposée dans ce chapitre peut être téléchargée à partir du lien suivant : <http://www.mathworks.com/matlabcentral/fileexchange/35130-award-winning-fdi-solution-in-wind-turbines>

Le scénario des défauts considérés dans ce travail est le suivant :

### Défauts capteurs

Défaut 1 (1a) : la mesure de capteur  $\beta_{1,m1}$  fixé à  $4^\circ$  entre les instants 100s et 200s

Défaut 2 (1b) : la mesure  $\beta_{2,m2}$  multipliée par le facteur 1.8 entre les instants 2500s et 2600s

Défaut 3 (1a) : la mesure  $\beta_{3,m1}$  fixée à  $12^\circ$  entre les instants 900s et 1000s

Défaut 4 (2a) :  $w_{r,m1}$  fixé à 1.2 rad/s entre les instants 1200s et 1300s

Défaut 5 (2b et 3b) :  $w_{r,m2}$  multiplié par le facteur 0.7 et  $w_{g,m1}$  multiplié par le facteur 1.7 entre les instants 1700s et 1800s.

Les Figure III.11 et Figure III.12 montrent respectivement l'évolution en fonction du temps de la vitesse du vent (l'entrée) et la puissance électrique générée (la sortie). Il apparaît clairement qu'il ya une forte liaison entre ces deux variables (la valeur de l'intercorrélation calculée 0.8643 démontre bien cela). Ainsi, la puissance est maximum si la vitesse du vent dépasse une certaine valeur.

Cependant la production de l'énergie est aussi influencée par la commande ainsi que par les défauts. Cela est assez clair dans l'intervalle de temps entre 2000 s et 2200s avec l'introduction du défaut sur le système d'entraînement.

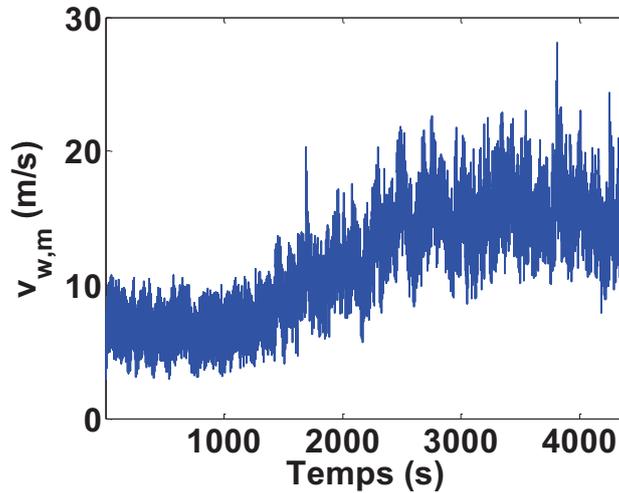


Figure III.11 Vitesse du vent mesurée

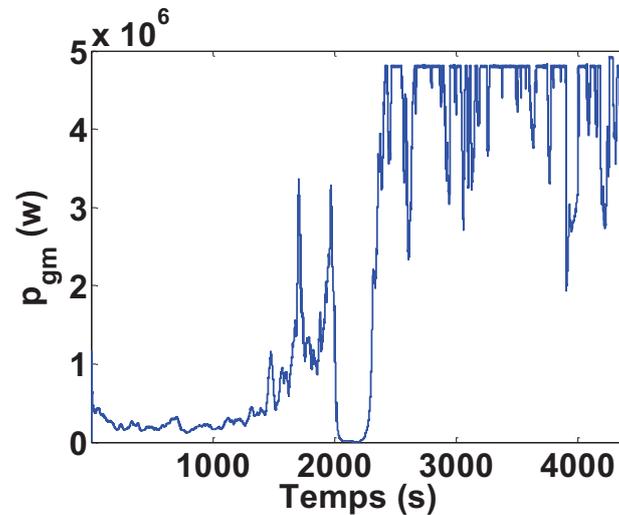


Figure III.12 Puissance électrique générée

### Défauts actionneurs

Défaut 6 (5a) : chute de pression hydraulique dans l'actionneur de «pitch» 2, ce défaut se manifeste par un changement brutal de la dynamique de l'actionneur «pitch» 2 avec  $w_{n2} = 5.73$  et  $\zeta_2 = 0.45$  entre 3200s et 3300s.

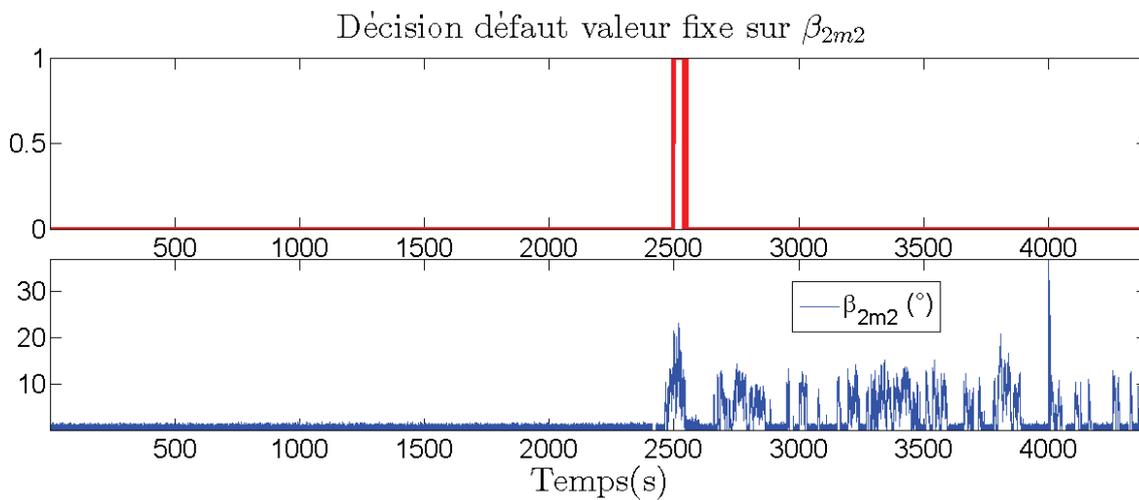
Défaut 7 (5b) : augmentation de la teneur de l'air dans l'huile hydraulique dans l'actionneur 3 entre 3400s-3500s. Ce défaut est introduit par le changement des paramètres du modèle avec les valeurs  $w_{n3} = 3.42$  et  $\zeta_2 = 0.9$ .

Défaut 8 (4a) : le défaut d'actionneur sur le convertisseur est noté  $\Delta\tau_g$  de type offset causé par un décalage dans les boucles internes de contrôle. Ce défaut est introduit entre 4200s et 4300s.

### Défaut sur le système

Défaut 9 (6) : Changement des dynamiques du système d'entraînement entre 2000s et 2200s.

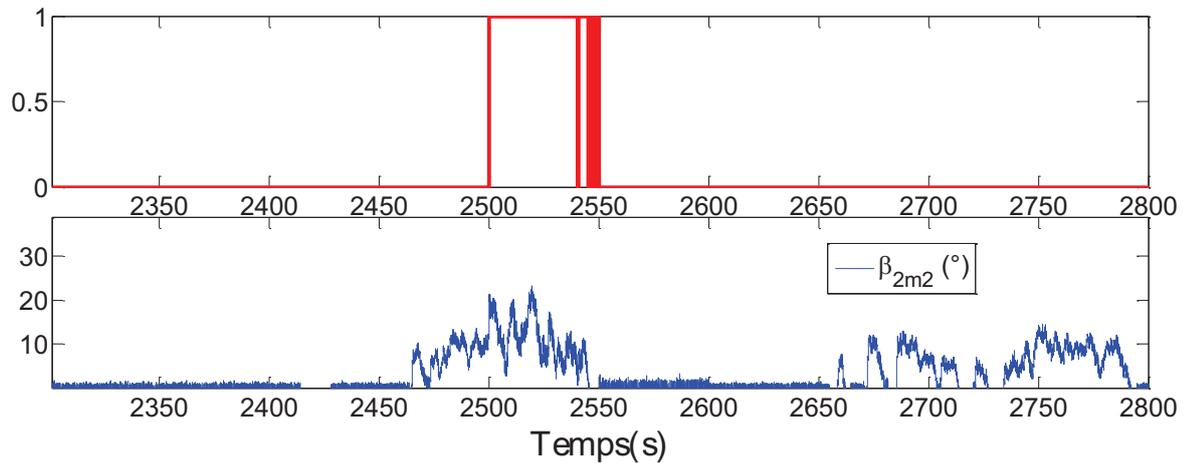
Les défauts (1 et 3) de la position pitch de type valeur fixe peuvent être détectés après seulement 2 périodes d'échantillonnage satisfaisant ainsi le cahier de charge (voir annexe C). La Figure III.13 montre les résultats de décision du défaut de type (gain) ainsi que la mesure du capteur  $\beta_{2,m2}$  qui sont assez satisfaisantes puisque ce défaut est bien détecté et isolé sans fausses alarmes malgré la difficulté liée à ce type de défauts qui peut être confondu facilement avec une situation normale à cause de bruit, perturbations et les oscillations. Cependant, seuls les défauts avec un offset  $\Delta\beta \geq 2^\circ$  peuvent être détectés durant le temps requis ( $< 10.T_e$ ) en fonction de la zone de commande. Si le gain est appliqué plutôt qu'un offset, les oscillations peuvent augmenter si l'ordre de gain est élevé mais la capacité de détection reste la même.



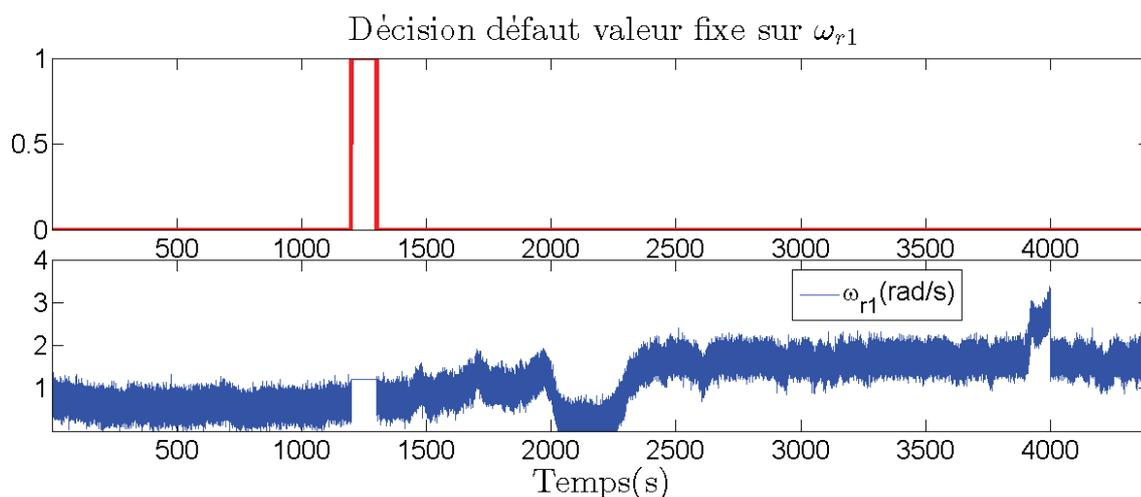
**Figure III.13 Détection et isolation du défaut 2 de type gain (1b) sur le capteur de la position pitch  $\beta_{2,m2}$**

Le zoom ci-dessous sur la Figure III.13 permet de voir plus de détails. Ainsi il est clair que le défaut qui correspond à la multiplication de la mesure  $\beta_{2,m2}$  par le facteur 1.8 entre les instants 2500 s et 2600 s est détecté rapidement à l'instant 2500.03 s (juste après  $3T_e$ ), et bien isolé

puisque les autres signaux de décisions sur les défauts affichent une valeur nulle. Cependant, cette décision sur le défaut 2 indique la fin de ce dernier à l'instant 2550.17 s alors que ce défaut se termine à 2600 s, ce résultat pouvant se comprendre puisque de 2550.17 s à 2600 s comme indique la Figure ci-dessous la mesure  $\beta_{2,m2}$  est proche de zéro et donc ce défaut de type gain va seulement amplifier le bruit.

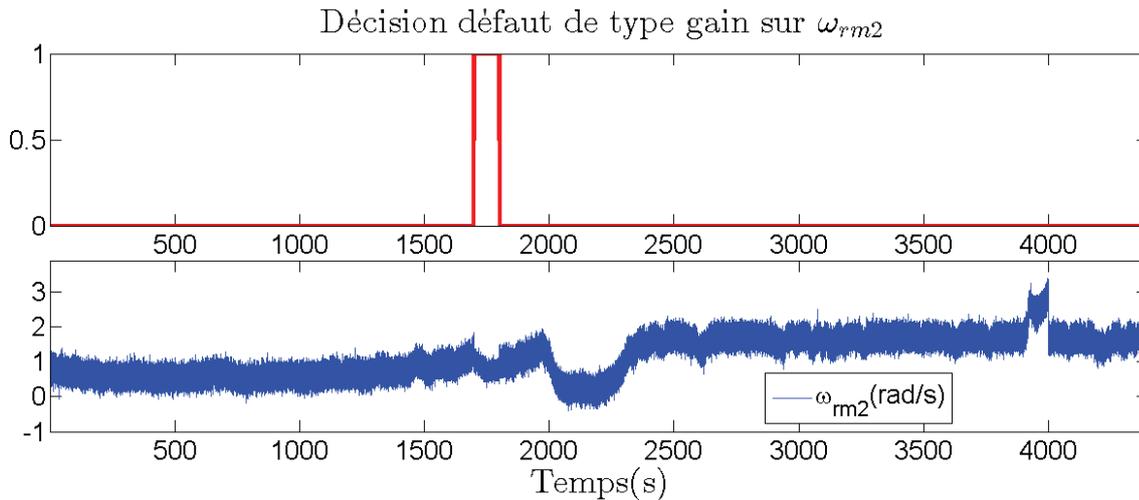


La Figure III.14 indique l'apparition du défaut 4 (2a) de type valeur fixe dans le capteur de la vitesse de rotation du rotor entre 1200 s et 1300 s. On remarque que la détection est réalisée instantanément à l'instant 1200.02s et prend rapidement la valeur zéro après la disparition de ce défaut à l'instant 1300.01 s. Notez toujours que les signaux de décision sur les autres défauts affichent une valeur nulle, ainsi, que la non-présence des fausses alarmes dans le résultat de détection, malgré les bruits et les perturbations.



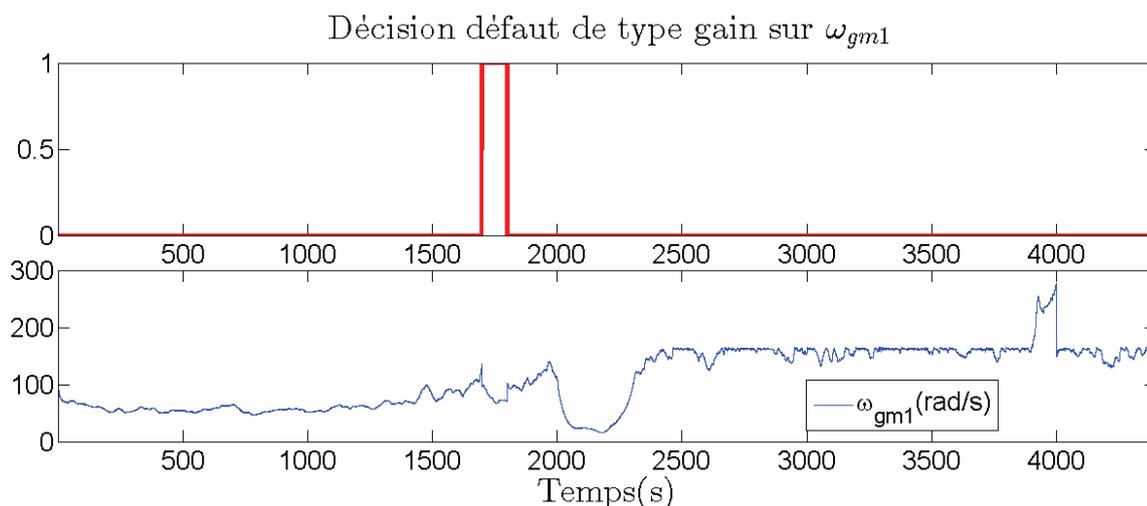
**Figure III.14 Détection et isolation du défaut 4 valeur fixe (type 2a) sur la vitesse du rotor**

Le défaut 5 (2b) de type gain de valeur  $\Delta\omega_{r,m2} = 0.7 \text{ rad/s}$  dans le capteur de vitesse du rotor a pu être détecté à l'instant 1700.37 s après 37 périodes d'échantillonnage de son apparition comme le montre la Figure III.15. La satisfaction des exigences demandées concernant ce type de défaut ( $<10.T_e$ ) ne peut se faire si la valeur de  $\omega_r$  n'est pas proche de zéro et si le facteur de multiplication du défaut a une valeur conséquente qui permet de le différencier du bruit et des perturbations rapidement et sans fausse alarme. Notez que la décision sur le défaut 5 prend la valeur zéro (pas de défaut) à l'instant 1800.5s après 50  $T_e$  de la disparition de ce défaut.



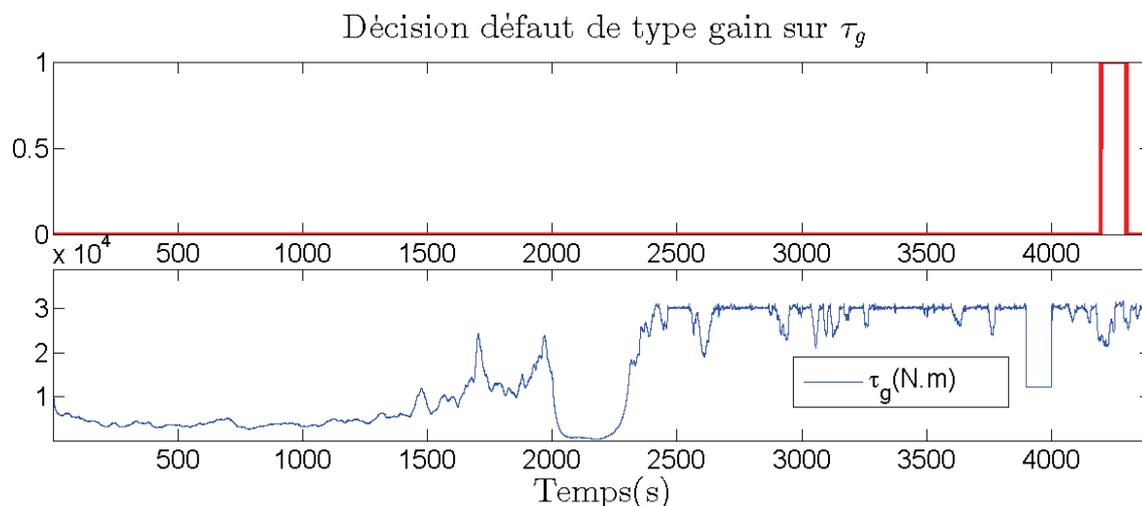
**Figure III.15** Détection et isolation du défaut 5 de type gain (2b) sur le capteur de la vitesse rotationnelle du rotor  $\omega_{r,m2}$

La Figure III.16 montre les résultats de détection et isolation de défaut 5 (3b) du capteur vitesse rotationnel de la génératrice introduit entre les instants 1700s et 1800s. Dans le cas d'une erreur de facteur de gain conduisant à  $\omega_g > 50 \text{ rad/s}$ , un changement important dans le résidu se produit rapidement menant à une détection facile de ce défaut. Cependant, la mesure  $\omega_{g,m1}$  est influencée aussi par des perturbations et d'autres types de défauts (comme le défaut sur le système entre 2000 s et 2200 s). La tâche d'isolation est donc plus compliquée. Malgré tout cela le défaut est bien isolé et sans fausses alarmes. Le résultat sur ce type de défaut est assez bon dû à la rapidité et la justesse de décision qui commence à 1700.03 s après juste 3  $T_e$  de l'apparition de ce défaut et se termine à 1800.06s (après 6  $T_e$  de la disparition du défaut).



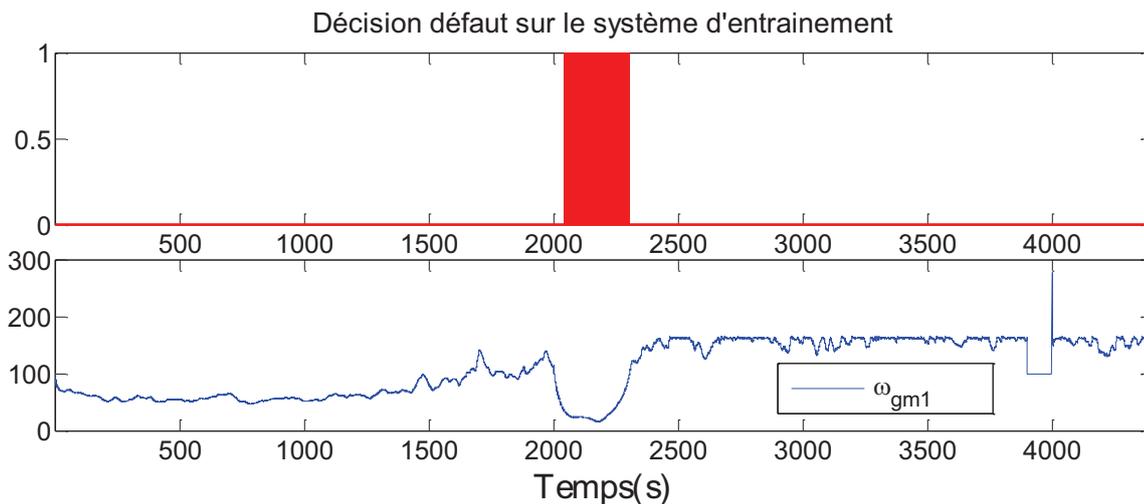
**Figure III.16 Détection et isolation du défaut 5 de type gain (3b) sur le capteur de la vitesse rotationnelle de la génératrice  $\omega_{g,m1}$**

Le défaut 8 (4a) sur l'actionneur du convertisseur a pu être détecté comme désiré en termes d'amplitude de défaut et de la rapidité comme montré dans la Figure III.17. On rappelle que  $x$  utilise notamment la valeur de consigne du couple qui est comparée à celle mesurée à l'aide d'un retard de 2 périodes d'échantillonnage. On remarque que la détection est réalisée instantanément après juste deux période d'échantillonnage à l'instant 4200.02s et prend rapidement la valeur zéro après la disparition ce défaut à l'instant 4300.02s, avec un pourcentage nul de fausses alarmes.



**Figure III.17 Détection et isolation du défaut 8 de type offset (4a) sur le convertisseur**

En ce qui concerne les actionneurs des systèmes pitch, leurs défauts ne pouvaient pas être détectés par le vecteur  $x$  proposé. Ces types de défauts seront traités par une approche hybride combinant SVM avec un observateur de type Kalman (voir section III.5).



**Figure III.18 Détection et isolation du défaut 9 (6) sur le système d'entraînement**

Finalement, la Figure III.18 montre la détection de défaut 9 sur le système de transmission (entraînement). Ce défaut est modélisé par le changement du paramètre  $\eta_{dt} = 0.22 \times \eta_{dt}$ . Cependant, cette erreur n'a pu être détectée qu'avec un niveau de défaut beaucoup plus élevé dans ce paramètre  $\eta_{dt} = 0.95 \times \eta_{dt}$ . Ce résultat est un peu inférieur à celui requis. Une amélioration du modèle SVM décisionnel peut être réalisée avec une enquête plus approfondie sur le vecteur des caractéristiques, l'introduction de plus de données dans l'étape d'apprentissage et l'adoption d'une méthode appropriée de filtrage.

La plupart des exigences relatives à la détection des défauts ont été réalisées. Les différents types de défauts ont été détectés et isolés le plus rapidement possible et sans fausses alarmes.

### **III.4. Comparaison entre différentes méthodes de diagnostic**

Il s'agit de comparer différentes méthodes de diagnostic proposées dans le cadre de la compétition internationale posée par kk-electronic. Les méthodes considérées pour cette comparaison sont celles ayant obtenues les meilleurs résultats :

**L'approche proposée par l'auteur de cette thèse basée sur SVM (Laouti, et al., 2011a) *MBSVM*.**

**Une Méthode Basée sur l'Estimation (Zhang et al., 2011) *MBE***

Solution basée sur l'estimation, un estimateur basé sur les modèles et les paramètres associés fournis dans (Odgaard and Stoustrup, 2010) est utilisé pour la détection de défauts, ensuite un banc d'estimateurs est utilisé pour l'isolation.

### **Une solution basée sur un Compteur réversible (Ozdemir et al., 2011) *MBC***

Solution FDI basée sur un compteur réversible (*Up-down counter*) qui permet de compter ou décompter selon le choix du sens de comptage. En effet, la génération de résidus est obtenue par la redondance et les connaissances physiques, les équations de parité et différents filtres. Par la suite, le compteur réversible utilise les résidus et les dynamiques discrètes dans le temps pour la décision sur la détection et l'isolation de chaque défaut. A l'inverse du seuil fixe qui trouve une difficulté pour régler le compromis entre fausses alarmes et détection manquée, la solution proposée dans ce travail est basée sur un seuil variable basé sur un compteur.

### **Méthode Basée sur les Observateurs (Chen et al., 2011) *MBO***

Une approche basée sur les observateurs est conçue pour le diagnostic de défauts dans l'éolienne. Les résidus sont générés par un filtre de Kalman et des observateurs dont les paramètres sont choisis sur un compromis entre détection rapide et fausses alarmes.

Le test de rapport de vraisemblance généralisée (*Generalized Likelihood ratio test*) et l'index de variance cumulative sont utilisés pour évaluer les résidus générés. L'isolation est effectuée par l'aide de la redondance.

### **Méthode Automatique Générale de Défauts (Svärd and Nyberg, 2011) *MAGD***

Cette approche se décompose en 3 étapes :

- 1) Génération de résidus candidats
- 2) Sélection des résidus les plus adaptés
- 3) Construction d'un test de diagnostic sur les résidus choisis avec la méthode de divergence de Kullback-Leibler

Une comparaison entre les distributions de probabilités estimées de résidus est utilisée pour les tests de diagnostic et l'évaluation des données actuelles et sans faute.

Plusieurs scénarios ont été testés aussi par le changement du temps de l'occurrence du défaut. Les tableaux suivants montrent les résultats de comparaison :

Il apparaît clair d'après le Tableau III.3 que la solution MBSVM est meilleure que les autres solutions de détection du défaut 1 car elle permet de détecter rapidement ( $2T_e$ ) sans fausses alarmes ni détections manquées.

**Tableau III.3 Résultats de diagnostic du défaut 1**

Défaut 1	Temps de Détection			Fausses Alarmes			Détection Manqués		
	moyenne	minimum	Maximum	moyenne	minimum	Maximum	Moyenne	minimum	Maximum
<i>MBSVM</i>	0.02	0.02	0.02	0	0	0	0	0	0
MBE	0.02	0.01	0.02	0	0	0	3%	0%	20%

MBC	0.03	0.02	0.03	0	0	0	0	0	0
MBO	10.32	10.23	10.33	0.89	0	1	0	0	0
<i>MAGD</i>	0.04	0.03	0.04	0	0	0	0	0	0

Le défaut 2 (capteur de type gain) ne peut pas être détecté si  $\beta_r = 0$  sauf pour des grandes valeurs de défaut gain puisque  $\beta$  sera juste le bruit de mesure.  $\beta$  prend la valeur zéro assez fréquemment puisqu'elle correspond à la position optimale dans la zone 2 c'est pour cette raison que la détection du défaut est un peu lente. Ici aussi la méthode MBSVM surpasse les autres dans le pourcentage nul des fausses alarmes, seule MBC réalise un résultat équivalent sauf que niveau de rapidité de détection MBSVM est meilleur que MBC comme le montre le Tableau III.4.

**Tableau III.4 Résultats de diagnostic du défaut 2**

Défaut 2	Temps de Détection			Fausses Alarmes			Détection Manqués		
	moyenne	minimum	Maximum	moyenne	minimum	Maximum	Moyenne	minimum	Maximum
<i>MBSVM</i>	47.24	3.23	95.09	0	0	0	56%	0	100%
MBE	44.65	0.63	95.82	22	16	28	56%	0%	100%
MBC	69.12	7.60	95.72	0	0	0	67%	0	100%
MBO	19.24	3.43	49.93	0.97	0	5	0	0	0
<i>MAGD</i>	13.70	0.38	25.32	3.08	1	18	0	0	0

Pour plus de détails sur les résultats de cette compétition voir (Odgaard and Stoustrup, 2012).

Il est à noter que la solution proposée par l'auteur de cette thèse a été classée comme la meilleure stratégie pour le diagnostic de défauts appliquée au modèle de référence de l'éolienne malgré qu'elle ne traite pas les défauts 6 et 7 d'actionneurs. Car cette stratégie se caractérise par une rapidité de détection, une justesse de décision sur les défauts et un pourcentage quasi nul de fausses alarmes.

### **III.5. Combinaison entre l'approche basée sur SVM et un observateur**

Comme on vient de le voir, la détection des défauts 6 et 7 n'a pas pu être faite uniquement par SVM. Dans ce qui suit, une amélioration de cette stratégie est faite par la combinaison entre la méthode de diagnostic basée sur les données « SVM » et une méthode basée sur le modèle « observateur de type Kalman ». Ceci a permis de détecter et d'isoler les défauts d'actionneurs 6 et 7.

Les valeurs  $w_n = 11.11 \text{ rad/s}$  et  $\zeta = 0.6$  correspondent aux paramètres d'un comportement normal. Les défauts d'actionneurs 6 et 7 sont introduits par le changement des valeurs nominales des paramètres du modèle d'actionneur  $w_n$  et  $\zeta$  :

Défaut 6 :  $w_{n2} = 5.73 \text{ rad/s}$  et  $\zeta_2 = 0.45$  ou Défaut 7 :  $w_{n3} = 3.42 \text{ rad/s}$  et  $\zeta_3 = 0.9$

La Figure III.19 illustre le schéma de combinaison entre SVM et un observateur de type Kalman. L'idée est de créer une sorte de communication entre eux. Le bloc *SVM FDI* envoie à l'observateur 2 un signal qui surveille la présence du défaut 6 appelé  $F_6$ , et à l'observateur 3 un signal qui surveille la présence du défaut 7 appelé  $F_7$ . Dans ce travail, le rôle des l'observateurs 2 et 3 est d'estimer en ligne les paramètres du modèle,  $w_{n2e}$  et  $\zeta_{2e}$  de l'actionneur 2 pour l'observateur 2 et  $w_{n3e}$  et  $\zeta_{3e}$  de l'actionneur 3 pour l'observateur 3. Il faut savoir que pour chaque actionneur on peut construire un observateur dédié à lui.

