



UNIVERSITE M'HAMED BOUGARA DE BOUMERDES
FACULTÉ DES HYDROCARBURES ET DE LA CHIMIE



Laboratoire d'Automatique Appliquée

THESE DE DOCTORAT

Présentée par

Abdelmalek KOUADRI

En vue de l'obtention du diplôme de
Docteur en Génie Electrique

Option : Automatique Appliquée et Traitement du Signal

Thème

***Contribution au diagnostic des systèmes automatiques complexes
basée sur une approche neuronale en composantes principales
non linéaires.***

Application sur un système hydraulique à trois cuves.

Devant le Jury composé de :

Larbi REFOUFI	Professeur	UMBB	Président
Mimoun ZELMAT	Professeur	UMBB	Rapporteur
Belkacem OULD- BOUAMMAMA	Professeur	Univ. Lille-France	Examineur
Abd Rezak LACHOURI	MC/A	Univ. Skikda	Examineur

Table des matières

Introduction générale	01
Chapitre I Détection et localisation de défauts par l'Analyse en Composantes Principales	
1. Introduction	05
2. Principe de l'ACP	06
3. Identification du modèle ACP	07
3.1. Critères empiriques	08
3.2. Critères théoriques	10
4. Détection et localisation de défauts	11
4.1. Détection de défauts	11
4.2. Localisation de défauts	13
5. Conclusion	14
Chapitre II Différentes approches de l'Analyse en Composantes Principales Non Linéaire	
1. Introduction	16
2. Analyse en Composantes Curvilignes	17
3. Courbes Principales	20
4. Approche neuronale de l'ACP NL	20
4.1. ACP NL par réseau à cinq couches	21
4.2. ACP NL par optimisation des entrées du réseau	25
4.3. Réseau à trois couches et courbes principales	26
5. Conclusion	31
Chapitre III Approche de l'ACP NL combinant la carte de Kohonen et le réseau de neurones de fonction à base radiale	
1. Introduction	32
2. Principe de la carte d'auto-organisation de Kohonen	33
3. Méthodologie de la technique	33
4. Procédure de calcul dans les cartes auto-organisatrices	35
5. Réseau de neurones de fonction à base radiale	37
6. Apprentissage du réseau de neurone à fonction de base radiale	39
7. Principe d'optimisation globale par application de l'algorithme DIRECT	42
8. Conclusion	47
Chapitre IV Application de l'ACP NL proposée au diagnostic de défauts dans un système à trois réservoirs	
1. Présentation du système	48
2. Extraction des composantes principales	49

3. Reconstruction de données initiales	52
4. Détection de défauts	57
5. Localisation de défauts	62
6. Conclusion	64
Conclusion générale	66
Références bibliographiques	70

Résumé

Dans cette thèse, nous avons présenté une extension de l'ACP dans le cas non linéaire par l'utilisation de deux types de réseaux de neurones. Le premier réseau, représenté par la carte d'auto-organisation de Kohonen, sert à extraire les composantes principales, tandis que le deuxième de fonction à base radiale est utilisé pour la reconstruction des données initiales. La connaissance des poids référents de la carte de Kohonen détermine une quantification vectorielle de la projection dans l'espace de dimension réduit. La transformation inverse permettant de calculer l'estimation des données initiales à partir des composantes principales obtenues par la carte de Kohonen se fait par un réseau de neurones de fonction à base radiale. Ce réseau de neurones est utilisé comme outil de modélisation par apprentissage. L'apprentissage est réalisé par des algorithmes d'optimisation dans le but est de minimiser la fonction de coût qui définit l'écart entre les sorties du réseau et les sorties réelles, en modifiant les paramètres d'une manière itérative. Nous avons présenté un algorithme d'optimisation direct et globale, appelé 'DIRECT' dans sa forme standard. Cet algorithme est basé sur la division continue de l'espace de recherche de la fonction coût. Cette approche neuronale proposée a été utilisée au diagnostic de défauts dans un système à trois réservoirs DTS-200.

Mots clés

Analyse en Composantes Principales Non Linéaires, Détection et Localisation de Défauts, Système à trois réservoirs type DTS-200

ملخص:

إن اكتشاف حدوث العطب مبكرا و تشخيصه يعتبر من أهم انشغالات و اهتمامات المؤسسة الصناعية. هدفه ينصب في حماية الأشخاص و الممتلكات و كذا كمية و نوعية الإنتاج، و الذي له مدلوله من الناحية الاقتصادية، و الحفاظ على البيئة من الناحية الايكولوجية. في هذا الإطار اقترحنا طريقة جديدة تعمد أساسا على التحليل الإحصائي، و التي استخدمنا فيها نوعين مختلفين من أنواع الشبكات العصبية، لتمثيل نموذج التحليل بالإحداثيات الأساسية.

الشبكة الوصية الأولى، و المتمثلة في البطاقة الذاتية التنظيم لكوهونن، استعملت لاستخلاص الإحداثيات الأساسية، أما الثانية الممثلة بالدالة القطرية فدورها الأساسي إعادة بناء المعطيات الأولية. هذا البناء يتم عن طريق تغيير معاملات الشبكة الوصية بالتلقين، و الذي يهدف إلى تقليص الفارق بين المخارج الأصلية و مخارج الشبكة.