Le signal de défaut  $F_6$  (comme  $F_7$ ) est un signal binaire qui prend la valeur zéro quand il n'y a pas le défaut 6 et la valeur 1 dès l'occurrence de ce défaut.  $F_6$  est filtré avec un filtre passe-bas de constante de temps  $\tau_2 = 10s$ . Afin de diminuer un peu l'effet de commutations brusques dans le fonctionnement de l'observateur 2, l'estimation de  $w_{n2e}$  et  $\zeta_{2e}$  est mise en marche dès que le signal  $F_6$  filtré dépasse le seuil 0.02.

De la même façon que  $F_6$ ,  $F_7$  est filtré avec un passe-bas de constante de temps  $T_3=3s$  et l'estimation de  $w_{n3e}$  et  $\zeta_{3e}$  commence dès que le signal  $F_6$  filtré dépasse le seuil 0.01.

Les estimés  $w_{n2e}$ ,  $\zeta_{2e}$ ,  $w_{n3e}$  et  $\zeta_{3e}$  sont aussi filtrés avec une constante de temps de 1s pour rejeter les bruits de mesures ainsi que les perturbations.

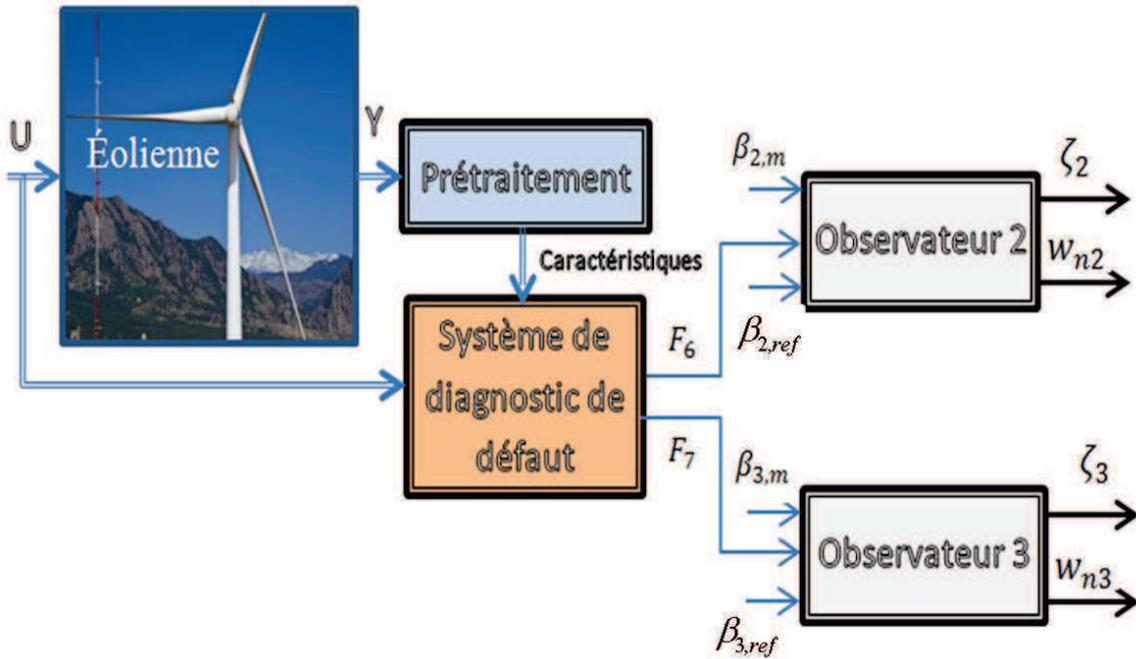


Figure III.19 Schéma de combinaison SVM et observateur

### III.5.1. Conception de l'observateur de type Kalman

$$\frac{y}{u} = \frac{\beta_m}{\beta_r} = \frac{1}{\frac{s^2}{w_n^2} + \frac{2\zeta}{w_n} \cdot s + 1} = \frac{1}{a \cdot s^2 + b \cdot s + 1} \quad \text{avec } a = \frac{1}{w_n^2}, b = \frac{2\zeta}{w_n} \text{ et } w_n \neq 0 \quad (\text{III. 15})$$

D'après la fonction de transfert du 2<sup>ème</sup> ordre de l'actionneur :

$$a \cdot \ddot{\beta} + b \cdot \dot{\beta} + \beta = \beta_r \Rightarrow \ddot{\beta} = -F \cdot \dot{\beta} - G \cdot \beta + H \cdot \beta_r$$

$$\begin{cases} F = \frac{b}{a} = 2\zeta \cdot w_n \Rightarrow \zeta = \frac{F}{2\sqrt{G}} \\ G = H = \frac{1}{a} = w_n^2 \Rightarrow w_n = \sqrt{G} \end{cases} \quad (\text{III. 16})$$

Avec la condition que  $G \neq 0$ . La construction de l'observateur se fait par le choix des états suivants :  $x_1 = \beta$ ,  $x_2 = \dot{\beta}$  et les états augmentés  $x_3 = F$ ,  $x_4 = G = H$  et  $X = (x_1, x_2, x_3, x_4)^T$

Le système d'état s'écrira alors :

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -F.x_2 - G.x_1 + H.u \\ \dot{x}_3 = 0 \\ \dot{x}_4 = 0 \end{cases} \quad (\text{III. 17})$$

Avec  $u = \beta_r$  et  $y = \beta$

On obtient le système d'état suivant

$$\dot{X} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -x_2 & (u - x_1) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} X + B.u = A.X + B.u \quad (\text{III. 18})$$

L'observateur est donné comme suit

$$\begin{cases} \dot{\hat{X}} = A(u)\hat{X} + G(u) - S^{-1}C^T(C\hat{X} - y) \\ \dot{S} = -\theta.S - A^T(u)S - SA(u) + C^TC \end{cases} \quad (\text{III. 19})$$

La valeur du paramètre de réglage  $\theta$  qui détermine la vitesse de convergence de l'observateur est réglée à 3. La matrice symétrique définie positive  $S$  de taille  $\mathfrak{R}^{4 \times 4}$  est initialisée à la matrice identité. L'état initial de l'observateur est défini comme suit :

$$X_0 = (0, 0, 2.\zeta.w_n, w_n^2)^T = (0, 0, 13.33, 123.43)$$

Cet observateur permet d'estimer en ligne les paramètres  $\zeta = \frac{x_{3e}}{2\sqrt{x_{3e}}}$ ,  $w_n = \sqrt{x_{4e}}$ , sachant qu'ils sont observables en présence de dynamique.

### III.5.2. L'approche basée SVM pour la détection/isolation de défaut actionneur sur le système pitch

Pour la détection de défauts 6 (5a) et 7(5b), le vecteur d'apprentissage utilisé est :

$$x = \begin{bmatrix} |\beta_{k,m_1}(t_j) - \beta_{k,m_2}(t_j)| \\ |\hat{w}_{g,m_1}(t_j - 250.T_e) - \hat{w}_{g,m_2}(t_j - 250.T_e)| \\ |\hat{\beta}_r(t_j) - \beta_{k,m_1}(t_j)| \\ |\hat{\beta}_r(t_j) - \beta_{k,m_2}(t_j)| \end{bmatrix} \quad (\text{III.20})$$

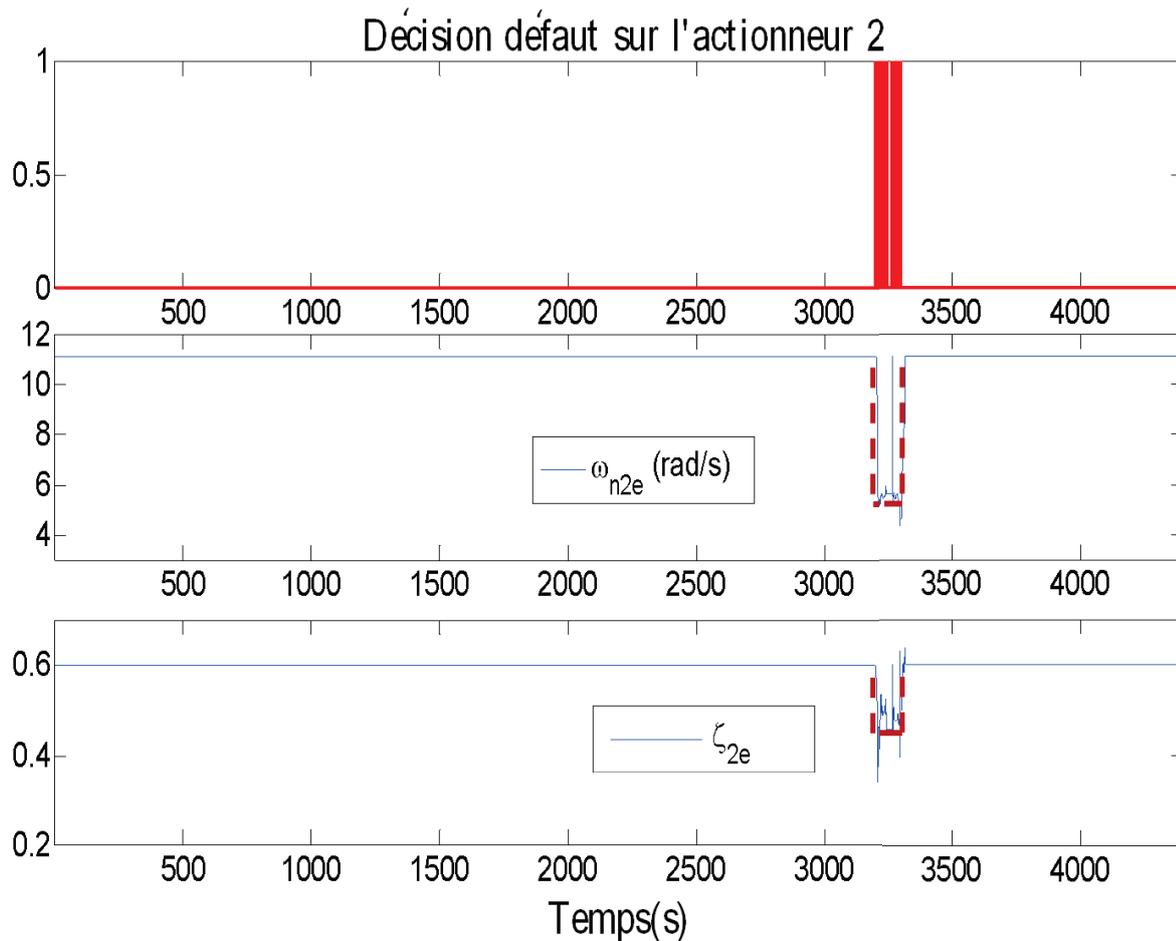
Où  $k=1, 2, 3$  correspond au numéro de l'actionneur de pitch. Pour le défaut 6 (5a) qui affecte l'actionneur 2,  $j=2$  et pour le défaut 7 (5b),  $j=3$ .  $\hat{\beta}_r$  est l'angle de pitch de référence filtré avec la fonction de 2<sup>ème</sup> ordre du modèle actionneur pitch  $\frac{w_n^2}{s^2 + 2\zeta \cdot w_n \cdot s + w_n^2}$ . L'objectif de ce filtre est de prendre en considération les dynamiques du système hydraulique pitch (actionneur) de telle sorte à avoir le comportement normal (sans défaut)  $\beta$  après une commande  $\beta_r$ . Il faut préciser que le rôle de ce filtre n'est pas de rejeter les bruits de mesures ou les perturbations.

Ce terme  $|\hat{w}_{g,m_1}(t_j) - \hat{w}_{g,m_2}(t_i)|$  est filtré avec un filtre passe-bas de constante de temps égal à 1.5 s pour enlever le bruit, et aussi retardé de 2.5s (soit 250  $T_e$ ) pour des raisons de synchronisation avec les autres signaux caractéristiques des défauts 6 et 7.

La Figure III.20 montre que cette approche hybride a permis non seulement de détecter la présence du défaut 6 qui apparaît entre 3200s et 3300s sur l'actionneur pitch 2 mais aussi de bien estimer le changement de la valeur du coefficient d'amortissement.

La difficulté liée à la détection du défaut d'actionneur 6 est due au fait qu'il n'est pas possible de le détecter si la consigne  $\beta_r$  est égale à 0. Dans ce cas, on oblige l'actionneur à ne pas réagir. Ce défaut est détecté entre 3200.02 s et 3300.03 s mais avec des intermittences. L'utilisation d'un filtre passe bas permet de diminuer ces effets sur l'estimation du coefficient d'amortissement qui prend une valeur aux alentours de 0.45 comme prévu.

La Figure III.20 illustre également la bonne estimation par l'approche du changement de la pulsation naturelle vers une valeur proche de 6 rad/s. Il faut noter aussi l'absence de fausses alarmes.

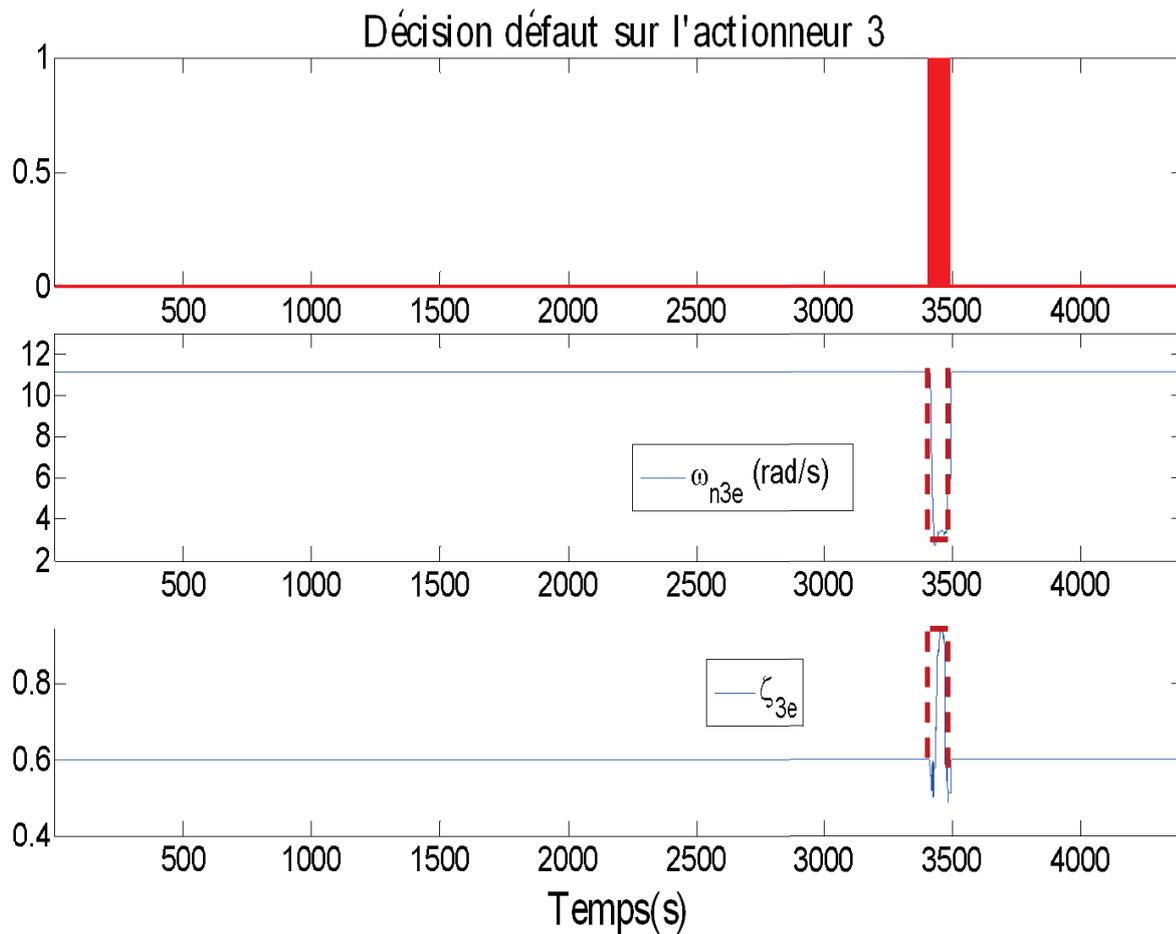


**Figure III.20** Détection et isolation du défaut 6 (5a) sur l'actionneur2 et estimation du coefficient d'amortissement  $\zeta_{2e}$  et de la pulsation naturelle  $w_{n2}$  avec l'observateur

### Défaut 7 sur l'actionneur 3

La Figure III.21 montre qu'avant l'apparition du défaut 7 sur l'actionneur 3 à l'instant 3400 s, la valeur du coefficient d'amortissement est fixée à la valeur nominale (0.6). Après la détection du défaut par l'approche MBSVM, le coefficient d'amortissement prend une valeur proche de 0.9 qui indique bien un défaut de type 5b. On rappelle que le défaut 7 est introduit progressivement d'une façon linéaire sur 30 s (de 3400s jusqu'à 3430s). Il reste actif pendant 40 s (de 3430s jusqu'à 3470s), puis il va diminuer lentement à nouveau sur 30 s (de 3470s jusqu'à 3500s).

La détection de ce défaut est réalisée à l'instant 3411.49s. Elle n'est pas très rapide puisque les symptômes de ce défaut 7 apparaissent lentement sur les mesures. La Figure III.21 illustre la bonne estimation par l'approche du changement de la pulsation naturelle vers une valeur proche de 3.42 rad/s.



**Figure III.21 Détection et isolation du défaut 7 (5b) sur l'actionneur 3 et estimation du coefficient d'amortissement  $\zeta_{3e}$  et de la pulsation naturelle  $w_{n3}$  avec l'observateur**

La combinaison entre l'approche basée sur SVM et un filtre de type Kalman a permis non seulement de spécifier l'instant d'apparition du défaut et l'actionneur défectueux mais aussi de préciser de quel type de défaut il s'agit, cela par l'estimation en ligne des paramètres  $w_n$  et  $\zeta$ .

Ainsi, dans ce travail, toutes les exigences relatives à la détection et isolation des défauts dans le benchmark utilisé ont été respectées. Les différents types de défauts ont été détectés et isolés rapidement et avec une grande précision, sauf celui du système comme montré par le Tableau III. 5.

**Tableau III. 5 Résultats de détection des défauts**

N°	Défaut	Le niveau de défaut à détecter	Détection et isolation du défaut	Nombre de $T_e$	Nombre de $T_e$ exigé
1a)	Valeur fixe $\Delta\beta_{k,m_i}$	-	Oui	2	<10
1b)	Offset (ou gain) sur $\Delta\beta_{k,m_i}$	$1.1 \times \beta, (\Delta\beta \approx 1)$	Oui si $\Delta\beta \geq 0.5$	10 si $\Delta\beta = 1.5$	
2a)	Valeur fixe $\Delta w_{r,m_i}$	-	Oui	2	
2b)	Offset (ou gain) $\Delta w_{r,m_i}$	$0.1 \times w_r$	Oui si $\geq 0.2 \times w_r$	67	
3a)	Valeur fixe $\Delta w_{g,m_i}$	-	Oui	2	
3b)	Offset (ou gain) $\Delta w_{g,m_i}$	$0.1 \times w_g, (\approx 10 \text{ rad.s}^{-1})$	Oui si $\geq 0.1 \times w_r$	2	
4a)	Offset sur $\Delta\tau_g$	$\Delta\tau_g \geq 100$	Oui	2	<5
5a)	Brusque $\Delta\beta_k$	$\Delta w_n \geq 0.5, \Delta\zeta \geq 0.25$	Oui si $\beta_r \neq 0$	2	<8
5b)	Lent $\Delta\beta_k$	$\Delta w_n \geq 0.25, \Delta\zeta \geq 0.5$	Oui si $\beta_r \neq 0$	149	<100
6)	Système d'entraînement, $\Delta w_r, \Delta w_g$	$\Delta\eta_{dt} \geq 5\%$	Oui si $\Delta\eta_{dt} \geq 50\%$	2835	libre

### III.6. Conclusion

La stratégie de diagnostic de défaut proposée dans ce chapitre fait appel à plusieurs disciplines allant du traitement de signal, physique, apprentissage automatique, électricité jusqu'à l'automatique. Les Séparateurs à Vaste Marge sont utilisés pour apprendre un modèle de décision capable de détecter et d'isoler des défauts de type capteurs, actionneurs et système dans une éolienne. Définir le vecteur caractéristique ainsi que le réglage des paramètres du modèle sont primordiaux afin de détecter et d'isoler ces défauts. Un compromis réalisé manuellement entre la sensibilité de la détection au bruit et au défaut doit être déterminé.

L'observateur de type Kalman a permis de cerner le type défaut sur l'actionneur système pitch par une estimation assez précise du changement de la valeur du coefficient d'amortissement ainsi que la pulsation naturelle.

L'approche proposée basée sur SVM permet d'avoir une capacité de généralisation intéressante même avec un nombre réduit d'échantillons d'apprentissage, et aussi de manipuler facilement la non linéarité du système éolienne.

La comparaison entre différentes méthodes de diagnostic de défauts n'est pas toujours facile à réaliser dans la littérature puisque certaines données disponibles pour certains chercheurs ne le

sont pas pour les autres. Donc le fait d'avoir une compétition internationale où tout le monde est confronté au même problème avec les mêmes données et aux mêmes exigences à respecter, me permet de dire d'une façon objective que mon approche a montré son efficacité et sa supériorité dans la résolution du problème de diagnostic de défauts dans les éoliennes par rapport à d'autres méthodes utilisées dans cette compétition.



# **CHAPITRE IV : Réacteur chimique parfaitement agité**



Dans ce chapitre, la stratégie de diagnostic de défaut proposée dans cette thèse est appliquée à un réacteur chimique parfaitement agité. Même si le modèle de ce procédé contient moins de variables que le procédé de traitement des eaux usées ou le système éolien, il peut néanmoins être très non linéaire dans certains cas ce qui rend le problème de diagnostic plus difficile. Pour cette raison, de nombreuses méthodes de diagnostic ont été proposées pour ce système qui ne pourront pas toutes être rapportées ici. La revue bibliographique suivante se limitera donc aux applications de SVMs aux réacteurs chimiques.

## IV.1. Revue bibliographique

La recherche de nouvelles méthodes pour le diagnostic de pannes dans les installations chimiques industrielles est nécessaire afin d'assurer la sécurité des opérations, la conformité environnementale et afin d'éviter la réduction de la qualité du produit et de la productivité des procédés.

Seuls quelques articles de revues concernent l'utilisation des SVMs pour la détection de défauts dans les applications chimiques, et ce principalement à travers des études de simulation. Le cas d'étude le plus généralement traité est celui du *Tennessee Eastman benchmark* (Monroy et al., 2010, Hsu et al., 2010, Mahadevan and Shah, 2009, Yélamos et al., 2009, Zhang, 2009, Zhang, 2008, Mao et al., 2007, Yélamos et al., 2007, Kulkarni et al., 2005, Chiang et al., 2004). Ce procédé non linéaire (Downs and Vogel, 1993) permet de simuler des conditions de fonctionnement normales ainsi que 21 situations de défauts, et est assez représentatif des réacteurs chimiques, faisant ainsi intervenir un réacteur agité, un condensateur, un compresseur, un séparateur et un décapeur. Peu d'autres applications des SVMs aux procédés chimiques ou biochimiques peuvent être trouvées. On peut cependant citer le traitement de minéraux (Jemwa and Aldrich, 2006), la production de pénicilline (Li et al., 2006), le réacteur à eau bouillante dans une centrale nucléaire (Rocco S. and Zio, 2007), la distillation de butadiène (Bo et al., 2010), un réacteur chimique parfaitement agité (Saludes Rodil and Fuente, 2010) et la prédiction de la vitesse critique d'un flux de suspension solide-liquide (Lahiri and Ghanta, 2008). Dans la plupart des travaux mentionnés, les SVMs ont montré une supériorité dans la précision et un effort réduit de calcul par rapport à d'autres techniques, telles que les réseaux de neurones. Ceci est cohérent avec la théorie de SVM basée sur la recherche de séparateurs entre les classes par minimisation de risque structurel tandis que les réseaux de neurones sont basés sur la minimisation du risque empirique (Ravikumar et al., 2008, Ge et al., 2004).

Bien que les travaux utilisant les SVMs pour le diagnostic de défauts dans les procédés chimiques soient encore limités, ils montrent que les SVMs représentent une alternative intéressante aux approches basées sur les modèles qui sont largement étudiées dans ce domaine.

Lors de l'application des SVMs aux réacteurs chimiques, certains chercheurs ont trouvé nécessaire d'adapter la méthode d'optimisation, le choix de noyaux, le réglage des paramètres, le prétraitement des données ou la méthode de classification (mono- ou multi-classe). Tout d'abord, la méthode SVM traditionnelle est un classificateur binaire qui ne peut différencier que deux classes. (Yélamos et al., 2009) a utilisé les méthodes « one versus all » pour généraliser son application au cas multi-classe tout en assurant un effort réduit de calcul. Comme décrit dans le chapitre II, cette méthode nécessite  $k$  classificateurs pour  $k$  classes. De même, (Chiang et al., 2004) ont utilisé la méthode « one versus one » qui exige  $k(k-1)/2$  classificateurs qui est très élevé, réduisant donc la vitesse d'apprentissage et de test. Enfin, (Yuan and Chu, 2006) a utilisé la méthode « one to others » qui nécessite seulement  $(k-1)$  classificateurs. En contraste avec le cas multi-classe, le cas mono-classe a été utilisé (Mahadevan and Shah, 2009) pour la détection de défauts dans l'industrie chimique. Un deuxième paramètre qui peut affecter le temps de calcul est la stratégie d'optimisation sous contrainte quadratique nécessaire pour l'apprentissage d'un grand nombre de données. Différentes techniques d'optimisation ont été utilisées pour résoudre le critère d'optimisation des SVMs dans les procédés chimiques. Parmi ces algorithmes se trouvent les multiplicateurs de Lagrange, « *particle swarm optimisation* » (Tang et al., 2010, Wu, 2010, Samanta and Nataraj, 2009), les algorithmes d'optimisation séquentielle (Davy et al., 2006) et les algorithmes génétiques (Zeng et al., 2006). Dans (Saravanan et al., 2010), le critère de SVM a été modifié pour optimiser le temps de calcul. Une variante nommée « *proximal SVM* » a été utilisée pour détecter les défauts dans une boîte à vitesse. L'algorithme d'optimisation peut également être utilisé pour le réglage des paramètres. Parmi les techniques utilisées pour le réglage des paramètres de SVM pour le diagnostic des pannes dans les procédés chimiques apparaissent la validation croisée, l'évolution différentielle (Lahiri and Ghanta, 2008), « *The artificial immune algorithm* » (Yuan and Chu, 2006) ou encore des algorithmes génétiques (Fei and Zhang, 2009, Kulkarni et al., 2004). En plus du réglage des paramètres, le choix du noyau est également déterminant dans SVM car il doit décrire la relation entre les entrées et les sorties (linéaire / non linéaire). Enfin, la combinaison de SVM avec des outils basés sur les données dans l'objectif de pré traitement des données s'est avéré utile afin de décorréler les variables et de les projeter dans un espace inférieur en dimensions avec une variance maximale. Les principales méthodes statistiques qui ont été combinées aux SVMs pour le prétraitement des

données sont ACP, ACP à noyau, l'analyse en composantes indépendantes avec et sans noyau (Hsu et al., 2010, Monroy et al., 2010, Widodo et al., 2009, Zhang, 2008).

Alors que l'attention des recherches a été axée principalement sur la détection et l'isolement de défaut, moins d'effort de recherche a été consacré à l'identification des défauts à l'aide des SVMs. Cependant, l'identification des défauts pourrait être d'une grande utilité pour la reconfiguration de la commande. En effet, les dérives dans les capteurs peuvent être corrigées par le logiciel si le niveau de la dérive est connu. En outre, si un actionneur est partiellement défectueux, il peut encore être utilisé dans le système de contrôle si le niveau de la dérive est évalué. Cela représente une des principales contributions de ce chapitre. Les différentes étapes de détection, isolement et d'identification des défauts sont réalisées par SVM.

La deuxième contribution dans cette application consiste en la combinaison de la méthode SVM avec un observateur à grand gain pour la détection et l'isolation de défaut dans une partie du procédé où la performance des SVMs a été limitée. Il s'agit des erreurs de capteurs dans le milieu réactionnel qui est le plus influencé par la non-linéarité du procédé.

Dû à cette non-linéarité, afin de garantir des résultats acceptables par SVM, l'apprentissage doit contenir un grand nombre de données dans des conditions variées. Dans l'absence de ce grand nombre de données, un modèle de cette partie de système peut aider au diagnostic de défaut s'il est couplé aux SVMs. L'idée proposée est la suivante : Les SVMs sont utilisés en premier lieu pour le diagnostic de tous les défauts dans le réacteur chimique. Là où cette méthode montre des faiblesses avec le jeu de données d'apprentissage choisi par défaut, un observateur est ajouté afin de rendre le diagnostic plus robuste.

Ce chapitre est organisé comme suit : Tout d'abord, le procédé considéré est présenté ainsi que son modèle. Dans un deuxième temps, l'application des SVMs au diagnostic des différents défauts est considérée. Ensuite, un observateur est développé pour une partie du système. La méthodologie est évaluée en premier par simulation puis expérimentalement sur un réacteur de polymérisation à l'échelle du laboratoire. Ensuite, des données d'un réacteur de polymérisation pilote de 250 L sont utilisées pour valider les défauts de capteurs dans le milieu réactionnel, ce qui semblait être la difficulté principale. Le chapitre est terminé par des conclusions partielles.

## IV.2. Présentation du système

### IV.2.1. Le procédé

Deux réacteurs chimiques parfaitement agités à double enveloppe (*Stirred Tank Reactor*), ont été utilisés : un réacteur d'un litre, à l'échelle du laboratoire (Pignat ®) et un réacteur pilote de 250 litres (de la société Arkema (Guinot et al., 2000)). Le système de chauffe/refroidissement (*heating/cooling HC*) utilisé pour chauffer la double enveloppe est un bain d'huile thermostaté dans le réacteur de laboratoire (Figure IV.1) et de la vapeur d'eau dans le réacteur pilote.