استعملنا التقسيم المستمر لمجال البحث لمعاملات الشبكة أين أثبتت هذه الخوارزمية المعدلة نجاعتها في سرعة تقاربها إلى الحد الأدنى للدالة موضوع التحسين.

استخدمنا هذه الدالة كدليل لحدوث العطب من عدمه. أي عطب في نظام معين يكتشف عند اجتياز الدالة موضوع التحسين خط الثقة.

بعد التحليل الإحصائي لهذه الإشارة و باستخدام اختبار كلموغوروف و سميرونوف حدد عتبة الثقة. عدة تجارب قمنا بها باستحداث اعطاب مختلفة على الجهاز التجريبي ذو ثلاث خزانات مرتبطة، أين أثبتت الطريقة المقترحة نجاعتها في اكتشاف و تشخيص العطب.

الكلمات المفتاحية

التحليل بالإحداثيات الأساسية الغير خطية؛ اكتشاف حدوث العطب و تشخيصه؛ الجهاز التجريبي ذو ثلاث خزانات .

Abstract

The prompt detection of anomalous situations (fault detection) and the fast identification of the most probable causes (fault isolation) are significant aspects of the condition monitoring of complex dynamic systems in order to increase their reliability and safety. Artificial Neural Networks (ANN) can be used perfectly to detect and identify faults. In this context, Principal Component Analysis (PCA) based on the neural networks is investigated in this thesis. This technique is employed to detect and isolate faults in Three Tank System pilot plant. To extract the principal components, Self-Organising Map (SOM) neural network is considered. Radial Basis Function (RBF) neural network is used to reconstruct observed system features. SOM and RBF neural networks in cascade are known as NonLinear Principal Component Analysis (NLPCA) which is an interesting approach for the identification of significant correlations between variables of the process to be monitored without an explicit mathematical model of the system. Stat of Squared Predictive Error (*SPE*) is exploited as a tool for Fault Detection and Isolation (FDI). The experimental results demonstrate the effectiveness of this approach and its ability to detect and classify the various types of anomalies that are either on sensors or actuators in the three interconnected water tank system DTS-200.

Keywords

Nonlinear Principal Component Analysis, Fault Detection and Isolation, Three Tank System DTS-200.

1. Introduction générale

Au cours de la dernière décennie, la vision des scientifiques est orientée, dans certains domaines technologiques, vers une conception particulièrement intéressante, dans le sens d'un développement de la théorie du diagnostic et de ses applications. Il est donc naturel de rechercher, un mécanisme susceptible de faire apparaître des caractéristiques intelligentes, ceci face à des variations intensives dans un environnement plus au moins complexe.

Il est rare que les grandes conquêtes technologiques s'inspirent du vivant. La nature n'est pourtant pas avare de réalisations qui nous étonnent ou tiennent en échec notre curiosité. Parmi elles, le cerveau occupe évidemment une place de choix. Cette « machine » surprenante est en effet composée de plus de dix milliards d'unités élémentaires, les neurones, reliés entre eux par une multitude de connexions qui travaillent de façon concertée à l'élaboration des tâches les plus complexes. L'idée de s'inspirer du cerveau pour construire une machine intelligente, douée de propriétés cognitives et capable d'apprentissage, est un vieux et ambitieux projet humain. Ce projet, s'il n'est pas encore en passe de se réaliser, s'est, en tout cas précisé au cours des cinquante dernières années, grâce au progrès de nos connaissances en neurosciences et à l'apport de théories mathématiques et physiques. L'objectif n'est pas de copier aveuglement une réalité encore confuse ou mal comprise, au contraire, par l'introduction de ce concept de « cybernétique » qui signifie littéralement « science de la commande », nous tentons simplement à améliorer certaines techniques traditionnelles qui se trouvent incapables de s'adapter à des environnements complexes, incertains ou variants dans le temps et à des comportements à caractère non linéaire.

En particulier, dans le domaine du diagnostic, des méthodes sont développées en générale sur un test de cohérence entre un comportement observé du processus fourni par les capteurs et un comportement prévu fourni par une modélisation mathématique du processus. Ces méthodes analytiques nécessitent un modèle du système à surveiller. Ce modèle comprend un certain nombre de paramètres dont les valeurs sont supposées connues lors du fonctionnement nominal. En dépit de sa simplicité, quoiqu'il ne tient pas compte des bruits aléatoires qui sont à l'origine des variations paramétriques dans et sur le processus. Ceci impose de réaliser un

modèle fiable et capable de s'adapter à un certain degré de fiabilité aux éventuelles variations des paramètres du processus et des conditions de son fonctionnement, ce qui n'est pas toujours une tâche aisée. De plus, la complexité de l'élaboration de ce modèle augmente avec la dimension du système.