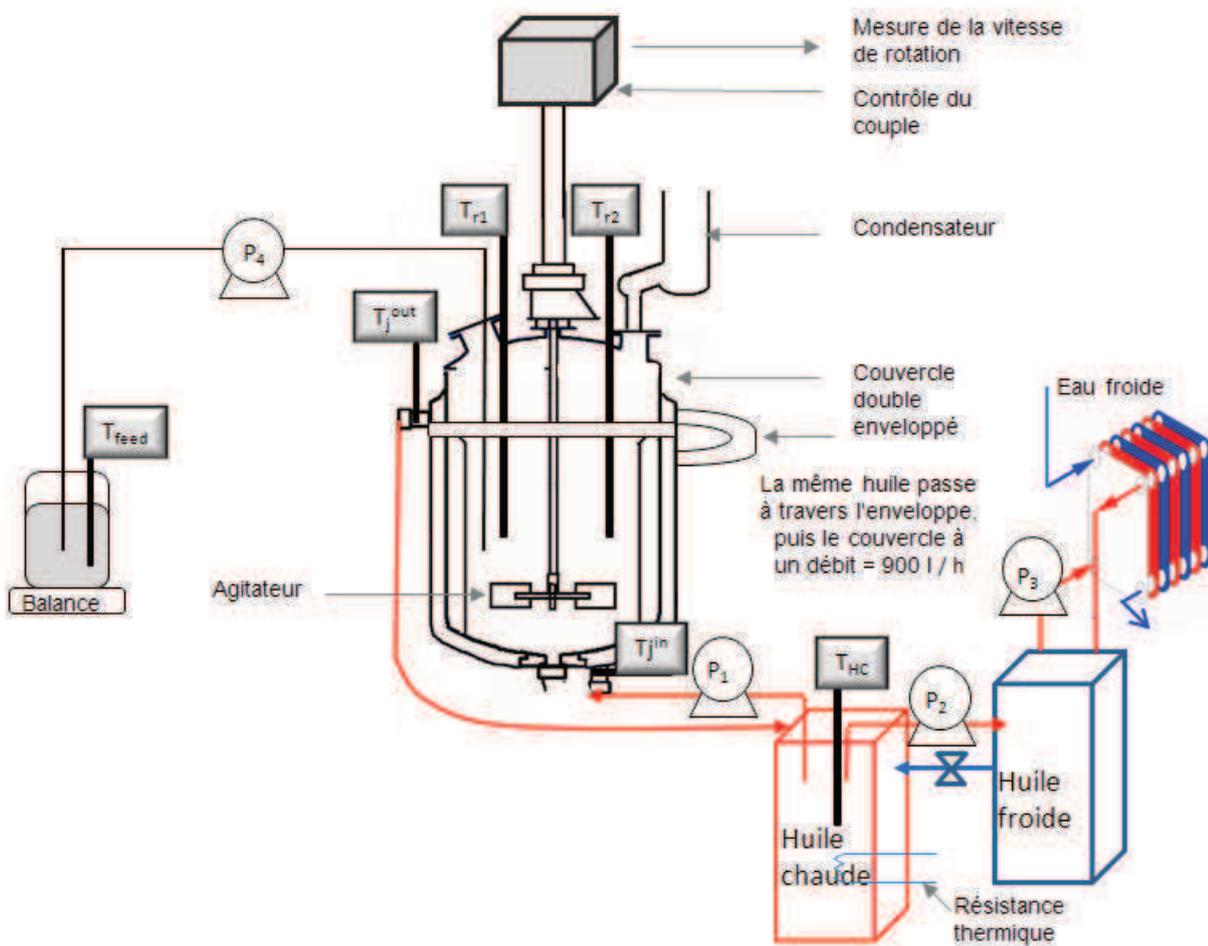


Figure IV.1 Schéma simplifié du réacteur parfaitement agité du laboratoire ( $T_{r1}$ ,  $T_{r2}$ ,  $T_{j1}$ ,  $T_{j2}$  et  $T_{HC}$  sont des sondes de température  $P_1$ ,  $P_2$ ,  $P_3$  et  $P_4$  sont des pompes)

Le réacteur est équipé de deux sondes de température dans le réacteur (noté  $T_{r1}$  et  $T_{r2}$ ), une sonde de température dans le bain thermostaté ( $T_{HC}$ ), une sonde de mesure de la température d'alimentation ( $T_{feed}$ ), une sonde de mesure de la température d'entrée de la double enveloppe ( $T_{j}^{in}$ ) et une pour la sortie de la double enveloppe ( $T_{j}^{out}$ ). Le mélange d'alimentation est mis sur

une balance pour évaluer la masse pompée et calculer le débit d'alimentation précisément ( $F_{feed}$ ). La vitesse de rotation de l'agitateur est mesurée et utilisée pour contrôler le couple agitateur. Les programmes de modèle, de contrôle et SVM sont développés en utilisant Matlab/Simulink ® combinée à des algorithmes de SVM de (Canu et al., 2005). Des régulateurs PID locaux permettent de contrôler la température du bain thermostaté (**HC** : *heating/cooling*), le débit d'alimentation et la rotation de l'agitateur. La température du réacteur est contrôlée en utilisant une commande non linéaire par linéarisation entrée/ sortie (Sheibat-Othman et al., 2011, Sheibat-Othman and Othman, 2006).

Les défauts de capteurs considérés ici sont :  $T_{R1}$ ,  $T_{R2}$ ,  $T_j^{in}$ ,  $T_j^{out}$ ,  $T_{HC}$ , et le défaut d'actionneur consiste principalement en la capacité du bain thermostaté à chauffer ou à refroidir (**HC**). Les autres capteurs (balance, couple de l'agitateur) ou actionneurs (couple de l'agitateur, pompes d'introduction dans le réacteur et dans la double enveloppe) ne sont pas considérés.

Les mesures de procédés utilisés dans la suite sont donc :  $y = [T_{r1}, T_{r2}, T_j^{in}, T_j^{out}, T_{HC}, T_{feed}, F_{feed}]$ , obtenues avec une fréquence de 10 s.

#### IV.2.2. Le modèle

Le modèle considéré pour la simulation se base sur le bilan de matière et de chaleur d'un réacteur semi-continu parfaitement agité (valable pour les cas du réacteur du laboratoire et du réacteur pilote) :

$$\begin{cases} MC_p \dot{T}_r = Q_r + \underbrace{UA(T_j^{out} - T_r)}_{Q_j} + F_{feed} C_{pfeed} (T_{feed} - T_r) - Q_{lr} \\ M_j C_{pj} \dot{T}_j^{out} = F_j C_{pj} (T_j^{in} - T_j^{out}) - UA(T_j^{out} - T_r) - Q_{lj} \end{cases} \quad (IV.1)$$

Où  $T_r$ , est la température moyenne dans le milieu réactionnel,  $U$  est le coefficient de transfert de chaleur et  $A$  la surface de transfert de chaleur entre le réacteur et la double enveloppe.  $M$  est la masse totale du milieu réactionnel.  $C_p$ ,  $C_{pfeed}$  et  $C_{pj}$  sont les capacités thermiques du milieu réactionnel, de l'alimentation et de l'huile caloporteuse.  $Q_{lr}$  est la perte de chaleur du réacteur principalement par le condenseur et  $Q_{lj}$  la perte de chaleur de la double enveloppe.  $F_j$  est le débit massique dans la double enveloppe. La chaleur produite par la réaction est proportionnelle à la vitesse de réaction ( $R_p$ ), où  $(-\Delta H)$  est l'enthalpie de réaction.  $Q_r = (-\Delta H) \times R_p$

La polymérisation en émulsion de styrène est considérée pour la validation expérimentale dans le réacteur de laboratoire, et un mélange de 3 monomères est considéré dans le réacteur pilote.

## IV.3. Implémentation du diagnostic de défauts

Ce chapitre présente deux contributions originales :

- La combinaison de la méthode SVM avec une méthode basée sur le modèle pour le diagnostic de défaut, sans avoir à modéliser tout le procédé en détail.
- L'utilisation de SVM régression (SVR) pour l'identification du niveau du défaut.

### IV.3.1 Implémentation des SVMs

La classification bi-classe de SVM est considérée, afin de distinguer une situation normale d'une situation défectueuse pour chacun des défauts considérés.

J'ai essayé d'obtenir des mesures réelles du procédé dans des situations saines et défectueuses à chaque fois que c'était possible. Dans le cas où les mesures ne sont pas utilisées pour la commande, un simple ajout du défaut sur la mesure est réalisé hors ligne dans le logiciel. Dans le cas où la mesure est utilisée pour la commande ou bien dans le cas de défaut d'actionneur, le défaut est introduit de manière logicielle mais en ligne afin de mesurer ses effets sur les autres mesures.

La méthode SVM peut être directement appliquée à ces données  $y$ , ou éventuellement après filtrage, normalisation, dérivation ou en considérant un horizon de temps. Généralement, il est nécessaire de discerner les variables ou combinaisons de variables affectées par la faute afin d'accroître leur identifiabilité. Dans cette application, l'utilisation de la méthode *Analyse en composante principale* (ACP) ou « *Partial least square PLS* » pour la réduction de la dimension n'a pas été intéressante : La pré-utilisation de l'ACP a dégradé les estimations et la PLS était presque équivalente aux SVMs, mais avec un temps de calcul plus important et des performances moindres dans certains cas. Par conséquent, le vecteur  $x$  est obtenu en sélectionnant les éléments les plus représentatifs pour chaque faute en se basant sur un certain savoir-faire du réacteur donné dans le tableau IV.1 . Par exemple, le premier élément  $x_i$  pour les fautes de  $T_r$  permet la détection de défaut et les autres sont destinés à l'isolation de défaut. En ce qui concerne le système HC, on remarque qu'une faute éloigne  $T_{HC}$  anormalement de  $T_j^d$ . Cependant, on remarque que les résidus des capteurs  $T_j^{in}$ ,  $T_j^{out}$  et  $T_{HC}$  sont corrélés. Cette corrélation peut être découplée par l'application d'une décision logique, comme le suggère le tableau IV.2 pour isoler le capteur défectueux.

**Tableau IV.1 Vecteurs d'entrées SVM ( $x_i$ ), ( $i$  est l'instant de temps)**

Capteur/ actionneur	$T_j^x$ ( $x=in, out$ )	$T_{HC}$	$T_{r1}, T_{r2}$	HC
$x_i$	$ T_j^x(i) - T_{HC}(i) $	$\begin{vmatrix}  T_j^{in}(i) - T_j^{out}(i)  \\  T_j^{in}(i) - T_{HC}(i)  \end{vmatrix}$	$\begin{vmatrix} T_{r1}(i) - T_{r2}(i) \\ T_{r1}(i) - T_j(i) \\ T_{r2}(i) - T_j(i) \\ T_{r1}(i) - T_j(i-5) \\ T_{r2}(i) - T_j(i-5) \end{vmatrix}$	$T_j^d(i) - T_{HC}(i)$

En utilisant des collections des données avec situations normales et défectueuses pour chaque type de défaut, et en utilisant certaines valeurs initiales de  $C$  et  $\sigma$ , nous pouvons identifier les multiplicateurs de Lagranges  $\alpha_i$ , le biais  $b$  et les vecteurs de support  $x_{sup}$ . Le réglage de  $C$  et  $\sigma$  s'est fait par l'optimisation et validation croisée. La comparaison du noyau gaussien et le noyau polynômial de différents degrés a montré une meilleure généralisation avec le noyau gaussien. Pour le réglage des paramètres, (Cherkassky and Ma, 2004) ont proposé une relation pour calculer  $\sigma$  à partir des données d'apprentissage comme suit:  $\sigma=(0,2-0,5)\times plage(x)$ . Cette relation donne des valeurs de  $\sigma$  légèrement inférieures à celles optimisées dans ce travail soit manuellement soit par validation croisée.

La fonction de décision peut alors être évaluée pour des nouvelles données  $x$ . Le seuil du résidu est fixé à  $0,1$  pour toutes les situations (en théorie, il peut être  $= 0$ ), ce qui signifie que la situation est estimée normale si la valeur résiduelle est inférieure à  $|+/- 0,1|$ . Les résultats de l'optimisation ont donné  $n_{sv} = 30\%$ ,  $6\%$ ,  $42\%$  et  $57\%$  de  $N$  pour  $T_{HC}$ ,  $T_j$ ,  $T_r$  et  $HC$  respectivement, avec  $N = 810$ . Cette réduction systématique de la taille des données réduit le temps de calcul en ligne de manière importante ( $<5 \times 10^{-3}s$  pour tous les résidus), ce qui représente un avantage des SVMs par rapport aux autres méthodes statistiques. La valeur de  $T_j$  (et  $T_r$ ) est calculée par  $(T_j^{in}+T_j^{out})/2$  en l'absence de défaut. En présence de défaut,  $T_j$  prend la valeur de la sonde non défectueuse.

**Tableau IV.2 Correspondance entre résidu de la température de la double enveloppe et du bain thermostaté (R=résidu)**

	R $T_j^{in}$	R $T_j^{out}$	R $T_{HC}$
Fault $T_j^{in}$	×		
Fault $T_j^{out}$		×	
Fault $T_{HC}$	×	×	×

### IV.3.2 Implémentation de l'observateur de grand gain

La non-linéarité principale dans ce système est due à la vitesse de la réaction qui affecte directement les capteurs  $T_{r1}$  et  $T_{r2}$ . Pour cette raison, l'apprentissage SVM doit contenir un nombre important de données dans des situations variées afin de garantir des bons résultats. Afin d'améliorer le diagnostic de cette partie sans avoir à enrichir les données d'apprentissage de manière importante, un observateur de  $T_r$  est utilisé.

On note que les mesures  $T_{r1}$  et  $T_{r2}$  ne doivent pas être utilisées dans la conception d'observateur puisqu'elles sont susceptibles de subir le défaut que nous cherchons à détecter et à isoler. Pour ce faire, nous ne considérons que les mesures  $T_j^{in}$  et  $T_j^{out}$  et en estime  $T_r$ . Ensuite, nous considérons le système augmenté suivant où  $\varepsilon_1$  représente un certain manque de connaissances de la dynamique du bilan thermique de la double enveloppe (la dynamique de  $T_j^{out}$ ) et  $\varepsilon_2$  représente la dynamique de  $T_r$  supposée inconnue :

$$\begin{bmatrix} \dot{T}_j^{out} \\ \dot{T}_r \end{bmatrix} = \begin{bmatrix} 0 & \frac{UA}{M_j C_{pj}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} T_j^{out} \\ T_r \end{bmatrix} + \frac{1}{M_j C_{pj}} \begin{bmatrix} F_j C_{pj} (T_j^{in} - T_j^{out}) - UA T_j^{out} + Q_{ij} \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (IV.2)$$

$$y = T_j^{out}$$

Nous pouvons voir que  $T_r$  est observable à partir du système (IV.2) car la forme du système respecte la forme de système de (IV.1), si  $\varepsilon_1$  est connu. Dans ce cas, l'observateur donné par le système peut être utilisé pour estimer  $T_r$  comme suit :

$$\begin{bmatrix} \dot{\hat{T}}_j^{\text{out}} \\ \dot{\hat{T}}_r \end{bmatrix} = \begin{bmatrix} 0 & \frac{UA}{M_j C_{Pj}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{T}_j^{\text{out}} \\ \hat{T}_r \end{bmatrix} + \frac{1}{M_j C_{Pj}} \begin{bmatrix} F_j C_{Pj} (T_j^{\text{in}} - \hat{T}_j^{\text{out}}) - UA \hat{T}_j^{\text{out}} + Q_{lj} \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{2\theta}{M_j C_{Pj}} \\ \frac{1}{UA} \end{bmatrix} \theta^2 (T_j^{\text{out}} - \hat{T}_j^{\text{out}}) \quad (\text{IV.3})$$

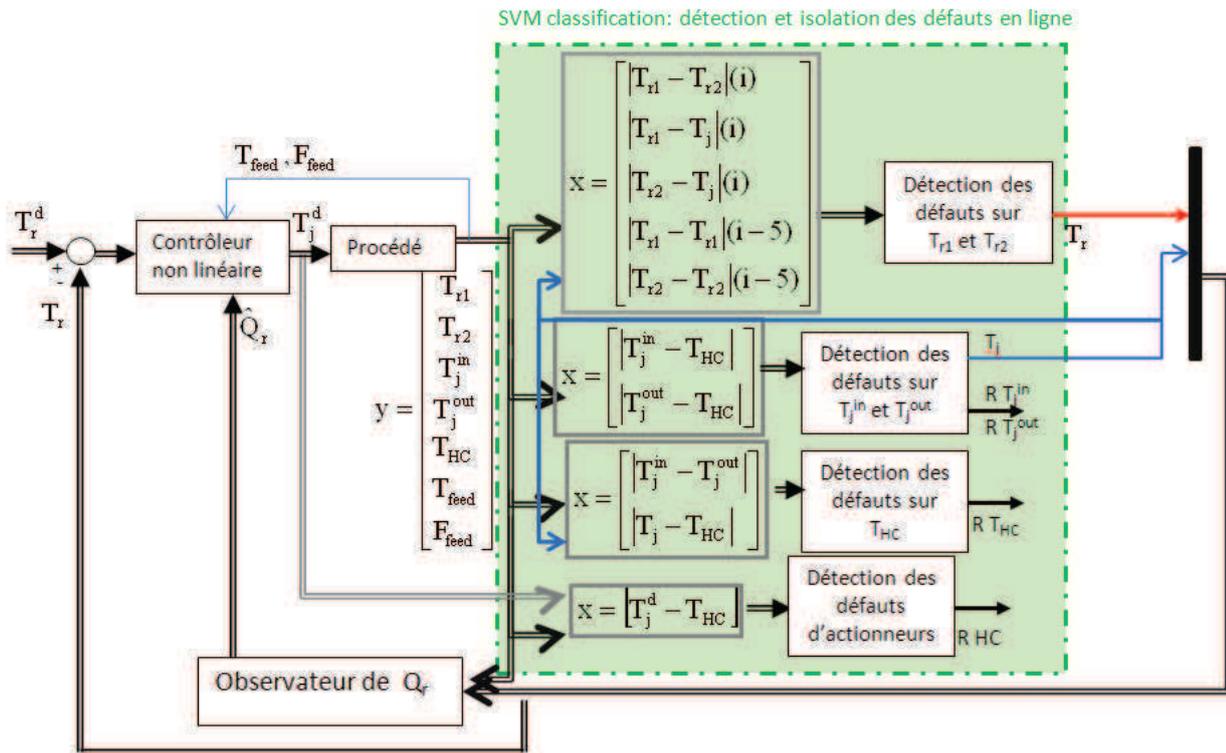
Cependant, le paramètre  $\varepsilon_1$  n'est pas connu, par exemple en raison d'un manque de connaissance de  $U$ ,  $A$ ,  $Q_{lj}$ ,  $F_j$ ,  $M_j$  ou  $C_{Pj}$ . L'objectif de ce travail est de proposer une méthodologie pour le diagnostic qui peut être appliquée rapidement sans étape de modélisation. Afin de rendre l'observateur robuste à ces incertitudes, le modèle de base qui représente le bilan thermique de la double enveloppe doit être corrigé en ligne basé sur les mesures de  $\mathbf{T}_{r1}$  et  $\mathbf{T}_{r2}$  dans des situations normales, c'est à dire quand il n'y a pas de défaut (qui peut être déterminé par SVM). La correction du modèle est effectuée en introduisant un filtre du premier ordre sur  $\mathbf{T}_r$  (à appliquer en l'absence de défaut) :

$$\dot{\tilde{T}}_r = \theta_{Tr} \left( \frac{T_{r1} + T_{r2}}{2} - \tilde{T}_r \right) \quad (\text{IV.4})$$

$\tilde{T}_r$  est une valeur filtrée de  $\mathbf{T}_r$  et  $\theta_{Tr}$  est un paramètre de filtrage.  $\tilde{T}_r$  permet de calculer un paramètre de correction du modèle  $\hat{\delta} = \tilde{T}_r - \hat{T}_r$ . Ensuite, l'estimation  $\hat{T}_r$  est obtenue par l'observateur (IV.3). Cette estimée peut être utilisée pour isoler la sonde défectueuse même en présence d'erreurs de modélisation importantes en utilisant les résidus suivants :

$$R_{Tr1} = |T_{r1} - \hat{T}_r - \hat{\delta}|, \quad R_{Tr2} = |T_{r2} - \hat{T}_r - \hat{\delta}| \quad (\text{IV.5})$$

On note que dans l'absence de faute l'équation (IV.5) se réduit à  $R_{Tri} = |T_{ri} - \tilde{T}_r|$ , mais en présence de défaut  $\hat{\delta}$  est une constante. Le schéma global de la détection et isolation des défauts utilisée dans ce chapitre est donnée dans la Figure IV.2 .



**Figure IV.2 Schéma de diagnostic de défauts par SVM en ligne**

### IV.3.3. Résultats et discussion

L'objectif des scénarios considérés ici est de démontrer la robustesse de la méthode SVM pour le diagnostic dans la plupart des parties du procédé et sa mise en œuvre rapide. Puis nous montrons la supériorité de la méthode couplant les SVMs à une méthode basée sur le modèle dans une partie qui subit une forte non-linéarité du procédé.

#### IV.3.3.1. Défaut de capteur dans $T_j^{in}$

La Figure IV.3 montre une validation expérimentale du modèle obtenu dans le cas d'apparition du défaut dans les capteurs de  $T_j$ . Une dérive croissante dans  $T_j^{in}$  a commencé à 55min. Elle a pu être détecté et isolé à 87 min lorsque le niveau de dérive atteint  $0,8^\circ\text{C}$  (seuil = 0,1). Notez que les résidus des autres actionneurs et capteurs ne sont pas affectés par la faute. La sonde de  $T_j$  estimée normale est correcte tout le temps. Basé sur 4000 échantillons de validation, le pourcentage de fausses alarmes a été estimé très faible pour  $T_j^{in}$ ,  $T_j^{out}$  et de  $T_{HC}$  (0,6%) et tous les défauts avec  $\Delta T > 2^\circ\text{C}$  ont pu être détectés. En outre, les modèles SVM obtenus (les fonctions de décision) étaient valables pour une grande variabilité dans les conditions de fonctionnement du procédé.

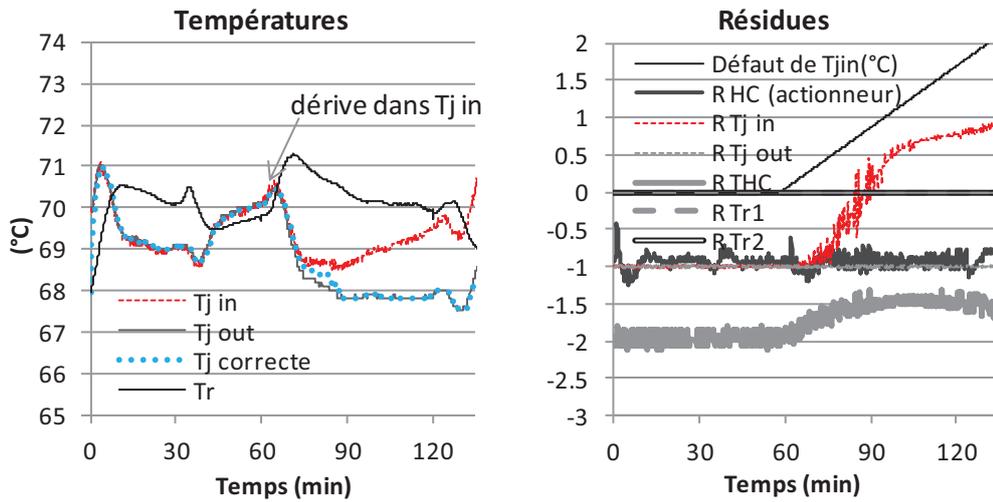


Figure IV.3 Dérive dans  $T_{j,in}$  à 55min

#### IV.3.3.2 Défaut dans $T_{HC}$

- Réglage de la variance :

L'apprentissage des défauts de  $T_{HC}$  est un bon exemple pour illustrer l'effet de réglage des paramètres sur les hyperplans car  $x_i$  est de dimension 2. La Figure IV.4 montre les hyperplans séparatifs obtenus en utilisant différentes valeurs pour la variance de la fonction noyau  $\sigma = 4, 20$ . On peut observer qu'une grande valeur de  $\sigma$  conduit à une séparation presque linéaire. En outre, la largeur de la marge augmente avec des valeurs élevées de  $\sigma$ , ce qui conduit à un plus grand nombre de vecteurs de support. En conséquence, des valeurs élevées de  $\sigma$  sont connues pour donner une généralisation améliorée, mais les très hautes valeurs de  $\sigma$  pourraient ne pas correspondre aux données d'apprentissage du tout. De l'autre côté, des petites valeurs de  $\sigma$  sont connues pour s'adapter parfaitement aux données d'apprentissage mais peuvent ne pas être en mesure d'évaluer des nouvelles données (surapprentissage, capacité réduite de la généralisation).

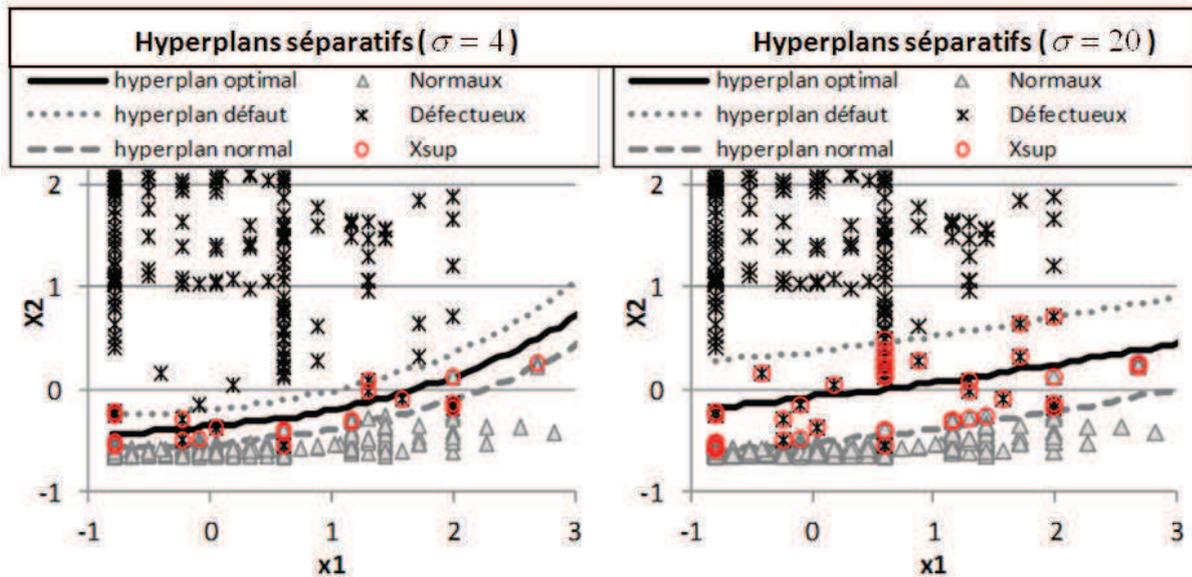
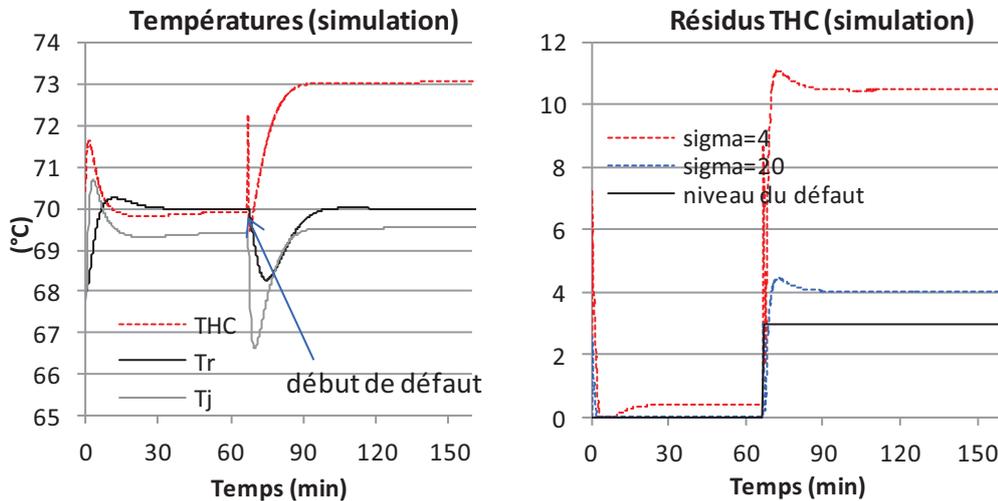


Figure IV.4 Classification des défauts de  $T_{HC}$ . Effet de  $\sigma$  (variance du noyau gaussien) sur les hyperplans optimaux (avec  $C=100$ )

Afin d'illustrer l'effet de  $\sigma$  dans le cas de détection de défaut, les fonctions de décision obtenues avec différentes valeurs de  $\sigma$  ont été évaluées par simulation. La Figure IV.5 à gauche montre qu'une faute se produit en  $T_{HC}$  à 66 min. Comme  $T_{HC}$  est utilisé dans un régulateur PID local du système de HC, un refroidissement rapide est effectué pour ramener  $T_{HC}$  vers  $T_j^d$  afin de maintenir  $T_r^d = 70^\circ\text{C}$ . En conséquence,  $T_r$  s'écarte de la consigne pendant environ 15 min. En raison de la présence d'une intégrale dans le contrôleur de  $T_r$ ,  $T_j^d$  change de manière à ramener  $T_r$  à  $T_r^d$  même en présence de défaut dans  $T_{HC}$ . Un régulateur proportionnel seul ne serait pas en mesure de compenser ce défaut. Par conséquent, le système de contrôle est robuste à des défauts dans cette sonde, mais nécessite une détection efficace dans la sonde de température de la réaction, qui sera examinée dans la section suivante. On note que même si le défaut dans  $T_{HC}$  est brutal, il n'est pas évident à détecter visuellement par l'opérateur car il apparaît comme une perturbation ponctuelle, ce qui justifie l'utilisation des SVMs.

La Figure IV.5 à droite montre que le défaut est instantanément détecté par les deux fonctions de décision (avec  $\sigma = 4$  et  $\sigma = 20$ ). L'amplitude du résidu est plus élevée avec  $\sigma = 4$ . Le seuil peut être fixé à 0,1 (en théorie, il peut être 0) ce qui signifie que la situation est estimée normale si la valeur du résidu est inférieure à  $|+/-0,1|$ . On remarque qu'une faible valeur de  $\sigma$  conduit à des fausses alarmes. Les fausses alarmes ont eu lieu avec les deux fonctions au début de la réaction due à la vitesse de réaction élevée dans cette période. En guise de compromis,  $\sigma$  a été fixé à 10. En utilisant la relation proposée par (Cherkassky and Ma, 2004),  $\sigma=(0,2-0,5)\times\text{plage}(x)$ , et avec la

normalisation zscore  $\frac{x - \text{moyenne}(x)}{\text{écart type}(x)}$ , la gamme de  $x$  est d'environ [8-10] en l'absence de défaut, ce qui donne  $\sigma \approx 2-5$ . On rappelle que les valeurs de  $\sigma$  optimisées dans ce travail sont en général légèrement supérieures à celles proposées par (Cherkassky and Ma, 2004).