Comme alternative, les méthodes basées sur le développement statistique des données, et plus particulièrement l'Analyse en Composantes Principales (ACP), sont très intéressantes pour la mise en évidence des corrélations linéaires significatives entre les variables du processus sans formuler d'une façon explicite le modèle du système. Plusieurs travaux ont été consacrés à son emploi dans le domaine du diagnostic, dont la plupart ont présenté des modèles rigoureux, avec identification des structures de dépendance entre les observations multivariées, afin d'obtenir une représentation compacte de celles-ci. En fait, il semble plus que nécessaire de disposer de connaissances suffisamment précises et complètes du processus à surveiller pour remplir les exigences de robustesse et de performance au sens d'améliorer les conditions d'exploitation et éventuellement de réduire et faciliter les interventions. En revanche et suivant la précision exigible, nous abordons dans le présent travail une approche neuronale du modèle ACP car elle est caractérisée par leur faculté de se « reconfigurer » -souvent par les principes d'auto-organisation- face à un environnement inconnu de façon incrémental et robuste. Le modèle de l'Analyse en Composantes Principales Non Linéaire (ACPNL) est obtenu à partir de deux types de réseaux : la carte auto-organisatrice de Kohonen pour l'extraction des composantes principales, et un deuxième réseau dit à fonction de base radiale utilisé pour la reconstruction des données initiales.

Quant à l'idée d'hybridation des techniques, elle consiste en la combinaison des cartes auto-organisatrices de Kohonen et les réseaux de neurones à fonction de base radiale, afin d'en marier les avantages respectifs et d'augmenter ainsi l'efficacité des méthodes développées. Cette combinaison se démarque des autres outils par sa capacité d'apprentissage et de généralisation. Il vient s'ajouter qu'elle semble nécessiter moins d'hypothèses restrictives et d'information a priori sur le processus ou, tout au moins, des besoins différents. Elle constitue donc une approche prometteuse pour le diagnostic de processus.

C'est dans cette optique que s'inscrit la présente thèse. Il s'agit de collecter les informations à partir des mesures des variables du processus qui permettent de traduire son

comportement. Dans une première étape, les variables, porteuses d'un maximum d'informations et obtenues par la carte d'auto-organisation de Kohonen, dites composantes principales sont retenues à la base de la mesure de leurs variances cumulées explicatives. La seconde étape consiste en la reconstruction des données initiales. Cette opération de reconstruction s'effectue par modification des paramètres du réseau de fonction à base radiale dans le sens d'une minimisation de l'erreur quadratique de l'estimation des données *EQE*. Cette modification est réalisée par une méthode d'optimisation directe et globale dont le but est d'adapter les paramètres du réseau en fonction des stimuli présentés à son entrée. L'algorithme d'apprentissage du second réseau, que nous avons élaboré dans le présent travail, s'articule principalement sur une technique de division continue de l'espace de recherche. Cet algorithme dit DIRECT plus au moins modifié est suffisant pour constituer une méthode puissante et intéressante afin de dépasser les difficultés et les inconvénients lors de la mise en œuvre de réseau de neurones de fonction à base radiale. La précision, la simplicité, l'efficacité et la robustesse du modèle neuronal de l'ACPNL reposent essentiellement sur la qualité de la stratégie proposée.

La qualité de la reconstruction, mesurée par *EQE*, sert à décider de la présence ou non d'anomalies qui se traduit par le franchissement du seuil de détection de défauts approprié. Les méthodes empiriques pour la détermination du seuil de détection présentent l'inconvénient d'une intervention continuelle sans interruption de l'opérateur par suite des fausses alarmes, ce qui a rendu impossible, leur application aux systèmes de productions importantes. Plus précisément, il est question de mettre en œuvre un algorithme de détection de défauts, visant à extraire convenablement les caractéristiques statistiques du signal *EQE* analysé afin de cerner au mieux les phénomènes aléatoires qui lui sont associés.

Pour la maîtrise des défaillances, il est également requis de faciliter les interventions par la recherche de la cause de défauts lorsqu'ils sont apparus. Le calcul des contributions de chaque variable à la valeur de *EQE* permet d'identifier la ou les variables qui sont en cause. La variable symptôme d'un défaut est le synonyme de la plus forte contribution de celle-ci. L'évaluation pratique de ces stratégies de détection et de localisation de défauts sera réalisée à partir de son application sur un stand expérimental de trois réservoirs interconnectés de type DTS-200.

Cette thèse, consacrée aux méthodes statistiques multivariées basées sur l'Analyse en Composantes Principales Non Linéaires (ACPNL) pour la détection et la localisation de défauts dans un système de trois réservoirs DTS-200, comporte quatre chapitres.

Dans le premier chapitre, nous synthétisons les modalités de l'analyse en composantes principales linéaires et les différents critères utilisés au choix du nombre des axes principaux à retenir lors de la projection des données dans un espace compacté. De plus, il est fait allusion à l'utilisation de cette technique de réduction de dimension à la détection et la localisation de défauts.

Le chapitre 2 est consacré à la description de quelques techniques à base d'ACPNL. On distinguera les méthodes de projection classiques telles que l'Analyse en Composantes Curvilignes (ACC), les courbes principales et les méthodes de projection à base de réseaux de neurones. On présentera les différents types, les architectures, et les algorithmes d'apprentissage utilisés dans cette dernière catégorie.

Le chapitre 3 décrit la présentation de l'approche proposée à la détection et la localisation de défauts. Cette technique à base de l'ACPNL utilise conjointement la carte auto-organisatrice de Kohonen et le réseau de neurones de fonction à base radiale. Le premier réseau est utilisé pour l'extraction des composantes principales tandis que le second réseau est employé pour reconstruire les données initiales. L'algorithme de recherche globale dénommé DIRECT, (acronyme en anglais pour, DIviding RECTangles) qui appartient à la famille des algorithmes de recherche directe a été introduit dans sa forme standard.

Le dernier chapitre a pour objectif d'examiner et d'illustrer l'aptitude du nouvel outil, les perspectives offertes pour le diagnostic, d'une part à la caractérisation et la représentation des systèmes complexes, et d'autre part à la synthèse de stratégies de détection et de localisation de défauts appropriées pour ce type de systèmes, en prenant l'exemple d'un processus expérimental à trois réservoirs DTS-200.

1. Introduction

La détection des défauts et leur diagnostic occupent une position stratégique dans les processus de surveillance et maintenance industriels modernes. Diverses approches sont proposées pour leur caractérisation [1,2]. Ces dernières sont en général développées sur un test de cohérence entre un comportement observé du processus, fourni par les capteurs, et un comportement prévu, issu d'une représentation mathématique du système. Ces méthodes nécessitent un modèle fiable du système à surveiller, ce qui n'est pas toujours une tâche aisée [3]. De plus, la complexité de l'élaboration de ce modèle augmente avec le nombre souvent élevé de variables dans une application à l'échelle industrielle. L'étude séparée de chacune de ces variables selon les techniques de la statistique descriptive unidimensionnelle est une phase indispensable dans le processus de dépouillement des données mais reste cependant insuffisante. Comme alternative, les méthodes basées sur l'Analyse en Composantes Principales (ACP), sont très intéressantes pour la mise en évidence des corrélations significatives entre les variables du processus sans formuler de façon explicite le modèle du système [4,5].

Le principe de base de l'ACP est de réduire les dimensions de la matrice des données, en conservant toutefois les variations présentes dans le jeu de données de départ. En effet, cette technique permet de privilégier les directions d'un espace de données porteuses du maximum d'informations au sens de la maximisation des variances de projection. L'amplitude des valeurs propres de la matrice de covariance des données quantifie pour chacune de ces directions la quantité d'information encodée.

Les variables initiales sont ainsi transformées en de nouvelles variables, appelées composantes principales. Elles sont obtenues par combinaison linéaire des composantes originales, ordonnées et non corrélées. L'ACP cherche à identifier les vecteurs propres et les valeurs propres de la matrice de corrélation des variables de départ.

Concernant l'application de l'ACP dans le domaine du diagnostic, un modèle ACP est défini globalement par la matrice des premiers vecteurs propres de la matrice de corrélation des données. Cette dernière permet de définir à la fois les projections permettant d'obtenir les composantes principales et la projection inverse pour estimer les données de départ. La différence entre les données de départ et celles estimées par l'ACP, représentée par la matrice

des résidus, sera utilisée comme paramètre de détection de défauts. Plus précisément, il est question de mettre en œuvre un algorithme de détection de défauts, visant à extraire convenablement les caractéristiques statistiques du signal analysé afin de cerner au mieux les phénomènes aléatoires qui lui sont associés. A chaque instant, la statistique *EQE* (Erreur Quadratique d'Estimation), représentant la somme des carrés des résidus, est évaluée. Un défaut quelconque serait mis en évidence si ce signal franchit le seuil de détection calculé sur une distribution de l'*EQE* et par rapport à un niveau de signification approprié. C'est le principe de détection qui sera adopté.

En présence d'un défaut suspecté, il devient nécessaire de déterminer la ou les variables qui en sont à l'origine. La contribution instantanée de la statistique *EQE*, évaluée pour chaque variable du système, permet d'identifier la cause du défaut. La variable admettant la plus grande valeur en terme de contribution, représentera le symptôme de défaut.

2. Principe de l'ACP

L'Analyse en Composantes Principales (ACP) a pour but de réduire la dimension de l'espace de la représentation naturelle d'un système donné par la détermination de nouvelles variables que nous qualifierons d'artificielles. Ainsi, les N mesures des m variables du processus représentant tous les modes en fonctionnement normal de ce dernier, sont collectées pour former une matrice des données \mathbf{X} .

Un prétraitement indispensable consiste à centrer et réduire les variables. Soit \mathbf{X} la matrice de données prenant la forme :

$$\mathbf{X} = \left[x_i^j \right]_{i=1, N}^{j=1, m} \quad (\text{I.1})$$

avec $x_i^j \in \mathfrak{R}$.