**Figure IV.5 Cas d'étude par simulation (avec  $C=100$ ) et différentes valeurs de  $\sigma$ : Occurrence de défaut dans  $T_{HC}$  (objective de commande  $T_r^d=70^\circ\text{C}$ )**

- **Réglage de  $C$  :**

En ce qui concerne le paramètre  $C$  qui borne les multiplicateurs de Lagrange, il permet de créer un compromis entre la pénalisation d'une mauvaise classification et la minimisation de la complexité de fonction. La Figure IV.6 montre que l'augmentation de la valeur de  $C$  permet de réduire la largeur du tube (similaire à la diminution de  $\sigma$ ), ce qui diminue le nombre de vecteurs de support. Ceci est supposé jusqu'à une certaine limite améliorer les résultats des estimations. Toutefois, cela pourrait réduire la capacité de généralisation. On note aussi que le changement de l'un des paramètres de réglage conduit à  $\mathbf{x}_{\text{sup}}$  et  $\alpha$  différents. Augmenter  $C$  pourrait conduire à une augmentation de la dimension de  $\mathbf{x}_{\text{sup}}$  ce qui augmente le temps de calcul en ligne et pourrait réduire la capacité de généralisation. Cette conséquence n'est cependant pas systématique.

De plus le vecteur  $\mathbf{x}_{\text{sup}}$  obtenu dépend de la marge et des points de données passant près de la fonction dans une configuration spécifique du problème d'optimisation. En guise de compromis, dans cette application,  $C$  a été fixé à 100.

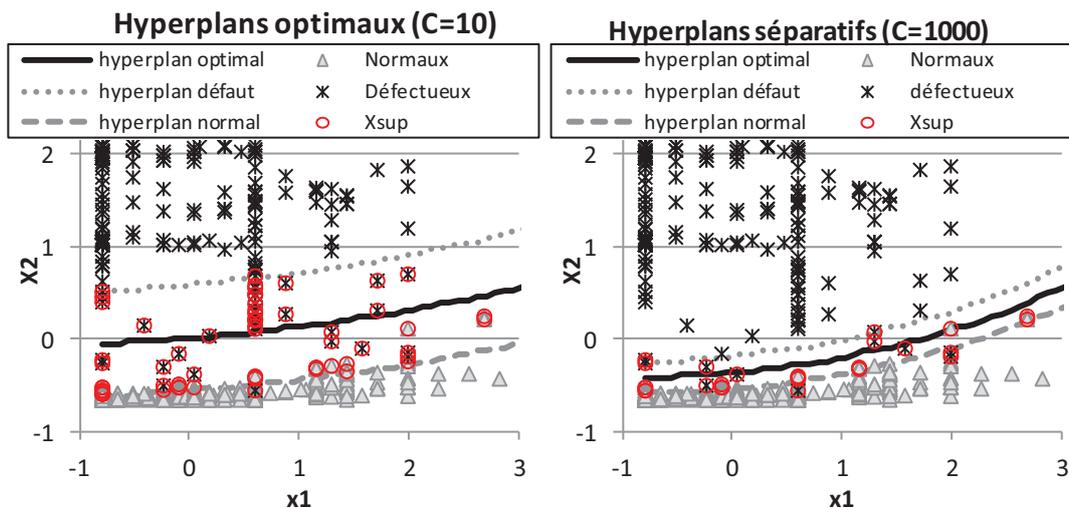
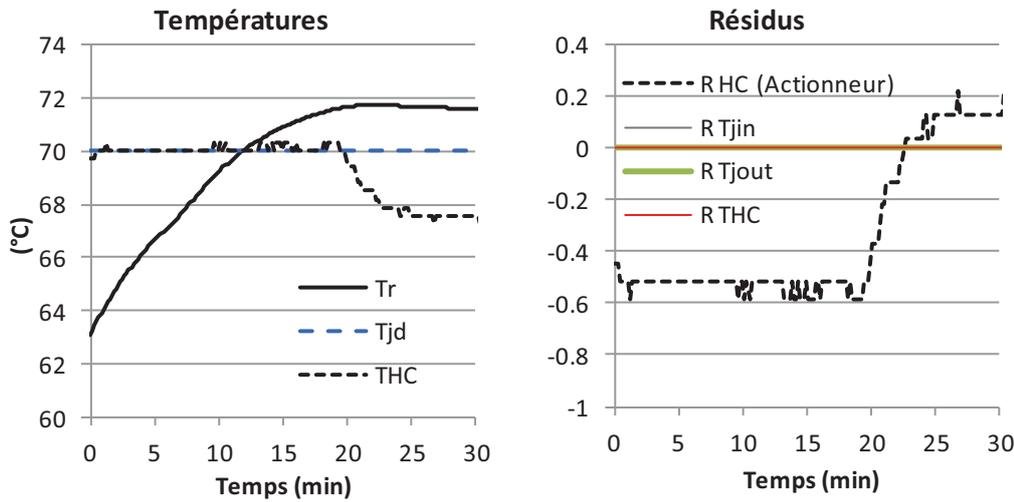


Figure IV.6 Classification des défauts de  $T_{HC}$ . Effet de  $C$  (paramètre de régularisation) sur les hyperplans optimaux (avec  $\sigma=10$ )

#### IV.3.3.3. Défaut d'actionneur dans HC

En ce qui concerne les défauts d'actionneurs (système **HC**), ils sont dus à différentes raisons : pompes, résistance thermique ou système de contrôle, logiciel. Certains défauts sont compensés par le régulateur PID local et ne donnent aucune incidence sur les mesures de processus. Seuls les défauts qui conduisent à l'éloignement de  $T_{HC}$  de la consigne peuvent être détectés et isolés. La Figure IV.7 montre une expérience réelle où les paramètres du PID du système **HC** ont été changés à environ 20 min. Cela a provoqué une légère dérive  $T_{HC}$  de  $T_j^d$ . Lorsque l'impact de défaut conduit à  $T_j^d - T_{HC} = 2,4^\circ\text{C}$ , le résidu de l'actionneur traverse 0,1 (à  $T_j^d - T_{HC} = 1,8^\circ\text{C}$ , le résidu traversé 0). Les résidus des sondes  $T_{HC}$ ,  $T_j^{in}$  et  $T_j^{out}$  sont restés égaux à zéro.



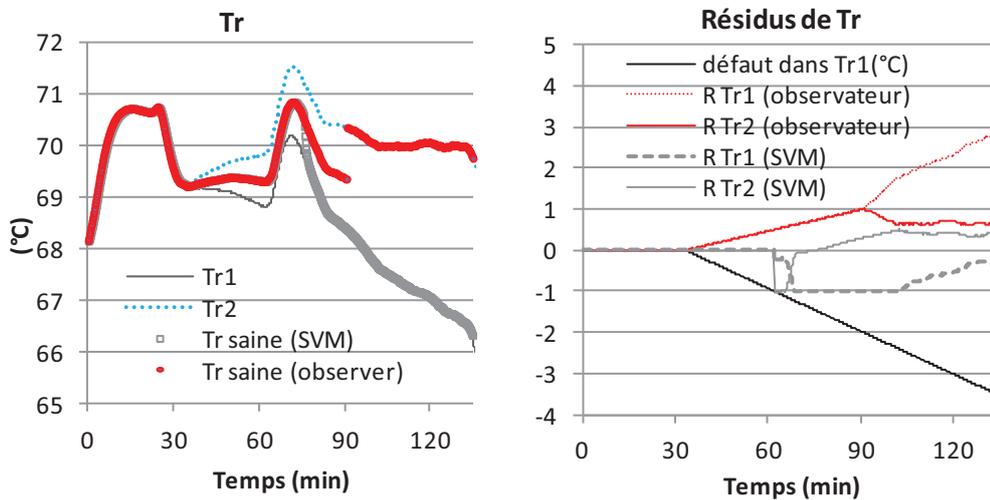
**Figure IV.7 Défaut dans le système de chauffe/refroidissement (HC : un actionneur) en changeant les paramètres du PID afin d'écartier  $T_{HC}$  de  $T_j^d$  ( $T_j^d=70^\circ\text{C}$ )**

#### ***IV.3.3.4. Défaut de capteur dans $T_r$ à l'échelle de laboratoire***

Dans ce scénario, nous utilisons à la fois la méthode SVM et l'observateur à grand gain pour la détection et l'isolation de défaut en raison de la non-linéarité élevée et de la variabilité du comportement à l'intérieur du réacteur. Afin d'appliquer l'observateur, certains paramètres du modèle sont nécessaires (voir équation IV.1). Beaucoup d'entre eux sont connus dans le réacteur de laboratoire (ex.  $A$ ,  $M_j$ ,  $C_{pj}$ ,  $F_j$ ). Le principal paramètre inconnu à l'échelle du laboratoire est la perte de chaleur de la double enveloppe à l'environnement  $Q_{ij}$  qui dépend de la différence entre la température de la double enveloppe et la température ambiante, qui varie. Le coefficient de transfert de chaleur à la double enveloppe peut également varier au cours de la réaction dû à l'encrassement de la paroi. Des approximations constantes de ces paramètres sont utilisées dans le modèle (obtenues par calibration).

La Figure IV.8 montre une évaluation expérimentale des deux méthodes dans le réacteur du laboratoire. Une dérive a été introduite artificiellement aux mesures réelles de  $T_{r1}$  à 33 min. Le défaut est détecté par SVM à 75 min, mais est identifié comme un défaut  $T_{r2}$ . Le cas est détecté et isolé correctement par l'observateur à 88 min (seuil = 1). La méthode SVM a pu détecter et correctement isoler le défaut dans d'autres scénarios, mais la robustesse de l'observateur était toujours plus élevée. On rappelle que la méthode SVM est utilisée pour la détection de défaut et que cette information est nécessaire pour la correction du modèle dans l'observateur. D'après les résultats de cette section, on peut conclure que l'application d'une méthode statistique nécessite l'utilisation de données exhaustives pour l'apprentissage en raison de la grande vitesse de

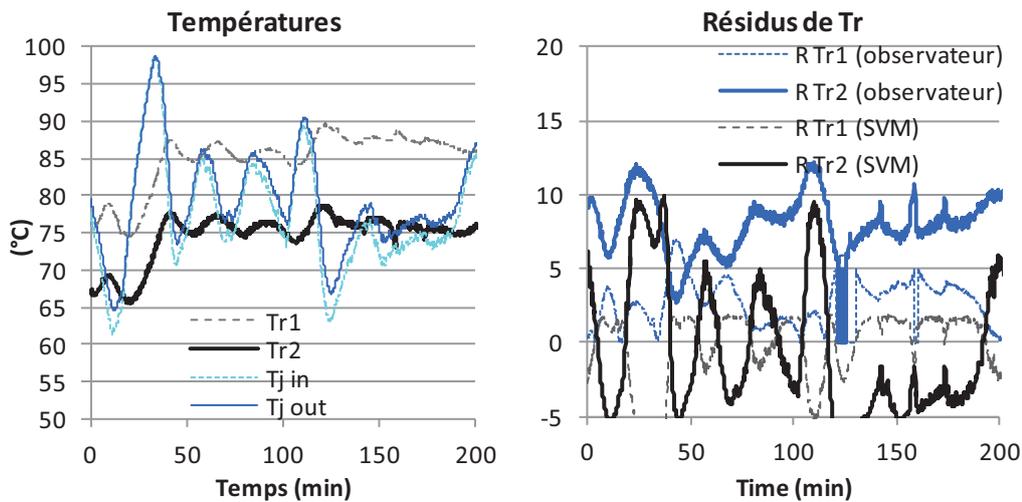
réaction et de la variation de la dynamique du procédé. L'utilisation d'un modèle du procédé simplifié peut renforcer l'isolation de défauts, qui est faite à l'aide d'un observateur. L'utilisation des SVMs devrait toutefois rester le premier choix, car ils sont faciles à mettre en œuvre et capables de détecter la plupart des défauts de capteurs et d'actionneurs presque instantanément avec peu de données d'apprentissage.



**Figure IV.8 Dérive dans  $T_{r1}$  à 33 min dans le réacteur du laboratoire**

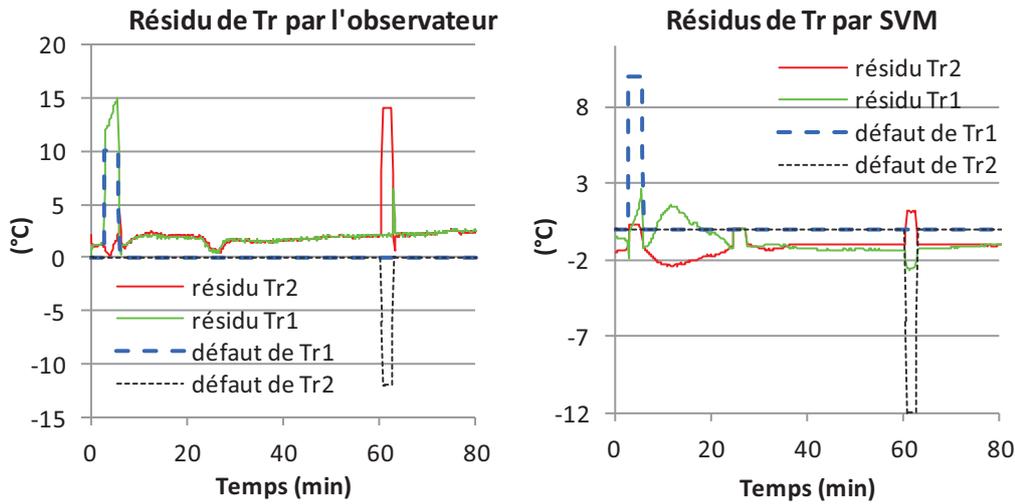
#### ***IV.3.3.5. Défaut de capteur dans $T_r$ à l'échelle de pilote***

Pour le réacteur pilote de 250 L, les différences entre  $T_{r1}$  et  $T_{r2}$  sont élevées (environ  $4^{\circ}\text{C}$  à l'état stationnaire) en raison du volume plus grand du réacteur. Ceci rend la détection de défaut plus difficile qu'à l'échelle du laboratoire. On note que l'une des sondes est placée à proximité de l'agitateur et une autre au fond du réacteur, plus proche de la paroi. Les deux sondes sont nécessaires pour évaluer l'homogénéité du réacteur, ce qui donne des informations indirectes sur la coagulation de latex sur l'agitateur ou sur la paroi du réacteur. Afin d'appliquer l'observateur dans ce système, des paramètres approximatifs sont utilisés par simple extrapolation approximative du réacteur du laboratoire. Bien évidemment, ces paramètres sont faux, mais justement, la méthode doit y faire face. Pour les autres capteurs et actionneurs, seule la méthode SVM doit être utilisée, ce qui évite la modélisation du procédé en entier. Notez que la méthode d'estimation est en partie basée sur la classification SVM pour la correction du modèle en l'absence de fautes.



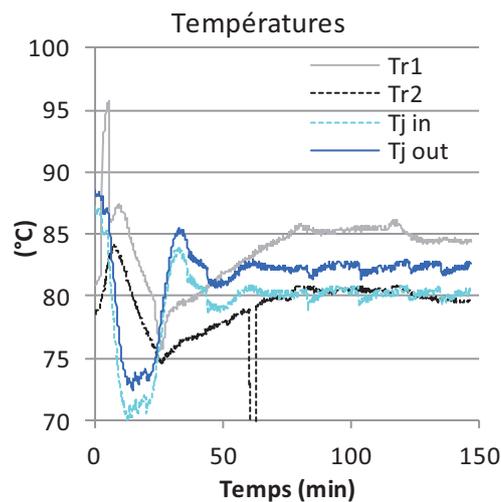
**Figure IV.9 Erreur de calibration de  $-5^{\circ}\text{C}$  est considérée dans la sonde  $T_{r2}$  dans le réacteur pilote de 250 L**

La Figure IV.9 montre les résultats obtenus par les deux méthodes dans le réacteur pilote au cours d'une réaction typique de copolymérisation en semi-continu. Une erreur de calibration de  $-5^{\circ}\text{C}$  est considérée dans  $T_{r2}$ . Une telle erreur ne peut pas être détectée par l'analyse visuelle des courbes de température où une multitude de situations est possible. En effet, la dynamique de la réaction varie surtout pendant les premières minutes et la différence entre  $T_{r1}$  et  $T_{r2}$  n'est pas constante dans toutes les expériences. Les oscillations dans ce scénario sont dues au système de contrôle, non traité dans ce projet. L'observateur parvient à détecter et isoler l'erreur statique instantanément. Le seuil de détection dans le réacteur pilote est fixé à environ 4. Le résidu de  $T_{r1}$  obtenu par l'observateur est inférieur au seuil. La méthode SVM parvient à détecter le défaut rapidement, mais elle ne parvient pas à l'isoler tout le temps car les résidus de  $T_{r1}$  et  $T_{r2}$  oscillent. Une plus grande confiance peut donc être attribuée à l'observateur. Notez que dans un tel scénario, la correction du modèle n'a pas lieu car une erreur est détectée pendant toute l'expérience. La correction du modèle de correction n'a lieu que dans des situations normales, par exemple dans les premières minutes du scénario précédent (Figure IV.9).



**Figure IV.10 Une évaluation expérimentale de la méthode SVM et de l'observateur dans un réacteur pilote de 250 L**

La Figure IV.10 montre les résultats obtenus dans un second scénario par les deux méthodes dans le réacteur pilote. Un défaut de niveau 10 est introduit dans ce cas dans  $T_{r1}$  à 8 min (de manière artificielle) et un défaut de niveau  $-12^{\circ}\text{C}$  dans  $T_{r2}$  à 64 min. La méthode basée sur l'observateur parvient à détecter et isoler les deux défauts. La méthode SVM parvient à détecter et d'isoler la deuxième panne dans  $T_{r2}$ . Mais elle a un retard dans la détection du premier défaut, et provoque quelques fausses alarmes au cours des premières minutes de la réaction si le seuil est fixé à 0. On note que le premier défaut a lieu pendant de fortes variations de la dynamique du procédé comme le montrent les variations dans les températures du réacteur (Figure IV.11).



**Figure IV.11 Evolution des températures dans le réacteur pilote sans défaut**

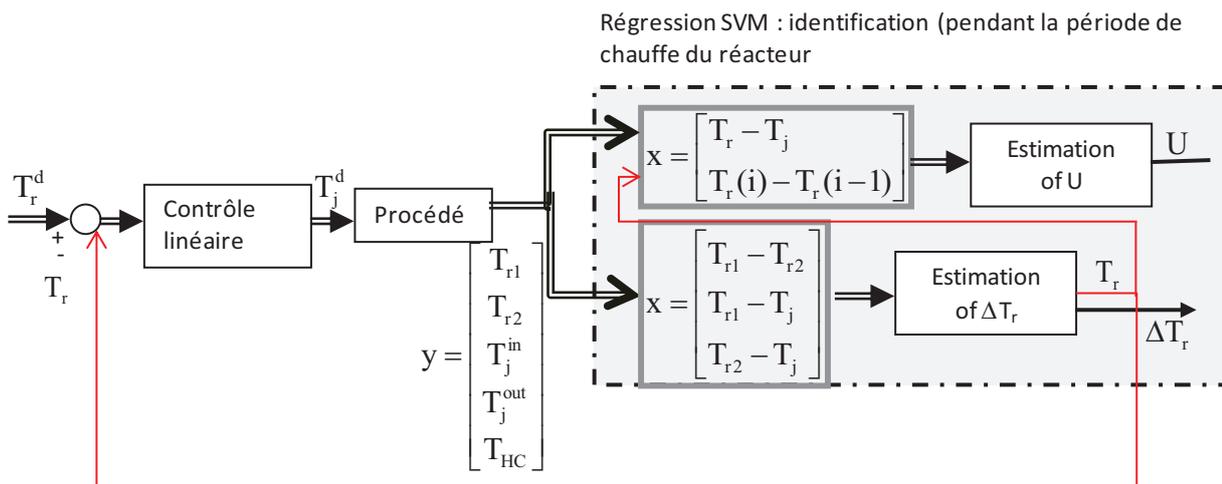
## IV.4. Implémentation de SVR

Pendant la phase de chauffage du réacteur, le modèle de procédé (éq. IV.1) est réduit à un modèle linéaire ( $Q_r = 0$  et  $U$  est supposé constant sur la plage de température considérée). Cette phase peut être utilisée pour des finalités de calibration :

- des sondes de températures, où des dérives peuvent avoir lieu dans le temps.
- du coefficient d'échange thermique  $U$ , qui peut évoluer au cours du temps dû à l'encrassement du réacteur.

Des dérives légères dans les températures du réacteur ( $T_{r1}$  et  $T_{r2}$ ) peuvent être détectées rien qu'en examinant la différence entre les deux mesures. Cependant, isoler la sonde défectueuse reste difficile et nécessite l'utilisation d'une méthode de classification. Au lieu d'appliquer d'abord une méthode de classification qui isole la sonde défectueuse et d'ensuite appliquer une analyse des données afin d'identifier le niveau du défaut, nous proposons d'utiliser la méthode de régression SVR directement. La détection du niveau de la dérive peut servir à la correction de la sonde avant la réaction.

L'estimation du coefficient d'échange de chaleur entre le réacteur et la double enveloppe ( $U$ ) est essentielle pour le suivi contrôle de la réaction. On utilisera la méthode SVR et on comparera les résultats à ceux obtenus avec un observateur à grand gain. La mesure saine de  $T_r$  sera utilisée en cas de défaut dans l'une des sondes  $T_{r1}$  ou  $T_{r2}$ .



**Figure IV.12 Schéma de SVR pour l'estimation de  $U$  et l'identification de niveaux de défauts pendant la période la chauffe du réacteur**

#### IV.4.1. Identification d'une dérive de température par régression SVM

La régression SVM (SVR) est utilisée pour la détection, l'isolation et l'identification du niveau de défaut dans les sondes de températures  $T_{r1}$  et  $T_{r2}$  dans le réacteur. L'algorithme employé (apprentissage et validation) peut être résumé comme suit :

- Recueillir des données réelles ( $y$ ) des différents capteurs au cours de deux périodes de chauffe du réacteur (30 min de chauffage de chaque période d'échantillonnage avec  $T_e=10$  s),
- La première expérience est utilisée telle quelle et la seconde est artificiellement modifiée en ajoutant diverses  $\Delta T_{r1}$  (de +/-1 à +/-4°C) à différents moments au cours d'une phase de chauffage. Les défauts simultanés dans  $T_{r1}$  et  $T_{r2}$  ne sont pas considérés. Cela a donné 50 points de données dans une situation défectueuse pour  $T_{r1}$  et 310 points de données normales. Le modèle de  $T_{r1}$  sera utilisé également pour  $T_{r2}$ , donc il n'y a pas d'apprentissage nécessaire pour  $T_{r2}$ . Les données n'ont pas été normées dans le cas de la régression des défauts de  $T_r$  et de l'estimation de  $U$ .
- Le vecteur suivant est utilisée comme entrée pour le développement du modèle SVM car le premier terme donne des informations sur le niveau de la dérive (pour la détection et l'identification) et le second terme est utile pour l'isolement (la taille de  $x=360$ ) :

$$x_i = \left( T_{r1}(i) - T_{r2}(i) \quad T_{r1}(i) - T_j(i) \quad T_{r2}(i) - T_j(i) \right)^T$$

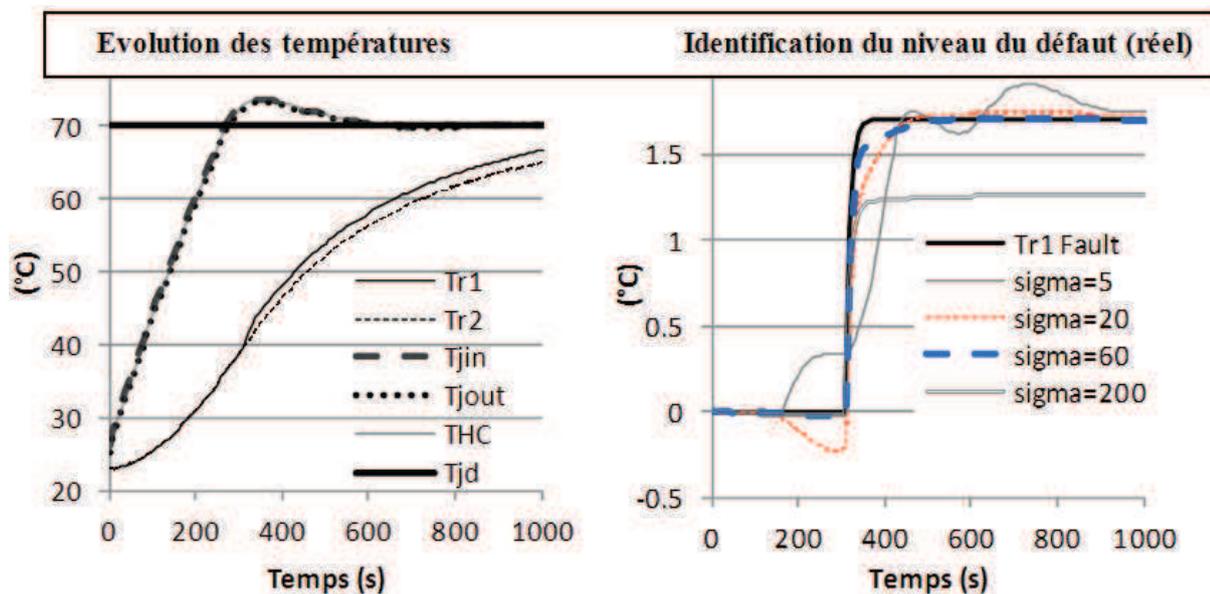
- Le développement du modèle (ou de la fonction de décision) pour l'identification du niveau de défaut est ensuite réalisé en utilisant la méthode SVR. Cela se fait par l'optimisation de l'équation (IV.6) qui donne les paramètres  $\alpha_i$ ,  $\mathbf{b}$  et les vecteurs de support ( $\mathbf{x}_{sup}$ ) pour un réglage donné de  $\mathbf{C}$ ,  $\sigma$ , choix du noyau et les paramètres du noyau. On note que dans le cas de la régression, il faut également régler la largeur du tube  $\varepsilon$  (par rapport au cas de la classification).
- Le modèle obtenu est ensuite évalué dans des situations nouvelles par simulation. Pour ce faire, de nouveaux vecteurs  $\mathbf{x}$  sont introduits dans le noyau sélectionné, puis le résidu (donnant ici le niveau de défaut) est calculé à partir de  $\alpha_i$ ,  $\mathbf{b}$  et  $\mathbf{x}_{sup}$  en utilisant l'équation(IV.6). En utilisant le noyau gaussien, la fonction de décision pour des nouvelles mesures  $\mathbf{x}$  à temps  $\mathbf{t}$  est donnée par :

$$f(x) = \sum_{i=1}^{n_{sp}} (\alpha_i^- - \alpha_i^+) \underbrace{e^{-\frac{\|x_{sup i} - x\|^2}{2\sigma^2}}}_{K(x_{sup i}, x)} + b \quad (IV.6)$$

- Valider le modèle SVM expérimentalement pour des nouvelles  $x$  en utilisant l'équation(IV.6).

- **Réglage de la variance du noyau gaussien:**

La Figure IV.13 compare les résultats de trois modèles obtenus en utilisant différentes variances de la fonction noyau  $\sigma = 5, 20, 60, 200$  (pour des données non normées). Un défaut du niveau de  $1,7^\circ\text{C}$  est introduit à  $300\text{ s}$ . La détection de défauts est instantanée. Comme décrit précédemment dans le cas de la classification, on remarque que l'augmentation de  $\sigma$  permet de mieux identifier le niveau de défaut jusqu'à une certaine limite dans le cas de la régression aussi. Le seuil de résidu d'identification de  $\Delta T_r$  peut être fixé à  $0,5$ , ce qui signifie que la situation est estimée correcte si le niveau de défaut (résidu) est inférieur à  $|\pm 0,5^\circ\text{C}|$ . En appliquant la relation proposée par (Cherkassky and Ma, 2004) pour régler  $\sigma$  [ $\sigma = (0,2-0,5) \times \text{marge}(x)$ ], on trouve dans cet exemple  $\sigma \approx 12$ , qui est un peu faible (données non normées dans le cas de la régression).

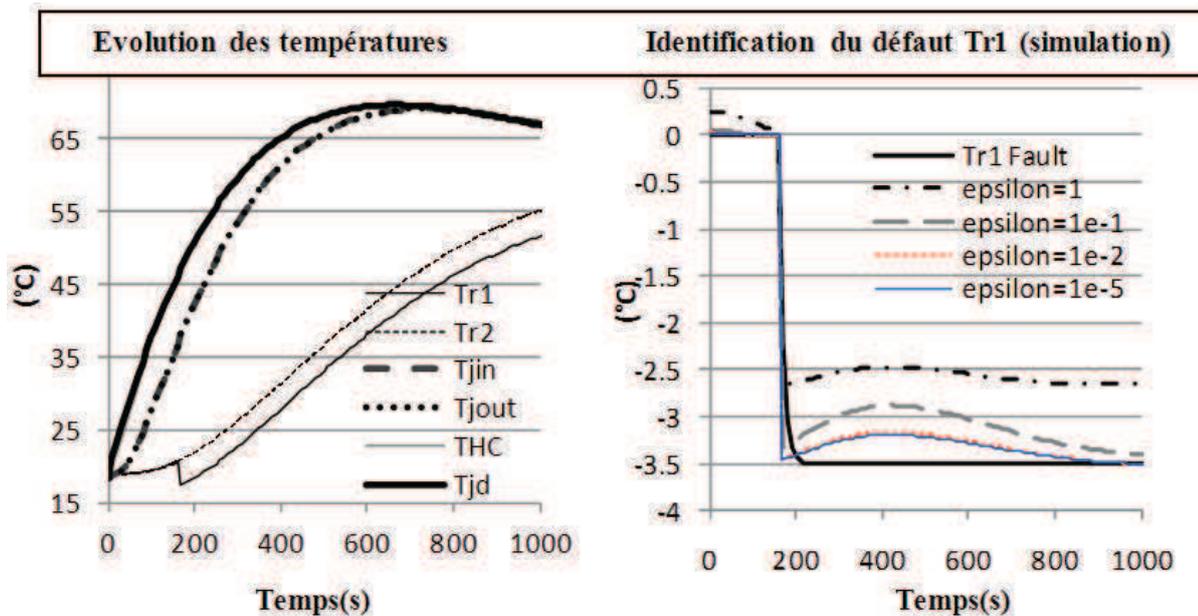


**Figure IV.13 Validation expérimentale de l'identification d'une dérive dans la sonde de température  $T_{r1}$  avec différentes valeurs de  $\sigma$  ( $\epsilon=0.01$ ,  $C=100$ , objectif de commande  $T_j^d=70^\circ\text{C}$ )**

- **Réglage de la largeur du tube :**

Dans le cas de la régression, un paramètre supplémentaire de réglage du modèle SVR par rapport à la classification est ajouté. Il s'agit de la largeur du tube autour des données d'apprentissage, qui pourrait affecter les estimations. L'augmentation de  $\epsilon$  réduit le nombre de vecteurs de support et donc la complexité du modèle, ce qui peut entraîner des estimations plus mauvaises.

La Figure IV.14 montre que la réduction de la valeur de  $\epsilon$  jusqu'à une certaine limite permet d'améliorer la qualité de l'estimation. Un défaut du niveau de  $-3,5^\circ\text{C}$  est introduit dans  $T_{r1}$  à 150s. Sur la base des observations,  $\epsilon$  a été fixé à 0,01. En utilisant la relation empirique proposée par (Cherkassky and Ma, 2004) nous obtenons  $\epsilon = 0,0038$ , ce qui est approprié. Il est toutefois important de noter que la qualité (variabilité, richesse) des données de formation est également importante et non seulement  $N$ . Par conséquent, leur équation peut être utilisée pour trouver une estimation initiale de  $\epsilon$  mais qui doit ensuite être affinée par l'utilisateur.

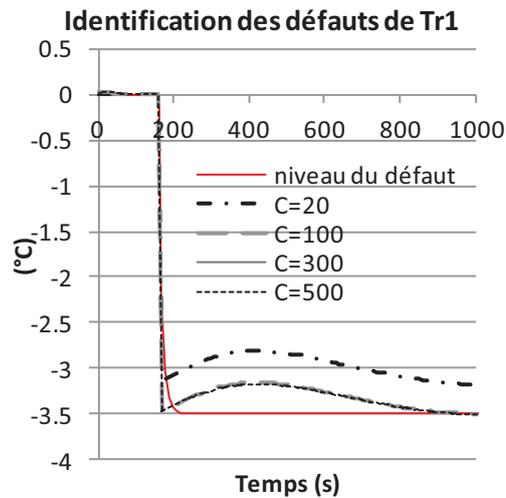


**Figure IV.14 Identification de la dérive dans la sonde  $T_{r1}$  avec différentes valeurs de  $\epsilon$**   
 ( $\sigma=60$ ,  $C=100$ , objectif de commande  $T_r^d=60^\circ\text{C}$ )

- **Réglage de C :**

Le réglage du paramètre  $C$  en régression est similaire à la classification. La Figure IV.15 montre que l'augmentation de la valeur de  $C$  jusqu'à une certaine limite permet d'améliorer les résultats des estimations. Dans cette application,  $C$  a été fixé à 100, en utilisant la relation proposée par (Cherkassky and Ma, 2004), comme  $z$  varie de  $\pm 1$  à  $\pm 4^\circ\text{C}$  et que la précision de la sonde est

0,1°C,  $C$  devrait être légèrement supérieure à 0,4. Comme on peut le voir sur la Figure IV.15, cette valeur ne donne pas de précision acceptable. Nous utilisons donc la valeur  $C=100$ .



**Figure IV.15 Identification de la dérive du capteur  $T_{r1}$  avec différentes valeurs de  $C$  ( $\epsilon=0.01$ ,  $\sigma=60$ )**

Avec les paramètres calculés dans les derniers paragraphes, l'optimisation donne  $x_{sup}$  de dimension 114, tandis que  $x$  était de taille 358, ce qui représente un bon compromis (réduction de la taille initiale à 32%). On note que le réglage de ces paramètres aurait pu se faire également par validation croisée (comme dans les autres chapitres), ce qui exige cependant un temps de calcul supérieur.

#### IV.4.2. Identification des $U$ par régression SVM

L'identification de  $U$  est importante pour la supervision et la commande de procédé, car il intervient dans le bilan thermique du réacteur. L'estimation de la valeur initiale de  $U$  peut être faite au cours de la phase de chauffe du réacteur, où les paramètres du modèle sont connus (vitesse de réaction nulle).

L'apprentissage de  $U$  est réalisé entièrement par simulation, car autrement, on aurait besoin de réacteurs avec différents matériaux. Différentes valeurs de  $U$  ont été considérées et pour chacune une simulation de chauffe de réacteur est réalisée sur 30 min ( $U=130-300$  W/m<sup>2</sup>/°C). Puis, comme pour l'identification d'une dérive  $\Delta T_r$ , les données collectées sont utilisées pour développer un modèle en utilisant la régression SVM. En s'inspirant de la structure du modèle du procédé (eq. IV.1), le vecteur d'entrée suivant est utilisé pour la régression pour l'identification de  $U$  (taille de  $x=450$ ):

$$x_i = \begin{pmatrix} T_r(i) - T_j(i) & T_r(i) - T_r(i-1) \end{pmatrix}^T$$

Bien évidemment, des mesures saines de  $T_r$  doivent être utilisées pour évaluer  $U$  en cas de détection de dérive dans  $T_{r1}$  ou  $T_{r2}$  pendant la phase de chauffe. En l'absence de défaut,  $T_r$  est calculée comme la moyenne entre  $T_{r1}$  et  $T_{r2}$ . Les paramètres évalués comme optimaux sont:  $\varepsilon=1 \times 10^{-2}$ ,  $C=7000$  et  $\sigma=30$  (les paramètres calculés par (Cherkassky and Ma, 2004) sont  $\varepsilon \approx 0.34$ ,  $C \approx 200$  et  $\sigma \approx 17$ , en supposant  $\nu=1$ ). On note que des valeurs les plus élevées de  $z$  (ici  $U$ ) nécessitent une valeur plus élevée que  $C$  ( $C = 100$  pour l'identification de  $\Delta T_r$ ). Avec ces paramètres, le pourcentage des vecteurs de support est de 100%.

Afin de valider ce modèle expérimentalement, il a été nécessaire de filtrer  $T_r$  avec un filtre passe-bas ( $\tau = 5$  s). Ceci est dû au terme dérivatif  $T_r(i) - T_r(i-1)$  dans  $x$ . Les résultats expérimentaux sont validés par comparaison à deux autres méthodes d'identification : Tout d'abord, un observateur non linéaire est développé et implémenté au cours de la phase de chauffe ; Aussi, une identification de  $U$  est effectuée par optimisation à l'état stationnaire avec introduction d'une résistance thermique donnant différents niveaux de chaleur.

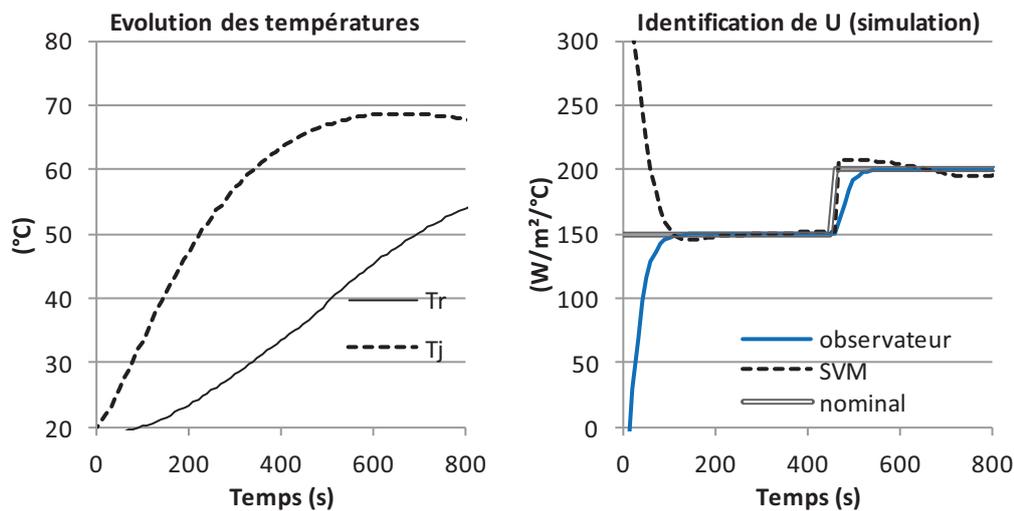
L'observateur de  $U$  utilisé est celui proposé par (Hammouri et al., 2002). Il est basé sur le modèle suivant, qui utilise les mesures  $T_r$  et  $T_j$  :

$$\begin{bmatrix} \dot{\hat{T}}_r \\ \dot{\hat{U}} \end{bmatrix} = \begin{bmatrix} 0 & (T_j - T_r)A / MC_p \\ 0 & 0 \end{bmatrix} \begin{bmatrix} T_r \\ U \end{bmatrix} \quad (IV.3)$$

En appliquant l'observateur non linéaire à grand gain proposé par (Gauthier et al., 1992), nous obtenons l'observateur suivant pour  $U$  :

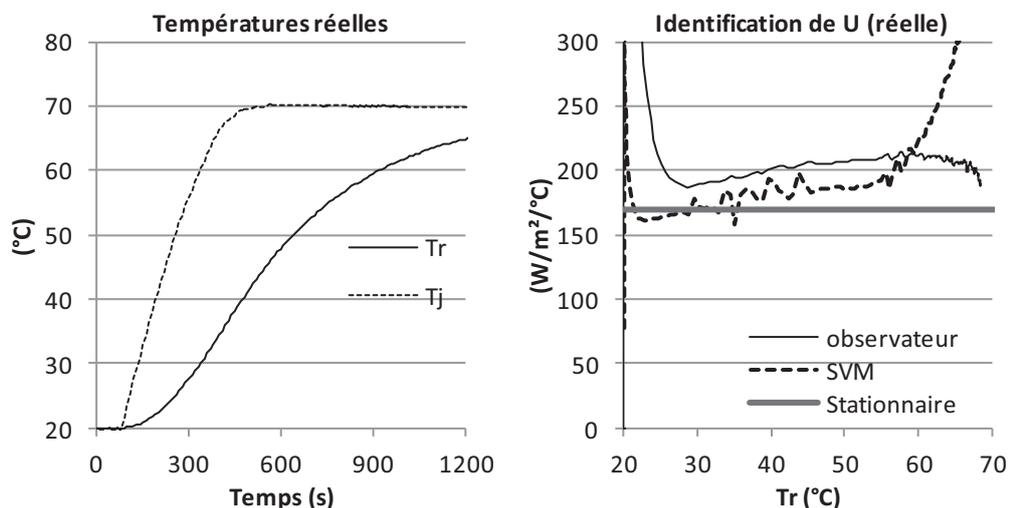
$$\begin{bmatrix} \dot{\hat{T}}_r \\ \dot{\hat{U}} \end{bmatrix} = \begin{bmatrix} 0 & \frac{A(T_j - T_r)}{MC_p} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{T}_r \\ \hat{U} \end{bmatrix} - \begin{bmatrix} 2\theta \\ MC_p \theta^2 / (A(T_j - T_r)) \end{bmatrix} (\hat{T}_r - T_r) \quad (IV.4)$$

La Figure IV.16 montre une simulation de la méthode SVM et de l'observation à grand gain, avec  $U = 150$  et  $200 \text{ W/m}^2/\text{°C}$ . L'observateur est initié à 0 et converge avec le temps vers la valeur nominale introduite dans la simulation. Même avec un bon réglage, on s'attend à ce qu'il converge un peu plus lentement que la méthode SVM. Le réglage du paramètre  $\theta$  a été choisi comme compromis entre la vitesse de convergence et la sensibilité au bruit de mesure. Le comportement des deux systèmes est acceptable. Lorsque  $T_r$  se rapproche de  $T_j$ , l'observabilité du système est réduite, ce qui cause une divergence des deux méthodes.



**Figure IV.16 Estimation du coefficient d'échange thermique U pendant la chauffe du réacteur (simulation avec différentes valeurs de U, paramètre de l'observateur  $\theta=0.04$ )**

Les deux méthodes ont ensuite été appliquées en ligne sur des données réelles de chauffe du réacteur afin d'estimer U.  $T_r$  a été filtrée avec un filtre passe-bas pour son utilisation dans la méthode SVM. La Figure IV.17 montre que l'observateur estime U à environ 207 W/m<sup>2</sup>/°C à  $T_r=50^\circ\text{C}$ . La valeur évaluée à l'état stationnaire par optimisation est égale à 170 W/m<sup>2</sup>/°C. Les estimations SVM se situent entre les deux valeurs (environ 190 W/m<sup>2</sup>/°C). La réponse de la méthode SVM est plus rapide pour la convergence et la divergence.



**Figure IV.17 Estimation expérimentale du coefficient d'échange thermique U pendant la chauffe du réacteur**

## IV.5. Conclusions

Dans ce chapitre, la méthode SVM a été combinée à un observateur non linéaire pour la détection de défaut. L'efficacité de la stratégie proposée a été démontrée par simulation et expérimentalement dans deux réacteurs de volumes différents. La mise en œuvre de la méthode SVM est rapide car elle ne nécessite pas de modélisation du procédé. Avec un bon réglage des paramètres et des données complètes, la méthode SVM peut assurer une bonne généralisation. En outre, le temps de calcul en ligne de SVM est réduit en raison de l'utilisation uniquement des vecteurs de support dans la fonction de décision. Il est démontré que lorsque la dynamique du procédé est très variable, l'utilisation d'un modèle simplifié dans la partie fortement non linéaire peut être utile pour améliorer l'isolation du défaut. Dans ce cas, l'information obtenue par SVM est utilisée pour corriger le modèle simplifié en l'absence de défaut. De cette manière, une méthode hybride FDI qui est robuste aux erreurs de modélisation et des variations de la dynamique est proposée. La méthode a pu être appliquée à un réacteur pilote pour lequel les paramètres du modèle ne sont pas connus et où les dynamiques varient en raison de la nature exothermique de la réaction.

# Conclusion générale

*« Le doute est le commencement de la sagesse » Aristote*

La Croissance des exigences en termes de productivité, qualité, économie, sécurité, maintenance et respect de l'environnement nous amènent à réfléchir à améliorer davantage les outils de surveillance et de diagnostic de défauts.

Le choix porté dans cette thèse sur les méthodes d'apprentissage automatique « SVMs » était, à mon avis, assez ambitieux et pas facile malgré le caractère solide de leur base théorique. Ces méthodes restent juste un outil qu'il faut bien comprendre, bien mettre dans le contexte de travail et l'accompagner par non seulement une bonne stratégie, mais aussi d'autres outils ou méthodes qui permettent de faciliter sa tâche de diagnostic de défauts.

La finalité de leur utilisation est d'exploiter de la meilleure façon la quantité des données disponibles avec les technologies modernes afin d'extraire les informations utiles et le maximum de connaissances.

Après les années de recherche dans le sujet de diagnostic de défauts, j'ai constaté que cette tâche doit être vue sous différents angles pour qu'elle soit réalisée de la meilleure façon possible, car les procédés réels sont complexes et regroupent différentes disciplines tels que le traitement de signal, automatique, statistique, mécanique, physique, chimie, biologie, probabilités, ... etc.

La réalisation d'un bloc de diagnostic de défauts précis et efficace est fondamentale pour donner une information précise de l'état du procédé à l'opérateur et aussi pour toute reconfiguration de commande dans le cadre d'une commande tolérante aux défauts, ainsi que pour toute procédure d'automatisation d'un système.

Je crois que les recherches réalisées dans cette thèse apportent une contribution relative à l'estimation et la modélisation, et aussi aux techniques de diagnostic de défauts par une méthode assez originale qui a montré de bons résultats en termes de précision, temps de détection et surtout le pourcentage presque nul des fausses alarmes.

La stratégie de détection et de diagnostic de défauts proposée dans cette thèse se base globalement sur les Machines à Vecteurs de Support (SVMs). Elle permet de manipuler facilement le bruit, les perturbations et les valeurs aberrantes, et offre une grande capacité de généralisation sur de nouvelles situations non utilisées dans l'apprentissage.

J'ai pu démontrer la puissance de prédiction de SVR dans ce travail. Ces bons résultats obtenus ne pouvaient pas être atteints si on avait négligé la mise à l'échelle des données, l'introduction des décalages temporels ainsi que le choix des régresseurs. Je peux conclure que la technique proposée dans cette thèse utilisant SVR fournit une façon simple et efficace de modéliser les systèmes réels tout en captant la non-linéarité et la variabilité qui décrivent le comportement dynamique du système.

La décision par rapport au défaut obtenue par l'approche basée sur SVMs proposée offre une solution nette, rapide, précise et sans fausses alarmes. Ce résultat est confirmé sur plusieurs systèmes non linéaires complexes : éolienne, procédé de traitement des eaux usées et réacteur chimique parfaitement agité.

A l'issue de ce travail, la combinaison entre Séparateurs à vaste Marge (SVM) et un observateur a permis la détection et l'isolation en déterminant le changement des paramètres estimés en ligne par l'observateur de type Kalman.

L'objectif de cette thèse était de proposer une stratégie de diagnostic de défauts assez générale et efficace pouvant être appliquée par la suite aux différents procédés réels. Les résultats obtenus dans ce travail me permettent de confirmer que l'objectif a été atteint en grande partie. Néanmoins, comme tout travail de recherche qui opte pour la perfection, la poursuite de ces travaux apparaît une nécessité afin d'améliorer encore l'estimation, la surveillance et le résultat de diagnostic de ces systèmes complexes.

Les travaux entrepris dans cette thèse constituent les fondations de base pour plusieurs recherches en perspective telles que :

- L'intégration à la stratégie de diagnostic de défauts, un modèle appris en ligne et non pas hors ligne comme proposé dans la thèse.
- Un des avantages du SVR est qu'il ne dépend pas des conditions initiales et qu'il apprend rapidement. Cependant, il doit être modifiable et adaptable, d'une manière efficace, pour faire face aux changements dans les statistiques du procédé. Par conséquent, j'envisage dans le futur d'améliorer l'estimateur non-linéaire SVR. Dans cette perspective, j'estime que ceci peut être fait en rajoutant un aspect d'adaptation par un terme correctif (comme celui de l'observateur). Ce qui peut ouvrir un nouvel axe de recherche.
- J'envisage d'ajouter à la stratégie de diagnostic de défaut actuelle la variante *SVM une classe* au début. Cela va permettre de déterminer si les nouvelles données acquises en ligne correspondent

a une situation complètement nouvelle et différente qui n'appartient pas au cas de défauts connus ni au cas de fonctionnement normal.

- La combinaison de SVMs avec la transformée d'ondelettes va également permettre à mon avis d'améliorer les résultats puisque l'ondelette est un outil puissant de traitement de signal capable de montrer les tendances et les discontinuités des mesures qui sont utilisées dans un schéma de diagnostic de défauts.

Par ailleurs, afin de rendre le processus d'apprentissage un peu plus automatisé, l'utilisation des algorithmes génétiques peut aider dans ce sens, et plus précisément pour le choix des caractéristiques pertinentes descriptives de défaut dans le cas de diagnostic ou celles descriptives d'une variable à estimer (régresseurs).

Une autre perspective de recherche consiste à envisager de réaliser par la suite une commande tolérante aux défauts pour les éoliennes et le procédé de traitement des eaux usées.



# Notations

## Notations relatives aux Machines à Vecteurs Supports

$\  \cdot \ $	Norme d'un vecteur
$\langle \cdot, \cdot \rangle$	Produit scalaire
$f(x_i)$	Fonction de décision
$R_{emp}[f]$	Fonction risque empirique
$R[f]$	Fonction risque réel
$\alpha_i, \mu_i$	Multiplicateurs de Lagrange
$\zeta_i$	Variabes de pénalité
$C$	Paramètre de régularisation
$\Phi$	Fonction non linéaire de projection
$\lambda_i$	Les valeurs propres
$K(x, x_i)$	Fonction noyau
$\sigma$	La largeur de distribution du noyau gaussien
$P(x)$	Probabilité de distribution de $x$
$P(x, y)$	Probabilité de distribution conjointe de $x, y$
$x_i$	Exemple d'apprentissage
$B$	Le biais
$D$	Ensemble de données d'apprentissage

D	Degré du noyau polynomial
dVC	Dimension de Vapnik et Chervonenkis
H	Hyperplan
K	Nombre de classes
L	Lagrangien
N	Nombre d'exemples d'apprentissage
<i>Sgn</i>	Fonction signe
w, b	Paramètres de la fonction de décision linéaire $f_w(x)$
$X_{sup}$	Les vecteurs supports

## Notations relatives au procédé de traitement des eaux

$NH_3$	Ammoniac
$NH_4^+$	Ion ammonium
$NO_3^-$	Nitrates
$NO_2^-$	Nitrites
$S_O^*$	La concentration de saturation pour l'oxygène
$\mu_A$	Taux de croissance maximal de la biomasse autotrophe
$f_P$	Fraction de biomasse recyclée en produits particuliers
$\mu_H$	Taux de croissance maximal de la biomasse hétérotrophe
$\eta_g$	Facteur de correction pour la croissance anoxique des hétérotrophes
$\eta_h$	Facteur de correction pour l'hydrolyse sous condition anoxie
$v_0'$	Vitesse maximale de décantation
$v_0$	Vitesse maximale de décantation (Vitesse de Vesilind)
$r_h$	La décantation freinée
$r_P$	La décantation de particules isolées

$f_{ns}$	Fraction non décantable
$J_{s,j}$	Flux de particules qui décantent de la couche j vers la couche j+1
$X_{min}$	Concentration en solides non décantables
$V_{s,j}$	Vitesse de décantation dans la couche j
$v_{ij}$	Les coefficients stœchiométriques
$b_A$	Taux de décès autotrophe de la biomasse autotrophe
$b_H$	Taux de décès hétérotrophe
DBO <sub>5</sub> ou BOD <sub>5</sub>	Demande Biologique en Oxygène sur 5 jours ( <i>Biological Oxygen Demand</i> )
DCO ou COD	La Demande Chimique en Oxygène ( <i>Chemical Oxygen Demand</i> )
HO <sup>-</sup>	Ion hydroxyde
$i_{XB}$	Fraction d'azote dans la biomasse
$i_{XP}$	Fraction d'azote dans les produits de lyse
$k_a$	Taux d'ammonification
$k_h$	Taux d'hydrolyse spécifique maximal
$K_{la5}$	Coefficient de transfert d'oxygène
$K_{NH}$	Constante de demi-saturation de l'azote ammoniacal pour la biomasse autotrophe
$K_{NO}$	Constante de demi-saturation du nitrate pour la biomasse hétérotrophe
$K_{O,A}$	Constante de demi-saturation de l'oxygène pour la biomasse autotrophe
$K_{O,H}$	Constante de demi-saturation de l'oxygène pour la biomasse hétérotrophe
$K_S$	Constante de demi-saturation hétérotrophe
$K_X$	Constante de demi-saturation pour l'hydrolyse du substrat lentement biodégradable $X_S$
MES ou TSS	Les Matières En Suspension ( <i>Total Suspended Solids</i> )
NH <sub>3</sub>	Ammoniac
NH <sub>4</sub> OH	Ammoniaque
NTK ou TKN	Azote total ( <i>Total Kjeldahl Nitrogen</i> )
$Q_{intr}$ ou $Q_a$	Débit de recyclage interne
$r_j$	Vitesses des réactions
$S_{ALK}$	Alcalinité
$S_I$	Matière soluble inerte
$S_{ND}$	Azote organique soluble et biodégradable
$S_{NH}$	Azote ammonium (NH <sub>4</sub> <sup>+</sup> ) et ammoniac (NH <sub>3</sub> )

$S_{NO}$	Azote sous forme de nitrite ( $NO_2^-$ ) et de nitrate ( $NO_3^-$ )
$S_O$	Oxygène dissous
$S_S$	Substrat soluble facilement biodégradable
$X_{B,A}$	Biomasse active autotrophe
$X_{B,H}$	Biomasse active hétérotrophe
$X_I$	Matière organique particulaire inerte
$X_{ND}$	Azote organique particulaire et biodégradable
$X_P$	Produits particuliers résultant de décomposition de la biomasse
$X_S$	Substrat lentement biodégradable
$Y_A$	Rendement de croissance autotrophe
$Y_H$	Rendement de croissance hétérotrophe

## Notations relatives à l'éolienne

$\beta$	Angle de calage des pales « <i>pitch</i> »
$\beta_r$	Position du pitch de référence
$\beta_m$	Position du pitch mesurée
$\rho$	La densité de l'air
$\lambda$	Rapport de vitesse à la pointe de l'éolienne
$v_w$	Vitesse du vent qui agit sur l'éolienne
$\tau_w$	Le couple du vent qui agit sur les pales
$\tau_{w,m}$	L'estimation du couple du vent basée sur la mesure de la vitesse du vent
$\tau_r$	Couple du rotor
$\tau_g$	Couple du générateur
$\eta_g$	Le rendement de la génératrice
$\tau_{g,m}$	Couple de génératrice mesurée

$\tau_{g,r}$	Couple de référence de la génératrice
$\theta_{\Delta}$	L'angle de torsion du système d'entraînement « drive train »
$\zeta$	Le coefficient d'amortissement d'un système de second ordre
$C_q(\lambda, \beta)$	Le coefficient de couple aérodynamique
$w_r$	Vitesse rotationnelle du rotor
$w_{r,m}$	Vitesse de la rotation mesurée du rotor
$w_g$	Vitesse rotationnelle de la génératrice
$w_{g,m}$	Vitesse de la rotation mesurée de la génératrice
$B_{dt}$	Coefficient de torsion amortie du train d'entraînement
$B_g$	Frottement visqueux de l'arbre à grande vitesse
$J_g$	Le moment d'inertie de l'arbre à grande vitesse
$J_r$	Le moment d'inertie de l'arbre à faible vitesse
$K_{dt}$	La rigidité de torsion du train d'entraînement
$N_g$	Le rapport de transmission ou d'engrenage
$P_g$	Puissance produite par la génératrice
$P_r$	Puissance de référence de l'éolienne
$R$	Rayon des pales de l'éolienne
$T_D$	Temps de détection
$T_e$	Temps d'échantillonnage
$w$	Vitesse angulaire
$w_n$	La pulsation naturelle

# Notations relatives au réacteur chimique

$(-\Delta H)$	L'enthalpie de réaction
A	La surface de transfert de chaleur entre le réacteur et la double enveloppe
$C_p$	Capacité thermique du milieu réactionnel
$C_{pfeed}$	Capacité thermique de l'alimentation
$C_{pj}$	Capacité thermique de l'huile caloporteur
$F_{feed}$	Le débit d'alimentation
$F_j$	Le débit massique dans la double enveloppe
HC	Le système de chauffe/refroidissement ( <i>Heating/Cooling</i> )
M	La masse totale du milieu réactionnel
$Q_{lj}$	La perte de chaleur de la double enveloppe
$Q_{lr}$	La perte de chaleur du réacteur principalement par le condenseur
$R_p$	Vitesse de réaction
$T_{feed}$	Sonde de mesure de la température d'alimentation
$T_{HC}$	Sonde de température dans le bain thermostaté
$T_{jin}$	Sonde de mesure de la température d'entrée de la double enveloppe
$T_{jout}$	Sonde de mesure de la température de sortie de la double enveloppe
$T_r$	Température moyenne dans le milieu réactionnel
$T_{r1}$	Sonde 1 de température dans le réacteur
U	Le coefficient de transfert de chaleur

# Acronymes

ACP	L'analyse en composantes principales « <i>Principal Component Analysis</i> »
CSTR	Réacteurs chimiques parfaitement agités « <i>Continuous Stirred tank Reactor</i> »
dVC	<b>d</b> imension de <b>V</b> apnik et <b>C</b> hervonenkis
FDI	Diagnostic de défauts ( <i>Fault Detection and Isolation</i> )
KKT	<b>K</b> arush <b>K</b> uhn <b>T</b> ucker
LPV	Linéaire à Paramètre Variant
LQG	Linéaire Quadratique Gaussienne
LTI	Linéaire à Temps Invariant
MRE	<b>M</b> inimisation du <b>R</b> isque <b>E</b> mpirique
MRS	<b>M</b> inimisation du <b>R</b> isque <b>S</b> tructurel
PLS	Moindres carrés partiels ou Projection aux Structures Latentes « <i>Partial Least Squares</i> »
RBF	<b>R</b> adial <b>B</b> asis <b>F</b> unction (Fonction de base radiale)
SMO	<b>S</b> equential <b>M</b> inimal <b>O</b> ptimisation
SVM	Machines à vecteurs de supports pour la classification
SVMs	Machines à vecteurs de supports « <i>Support Vector Machines</i> »
SVR	Machines à vecteurs de supports pour la régression « <i>Support Vector Regression</i> »
SVs	<b>S</b> upports <b>V</b> ector (Vecteurs de Support)
VCP	Variance Cumulative en Pourcentage « <i>Cumulative Percent Variance</i> »
WWTP	Procédé de traitement des eaux usées « <i>Waste Water Treatment Process</i> »



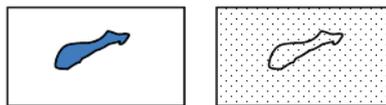
# Annexes



# Annexe A

## A.1 Méthode Monte Carlo

La méthode Monte Carlo (Kroese et al., 2011, Krauth, 2006) utilise un grand nombre de tirages aléatoires pour calculer une valeur numérique. Elle est aujourd'hui rendue possible par le fait que les ordinateurs modernes offrent un calcul rapide et précis d'un grand nombre de données avec un faible coût. On note que les résultats dépendent fortement du type de générateur des nombres aléatoires, qui ne doit pas être biaisé. Cette méthode est utilisée dans plusieurs domaines comme la simulation des phénomènes physiques, l'analyse du risque en finances, l'analyse de la fiabilité des composantes et la modélisation des molécules,... etc.



### Exemple Monte Carlo

A titre d'exemple, afin d'appliquer cette méthode pour déterminer la superficie d'un lac en bleu en connaissant la superficie du terrain (tout le rectangle), on procède de la façon suivante : on demande à un tireur de jeter  $N$  coups de canon de manière aléatoire ; on compte ensuite le nombre  $n$  de boules à l'intérieur du lac ; ensuite on réalise un rapport entre les valeurs :

$$\text{superficie lac} = (N - n) \times \frac{\text{superficie terrain}}{N}$$

Le processus d'application de Monte Carlo peut être résumé par les points suivants :

1. Générer des nombres aléatoires soit directement dans la loi de distribution ou à partir d'une loi uniforme (permet d'utiliser un générateur pseudo aléatoire).
2. Appliquer le modèle.
3. Calculer la moyenne empirique.
4. Ce processus (de 1 à 3) doit être répété où on applique une autre méthode, par exemple « *Bootstrapping* » pour calculer les intervalles de confiance.

Malgré que ces méthodes soient simples et faciles à implémenter, elles présentent des inconvénients tels que la convergence lente et la nécessité d'un grand nombre de répétitions. Il est possible de résoudre ce type de problème avec l'utilisation des méthodes de réduction de variance telles que la méthode des variables antithétiques, variable de contrôle, quasi Monte Carlo (séquences Halton, Sobol). On peut citer aussi la réduction de la variance par conditionnement, échantillonnage préférentiel « *Importance sampling* ».