On calcule ensuite la moyenne et la variance pour chaque variable. On obtient deux matrices :

$$\mathbf{M} = [m_1, m_2, \dots, m_m] \quad (\text{I.2})$$

et

$$\mathbf{S} = [s_1, s_2, \dots, s_m] \quad (\text{I.3})$$

avec les éléments identiques de chaque colonne de la matrice des moyennes \mathbf{M} et la matrice des inverses des écarts types \mathbf{S} de la matrice des données \mathbf{X} sont exprimés par :

$$m_i^j = \frac{1}{N} \sum_{k=1}^N x_k^j \quad (\text{I.4})$$

$$s_i^j = \left(\frac{1}{N} \sum_{k=1}^N (x_k^j - m_i^j)^2 \right)^{\frac{1}{2}} \quad (\text{I.5})$$

Les données peuvent alors être centrées et réduites comme suit :

$$\mathbf{X}^{(n)} = (\mathbf{X} - \mathbf{M})\mathbf{S} \quad (\text{I.6})$$

$\mathbf{X}^{(n)}$ représente la nouvelle matrice des données normalisées. Ceci permet de rendre les résultats indépendants des unités pour les différentes données.

La matrice de covariance ou de corrélation $\mathbf{\Sigma}$, exprime le lien entre les différentes variables, est donnée par :

$$\mathbf{\Sigma} = \frac{1}{(N-1)} (\mathbf{X}^{(n)})^T \mathbf{X}^{(n)} \quad (\text{I.7})$$

Une décomposition spectrale de cette dernière permet d'écrire :

$$\mathbf{\Sigma} = \sum_{i=1}^m \lambda_i \mathbf{p}_i \mathbf{p}_i^T \quad (\text{I.8})$$

où \mathbf{p}_i est le $i^{\text{ème}}$ vecteur propre de $\mathbf{\Sigma}$ et λ_i est la valeur propre correspondante.

La projection de la matrice des observations représente la nouvelle matrice des variables transformées indépendantes et dont les variances sont les valeurs propres de la matrice de covariance, ordonnées dans l'ordre décroissant. Cette transformation est donnée par :

$$\mathbf{t} = \mathbf{X}^{(n)} \mathbf{P} \quad (\text{I.9})$$

Celle-ci regroupe ce qu'on appelle les composantes principales et qui sont statistiquement non corrélées.

3. Identification du modèle ACP

Le principal intérêt de l'ACP consiste à réduire la dimension de l'espace des variables du processus. Le choix du nombre d'axes à retenir est un point essentiel qui n'a pas trouvé encore de solution rigoureuse [6]. Le sous-espace de dimension l recherché est tel que la moyenne des carrés des distances entre projections soit la plus grande possible. Soit \hat{P} l'opérateur de projection, tel que :

$$\hat{P}^T \hat{P} \approx I \quad (\text{I.10})$$

où \hat{P} représente la matrice des vecteurs propres correspondants aux l valeurs propres dominantes de la matrice de corrélation. Les valeurs propres de la matrice de corrélation sont les mesures de la variance et peuvent donc être utilisées dans la sélection du nombre de composantes principales. La qualité de la réduction de dimension est directement liée au choix du nombre l d'axes principaux à retenir, Joliffe énumère une série de règles couramment utilisées pour la détermination du nombre l [7] ; elles peuvent être regroupées en deux classes :

1. Les règles intuitives plausibles sans fondement théoriques
2. Les règles basées sur des hypothèses probabilistes.

3.1. Critères empiriques

Ce sont en réalité les seuls applicables pour la détermination du nombre de composantes principales à retenir et ainsi définir le modèle ACP.

a. Critère de Kaiser

On retient les composantes principales correspondant à des valeurs propres supérieures de module à 1. En effet les composantes principales t étant des combinaisons linéaires des x^j de variance maximale $V(t) = \lambda$, seules les composantes de variance supérieure à celle des variables initiales présentent un intérêt.

Le scree-test de Cattell en est la version analytique. On calcule les différences premières :

$$\lambda_1 - \lambda_2 = \varepsilon_1, \lambda_2 - \lambda_3 = \varepsilon_2, \lambda_3 - \lambda_4 = \varepsilon_3, \dots$$

puis les différences secondes :

$$\varepsilon_1 - \varepsilon_2 = \delta_1, \varepsilon_2 - \varepsilon_3 = \delta_2, \varepsilon_3 - \varepsilon_4 = \delta_3, \dots$$

On retient alors les valeurs propres telle que $\delta_1, \delta_2, \dots, \delta_l$ soient tous positifs.

b. Moyenne des valeurs propres

Le principe de cette règle consiste à ne prendre en considération que les composantes principales dont la valeur propre associée est de module supérieur à la moyenne arithmétique des modules de toutes les valeurs propres, car elles sont jugées plus informatives que les variables initiales.

c. Pourcentage de la variance cumulée

L'objectif est de garder un maximum d'information des données initiales, la mesure de cette information peut être expliquée par la variance. Les valeurs propres de la matrice de corrélation sont les mesures de cette variance. Le nombre de composantes principales à retenir et qui constitue le modèle ACP, est le plus petit nombre pris de telle sorte que le pourcentage de la variance cumulée soit atteint ou dépasse une valeur seuil fixée a priori. Les composantes sont choisies successivement dans l'ordre des variances décroissantes.