La méthode de Monte Carlo a déjà été utilisée dans le diagnostic de défaut (Marseguerra and Zio, 2009) . Dans cette thèse, elle n'a pas été utilisée pour le diagnostic de défaut mais pour évaluer l'efficacité du schéma de diagnostic (évaluer le taux de fausses alarmes, ou de non détection) ou la qualité d'un modèle de prédiction face à des incertitudes. Ceci est fait en réalisant un grand nombre de simulations avec des conditions aléatoires (1000 fois par exemple) en changeant par exemple le bruit et les perturbations d'entrée.

## A.2 L'entropie de Shanon

Une variante de l'**information mutuelle** appelée **la variation de l'information** peut être utilisée comme une distance métrique :

$$\begin{aligned} d(X,Y) &= H(X,Y) - I(X,Y) \\ &= H(X) + H(Y) - 2I(X,Y) = H(X/Y) - H(Y/X) \end{aligned}$$

Où **H** est l'entropie de Shanon définie comme la prédiction de la valeur de l'information attendue. Soit **x** une variable aléatoire, l'entropie de **X** appelée **H(X)** est donnée comme suit :

$$H(X) = E(I(X)) = E(-\ln(p(x)))$$

**H(X)** permet de quantifier l'incertitude (l'imprévisibilité) associée à une variable aléatoire **X**. L'information est grande quand elle est inattendue. Donc, l'entropie indique la quantité de l'information qu'il faut obtenir pour lever le doute sur un événement probable comme un défaut par exemple :

$$H(X) = -\sum_{i=1}^N p(x_i) \log(p(x_i))$$

### A.3 L'extraction des caractéristiques dans le domaine fréquentiel

Beaucoup d'informations peuvent être obtenues par l'analyse des données dans le domaine fréquentiel qui fournit une idée sur les perturbations par exemple. La sélection des caractéristiques peut se faire dans le domaine fréquentiel avec **la transformée S** « *S-transform* », qui est une technique récente de traitement de signal produisant une représentation temps/fréquence. Elle a été utilisée pour extraire les caractéristiques à partir des données de perturbation de formes d'ondes « *data disturbance waveforms* » afin de déduire par SVR les classes simples et multiples de perturbation de qualité de puissance dans un système triphasé de puissance industrielle.

Analyse spectrale :

$$y(f) = \sum_{k=0}^{m-1} y(k) e^{-i2\pi fk}$$

La densité spectrale estimée d'un signal (periodogramme) :

$$R_{per} = \frac{1}{m} |y(f)|^2$$

Peut être calculée par l'autocorrélation :

$$R = abs(fft(xcorr(y(t))))$$

L'apparition de haute fréquence dont un signal normalement de basse fréquence peut indiquer que le signal est soumis à une perturbation ou à un défaut sur une composante.

$$E_f(k) = \sum_{j=k}^{k-l} R_{per}(k)$$

## A.4 Validation croisée

Une façon simple et efficace pour déterminer les paramètres optimaux (garantissant une corrélation maximale entre les variables et minimisent le critère EQM) est d'utiliser la validation croisée. C'est une méthode d'estimation de la fiabilité et la prédiction d'un modèle avec l'utilisation d'échantillonnage (Witten et al., 2011, Lendasse et al., 2003). Elle évalue le degré de généralisation du modèle face à des données indépendantes utilisées pour la prédiction. Elle permet d'éviter donc le problème de surapprentissage. L'utilisation de la validation croisée fournit un estimateur un peu biaisé du modèle car l'ensemble d'entraînement est inférieur aux cas observés.

La validation croisée consiste à répéter la division de l'ensemble de données principales en deux : un ensemble de données d'apprentissage et un de test ou de validation : l'échantillon de dimension  $n$  est divisé en  $m$  parties égales disjointes. Le modèle est construit avec les données d'apprentissage (en utilisant  $m-1$  parties) et validé avec les données de test (la partie restante). Cette opération est répétée  $m$  fois. L'erreur est estimée par le calcul de l'erreur quadratique moyenne.

Les variantes les plus connues de la méthode de validation croisée sont :

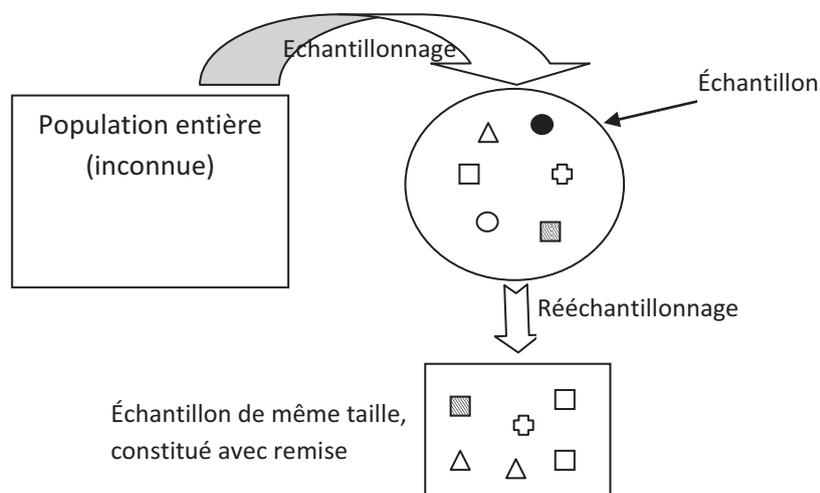
1. **Validation croisée à k plis ou replis « k-fold cross validation »** : l'ensemble de données sera partitionné aléatoirement en  $k$  sous-ensembles. Un seul sous-échantillon est utilisé pour la validation (test) alors que les  $k-1$  sous-ensembles restants sont utilisés pour l'apprentissage. Ce processus est répété  $k$  fois où à chaque fois un seul sous-ensemble est utilisé pour la validation. Les  $k$  résultats des  $k$  groupes peuvent être moyennés ou combinés pour produire une seule estimation ou choisir le modèle avec l'erreur la moins faible.
2. **Validation croisée avec un laissé en sortie « leave-one-out cross validation »** : utilise toutes les données sauf une. L'exemple non utilisé dans l'apprentissage sera utilisé dans le test. Cette procédure est répétée pour toutes les observations. Cette méthode est très exigeante en termes de temps de calcul.
3. **Validation croisée de Monte Carlo « The Monte Carlo cross validation »** ou « *repeated random sub-sampling validation* » : cette méthode divise l'ensemble des données de façon aléatoire en deux groupes : apprentissage et validation. Le modèle est adapté aux données d'apprentissage et la précision est évaluée en utilisant les données de validation.

Les résultats sont ensuite moyennés sur toutes les scissions. L'avantage de cette méthode par rapport à celle de k replis est que la proportion de la répartition (scission ou division) apprentissage/validation n'est pas dépendante du nombre d'itérations (plis). Son inconvénient réside dans le fait que certaines observations peuvent ne jamais être sélectionnées dans l'ensemble de validation.

## A.5 La méthode de Bootstrapping « Bootstrap »

Le mot Bootstrap provient de l'expression anglaise « *to pull one self up ones* » qui signifie littéralement « se soulever en tirant sur les languettes de ses bottes ».

Proposée à la fin des années 70 par Bradley Efron (Efron, 1988), Bootstrap est une méthode d'estimation basée sur le rééchantillonnage qui consiste à faire un tirage aléatoire d'un nombre assez grand d'échantillons avec remplacement à partir de l'échantillon initial. L'échantillon peut jouer le rôle d'une population mère. On tire un nouvel échantillon, c'est le rééchantillonnage, avec remise de manière à conserver le même nombre de données. Le Bootstrap ressemble à la validation croisée dans sa mise en œuvre mais diffère par la façon de détermination des ensembles d'apprentissages.



### Le rééchantillonnage dans le *Bootstrap*

L'inférence statistique est basée sur les résultats des échantillons ainsi obtenus. Le but de cette technique est de fournir des indications sur une statistique autre que sa valeur (dispersion, distribution, intervalle de confiance) afin de connaître la stabilité et la précision des estimations réalisées (ou modèle construit) sans recours à des nouvelles observations (données). Cette méthode permet entre autres d'estimer la confiance accordée à un estimateur (modèle, fonction

ou variable aléatoire) par le calcul des intervalles de confiance. *Bootstrap* peut être utilisé aussi comme une mesure générale d'erreur (biais d'un estimateur, erreur de prédiction, tests...).

Dans le cadre de l'apprentissage automatique artificiel « *Machine Learning* », cette technique permet de réduire l'écart entre le risque empirique et le risque réel fonctionnel et donc d'améliorer la robustesse du nouveau modèle prédictif (cela veut dire avoir des performances d'un niveau sensiblement constant face à des données inconnues).

## A.6 Autres variantes SVMs

- $\nu$  - SVR «  $\nu$  *Support Vector Classification* »

Cette variante utilise un autre paramètre  $\nu \in [0,1]$  pour remplacer le paramètre de régularisation  $C$  dans les SVMs classiques (Chen et al., 2005, Schölkopf et al., 2000). Ce paramètre est plus facile à interpréter et à régler que  $C$ . Il permet de contrôler le nombre maximal des points Supports et les erreurs d'apprentissage. Le problème primaire dans ce cas est formulé comme suit :

$$\left\{ \begin{array}{l} \min_{w,b,\zeta} \left( \frac{1}{2} \|w^2\| - \nu\rho + \frac{1}{N} \sum_{i=1}^N \zeta_i \right), \quad \rho \geq 0 \\ \text{sous les contraintes} \\ z_i [w^T \cdot \varphi(x) + b] \geq \rho - \zeta_i \quad \text{et} \quad \zeta_i \geq 0 \quad \text{pour} \quad i=1, \dots, N \end{array} \right.$$

Le problème dual donc devient :

$$\left\{ \begin{array}{l} \max_{\alpha \in \mathfrak{R}^N} \left( \frac{1}{2} \alpha^T \cdot G \cdot \alpha \right) \\ \text{sous les contraintes} \\ \alpha^T z = 0, 1^T \alpha \geq \nu, \text{ et } 0 \leq \alpha_i \leq \frac{1}{N} \quad \text{pour} \quad i=1, \dots, N \end{array} \right.$$

La marge de  $\nu$ -SVM est liée au paramètre  $C$  des  $C$ -SVM par :  $C = \frac{1}{\rho}$

- SVM à une classe: « *one class SVM* »

Dans le cas mono classe, le problème posé change puisqu'il n'y a pas d'information de sortie  $z_i$  proprement dite car tous les  $x_i$  appartiennent à une même classe (la classe normal +1).

$$D = \{x_i\}, i = 1, \dots, N$$

Il s'agit d'une extension de l'algorithme SVM classique binaire dans le cas des données non étiquetées (Mahadevan and Shah, 2009, Schölkopf et al., 2001). Cet algorithme cherche à trouver une fonction qui prend la valeur +1 dans une petite région capturant la plupart des données d'apprentissage et -1 dans le cas contraire .c.à.d. pour les valeurs qui se trouvent à l'extérieur du cluster de données d'apprentissage. L'objectif de cette variante est d'accueillir la plupart de données d'apprentissage dans un modèle en maximisant la séparation entre l'origine considérée comme la classe négative et le groupe d'échantillons normaux dans l'espace des caractéristiques. L'utilité de cette technique est la détection de nouveautés ou la détection des valeurs aberrantes.

Le problème primaire dans ce cas est donné comme suit :

$$\left\{ \begin{array}{l} \min_{w,b,\zeta} \left( \frac{1}{2} \|w^2\| + \frac{1}{\nu N} \sum_{i=1}^N \zeta_i + b \right), \quad \rho \geq 0 \\ \text{sous les contraintes} \\ w^T \cdot \phi(x) + b + \zeta \geq 0 \text{ et } \zeta_i \geq 0 \text{ pour } i = 1, \dots, N \end{array} \right.$$

Le problème dual devient donc :

$$\left\{ \begin{array}{l} \max_{\alpha \in \mathbb{R}^N} \left( \frac{1}{2} \alpha^T \cdot G \cdot \alpha \right) \\ \text{sous les contraintes} \\ 1^T \alpha = 1, \text{ et } 0 \leq \alpha_i \leq \frac{1}{\nu N}, \text{ pour } i = 1, \dots, N \end{array} \right.$$

Dans le cas de détection de défaut, le cas mono-classe est motivé par le fait que les données dans l'industrie ne sont pas riches en situations défectueuses.

- $\nu$ -SVR «  $\nu$  Support Vector Regression » :

Dans ce cas, le paramètre  $\nu \in [0,1]$  remplace la zone d'insensibilité  $\epsilon$  de  $\nu$ -SVR et contrôle le nombre des Vecteurs de supports (Chang and Lin, 2001, Schölkopf et al., 2000). Le problème à résoudre est :

$$\begin{aligned} & \text{Minimiser } \frac{1}{2} \|w\|^2 + C \left( \nu \epsilon + \frac{1}{N} \sum_{i=1}^l (\zeta_i^+, \zeta_i^-) \right) \\ & \text{sujet à } \begin{cases} z_i - \langle w, \phi(x_i) \rangle - b \leq \epsilon + \zeta_i^+ \\ \langle w, \phi(x_i) \rangle + b - y_i \leq \epsilon + \zeta_i^- \\ \zeta_i^+, \zeta_i^- \geq 0 \end{cases} \end{aligned}$$

La fonction de décision est alors de la même forme que le SVR classique :

$$f(x) = \sum_{i=1}^{N_{\text{sup}}} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b$$

(Chalimourda et al., 2004) traitent le problème du choix du paramètre  $\nu$  optimal.

# Annexe B

## Termes liés au traitement de l'eau

**L'eutrophisation** d'un écosystème aquatique se produit lorsque le milieu reçoit trop de matières nutritives assimilables par les algues (phosphore et azote). La multiplication de ces algues va limiter considérablement la respiration de l'eau et les échanges de ce dernier avec la lumière qui déséquilibre totalement le milieu [2008\_Belhani].

**Matière organique (MO)** : Désigne la matière d'origine humaine, animale ou végétale, capable d'être dégradée et décomposée par l'action d'organismes vivants. Elle constitue une faible proportion dans l'univers. Il s'agit par exemple des glucides, protides, lipides, en plus du carbone qui est le composant principal. Ces molécules contiennent des atomes d'hydrogène, d'oxygène et dans une moindre mesure d'azote, de phosphore, de soufre et de fer.

**Pollution** : Désigne l'introduction, généralement par suite de l'activité humaine, de substances ou de chaleur contaminantes dans un milieu naturel provoquant son instabilité, désordre et sa dégradation. Elle a un impact négatif sur la santé humaine ou sur la qualité des écosystèmes.

**Ammonification** : C'est l'activité biologique qui permet la génération d'ammonium ( $N_{H_4}^+$ ) ou d'ammoniac ( $N_{NH_3}$ ) à partir soit de matières organiques en décomposition ou par la réduction de nitrate.

**Aérobic** : Désigne la capacité des micro-organismes à vivre et se développer dans un environnement où l'oxygène libre est présent et joue son rôle.

**Ammoniac** : Composé gazeux d'azote et d'hydrogène, très soluble dans l'eau.

**Anaérobic** : Qualifie les milieux dans lesquels l'oxygène libre est absent sous toutes ses formes, et donc les nitrates nitrites sont absents aussi. Un dispositif anaérobic n'a pas besoin d'air ou de dioxygène pour fonctionner.

**Anoxie** : Qualifie les milieux dans lesquels l'oxygène n'est pas présent sous forme dissoute.

**Biomasse** : Masse totale de micro-organismes vivants, animaux et végétaux présente dans un milieu considéré pouvant devenir source d'énergie par combustion.

**Biomasse autotrophe** : Organisme qui utilise le dioxyde de carbone pour la synthèse cellulaire.

**Biomasse hétérotrophe** : Organisme qui utilise le carbone organique pour la synthèse cellulaire.

**Boues activées** : Flocons bactériens produits au cours de l'épuration biologique.

**Lyse** : Désigne la dissolution ou la destruction de l'intégrité physique de la cellule.

**Hydrolyse** : C'est une opération chimique utilisant l'eau pour décomposer de grosses molécules en plus petites molécules.

**La nitrification** consiste en l'oxydation de l'azote ammoniacal en nitrites ( $\text{NO}_2^-$ ) puis en nitrates ( $\text{NO}_3^-$ ) par l'intermédiaire de deux espèces de bactéries autotrophes. Ces bactéries sont dites « autotrophes » car elles utilisent du carbone minéral pour constituer leurs cellules. Leur taux de croissance est plus faible que celui des bactéries dégradant la pollution carbonée (bactéries hétérotrophes).

**Substrat** : Les matières formant l'alimentation requises au développement des micro-organismes.

**L'alcalinité** : L'alcalinité mesure la concentration équivalente de carbonate de calcium ( $\text{CaCO}_3$ ), et elle exprime la capacité de l'eau à résister à des variations du pH.

**Liqueur mixte** : Le mélange eaux usées-boues (flocs de micro-organismes).

**Biodégradable** : Se dit d'une substance capable d'être décomposée chimiquement par des micro-organismes ou d'autres moyens biologiques en éléments divers dépourvus d'effet dommageable sur le milieu naturel.

**Floculation** : Désigne le processus physico-chimique par lequel les particules fines et matières en suspension ayant le problème de ne pas sédimenter d'une manière naturelle, sont dues à se regrouper dans un floc par l'ajout d'un flocculant (chlorure ferrique et d'un polymère, etc.). Le floc peut alors soit remonter à la surface du liquide ou se déposer au fond.

**Les matières en suspension (MES)** (mg/L): c'est la fraction non dissoute de la pollution qui influe négativement sur le milieu aquatique, les animaux et les plantes (Gaïd, 2008).

**La demande biologique en oxygène (DBO)** est un paramètre représentatif de la quantité d'oxygène requise pour décomposer la somme des matières organiques biodégradables.

**L'azote** peut se trouver sous forme minérale (ammoniacal, nitrate) ou organique.

**Le phosphore** se trouve généralement sous forme de phosphate et il fait normalement partie de l'environnement. On peut le trouver dans les roches lorsqu'il y'a une décomposition des plantes ou d'animaux, mais aussi dans les engrais et les détergents. Il est considéré comme un élément indispensable à la vie des algues. Le phosphore est considéré comme un des grands responsables de l'eutrophisation des cours d'eau.

**Demande chimique en oxygène (D.C.O.) :** Quantité d'oxygène nécessaire à une dégradation purement chimique des composés organiques et minéraux contenus dans une eau polluée sans intervention de micro-organismes. Elle permet d'exprimer la teneur de matières organiques oxydables par voie chimique (Gaïd, 2008).



# Annexe C

## C.1 Le diplôme de première place dans la compétition sur la détection des défauts pour les éoliennes



## C.2 Termes liés à l'éolienne

### Éolienne aérogénératrice « *wind turbines* »

Dispositif fonctionnant sous l'action du vent permettant de convertir l'énergie cinétique du vent en énergie électrique.

### Les pales de l'éolienne « *blades en anglais* »

Permettre de capter l'énergie cinétique du vent

**Pitch** est l'angle de calage des pales

**Yaw** l'angle qui fait référence à l'orientation de la nacelle par rapport au vent.

### **La turbulence du vent**

Les variations de la vitesse du vent ne sont pas incluses dans la vitesse moyenne du vent, et sont causées par des facteurs multiples.

$$I = \frac{\sigma_w}{\bar{v}_w} \times 100 \%$$

Où I l'intensité de la turbulence,  $\sigma_w$  et  $\bar{v}_w$  sont respectivement l'écart type et la moyenne de la vitesse du vent.

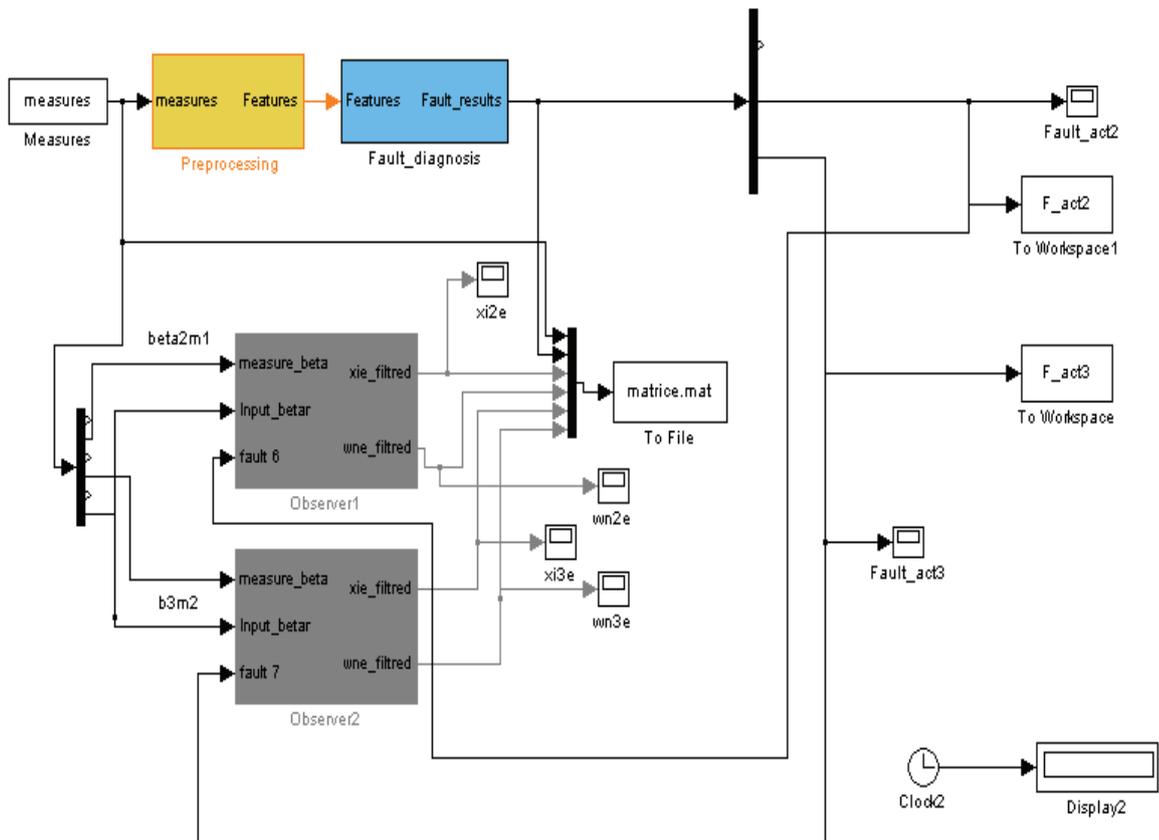
### **La portance de la pale**

C'est une force aérodynamique perpendiculaire au plan de la pale due à la différence de pression entre la partie bombée et la partie plate de la pale quand l'air s'écoule.

### **Remarque**

Il est à noter que la charge aérodynamique a un impact important sur la robustesse du système FDI à cause de ces caractéristiques stochastiques. Il est conseillé de découpler la charge du vent de la décision du système de surveillance d'état de l'éolienne.

### C.3 Implémentation de la solution de diagnostic de défaut basée sur SVM et un observateur de type Kalman dans l'éolienne





# Bibliographie

- Abbasion, S., Rafsanjani, A., Farshidianfar, A., Irani, N., 2007. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing* 21, 2933–2945.
- Adachi, S., Ogawa, T., Konno, R., 2001. A System Identification Method for Linear Regression Models Based on Support Vector Machine. *Transactions of the Society of Instrument and Control Engineers* 37, 1189–1191.
- Aggarwal, S., Aggarwal, N., 2011. Classification of Audio Data using Support Vector Machine. *International Journal of Computer Science and Technology* 2, 398–405.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Alcorta García, E., Frank, P.M., 1997. Deterministic nonlinear observer-based approaches to fault diagnosis: A survey. *Control Engineering Practice* 5, 663–670.
- Alex, J., Benedetti, L., Copp, J., Gernaey, K.V., Jeppsson, U., Nopens, I., Pons, M.N., Steyer, J.P., Vanrolleghem, P., 2008. Benchmark Simulation Model no. 1 (BSM1).
- Amirat, Y., Benbouzid, M.E.H., Al-Ahmar, E., Bensaker, B., Turri, S., 2009. A brief status on condition monitoring and fault diagnosis in wind energy conversion systems. *Renewable and Sustainable Energy Reviews* 13, 2629–2636.
- Amirat, Y., Choqueuse, V., Benbouzid, M.E., 2010. Failure Detection and Diagnosis in DFIG-based Wind Turbines using generator current amplitude demodulation. Presented at the International Conference on Industrial Engineering and Manufacturing ICIEM'10, Batna, Algeria, pp. 444–449.
- An, X.L., Jiang, D.X., Li, S.H., Chen, J., 2011. Fault diagnosis of direct-drive wind turbine based on support vector machine. *Journal of Physics: Conference Series* 305, 012030.
- Ao, S.-I., Rieger, B.B., Amouzegar, M.A., 2010. *Machine learning and systems engineering*. Springer, Dordrecht.
- Arnanz, R., de Santiago, M.A.G., Reones, A., Rodriguez, J., Saludes, S., 2011. Monitoring and Fault Diagnosis in Manufacturing Processes in the Automotive Industry, in: Chiaberge, M. (Ed.), *New Trends and Developments in Automotive Industry*. InTech.

- Artan, N., Orhon, D., Choi, E., 2004. Appropriate Design of Activated Sludge Systems for Nitrogen Removal from High Strength Wastewaters. *Journal of Environmental Science and Health, Part A* 39, 1913–1924.
- Baccarini, L.M.R., Rocha e Silva, V.V., de Menezes, B.R., Caminhas, W.M., 2011. SVM practical industrial application for mechanical faults diagnostic. *Expert Systems with Applications* 38, 6980–6984.
- Baggiani, F., Marsili-Libelli, S., 2009. Real-time fault detection and isolation in biological wastewater treatment plants. *Water Science & Technology* 60, 2949–2961.
- Barker, M., Rayens, W., 2003. Partial least squares for discrimination. *Journal of Chemometrics* 17, 166–173.
- Barszcz, T., Randall, R.B., 2009. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. *Mechanical Systems and Signal Processing* 23, 1352–1365.
- Basu, A., Walters, C., Shepherd, M., 2003. Support vector machines for text categorization. *Proceedings of the 36th Hawaii International Conference on System Sciences*.
- Baylar, A., Hanbay, D., Batan, M., 2009. Application of least square support vector machines in the prediction of aeration performance of plunging overfall jets from weirs. *Expert Systems with Applications* 36, 8368–8374.
- Belia, E., Amerlinck, Y., Benedetti, L., Johnson, B., Sin, G., Vanrolleghem, P.A., Gernaey, K.V., Gillot, S., Neumann, M.B., Rieger, L., Shaw, A., Villez, K., 2009. Wastewater treatment modelling: dealing with uncertainties. *Water Science & Technology* 60, 1929–1941.
- Beltran, B., Ahmed-Ali, T., Benbouzid, M., 2009. High-Order Sliding-Mode Control of Variable-Speed Wind Turbines. *IEEE Transactions on Industrial Electronics* 56, 3314–3321.
- Benhalla, A., Houssou, M., Charif, M., 2010. Linearization of the full activated sludge model No 1 for interaction analysis. *Bioprocess Biosyst Eng* 33, 759–771.
- Bermolen, P., Rossi, D., 2009. Support vector regression for link load prediction. *Computer Networks* 53, 191–201.
- Biegel, B., Juelsgaard, M., Kraning, M., Boyd, S., Stoustrup, J., 2011. Wind Turbine Pitch Optimization. Presented at the Proceedings 2011 IEEE Multi-conference on Systems and Control, pp. 1327–1334.
- Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M., 2003. *Diagnosis and Fault-Tolerant Control*, 1st ed. Springer.

- Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M., 2006. *Diagnosis and Fault-Tolerant Control*, 2nd ed. Springer.
- Bo, C., Qiao, X., Zhang, G., Bai, Y., Zhang, S., 2010. An integrated method of independent component analysis and support vector machines for industry distillation process monitoring. *Journal of Process Control* 20, 1133–1140.
- Bornard, G., Couenne, N., Celle, F., 1989. Regularly persistent observers for bilinear systems, in: Descusse, J., Fliess, M., Isidori, A., Leborgne, D. (Eds.), *New Trends in Nonlinear Control Theory*, Lecture Notes in Control and Information Sciences. Springer Berlin / Heidelberg, pp. 130–140.
- Bornard, G., Hammouri, H., 1991. A high gain observer for a class of uniformly observable systems, in: *Decision and Control, 1991., Proceedings of the 30th IEEE Conference On.* pp. 1494–1496 vol.2.
- Boser, B., Guyon, I., Vapnik, V., 1992. A training algorithm for optimal margin classifiers, in: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory.* ACM, Pittsburgh, Pennsylvania, United States, pp. 144–152.
- Boukhezzar, B., 2006. *Sur les stratégies de commande pour l'optimisation et la régulation de puissance des éoliennes à vitesse variable.* Thèse de Doctorat de l'Université Paris-Sud.
- Bozdogan, H., 2000. Akaike's information criterion and recent developments in information complexity. *J. Math. Psychol.* 44, 62–91.
- Brandao, R.F.M., Carvalho, J.A.B., Barbosa, F.P.M., 2010. Neural networks for condition monitoring of wind turbines, in: *Modern Electric Power Systems (MEPS), Proceedings of the International Symposium.* pp. 1-4.
- Brereton, R.G., Lloyd, G.R., 2010. Support Vector Machines for classification and regression. *Analyst* 135, 230–267.
- Caccavale, F., Digiulio, P., Iamarino, M., Masi, S., Pierri, F., 2010. A neural network approach for on-line fault detection of nitrogen sensors in alternated active sludge treatment plants. *Water Science & Technology* 62, 2760- 2768.
- Caccavale, F., Pierri, F., Iamarino, M., Tufano, V., 2011. *Control and Monitoring of Chemical Batch Reactors.* Springer.
- Camci, F., Chinnam, R.B., Ellis, R.D., 2008. Robust kernel distance multivariate control chart using support vector principles. *International Journal of Production Research* 46, 5075–5095.
- Canu, S., Grandvalet, Y., Guigue, V., Rakotomamonjy, A., 2005. *SVM and Kernel Methods Matlab toolbox, Perception Systèmes et Information.* INSA de Rouen, Rouen, France.