Le pourcentage de la variance justifiée par les l premières composantes est donné par :

$$PVC(l)\% = 100 \left(\frac{\sum_{i=1}^l \lambda_i}{\sum_{i=1}^m \lambda_i} \right) \% \quad (\text{I.11})$$

d. Critère de validation croisée

La base de cette méthode est d'estimer les mesures d'un jeu de données de validation à partir d'un modèle qui a été calculé sur la base d'un autre jeu de données d'identification puis de comparer ces estimés avec les valeurs mesurées. Cette procédure dite de validation croisée est basée sur la minimisation de la somme des carrés des résidus, elle s'écrit :

$$SCR(l) = \frac{1}{Nm} \sum_{j=1}^m \sum_{i=1}^N (\hat{x}_i^{(l)j} - x_i^j)^2 \quad (\text{I.12})$$

3.2. Critères théoriques

Ils consistent à déterminer si les valeurs propres sont significativement différentes entre elles à partir d'un certain rang. Dans le cas affirmatif, la moyenne arithmétique a des $m-l$ dernières valeurs propres et leur moyenne géométrique g doivent être peu différentes. On admet que :

$$c = \left(N - \frac{2m+11}{6} \right) (m-l) \ln \left(\frac{a}{g} \right) \quad (\text{I.13})$$

suit alors une loi du χ^2 de degré de liberté $\frac{(m-l+2)(m-l-1)}{2}$; on rejettera l'hypothèse d'égalité des $m-l$ valeurs propres pour c est assez élevé.

On peut également construire des intervalles de confiance pour les différentes valeurs propres sachant que l'intervalle de confiance à 95% est donné par :

$$\left[\lambda_i \exp \left(-1.96 \sqrt{\frac{2}{N-1}} \right) \quad \lambda_i \exp \left(1.96 \sqrt{\frac{2}{N-1}} \right) \right] \quad (\text{I.14})$$

Ces propriétés ne sont malheureusement utilisables que pour des matrices de variance dans le cas gaussien m -dimensionnel. Elles ne s'appliquent absolument pas pour les matrices de corrélation ce qui est le cas le plus fréquent en pratique.

Une fois le nombre de composantes à retenir déterminé, la matrice \mathbf{X} des données peut être approximée à partir des l premières composantes principales correspondant aux l plus grandes valeurs propres de la matrice Σ . La matrice $\hat{\mathbf{X}}$ des données estimée est alors décrite par :

$$\hat{\mathbf{X}} = \mathbf{t}\hat{\mathbf{P}}^T = \mathbf{X}^n \hat{\mathbf{P}}\hat{\mathbf{P}}^T \quad (\text{I.15})$$

Rappelons que les valeurs propres de la matrice Σ sont non négatives par définition. La direction de l'espace des données est concrétisée par le vecteur propre associé à la plus grande valeur propre, la plus riche d'information. Inversement, la direction du vecteur propre directeur associé à la plus petite valeur propre est celle qui présente le minimum d'information. La perte d'information induite par la réduction de dimension de la

représentation de données est mesurée par la différence e entre ses représentations exactes et celles approchées, et y donnée par :

$$e = X^n - \hat{X} = \sum_{i=l+1}^m t_i P_i^T \quad (I.16)$$

Donc, on suppose que le meilleur modèle ACP est celui qui présente l'erreur quadratique d'estimation EQE la plus faible. Cette erreur est donnée par :

$$EQE = e^T e \quad (I.17)$$

4. Détection et localisation de défauts par l'ACP

4.1. Détection de défauts

Dans la plupart des travaux publiés utilisant l'analyse en composantes principales pour la détection de défauts, deux indices sont souvent utilisés. La statistique de Hotelling est calculée à partir des premières composantes principales et l'erreur quadratique d'estimation. Cependant la statistique de Hotelling ne présente pas vraiment un résidu puisqu'elle est calculée dans le sous espace principal. Généralement, cette statistique est représentée par la somme des carrés des composantes principales retenues pour la structuration du modèle ACP. Donc, il faut trouver un indice de détection permettant de s'affranchir du problème de la variation significative des variables du processus et d'être plus sensible aux défauts. L'erreur quadratique d'estimation EQE se trouve moins sensible aux erreurs de modélisation et plus sensible aux défauts. Le dysfonctionnement d'un processus à l'instant k se traduit par le franchissement de seuil de détection:

$$EQE(k) > \sigma_\alpha^2 \quad (I.18)$$

où σ_α^2 , représente le seuil de détectabilité à un niveau de confiance α ; il est obtenu à partir de l'expression suivante [8] :

$$\sigma_\alpha^2 = \theta_1 \left[1 + \frac{\sqrt{2C_\alpha h_0}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \quad (I.19)$$

avec

$$\theta_i = \sum_{j=l+1}^m \lambda_j^i, \quad i=1, 2, 3 \quad (\text{I.20})$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (\text{I.21})$$

et

$$C_\alpha = \frac{\theta_1 \left[\left(\frac{\|e\|^2}{\theta_1} \right)^{h_0} - 1 - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]}{\sqrt{2\theta_2 h_0^2}} \quad (\text{I.22})$$

La position d'un tel problème impose d'assurer une détection précise des défauts, en présence d'un grand nombre de facteurs aléatoires. Plus précisément, il convient de construire, à partir d'un prélèvement donné, une limite supérieure telle que la probabilité pour qu'il contienne une valeur $EQE(k)$ acceptable. Cette probabilité définira un niveau de confiance ou d'accessibilité de l'estimation. Il est facile de considérer que le calcul probabiliste de l'indice de détection et son intervalle de confiance associé soient basés sur le fait que la variable aléatoire EQE suit une telle distribution. Néanmoins, une telle condition n'est pas toujours remplie, ce qui nécessite le recours à un test statistique approprié afin de confirmer la validité des résultats.