- Chachuat, B., Roche, N., Latifi, M.A., 2003. Réduction du modèle ASM 1 pour la commande optimale des petites stations d'épuration à boues activées. *Revue des sciences de l'eau* 16, 5–26.
- Chakraborty, U.K., 2008. *Advances in differential evolution*. Springer Verlag, Berlin.
- Chang, C., Lin, C.-J., 2001. LIBSVM: a Library for Support Vector Machines.
- Chang, C., Lin, C., 2011. LIBSVM: a library for support vector machines. *Science* 2, 1–39.
- Chen, S., Wang, W., van Zuylen, H., 2009. Construct support vector machine ensemble to detect traffic incident. *Expert Systems with Applications* 36, 10976–10986.
- Chen, W., Ding, S.X., Haghani, A., Naik, A., Khan, A.Q., Yin, S., 2011. Observer-Based FDI Schemes for Wind Turbine Benchmark, in: IFAC. Presented at the Proceedings of the 18th IFAC World Congress, pp. 7073–7078.
- Cheng, C.-S., Cheng, H.-P., 2008. Identifying the source of variance shifts in the multivariate process using neural networks and support vector machines. *Expert Systems with Applications* 35, 198–206.
- Cheng, C.-S., Chen, P.-W., Huang, K.-K., 2011. Estimating the shift size in the process mean with support vector regression and neural networks. *Expert Systems with Applications* 38, 10624–10630.
- Cherkassky, V., Ma, Y., 2002. Selection of Meta-parameters for Support Vector Regression, in: *Proceedings of the International Conference on Artificial Neural Networks, ICANN '02*. Springer-Verlag, London, UK, UK, pp. 687–693.
- Cherkassky, V., Ma, Y., 2004. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* 17, 113–126.
- Chetouani, Y., 2006. Fault Detection in a Chemical Reactor by Using the Standardized Innovation. *Process Safety and Environmental Protection* 84, 27–32.
- Chiang, L.H., Russell, E.L., Braatz, R.D., 2000. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 50, 243–252.
- Chiang, L.H., Kotanchek, M.E., Kordon, A.K., 2004. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Computers & Chemical Engineering* 28, 1389–1401.
- Chitralakha, S.B., Shah, S.L., 2010. Application of support vector regression for developing soft sensors for nonlinear processes. *The Canadian Journal of Chemical Engineering* 88, 696–709.

- Choi, S.W., Lee, I.-B., 2004. Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chemical Engineering Science* 59, 5897–5908.
- Choi, Lee, I., 2005. Multiblock PLS-based localized process diagnosis. *Journal of Process Control* 15, 295–306.
- Chong, H.G., Walley, W.J., 1996. Rule-based versus probabilistic approaches to the diagnosis of faults in wastewater treatment processes. *Artificial Intelligence in Engineering* 10, 265–273.
- Chow, E.Y., Willsky, A.S., Systems, M.I. of T.L. for I. and D., 1982. Analytic redundancy and the design of robust failure detection systems. <http://dspace.mit.edu/handle/1721.1/2828>
- Copp, J.B., 2002. The COST simulation benchmark: description and simulator manual. A Product of COST Action 624 and COST Action 682, Directorate-General for Research. Luxembourg: Office for Official Publications of the European Communities.
- Corominas, L., Villez, K., Aguado, D., Rieger, L., Rosén, C., Vanrolleghem, P.A., 2011. Performance evaluation of fault detection methods for wastewater treatment processes. *Biotechnology and Bioengineering* 108, 333–344.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Machine Learning* 20, 273–297.
- Crone, S.F., Kourentzes, N., 2010. Naive Support Vector Regression and Multilayer Perceptron benchmarks for the 2010 neural network grand competition (NNGC) on time series prediction. *IEEE*, pp. 1–8.
- Das, S., Karnik, N., Santoso, S., 2011. Time-Domain Modeling of Tower Shadow and Wind Shear in Wind Turbines. *ISRN Renewable Energy* 2011, 1–11.
- David, R., Saucez, P., Vassel, J.-L., Vande Wouwer, A., 2009. Modeling and numerical simulation of secondary settlers: A Method of Lines strategy. *Water Research* 43, 319–330.
- David, R., Vande Wouwer, A., Vassel, J.-L., Queinnec, I., 2009. Robust control of the activated sludge process. *Biotechnology Progress* 25, 701–708.
- Davy, M., Desobry, F., Gretton, A., Doncarli, C., 2006. An Online Support Vector Machine for Abnormal Events Detection. *Signal Processing* 86, 2009 – 2025.
- De Jong, S., 1993. SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* 18, 251–263.
- De Persis, C., Isidori, A., 2001. A geometric approach to nonlinear fault detection and isolation. *Automatic Control, IEEE Transactions on* 46, 853 –865.

- Deswal, S., 2011. Modeling Oxygen-transfer by Multiple Plunging Jets using Support Vector Machines and Gaussian Process Regression Techniques. *International Journal of Civil and Environmental Engineering* 28–33.
- Detroja, K.P., Gudi, R.D., Patwardhan, S.C., 2007. Plant-wide detection and diagnosis using correspondence analysis. *Control Engineering Practice* 15, 1468–1483.
- Diehl, S., Jeppsson, U., 1998. A model of the settler coupled to the biological reactor. *Water Research* 32, 331–342.
- Dinesh Kumar, J., Harikrishan, N., Karupiah, S., Manoj, N., Satish, S., Vishal, L., Sunil Nag, P.V., 2011. Wavelet based Fault Detection for Wind Turbine. Presented at the International Symposium on Devices MEMS, Intelligent Systems & Communication (ISDMISC) 2011, New York, Etats-Unis, pp. 7–10.
- Ding, S.X., 2008. *Model-based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*, 1st ed. Springer.
- Dispan, J., 2003. Biodégradabilité de la matière organique dans le continuum aquatique réseau d'assainissement -station d'épuration- milieu naturel récepteur. Développement d'une méthodologie pour le fractionnement de la matière organique en classes de biodégradabilité. Thèse de Doctorat de l'école nationale des ponts et chaussées.
- Dolan, B., 2010. *Wind Turbine Modelling, Control and Fault Detection*.
- Donders, S., 2002. *Fault Detection and Identification for Wind Turbine Systems: a closed-loop analysis*. Master's Thesis, university of Twente.
- Dong, J., Verhaegen, M., 2011. Data Driven Fault Detection and Isolation of a Wind Turbine Benchmark. Presented at the Proceedings of the 18th IFAC World Congress, pp. 7086–7091.
- Downs, J.J., Vogel, E.F., 1993. A plant-wide industrial process control problem. *Computers & Chemical Engineering* 17, 245–255.
- Dumais, S., Platt, J., Heckerman, D., Sahami, M., 1998. Inductive learning algorithms and representations for text categorization, in: *Proceedings of the Seventh International Conference on Information and Knowledge Management, CIKM '98*. ACM, New York, NY, USA, pp. 148–155.
- Dunia, R., Qin, S.J., Edgar, T.F., McAvoy, T.J., 1996. Identification of faulty sensors using principal component analysis. *AIChE Journal* 42, 2797–2812.
- Echavarria, E., Tomiyama, T., Bussel, G.J.W. van, 2007. Fault Diagnosis approach based on a model-based reasoner and a functional designer for a wind turbine. An approach towards self-maintenance. *Journal of Physics: Conference Series* 75, 012078.

- Edwards, C., Spurgeon, S.K., Patton, R.J., 2000. Sliding mode observers for fault detection and isolation. *Automatica* 36, 541–553.
- Efron, B., 1988. *Jackknife: the Bootstrap and Other Resampling Plans*. Soc. Ind. Appl. Math., Philadelphia.
- Ekman, M., 2003. *Urban Water Management - Modelling, Simulation and Control of the Activated Sludge Process* (dissertation). Uppsala University.
- Ekman, M., Björlenius, B., Andersson, M., 2006. Control of the aeration volume in an activated sludge process using supervisory control strategies. *Water Research* 40, 1668–1676.
- Elhor, N., Bertrand, R., Postaire, J., Hamad, D., 1999. Neural networks for wind turbine supervision. *e & i Elektrotechnik und Informationstechnik* 116, 366–369.
- Entezami, M., Hillmansén, S., Weston, P., Papaélías, M.P., 2012. Fault detection and diagnosis within a wind turbine mechanical braking system using condition monitoring. *Renewable Energy* 47, 175–182.
- Esbensen, T., Sloth, C., 2009. *Fault Diagnosis and Fault-Tolerant Control of Wind Turbines*. Master's Thesis, Aalborg University.
- Fang, F., Ni, B., Li, W., Sheng, G., Yu, H., 2011. A simulation-based integrated approach to optimize the biological nutrient removal process in a full-scale wastewater treatment plant. *Chemical Engineering Journal* 174, 635–643.
- Farza, M., Hammouri, H., Othman, S., Busawon, K., 1997. Nonlinear observers for parameter estimation in bioprocesses. *Chemical Engineering Science* 52, 4251–4267.
- Fei, S., Zhang, X., 2009. Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Systems with Applications* 36, 11352–11357.
- Ferrer, A., Aguado, D., Vidal-Puig, S., Prats, J.M., Zarzo, M., 2008. PLS: A versatile tool for industrial process improvement and optimization. *Applied Stochastic Models in Business and Industry* 24, 551–567.
- Flake, G.W., Lawrence, S., 2002. Efficient SVM Regression Training with SMO. *Machine Learning* 46, 271–290.
- Fragkoulis, D., Roux, G., Dahhou, B., 2011. Detection, isolation and identification of multiple actuator and sensor faults in nonlinear dynamic systems: Application to a waste water treatment process. *Applied Mathematical Modelling* 35, 522–543.
- Frank, P.M., 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. *Automatica* 26, 459–474.

- Frank, P.M., Köppen-Seliger, B., 1997. New developments using AI in fault diagnosis. *Engineering Applications of Artificial Intelligence* 10, 3–14.
- Frank, P.M., García, E.A., Köppen-Seliger, B., 2000. Modelling for fault detection and isolation versus modelling for control. *Math. Comput. Simul.* 53, 259–271.
- Fuente, M., Vega, P., Zarrop, M., Poch, M., 1996. Fault detection in a real wastewater plant using parameter-estimation techniques. *Control Engineering Practice* 4, 1089–1098.
- Ganyun, L.V., Haozhong, C., Haibao, Z., Lixin, D., 2005. Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electric Power Systems Research* 74, 1–7.
- Gauthier, J.P., Hammouri, H., Othman, S., 1992. A simple observer for nonlinear systems applications to bioreactors. *Automatic Control, IEEE Transactions on* 37, 875 –880.
- Ge, M., Du, R., Zhang, G., Xu, Y., 2004. Fault diagnosis using support vector machine with an application in sheet metal stamping operations. *Mechanical Systems and Signal Processing* 18, 143–159.
- Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta* 185, 1–17.
- Gernaey, K., Bogaert, H., Vanrolleghem, P., Van Vooren, L., Verstraete, W., 1998. Sensors for Nitrogen Removal Monitoring in Wastewater Treatment. In: *Environmental Biomonitoring: The Ecotoxicology Biotechnology Interface*. Eds. Lynch J.M. and Wiseman A., Cambridge University Press, UK. 180-207.
- Gernaey, K.V., van Loosdrecht, M.C., Henze, M., Lind, M., Jørgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling & Software* 19, 763–783.
- Gernaey, K.V., Rosen, C., Jeppsson, U., 2006. WWTP dynamic disturbance modelling--an essential module for long-term benchmarking development. *Water Sci. Technol.* 53, 225–234.
- Gertler, J., 1998. *Fault detection and diagnosis in engineering systems*. Marcel Dekker, New York.
- Gestel, T.V., Suykens, J.A.K., Baestaens, D., Lambrechts, A., Lanckriet, G., Vandaele, B., Moor, B.D., Vandewalle, J., 2001. Financial Time Series Prediction using Least Squares Support Vector Machines within the Evidence Framework. *IEEE Transactions on Neural Networks* 12, 809–821.
- Goethals, I., Pelckmans, K., Suykens, J.A.K., Moor, B.D., 2005. Subspace Identification of Hammerstein Systems using Least Squares Support Vector Machines. *IEEE Trans. on Automatic Control* 50, 1509–1519.

- Gretton, A., Doucet, A., Herbrich, R., Rayner, P., Schoelkopf, B., 2001. Support Vector Regression for Black-Box System Identification. In: Proceedings of the 11th IEEE Workshop on Statistical Signal Processing, pp. 341 - 344.
- Guermeur, Y., 2007. SVM Multiclasses, Théorie et Applications. Habilitation à diriger des recherches de l'Université Nancy I.
- Guinot, P., Sheibat-Othman, N., Févotte, G., Mckenna, T., 2000. On-line monitoring of emulsion copolymerisations using hardware sensors and calorimetry. *Polymer Reaction Engineering* 8, 115–134.
- Gujer, W., Henze, M., Mino, T., Van Loosdrecht, M., 1999. Activated Sludge Model No. 3. *Water and Science Technology* 39, 183–193.
- Gujer, W., 2010. Nitrification and me – A subjective review. *Water Research* 44, 1–19.
- Gunn, S., 1998. Support vector machines for classification and regression. Technical report, University of Southampton.
- Gunter, L., Zhu, J., 2007. Efficient Computation and Model Selection for the Support Vector Regression. *Neural Comput.* 19, 1633–1655.
- Gupta, S.M., 2007. Support Vector Machines based Modelling of Concrete Strength. Presented at the Proceedings of world academy of science, engineering and technology, pp.305-311
- Gustafsson, F., 2007. Statistical signal processing approaches to fault detection. *Annual Reviews in Control* 31, 41–54.
- Haimi, H., Mulas, M., Sahlstedt, K., Vahala, R., 2009. Advanced operation and control methods of municipal wastewater treatment processes in Finland. Helsinki University of Technology, Water and Wastewater Engineering.
- Hajiyev, C., 2012. Tracy–Widom distribution based fault detection approach: Application to aircraft sensor/actuator fault detection. *ISA Transactions* 51, 189–197.
- Hameed, Z., Hong, Y.S., Cho, Y.M., Ahn, S.H., Song, C.K., 2009. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable Energy Reviews* 13, 1–39.
- Hammouri, H., de Leon Morales, J., 1990. Observer synthesis for state-affine systems, in: *Decision and Control, 1990.*, Proceedings of the 29th IEEE Conference On. pp. 784 –785 vol.2.
- Hammouri, H., Kinnaert, M., El Yaagoubi, E.H., 1999. Observer-based approach to fault detection and isolation for nonlinear systems. *IEEE Transactions on Automatic Control* 44, 1879 –1884.

- Hammouri, H., Kabore, P., Othman, S., Biston, J., 2002. Failure diagnosis and nonlinear observer. Application to a hydraulic process. *Journal of the Franklin Institute* 339, 455–478.
- Han, H.-G., Qiao, J.-F., 2011. Adaptive dissolved oxygen control based on dynamic structure neural network. *Applied Soft Computing* 11, 3812–3820.
- Hauduc, H., 2010. Modèles biocinétiques de boues activées de type ASM: analyse théorique et fonctionnelle, vers un jeu de paramètres par défaut (Thèse de doctorat). L'Université Laval, Québec.
- He, W., Wang, Z., Jiang, H., 2008. Model optimizing and feature selecting for support vector regression in time series forecasting. *Neurocomputing* 72, 600–611.
- Henriksen, L.C., Poulsen, N.K., Hansen, M.H., 2011. Nonlinear Model Predictive Control of a Simplified Wind Turbine. Presented at the the 18th IFAC World Congress, Milan, Italy, pp. 551–556.
- Henze, M., Grady Jr, C.P.L., Gujer, W., Marais, G.V.R., Matsuo, T., 1987. A general model for single-sludge wastewater treatment systems. *Water Research* 21, 505–515.
- Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M.C., Marais, G.v.R, 1995. Activated Sludge Model No. 2. IAWQ Scientific and Technical Reports, No.3, London.
- Henze, M., Gujer, W., Mino, T., Van Loosdrecht M.C.M., 2000. Activated Sludge Models Asm1, Asm2, Asm2d and Asm3. IWA Publishing, London.
- Heymann, Y.T., 2010. A theory of the lifecycle of bacteria. *Nature and Science*, 121–131.
- Hsu, C.-C., Chen, M.-C., Chen, L.-S., 2010. Integrating independent component analysis and support vector machine for multivariate process monitoring. *Computers & Industrial Engineering* 59, 145–156.
- Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23, 725–749.
- Huang, Y., Reklaitis, G.V., Venkatasubramanian, V., 2003. A Heuristic Extended Kalman Filter Based Estimator for Fault Identification in a Fluid Catalytic Cracking Unit. *Industrial & Engineering Chemistry Research* 42, 3361–3371.
- Hugo, G., Sbarbaro, D.G., 2011. Outliers Detection in Industrial Databases: An Example Sulphur Recovery Process. *IFAC World Congress* 18, 1652–1657.
- Humberstone, M.J., 2010. An Adaptive Nonparametric Modeling Technique for Expanded Condition Monitoring of Processes (PhD diss). University of Tennessee - Knoxville.

- Humberstone, M., Wood, B., Henkel, J., Hines, J., 2012. Differentiating between expanded and fault conditions using principal component analysis. *Journal of Intelligent Manufacturing* 23, 179–188.
- Ingildsen, P., 2002. Realising full-scale control in wastewater treatment systems using in situ nutrient sensors. Doctoral Dissertation in Industrial Automation, Lund University, Sweden.
- Isermann, R., Ballé, P., 1997. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice* 5, 709–719.
- Isermann, R., 2005. Model-based fault-detection and diagnosis – status and applications. *Annual Reviews in Control* 29, 71–85.
- Isermann, R., 2006. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer.
- Isermann, R., 2011. Fault-Diagnosis Applications Model-Based Condition Monitoring: Actuators, Drives, Machinery, Plants, Sensors, and Fault-Tolerant Systems. Springer, Berlin; Heidelberg; New York.
- Ivanciuc, O., 2007. Applications of Support Vector Machines in Chemistry, in: *Reviews in Computational Chemistry*. John Wiley & Sons, Inc., pp. 291–400.
- Jemwa, G.T., Aldrich, C., 2006. Kernel-based fault diagnosis on mineral processing plants. *Minerals Engineering* 19, 1149–1162.
- Jeppsson, U., 1996. Modelling aspects of wastewater treatment processes. Doctoral thesis, IEA, Sweden.
- Jeppsson, U., Alex, J., Pons, M.N., Spanjers, H., Vanrolleghem, P.A., 2002. Status and future trends of ICA in wastewater treatment--a European perspective. *Water Sci. Technol.* 45, 485–494.
- Jeppsson, U., Pons, M.-N., Nopens, I., Alex, J., Copp, J.B., Gernaey, K.V., Rosen, C., Steyer, J.-P., Vanrolleghem, P.A., 2007. Benchmark simulation model no 2: general protocol and exploratory case studies. *Water Science & Technology* 56, 67-78.
- Jianxu, L., 2011. Dynamic Optimization of Aeration Operations for a Benchmark Wastewater Treatment Plant. Presented at the Proceedings of the 18th IFAC World Congress, 2011, Milano, Italy, pp. 14189–14194.
- Joachims, T., 1999. Making large-scale support vector machine learning practical, in: *Advances in Kernel Methods: Support Vector Learning*. MIT Press, pp. 169–184.

- Johnson, K.E., Pao, L.Y., Balas, M.J., Fingersh, L.J., 2006. Control of variable-speed wind turbines: standard and adaptive techniques for maximizing energy capture. *IEEE Control Systems Magazine* 26, 70–81.
- Jonasson, M., 2007. Energy Benchmark for Wastewater Treatment Processes - a Comparison between Sweden and Austria, TEIE-5247, IEA.
- Jonkman, J.M., 2009. Dynamics of offshore floating wind turbines-model development and verification. *Wind Energy* 12, 459–492.
- Jun, B.H., 2011. Fault detection using dynamic time warping (DTW) algorithm and discriminant analysis for swine wastewater treatment. *Journal of Hazardous Materials* 185, 262–268.
- Juricek, B.C., Seborg, D.E., Larimore, W.E., 2004. Fault Detection Using Canonical Variate Analysis. *Industrial & Engineering Chemistry Research* 43, 458–474.
- Kabore, P., Othman, S., Mckenna, T.F., Hammouri, H., 2000. Observer-based fault diagnosis for a class of non-linear systems Application to a free radical copolymerization reaction. *International Journal of Control* 73, 787–803.
- Katipamula, S., Brambley, M.R., 2005. Review Article: Methods for Fault Detection, Diagnostics, and Prognostics for Building Systems—A Review, Part I. *HVAC&R Research* 11, 3–25.
- Khushaba, R.N., Al-Ani, A., Al-Jumaily, A., 2011. Feature subset selection using differential evolution and a statistical repair mechanism. *Expert Systems with Applications* 38, 11515–11526.
- Kiasi, F., Prakash, J., Shah, S.L., Lee, J.M., 2011. Fault Detection and Isolation of Benchmark Wind Turbine Using the Likelihood Ratio Test. Presented at the Proceedings of the 18th IFAC World Congress, 2011, Milano, Italy, pp. 7079–7085.
- Kim, K., Lee, J.-M., Lee, I.-B., 2005. A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction. *Chemometrics and Intelligent Laboratory Systems* 79, 22–30.
- Kourti, T., 2005. Application of latent variable methods to process control and multivariate statistical process control in industry. *International Journal of Adaptive Control and Signal Processing* 19, 213–246.
- Kulkarni, A., Jayaraman, V.K., Kulkarni, B.D., 2004. Support vector classification with parameter tuning assisted by agent-based technique. *Computers & Chemical Engineering* 28, 311–318.

- Kulkarni, A., Jayaraman, V.K., Kulkarni, B.D., 2005. Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process. *Computers & Chemical Engineering* 29, 2128–2133.
- Kusiak, A., Li, W., 2011. The prediction and diagnosis of wind turbine faults. *Renewable Energy* 36, 16–23.
- Kwok, J.T., Tsang, I.W., 2003. Linear dependency between  $\Sigma$  and the input noise in  $\Sigma$ -support vector regression. *IEEE Transactions on Neural Networks* 14, 544–553.
- Lahiri, S.K., Ghanta, K.C., 2008. The Support Vector Regression with the parameter tuning assisted by a differential evolution technique: Study of the critical velocity of a slurry flow in a pipeline. *Chemical Industry and Chemical Engineering Quarterly* 14, 191–203.
- Laouti, N., Sheibat-othman, N., Othman, S., 2011a. Support Vector Machines for Fault Detection in Wind Turbines. Presented at the IFAC World Congress, Milano, Italy, pp. 7067–7072.
- Laouti, N., Sheibat-Othman, N., Othman, S., 2011b. Fault diagnosis for wastewater treatment plants using Support Vector Machines. Presented at the STIC & ENVIRONNEMENT, France.
- Lauer, F., Bloch, G., 2006. Méthodes SVM pour l'identification, in: *Journées Identification Et Modélisation Expérimentale, JIME'2006*, Poitiers, France, p. CDROM.
- Lee, C., Choi, S.W., Lee, I.-B., 2004. Sensor fault identification based on time-lagged PCA in dynamic processes. *Chemometrics and Intelligent Laboratory Systems* 70, 165–178.
- Lee, D.S., Vanrolleghem, P.A., Park, J.M., 2005. Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. *Journal of Biotechnology* 115, 317–328.
- Lee, D.S., Lee, M.W., Woo, S.H., Kim, Y.-J., Park, J.M., 2006a. Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant. *Process Biochemistry* 41, 2050–2057.
- Lee, H.W., Lee, M.W., Park, J.M., 2007. Robust Adaptive Partial Least Squares Modeling of a Full-Scale Industrial Wastewater Treatment Process. *Industrial & Engineering Chemistry Research* 46, 955–964.
- Lee, J., Qin, S.J., Lee, I., Lee, J., Qin, S.J., Lee, I., 2006. Fault detection and diagnosis based on modified independent component analysis, Fault detection and diagnosis based on modified independent component analysis. *AIChE Journal* 52, 3501–3514.
- Lee, J.-M., Yoo, C., Lee, I.-B., 2004. Statistical process monitoring with independent component analysis. *Journal of Process Control* 14, 467–485.

- Lee, M.W., Joung, J.Y., Lee, D.S., Park, J.M., Woo, S.H., 2005. Application of a Moving-Window-Adaptive Neural Network to the Modeling of a Full-Scale Anaerobic Filter Process. *Industrial & Engineering Chemistry Research* 44, 3973–3982.
- Lee, M.W., Hong, S.H., Choi, H., Kim, J.-H., Lee, D.S., Park, J.M., 2008. Real-time remote monitoring of small-scaled biological wastewater treatment plants by a multivariate statistical process control and neural network-based software sensors. *Process Biochemistry* 43, 1107–1113.
- Li, G., Qin, S.J., Zhou, D., 2010. Geometric properties of partial least squares for process monitoring. *Automatica* 46, 204–210.
- Li, H., Liang, Y., Xu, Q., 2009. Support vector machines and its applications in chemistry. *Chemometrics and Intelligent Laboratory Systems* 95, 188–198.
- Li, J., Cui, P., 2009. Improved kernel fisher discriminant analysis for fault diagnosis. *Expert Systems with Applications* 36, 1423–1432.
- Li, S., Wunsch, D.C., O’Hair, E.A., Giesselmann, M.G., 2001. Using neural networks to estimate wind turbine power generation. *Energy Conversion, IEEE Transactions on* 16, 276–282.
- Li, W., Shah, S.L., Xiao, D., 2008. Kalman filters in non-uniformly sampled multirate systems: For FDI and beyond. *Automatica* 44, 199–208.
- Li, Y., Wang, Z., Yuan, J., 2006. On-line Fault Detection Using SVM-based Dynamic MPLS for Batch Processes. *Chinese Journal of Chemical Engineering* 14, 754–758.
- Liang, J., Du, R., 2007. Model-based Fault Detection and Diagnosis of HVAC systems using Support Vector Machine method. *International Journal of Refrigeration* 30, 1104–1114.
- Liu, Y., Zheng, Y.F., 2005. One-against-all multi-class SVM classification using reliability measures. *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, 2, pp. 849 – 854.
- Lootsma, T.F., 1999. Observer-based Fault Detection and Isolation for Nonlinear Systems. *IEEE Transactions on Automatic Control* 44, 1879–1884.
- Lootsma, T.F., 2001. Observer-based fault detection and isolation for nonlinear systems: Ph.D. thesis. Department of Control Engineering, Aalborg University, Aalborg.
- Lourenço, N.D., Menezes, J.C., Pinheiro, H.M., Diniz, D., 2008. Fuel park wastewater monitoring with UV-VIS spectra and partial. *Macedonian Journal of Chemistry and Chemical Engineering* 27, 19–23.
- Lu, B., Li, Y., Wu, X., Yang, Z., 2009. A review of recent advances in wind turbine condition monitoring and fault diagnosis, in: *IEEE Power Electronics and Machines in Wind*

- Applications, 2009. PEMWA 2009. Presented at the IEEE Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009, IEEE, pp. 1–7.
- Lu, N., Wang, X., 2008. Fault Diagnosis Based on Signed Digraph Combined with Dynamic Kernel PLS and SVR. *Industrial & Engineering Chemistry Research* 47, 9447–9456.
- Lu, Y., 2012. Wind Turbine Gearbox Fault Detection Using Multiple Sensors With Features Level Data Fusion. *Transactions of the ASME - A - Engineering for Gas Turbines and Power* 134, 042501.
- Ma, J., Theiler, J., Perkins, S., 2003. Accurate On-line Support Vector Regression. *Neural Computation* 15, 2683–2703.
- Maalouf, M., Khoury, N., Trafalis, T.B., 2008. Support vector regression to predict asphalt mix performance. *International Journal for Numerical and Analytical Methods in Geomechanics* 32, 1989–1996.
- Mahadevan, S., Shah, S.L., 2009. Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control* 19, 1627–1639.
- Mao, Y., Xia, Z., Yin, Z., Sun, Y., Wan, Z., 2007. Fault Diagnosis Based on Fuzzy Support Vector Machine with Parameter Tuning and Feature Selection. *Chinese Journal of Chemical Engineering* 15, 233–239.
- Martínez-Ramón, M., Rojo-Alvarez, J.L., Camps-Valls, G., Muñoz-Marí, J., Navia-Vázquez, A., Soria-Olivas, E., Figueiras-Vidal, A.R., 2006. Support vector machines for nonlinear kernel ARMA system identification. *IEEE Trans Neural Netw* 17, 1617–1622.
- Matha, D., 2009. Model Development and Loads Analysis of an Offshore Wind Turbine on a Tension Leg Platform, with a Comparison to Other Floating Turbine Concepts. National Renewable Energy Laboratory, United States.
- Mattera, D., Haykin, S., 1999. Advances in kernel methods, in: Schölkopf, B., Burges, C.J.C., Smola, A.J. (Eds.), MIT Press, Cambridge, MA, USA, pp. 211–241.
- Melssen, W., Üstün, B., Buydens, L., 2007. SOMPLS: A supervised self-organising map--partial least squares algorithm for multivariate regression problems. *Chemometrics and Intelligent Laboratory Systems* 86, 102–120.
- Mjalli, F.S., Al-Asheh, S., Alfadala, H.E., 2007a. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. *Journal of Environmental Management* 83, 329–338.
- Mohamed, E.A., Rao, N.D., 1995. Artificial neural network based fault diagnostic system for electric power distribution feeders. *Electric Power Systems Research* 35, 1–10.

- Mohsen, A.K.A., El-Yazeed, M.F.A., 2004. Selection of Input Stimulus for Fault Diagnosis of Analog Circuits Using ARMA Model. *AEU - International Journal of Electronics and Communications* 58, 212–217.
- Monod, J., 1949. The Growth of Bacterial Cultures. *Annual Review of Microbiology* 3, 371–394.
- Monroy, I., Benitez, R., Escudero, G., Graells, M., 2010. A semi-supervised approach to fault diagnosis for chemical processes. *Computers & Chemical Engineering* 34, 631–642.
- Mujahed, A.-D., Westwick, D.T., 2011. Support Vector Machine Identification of Output Error Hammerstein Models, in: Sergio, B. (Ed.), Presented at the the 18th IFAC World Congress, Milano (Italy), pp. 13948–13953.
- Mukherjee, S., Vapnik, V., 1999. *Multivariate Density Estimation: an SVM Approach*. Massachusetts Institute of Technology, Cambridge, MA, USA.
- Mulas, M., 2006. *Modelling and Control of Activated Sludge Processes*. PhD dissertation, University of Gagliari, Italy.
- Musyafa, A., Negara, I.M.Y., Robandi, I., 2011. Design Optimal of Adaptive Control and Fuzzy Logic Control on Torque-Shaft Small Scale Wind Turbine. *Canadian Journal on Electrical and Electronics Engineering* 2, 202–208.
- Negre, P.L., Puig, V., Pineda, I., 2011. Fault Detection and Isolation of a Real Wind Turbine Using LPV Observers. Presented at the the 18th IFAC World Congress, Milano, Italy, pp. 12372–12379.
- Nelson, M.I., Sidhu, H.S., 2009. Analysis of the activated sludge model (number 1). *Applied Mathematics Letters* 22, 629–635.
- Nguyen, D.V., Roche, D.M., 2002. Tumor Classification by Partial Least Squares Using Microarray Gene Expression Data. *Bioinformatics* 18, 39–50.
- Niska, H., 2011. *Predictive data-driven modeling approaches in environmental management decision-making (Dissertation)*. University of Eastern Finland.
- Nyberg, M., 1999. *Model Based Fault Diagnosis: Methods, Theory, and Automotive Engine Applications*, PhDthesis. Linköping University, Sweden.
- Odgaard, P., Stoustrup, J., 2009. Unknown Input Observer Based Scheme for Detecting Faults in a Wind Turbine Converter, in: *Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, pp. 161-166.
- Odgaard, P., Stoustrup, J., Kinnaert, M., 2009. Fault Tolerant Control of Wind Turbines - a Benchmark Model, in: *Proceedings of Fault Detection, Supervision and Safety of Technical Processes*. Barcelona, pp. 155-160.