Pour ce faire, nous avons opté pour le test de Kolmogorov-Smirnov. Ce choix est étroitement lié à la nature des mesures manipulées, à la distribution des fréquences de la variable aléatoire EQE et au nombre d'échantillons dont on dispose. Plus précisément, il s'agit de créer, sur un échantillon donné, un vecteur de fréquences cumulées de la variable aléatoire EQE obéissant à une loi de distribution. Ces fréquences cumulées F_k , qui représentent la fonction de répartition, s'obtiennent à partir des probabilités p_k évaluées pour tout $EQE(k)$ inférieur à δ_α ; elles sont données par :

$$F_k = \sum_{EQE(k) < \delta_\alpha} p_k \quad (\text{I.23})$$

Ainsi, on définit une loi intégrale empirique \mathcal{F} et le test à effectuer se ramène à vérifier pour tout $\delta_\alpha \in \mathfrak{R}$ la condition suivante :

$$\sup_{\delta_\alpha \in \mathfrak{R}} |F_k - \mathcal{F}| \rightarrow 0 \quad (\text{I.24})$$

Ces considérations améliorent la qualité de détection et réduisent voire éliminent les fausses alarmes.

4.2. Localisation de défauts

La localisation de défauts implique la recherche de la ou des variables qui sont en cause. Elle se base sur diverses méthodes telle que :

- la structuration des résidus [9]
- le principe de reconstruction [10]
- le calcul des contributions [11]

La mise en œuvre de la première méthode est relativement aisée mais elle présente cependant un inconvénient majeur qui réside dans la fréquente dégradation des signatures expérimentales. La raison principale de cette dégradation est liée aux incertitudes de mesures, aux erreurs de modèles et aux effets des perturbations agissant sur le système et qui affectent les résidus.

La condition nécessaire de la localisation par reconstruction est donnée par :

$$m - l > 1 \quad (\text{I.25})$$

pour que la reconstruction annule l'erreur d'estimation de la variable reconstruite. Par conséquent, cette méthode se révèle peu performante car son choix est étroitement lié à la nature des mesures manipulées et au nombre de composantes principales à retenir dans le modèle ACP.

La localisation de défauts par le calcul des contributions est largement utilisée dans le diagnostic par ACP. Cette approche est basée sur le calcul des contributions de chaque variable \mathbf{x}_i à la valeur de l'indice de détection EQE . Cette contribution à l'instant k s'écrit comme suit :

$$EQE_i(k) = (\mathbf{x}_i(k) - \hat{\mathbf{x}}_i(k))^2 = (\mathbf{e}_i(k))^2 \quad (\text{I.26})$$

La variable ayant la plus forte contribution est considérée comme le symptôme de défaut.

5. Conclusion

Le principe de l'analyse en composantes principales linéaire a été présenté. Cette méthode permet d'étudier les relations existant entre les variables et d'identifier les dépendances entre elles afin d'obtenir une représentation compacte de ces dernières. Après avoir centré et réduit la matrice d'observation, l'estimation des paramètres du modèle ACP est effectuée par le calcul des valeurs et des vecteurs propres de la matrice de corrélation des données. Cependant, pour l'obtention de la structure du modèle, il faut déterminer le nombre de composantes principales à retenir dans celui-ci. Plusieurs critères ont été cités, notre choix s'est porté sur le critère de la variance cumulée dans le reste de cette thèse. Le modèle ACP obtenu est défini globalement par la matrice des premières directions orthogonales porteuses du maximum d'information au sens de la maximisation des variances de projection. Cette matrice permet de définir à la fois la projection permettant d'avoir les composantes principales et la projection inverse pour l'estimation des données originelles.

La méthode de détection fondée sur le modèle ACP utilise l'espace résiduel comme indice de détection. Cet indice révélateur de la présence de défauts, qui exprime l'erreur quadratique d'estimation, est généré à partir d'un modèle ACP de bon fonctionnement du processus et des mesures disponibles. Un défaut quelconque serait mis en évidence si EQE dépasse le seuil de confiance établi sous certaines considérations et propriétés statistiques du signal EQE dans le but de définir au mieux les phénomènes aléatoires qui lui sont associés. Lorsqu'un défaut abrupt est détecté, le calcul des contributions de chaque variable à la valeur de EQE permet d'identifier la ou les variables qui sont en cause. La variable symptôme d'un défaut est le synonyme de la plus forte contribution de celle-ci.

1. Introduction

Comme la plupart des approches utilisent les réseaux de neurones pour l'obtention du modèle ACPNL [30], il se pose souvent des problèmes d'optimisation à caractère non linéaire comme les problèmes de convergence, initialisation des poids synaptiques et la topologie de ce type de réseau [31]. Comme alternative, nous proposons une approche pour l'ACPNL obtenue à partir de deux types de réseaux : la carte auto-organisatrice de Kohonen pour l'extraction des composantes principales, et un deuxième réseau dit à fonction de base radiale utilisé pour la reconstruction des données initiales.