- Odgaard, P.F., Stoustrup, J., 2010. Unknown input observer based detection of sensor faults in a wind turbine. Presented at the 2010 IEEE International Conference on Control Applications (CCA), IEEE, pp. 310–315.
- Odgaard, P.F., Stoustrup, J., 2011. Orthogonal Bases used for Feed Forward Control of Wind Turbines. Presented at the In: Proceedings of the 18th IFAC World Congress, pp. 532-537.
- Odgaard, P.F., Stoustrup, J., 2012. Results of a wind turbine FDI competition. Presented at the IFAC Workshop Series.
- Olsson, G., Pappa, G.L., Mattoso, M., 2010. Adaptive Normalization: A novel data normalization approach for non-stationary time series, in: Neural Networks (IJCNN), The 2010 International Joint Conference On. pp. 1–8.
- Olsson, G., 2007. Automation Development in Water and Wastewater Systems. *Environ. Eng. Res.* 12, 197–200.
- Østergaard, K.Z., Brath, P., Stoustrup, J., 2007. Estimation of effective wind speed. *Journal of Physics: Conference Series* 75, 012082.
- Østergaard, K.Z., Stoustrup, J., Brath, P., 2009. Linear parameter varying control of wind turbines covering both partial load and full load conditions. *International Journal of Robust and Nonlinear Control* 19, 92–116.
- Ozdemir, A.A., Seiler, P., Balas, G.J., 2011. Wind Turbine Fault Detection Using Counter-Based Residual Thresholding. *Proceedings of the 18th IFAC World Congress, 2011* 18, 8289–8294.
- Pachet, F., Roy, P., 2009. Analytical Features: A Knowledge-Based Approach to Audio Feature Generation. *EURASIP Journal on Audio, Speech, and Music Processing* 2009, 1–23.
- Pao, L.Y., Johnson, K.E., 2009. A tutorial on the dynamics and control of wind turbines and wind farms, in: *American Control Conference, 2009. ACC '09*. pp. 2076–2089.
- Parrella, F., 2007. Online Support Vector Regression. PhD Thesis. University of Genoa, Italy.
- Patton, R.J., 1997. Fault tolerant control systems : The 1997 situation, in: *IFAC SAFEPROCESS'1997*. Presented at the IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes, pp. 1033–1054.
- Patton, R.J., Chen, J., 2000. On eigenstructure assignment for robust fault diagnosis. *International Journal of Robust and Nonlinear Control* 10, 1193–1208.
- Paviglianiti, G., Pierri, F., Caccavale, F., Mattei, M., 2010. Robust fault detection and isolation for proprioceptive sensors of robot manipulators. *Mechatronics* 20, 162–170.

- Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1226–1238.
- Pierri, F., Paviglianiti, G., Caccavale, F., Mattei, M., 2008. Observer-based sensor fault detection and isolation for chemical batch reactors. *Engineering Applications of Artificial Intelligence* 21, 1204–1216.
- Pisu, P., Ayalew, B., 2011. Robust Fault Diagnosis for a Horizontal Axis Wind Turbine. Presented at the 18th IFAC World Congress, pp. 7055–7060.
- Platt, J.C., 1999. Advances in kernel methods, in: Schölkopf, B., Burges, C.J.C., Smola, A.J. (Eds.), MIT Press, Cambridge, MA, USA, pp. 185–208.
- Platt, J.C., Cristianini, N., Shawe-taylor, J., 2000. Large Margin DAGs for Multiclass Classification, in: *Advances in Neural Information Processing Systems*. MIT Press, pp. 547–553.
- Pons, M.N., Belhani, M., Bourgois, J., Dupuit, E., 2008. Analyse du cycle de vie. Epuration des eaux usées urbaines. *Techniques de l'ingénieur. Environnement n° 820*, 22 pp.
- Pourmohammad, S., Fekih, A., 2011. Fault-Tolerant Control of Wind Turbine Systems - A Review, in: *Green Technologies Conference (IEEE-Green), 2011 IEEE*. pp. 1–6.
- Pöyhönen, S., Arkkio, A., Jover, P., Hyötyniemi, H., 2005. Coupling pairwise support vector machines for fault classification. *Control Engineering Practice* 13, 759–769.
- Prakash, J., Patwardhan, S.C., Narasimhan, S., 2002. A Supervisory Approach to Fault-Tolerant Control of Linear Multivariable Systems. *Industrial & Engineering Chemistry Research* 41, 2270–2281.
- Priyadarshin, D., Acharya, M., Prasad Mishra, A., 2011. Link Load Prediction using Support Vector Regression and Optimization. *International Journal of Computer Applications* 24, 22–25.
- Pusha, S., Gudi, R., Noronha, S., 2009. Polar classification with correspondence analysis for fault isolation. *Journal of Process Control* 19, 656–663.
- Queinnec, I., Gómez-Quintero, C.-S., 2009. Reduced modeling and state observation of an activated sludge process. *Biotechnology Progress* 25, 654–666.
- Ráduly, B., Gernaey, K.V., Capodaglio, A.G., Mikkelsen, P.S., Henze, M., 2007. Artificial neural networks for rapid WWTP performance evaluation: Methodology and case study. *Environmental Modelling & Software* 22, 1208–1216.

- Ram Maurya, M., Rengaswamy, R., Venkatasubramanian, V., 2004. Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets. *Engineering Applications of Artificial Intelligence* 17, 501–518.
- Ramaswami, M., Bhaskaran, R., 2009. A Study on Feature Selection Techniques in Educational Data Mining. *Journal of computing* 1.
- Ravikumar, B., Thukaram, D., Khincha, H.P., 2008. Application of support vector machines for fault diagnosis in power transmission system. *IET Generation, Transmission & Distribution* 2, 119-130.
- Reiter, R., 1987. A theory of diagnosis from first principles. *Artificial Intelligence* 32, 57–95.
- Ren, N., Liang, J., Gu, B., Han, H., 2008. Fault diagnosis strategy for incompletely described samples and its application to refrigeration system. *Mechanical Systems and Signal Processing* 22, 436–450.
- Rieger, L., Takács, I., Shaw, A., Winkler, S., Ohtsuki, T., Langergraber, G., Gillot, S., 2010. Editorial: Status and future of wastewater treatment modelling. *Water Science & Technology* 61, 821-823.
- Rieger, L., Thomann, M., Gujer, W., Siegrist, H., 2005. Quantifying the uncertainty of on-line sensors at WWTPs during field operation. *Water Research* 39, 5162–5174.
- Rivas, A., Irizar, I., Ayesa, E., 2008. Model-based optimisation of Wastewater Treatment Plants design. *Environmental Modelling & Software* 23, 435–450.
- Rocco S., C.M., Zio, E., 2007. A support vector machine integrated system for the classification of operation anomalies in nuclear components and systems. *Reliability Engineering & System Safety* 92, 593–600.
- Rosen, C., Jeppsson, U., Rieger, L., Vanrolleghem, P.A., 2008. Adding realism to simulated sensors and actuators. *Water Science & Technology* 57, 337-344.
- Rosen, C., Lennox, J.A., 2001. Multivariate and multiscale monitoring of wastewater treatment operation. *Water Research* 35, 3402–3410.
- Rosipal, R., Krämer, N., 2006. Overview and recent advances in partial least squares, in: In “Subspace, Latent Structure and Feature Selection Techniques”, *Lecture Notes in Computer Science*. Springer, pp. 34–51.
- Rosipal, R., Trejo, L.J., 2002. Kernel partial least squares regression in reproducing kernel hilbert space. *J. Mach. Learn. Res.* 2, 97–123.
- Rosipal, R., Trejo, L.J., 2003. Kernel pls-svc for linear and nonlinear classification, in: In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*. pp. 640–647.

- Ruiz, D., Nougues, J., Calderón, Z., España, A., Puigjaner, L., 2000. Neural network based framework for fault diagnosis in batch chemical plants. *Computers & Chemical Engineering* 24, 777–784.
- Russell, E.L., Chiang, L.H., Braatz, R.D., 2000. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 51, 81–93.
- Salahshoor, K., Kordestani, M., Khoshro, M.S., 2010. Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers. *Energy* 35, 5472–5482.
- Saludes Rodil, S., Arnanz Gómez, R., Bernárdez, J.M., Rodríguez, F., Miguel, L.J., Perán, J.R., 2009. Laser welding defects detection in automotive industry based on radiation and spectroscopical measurements. *The International Journal of Advanced Manufacturing Technology* 49, 133–145.
- Saludes Rodil, S., Fuente, M.J., 2010. Fault tolerance in the framework of support vector machines based model predictive control. *Engineering Applications of Artificial Intelligence* 23, 1127–1139.
- Samanta, B., 2004. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing* 18, 625–644.
- Samanta, B., Al-Balushi, K.R., Al-Araimi, S.A., 2003. Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Engineering Applications of Artificial Intelligence* 16, 657–665.
- Samanta, B., Nataraj, C., 2009. Use of particle swarm optimization for machinery fault detection. *Engineering Applications of Artificial Intelligence* 22, 308–316.
- Samantaray, S.R., Dash, P.K., Panda, G., 2007. Distance relaying for transmission line using support vector machine and radial basis function neural network. *International Journal of Electrical Power & Energy Systems* 29, 551–556.
- Saravanan, N., Kumar Siddabattuni, V.N.S., Ramachandran, K.I., 2008. A comparative study on classification of features by SVM and PSVM extracted using Morlet wavelet for fault diagnosis of spur bevel gear box. *Expert Systems with Applications* 35, 1351–1366.
- Saravanan, N., Ramachandran, K.I., 2009. A case study on classification of features by fast single-shot multiclass PSVM using Morlet wavelet for fault diagnosis of spur bevel gear box. *Expert Systems with Applications* 36, 10854–10862.

- Saravanan, N., Siddabattuni, V.N.S.K., Ramachandran, K.I., 2010. Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM). *Applied Soft Computing* 10, 344–360.
- Schittkowski, K., 2005. Optimal parameter selection in support vector machines. *Journal of Industrial and Management Optimization* 1, 465–476.
- Schlechtingen, M., Ferreira Santos, I., 2011. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. *Mechanical Systems and Signal Processing* 25, 1849–1875.
- Schölkopf, B., Burges, C.J.C., Smola, A.J., 1999. *Advances in kernel methods support vector learning*. MIT Press, Cambridge, Mass.
- Schütze, M., Campisano, A., Colas, H., Schilling, W., Vanrolleghem, P.A., 2004. Real time control of urban wastewater systems-where do we stand today? *Journal of Hydrology* 299, 335–348.
- Selvam, K., Kanev, S., van Wingerden, J.W., van Engelen, T., Verhaegen, M., 2009. Feedback-feedforward individual pitch control for wind turbine load reduction. *International Journal of Robust and Nonlinear Control* 19, 72–91.
- Shawe-Taylor, J., Cristianini, N., 2004. *Kernel methods for pattern analysis*. Cambridge University Press, Cambridge, UK; New York.
- Sheibat-Othman, N., Othman, S., 2006. Control of an Emulsion Polymerization Reactor. *Industrial & Engineering Chemistry Research* 45, 206–211.
- Sheibat-Othman, N., Othman, S., Boyron, O., Alamir, M., 2011. Multivariable control of the polymer molecular weight in emulsion polymerization processes. *Journal of Process Control* 21, 861–873.
- Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., Murthy, K.R.K., 2000. Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks* 11, 1188–1193.
- Shinzawa, H., Ritthiruangdej, P., Ozaki, Y., 2011. Kernel Analysis of Partial Least Squares (PLS) Regression Models. *Applied Spectroscopy* 65, 549–556.
- Shitong, W., Jiagang, Z., Chung, F.L., Qing, L., Dwen, H., 2004. Theoretically Optimal Parameter Choices for Support Vector Regression Machines with Noisy Input. *Soft Computing* 9, 732–741.
- Simani, S., Castaldi, P., Bonfé, M., 2011. Hybrid Model-Based Fault Detection of Wind Turbine Sensors. *Proceedings of the 18th IFAC World Congress, 2011* 18, 7061–7066.

- Simani, S., Patton, R., Fantuzzi, C., 2003. *Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- Singh, K.P., Basant, N., Malik, A., Sinha, S., Jain, G., 2009. Multi-way modeling of wastewater data for performance evaluation of sewage treatment plant—A case study. *Chemometrics and Intelligent Laboratory Systems* 95, 18–30.
- Sloth, C., Esbensen, T., Stoustrup, J., 2011. Robust and fault-tolerant linear parameter-varying control of wind turbines. *Mechatronics* 21, 645–659.
- Smola, A., Burges, C., Drucker, H., Golowich, S., Van Hemmen, L., Müller, K.-R., Schölkopf, B., Vapnik, V., 1996. *Regression Estimation with Support Vector Learning Machines*.
- Smola, A.J., Murata, N., Schölkopf, B., Müller, K.-R., 1998. Asymptotically Optimal Choice of  $\epsilon$ -Loss for Support Vector Machines, in: *Proceedings of the 8th International Conference on Artificial Neural Networks, perspectives in Neural Computing*, Springer Verlag, pp. 105–110.
- Solomatine, D.P., Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. *Journal of hydroinformatics* 10, 3–22.
- Sotomayor, O.A.Z., Odloak, D., 2005. Observer-based fault diagnosis in chemical plants. *Chemical Engineering Journal* 112, 93–108.
- Spérandio, M., Heran, M., Gillot, S., 2007. *Modélisation dynamique des procédés biologiques de traitement des eaux*. Techniques de l'Ingénieur.
- Stefatos, G., Ben Hamza, A., 2010. Dynamic independent component analysis approach for fault detection and diagnosis. *Expert Systems with Applications* 37, 8606–8617.
- Sugumaran, V., Sabareesh, G.R., Ramachandran, K.I., 2008. Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. *Expert Systems with Applications* 34, 3090–3098.
- Sun, X., Marquez, H.J., Chen, T., Riaz, M., 2005. An improved PCA method with application to boiler leak detection. *ISA Trans* 44, 379–397.
- Sun, Y., Liu, L., Popp, M., Farmerie, W., 2006. Estimation Of Cross-Hybridization Signals Using Support Vector Regression, in: *Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences - Volume 1 (IMSCCS'06) - Volume 01, IMSCCS '06*. IEEE Computer Society, Washington, DC, USA, pp. 17–21.
- Svard, C., Nyberg, M., 2010. Residual Generators for Fault Diagnosis Using Computation Sequences With Mixed Causality Applied to Automotive Systems. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 40, 1310–1328.

- Svärd, C., Nyberg, M., 2011. Automated Design of an FDI-System for the Wind Turbine Benchmark. Presented at the Proceedings of the 18th IFAC World Congress, 2011, Milano, Italy, pp. 8307–8315.
- Takács, I., 2008. Experiments in activated sludge modelling. PhD Thesis, Ghent University, Belgium, pp 267.
- Takács, I., Patry, G.G., Nolasco, D., 1991. A dynamic model of the clarification-thickening process. *Water Research* 25, 1263–1271.
- Tang, H.-S., Xue, S.-T., Chen, R., Sato, T., 2006. Online weighted LS-SVM for hysteretic structural system identification. *Engineering Structures* 28, 1728–1735.
- Tang, W.H., Wu, Q.H., 2011. Condition monitoring and assessment of power transformers using computational intelligence. Springer, London; New York.
- Tang, X., Zhuang, L., Cai, J., Li, C., 2010. Multi-fault classification based on support vector machine trained by chaos particle swarm optimization. *Knowledge-Based Systems* 23, 486–490.
- Tenenhaus, A., Giron, A., Viennet, E., Béra, M., Saporta, G., Fertil, B., 2007. Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification. *Computational Statistics & Data Analysis* 51, 4083–4100.
- Tharrault, Y., 2008. Diagnostic de fonctionnement par analyse en composantes principales: application à une station de traitement des eaux usées. Thèse de Doctorat de l'Institut National Polytechnique de Lorraine. Université de Nancy, France.
- Theilliol, D., Ponsart, J.-C., Harmand, J., Join, C., Gras, P., 2003. On-line estimation of unmeasured inputs for anaerobic wastewater treatment processes. *Control Engineering Practice* 11, 1007–1019.
- Tomita, R.K., Park, S.W., Sotomayor, O.A.Z., 2002. Analysis of activated sludge process using multivariate statistical tools—a PCA approach. *Chemical Engineering Journal* 90, 283–290.
- Tötterman, S., Toivonen, H.T., 2009. Support vector method for identification of Wiener models. *Journal of Process Control* 19, 1174–1181.
- Valle, S., Li, W., 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, 4389–4401.
- Vanrolleghem, P.A., Lee, D.S., 2003. On-line monitoring equipment for wastewater treatment processes: state of the art. *Water Sci. Technol.* 47, 1–34.
- Vapnik, V.N., 1998. *Statistical Learning Theory*. John Wiley & Sons Inc.

- Vapnik, V., 1999a. *The Nature of Statistical Learning Theory*, 2nd ed. Springer.
- Vapnik, V.N., 1999b. An overview of statistical learning theory. *IEEE Transactions on Neural Networks* 10, 988–999.
- Vapnik, V.N., 2000. *The nature of statistical learning theory*. Springer.
- Varol, Y., Oztop, H.F., Avci, E., 2008. Estimation of thermal and flow fields due to natural convection using support vector machines (SVM) in a porous cavity with discrete heat sources. *International Communications in Heat and Mass Transfer* 35, 928–936.
- Venkatasubramanian, V., 2003. A review of process fault detection and diagnosis Part III: Process history based methods. *Computers & Chemical Engineering* 27, 327–346.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., 2003. A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies. *Computers & Chemical Engineering* 27, 313–326.
- Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N., 2003. A review of process fault detection and diagnosis - Part I: Quantitative model-based methods. *Computers and Chemical Engineering* 27, 293–311.
- Verron, S., 2007. *Diagnostic et surveillance des processus complexes par réseaux bayésiens* Thèse doctorat, université d'Angers, France.
- Verron, S., Tiplica, T., Kobi, A., 2007. Procedure based on mutual information and bayesian networks for the fault diagnosis of industrial systems. *IEEE*, pp. 420–425.
- Verucchi, C.J., Acosta, G.G., Benger, F.A., 2008. A review on fault diagnosis of induction machines. *Latin American applied research* 38, 113–121.
- Waewsak, C., Nopharatana, A., Chaiprasert, P., 2010. Neural-fuzzy control system application for monitoring process response and control of anaerobic hybrid reactor in wastewater treatment and biogas production. *Journal of Environmental Sciences* 22, 1883–1890.
- Wang, D., Zhang, Y., Zhao, Y., 2007. Support Vector Machines for Photometric Redshift Estimation from Broadband Photometry - researchr publication bibtex. *Data Science Journal* 6, 474–480.
- Wang, J., Yu, T., Jin, C., 2006. On-line Estimation of Biomass in Fermentation Process Using Support Vector Machine. *Chinese Journal of Chemical Engineering* 14, 383–388.
- Wang, L., Mu, Z., Guo, H., 2006. Application of support vector machine in the prediction of mechanical property of steel materials. *Journal of University of Science and Technology Beijing, Mineral, Metallurgy, Material* 13, 512–515.

- Wang, L., Shi, H., 2010. Multivariate statistical process monitoring using an improved independent component analysis. *Chemical Engineering Research and Design* 88, 403–414.
- Wang, X., Du, Z., Chen, J., Pan, F., 2009. Dynamic Modeling of Biotechnical Process Based on Online Support Vector Machine. *Journal of Computers* 4, 251-258.
- Wang, X., Li, A., Jiang, Z., Feng, H., 2006. Missing value estimation for DNA microarray gene expression data by Support Vector Regression imputation and orthogonal coding scheme. *BMC Bioinformatics* 7:32.
- Wei, X., Verhaegen, M., van Engelen, T., 2009. Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques. *International Journal of Adaptive Control and Signal Processing* 687–707.
- Wei, X., 2010. Sensor fault detection and isolation for wind turbines based on subspace identification and Kalman filter techniques. *International Journal of Adaptive Control and Signal Processing* 24, 687–707.
- Widodo, A., Yang, B.-S., 2007. Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing* 21, 2560–2574.
- Widodo, A., Yang, B., 2008. Wavelet support vector machine for induction machine fault diagnosis based on transient current signal. *Expert Systems with Applications* 35, 307–316.
- Widodo, A., Kim, E.Y., Son, J.-D., Yang, B.-S., Tan, A.C.C., Gu, D.-S., Choi, B.-K., Mathew, J., 2009. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. *Expert Systems with Applications* 36, 7252–7261.
- Wimberger, D., Verde, C., 2008. Fault diagnosticability for an aerobic batch wastewater treatment process. *Control Engineering Practice* 16, 1344–1353.
- Wise, B.M., Gallagher, N.B., 1996. The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control* 6, 329–348.
- Witten, I.H., Frank, E., Hall, M.A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*, 3rd ed. Morgan Kaufmann.
- Wold, H., 1966. Estimation of Principal Components and Related Models by Iterative Least squares, in: *Multivariate Analysis*. Academic Press, pp. 391–420.
- Wu, Q., 2010. Fault diagnosis model based on Gaussian support vector classifier machine. *Expert Systems with Applications* 37, 6251–6256.
- Xiang, X., Zhou, J., An, X., Peng, B., Yang, J., 2008. Fault diagnosis based on Walsh transform and support vector machine. *Mechanical Systems and Signal Processing* 22, 1685–1693.

- Xue, C.X., Zhang, R.S., Liu, H.X., Liu, M.C., Hu, Z.D., Fan, B.T., 2004. Support Vector Machines-Based Quantitative Structure-Property Relationship for the Prediction of Heat Capacity. *Journal of Chemical Information and Modeling* 44, 1267–1274.
- Yang, B.-S., Widodo, A., 2008. Support Vector Machine for Machine Fault Diagnosis and Prognosis. *Journal of System Design and Dynamics* 2, 12–23.
- Yang, W., Tavner, P.J., Wilkinson, M., 2008. Condition monitoring and fault diagnosis of a wind turbine with a synchronous generator using wavelet transforms. *IET Conference Publications* 2008, 6–10.
- Yang, Y., 2005. Support vector machines for environmental informatics: application to biological nitrogen removal in wastewater treatment plants. Theses and dissertations.
- Yang, Y.H., 2006. Support Vector Machines for Environmental Informatics: Application to Modelling the Nitrogen Removal Processes in Wastewater Treatment Systems. *Journal of Environmental Informatics* 7, 14–23.
- Yélamos, I., Escudero, G., Graells, M., Puigjaner, L., 2006. Fault diagnosis based on support vector machines and systematic comparison to existing approaches, in: 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering. Elsevier, pp. 1209–1214.
- Yélamos, I., Graells, M., Puigjaner, L., Escudero, G., 2007. Simultaneous fault diagnosis in chemical plants using a multilabel approach. *AIChE Journal* 53, 2871–2884.
- Yélamos, I., Escudero, G., Graells, M., Puigjaner, L., 2009. Performance assessment of a novel fault diagnosis system based on support vector machines. *Computers & Chemical Engineering* 33, 244–255.
- Yin, S., Ding, S.X., Zhang, P., 2011. Study on Modifications of PLS Approach for Process Monitoring. Presented at the the 18th IFAC World Congress, Milano (Italy), pp. 12389–12394.
- Yoo, C.K., Choi, S.W., Lee, I.-B., 2002. Dynamic Monitoring Method for Multiscale Fault Detection and Diagnosis in MSPC. *Industrial & Engineering Chemistry Research* 41, 4303–4317.
- Yoo, C.K., Vanrolleghem, P.A., Lee, I.-B., 2006. Fault detection, monitoring and diagnosis of sequencing batch reactor for integrated wastewater treatment management system. *Environ. Eng. Res.* 11, 1-14.
- Yoo, C.K., Villez, K., Lee, I.B., Hulle, S.V., Vanrolleghem, P.A., 2006. Sensor validation and reconciliation for a partial nitrification process. *Water Science & Technology* 53, 513–521.

- Yoo, C.K., Villez, K., Van Hulle, S.W.H., Vanrolleghem, P.A., 2008. Enhanced process monitoring for wastewater treatment systems. *Environmetrics* 19, 602–617.
- Yu, J., 2012. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chemical Engineering Science* 68, 506–519.
- Yu, Z., Fu, X., Li, Y., 2005. Online Support Vector Regression for System Identification, in: Wang, L., Chen, K., Ong, Y.S. (Eds.), *Advances in Natural Computation*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 627–630.
- Yuan, S.-F., Chu, F.-L., 2006. Support vector machines-based fault diagnosis for turbo-pump rotor. *Mechanical Systems and Signal Processing* 20, 939–952.
- Yuan, S., Chu, F., 2007. Fault diagnosis based on support vector machines with parameter optimisation by artificial immunisation algorithm. *Mechanical Systems and Signal Processing* 21, 1318–1330.
- Zeng, G.M., Li, X.D., Jiang, R., Li, J.B., Huang, G.H., 2006. Fault Diagnosis of WWTP Based on Improved Support Vector Machine. *Environmental Engineering Science* 23, 1044–1054.
- Zhang, J., Sato, T., Iai, S., 2007. Novel support vector regression for structural system identification. *Structural Control and Health Monitoring* 14, 609–626.
- Zhang, X., Hoo, K.A., 2011. Effective fault detection and isolation using bond graph-based domain decomposition. *Computers & Chemical Engineering* 35, 132–148.
- Zhang, X., Zhang, Q., Zhao, S., Ferrari, R., Polycarpou, M.M., Parisini, T., 2011. Fault Detection and Isolation of the Wind Turbine Benchmark: An Estimation-Based Approach. Presented at the the 18th IFAC World Congress, Milano, Italy, pp. 8295–8300.
- Zhang, Y., 2008. Fault Detection and Diagnosis of Nonlinear Processes Using Improved Kernel Independent Component Analysis (KICA) and Support Vector Machine (SVM). *Industrial & Engineering Chemistry Research* 47, 6961–6971.
- Zhang, Y., Jiang, J., 2008. Bibliographical review on reconfigurable fault-tolerant control systems. *Annual Reviews in Control* 32, 229–252.
- Zhang, Y., 2009. Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM. *Chemical Engineering Science* 64, 801–811.
- Zhou, D., Li, G., Qin, S.J., 2010. Total projection to latent structures for process monitoring. *AIChE Journal* 56, 168–178.

Zumoffen, D., Basualdo, M., 2008. Improvements in Fault Tolerance Characteristics for Large Chemical Plants: 1. Waste Water Treatment Plant with Decentralized Control. *Industrial & Engineering Chemistry Research* 47, 5464–5481.

# **Titre Diagnostic de défauts par les Machines à Vecteurs Supports : application à différents systèmes multivariables non linéaires**

**Résumé** Les systèmes réels sont généralement de nature non-linéaire, et leurs modélisations et surveillance restent une tâche difficile à accomplir. Néanmoins, avec les progrès technologiques on dispose maintenant d'un atout de taille sur ces systèmes qui est les données.

Ce travail présente une technique de diagnostic de défaut et de modélisation basée en grande partie sur la méthode d'apprentissage automatique « Les Machines à Vecteurs de Support, SVM » qui est basée sur les données. La méthodologie proposée est appliquée à différents systèmes multivariables et non linéaires, à savoir : un procédé de traitement des eaux usées, un système éolien et un réacteur chimique parfaitement agité.

L'objectif de cette thèse de doctorat est d'examiner la possibilité d'extraire le maximum d'information à partir de données afin de surveiller efficacement le comportement de systèmes réels et de détecter rapidement tout défaut qui peut compromettre leur bon fonctionnement. La même méthode est utilisée pour la modélisation des différents systèmes. Plusieurs défis ont été relevés tels que la complexité du comportement des systèmes, le grand nombre de mesures variant à différentes échelles de temps, la présence de bruit et les perturbations. Une méthode générique de diagnostic de défauts est proposée par la génération des caractéristiques de chaque défaut suivie d'une étape d'évaluation de ces caractéristiques avec une amélioration du transfert de connaissances en modélisation.

Dans cette thèse ont a démontré l'utilité de l'outil Machines à Vecteurs de Support, en classification par la construction de modèles de décision SVM dédiés à l'évaluation des caractéristiques de défaut, et aussi en tant qu'estimateur non linéaire/ou pour la modélisation par l'utilisation des machines à vecteurs de support dédiés pour la régression (SVR).

La combinaison de SVM et d'une méthode basée sur le modèle "observateur" a été aussi étudiée et a été nécessaire dans certains cas pour garantir un bon diagnostic de défauts.

**Mots-clés :** Diagnostic de défauts, Machines à Vecteurs de Supports, Apprentissage automatique, Modélisation, identification paramétrique, Systèmes multivariables non linéaires, Observateur d'état, Procédé de traitement des eaux usées, Éoliennes, Réacteur chimique parfaitement agité.

## **Title** Fault diagnosis using Support Vector Machines: application to different multivariable nonlinear systems

**Abstract** Real systems are usually nonlinear and their modeling and monitoring remains a difficult task. However, with advances in technology and the availability of big amounts of data, we have a facility to operate these systems.

This work presents a methodology for fault diagnosis and modeling which is in large part based on the method of Support Vector Machines (SVM) which data-based. The proposed methodology is applied to various nonlinear multivariable systems including: wastewater treatment processes, wind turbines and stirred tank reactors.

The objective of this PhD is to examine the possibility of extracting the maximum of information from data to effectively monitor the behavior of real systems and rapidly detect any faults which may impair their proper functioning. The same method is used for modeling the different systems. Several challenges were identified and surmounted such as the complexity of the system behavior, large amount of data varying at different time scales, the presence of noise and disturbances. A generic method of fault diagnosis is proposed for the generation of the fault characteristics followed by an evaluation of these characteristics as well as an improved transfer of knowledge in modeling.

In this thesis the usefulness of the tool Support Vector Machines in Classification has been demonstrated by the construction of decision models dedicated to evaluating the characteristics of faults, and also its usefulness for modeling/ or as estimator for the nonlinear systems using support vector machines dedicated for regression (SVR).

The combination of SVM and a method based on models “observer” was also considered and was found to be interesting in some cases to ensure proper fault diagnosis.

**Keywords:** Fault diagnosis, Support Vector Machines, Machine learning, Modeling, System identification, Multivariable nonlinear systems, Observers, Wastewater treatment process, Wind Turbine, Stirred tank reactor.