L'efficacité de l'utilisation de la carte de Kohonen est liée à la non disponibilité des informations a priori sur les composantes principales et la simplicité de son apprentissage. Dans la présente étude, la carte choisie utilise comme critère d'apprentissage, l'apprentissage non supervisé et la maximisation de la variance.

Après avoir identifié l'opérateur de projection par le premier réseau, les composantes principales obtenues sont injectées dans le deuxième réseau pour l'obtention des données reconstruites. Cette opération de reconstruction s'effectue par modification des paramètres de réseau à fonction de base radiale dans le sens d'une minimisation de l'erreur quadratique de l'estimation des données *EQE*. Cette modification est réalisée par les algorithmes de calcul dont le but est d'adapter les paramètres du réseau en fonction des stimuli présentés à son entrée. Fondamentalement, les modifications peuvent se faire suivant deux mécanismes de rétroaction.

Dans le présent travail, nous nous sommes limités aux algorithmes basés sur la version qui ne nécessite pas de calcul du gradient *EQE*, car leurs mise en œuvre augmente le pouvoir de généralisation et la rapidité de l'apprentissage. De plus, en pratique l'information dont on dispose est souvent numérique ce qui augmente la difficulté de différentiation de la fonction objective. Cet algorithme de recherche globale dénommé DIRECT, (acronyme en anglais pour, DIviding RECTangles) qui appartient à la famille des algorithmes de recherche directe à

a été utilisé. Plus précisément, c'est une méthode déterministe relevant de la classe des méthodes de recherche à motifs. Cet algorithme a été proposé par Jones et al. [32] en 1993 pour l'optimisation de fonctions à plusieurs variables dans un espace continu borné.

L'algorithme évalue la fonction au centre du domaine, ce qui le rend facilement généralisable pour les grandes dimensions. De plus, l'algorithme DIRECT décompose l'espace de recherche en hyper-rectangles dont la dimension est donnée par le nombre de variables. Au cours de son exécution, DIRECT réalise un compromis entre la recherche globale et la recherche locale par son mode de sélection des hyper-rectangles à explorer. Les hyper-rectangles sélectionnés sont dits "optimaux". Les blocs qui ne sont pas optimaux à une itération donnée peuvent le devenir par la suite. Cette stratégie permet selon Jones et al. de converger rapidement dans un espace de recherche raisonnablement large.

2. Principe des cartes auto organisatrices de Kohonen

Les cartes auto-organisatrices présentent un exemple des méthodes à apprentissage non supervisé. Elles analysent les données et extraient l'information à partir de ces données en utilisant un algorithme de classification très sophistiqué. La visualisation par cartes topologiques proposées par Kohonen cherche, par apprentissage à partir des données, à partitionner l'ensemble des observations disponibles en groupements similaires. Les groupements obtenus possèdent la propriété caractéristique d'avoir une structure de voisinage.

La carte auto-organisatrice est habituellement composée d'une grille de neurones à une ou deux dimensions (voir figure (III.1)). Parfois les cartes à une seule dimension peuvent être employées pour résoudre certains problèmes [33]. Dans notre cas, elle offre la possibilité d'effectuer une réduction de dimensions.

3. Méthodologie de la technique

De la même façon que dans le cas de l'ACP linéaire, la matrice d'observation doit être centrée et réduite. La carte auto-organisatrice de Kohonen se compose d'une seule couche de neurones organisés dans une dimension. Chaque neurone admet un nombre d'entrées équivalent au nombre de signaux à utiliser dans l'opération de projection (voir figure (III.2)).

La procédure d'apprentissage consiste à trouver le neurone ayant le poids le plus proche du vecteur de données d'entrée et le déclarer comme neurone gagnant. Les poids de tous les neurones à proximité du neurone gagnant sont ajustés par une quantité inversement proportionnelle à leur distance. Au fur et à mesure que le nombre des itérations augmente, le rayon de voisinage admis sera réduit. Le processus d'apprentissage se termine si les variances de toutes les sorties sont à un niveau acceptable ou après un nombre d'itération prédéfini.

Il y a deux méthodes pour déterminer la similitude des réponses de données. Dans la première, chaque entrée est pondérée par le vecteur poids du neurone correspondant et les résultats sont additionnés. La sortie du neurone i , calculée à partir du produit scalaire d'un vecteur de données x^j pour un vecteur de poids $w_{i,j}$ associé, est donnée par la relation suivante :

$$t_i = \sum_{j=1}^m (x^j w_{i,j}) \quad (\text{III. 1})$$

Ce produit représente la projection du vecteur de données sur le vecteur de poids à chaque instant. Il représente le cosinus de l'angle délimité par ces deux vecteurs. La variance de projection maximale représenterait deux vecteurs colinéaires et leurs paramètres seront semblables. Par contre une valeur nulle ou minimale de la variance signifie que les deux vecteurs sont perpendiculaires ou quasi perpendiculaires, alors ils ne sont pas similaires.

La deuxième méthode pour mesurer la similitude de deux vecteurs est de calculer la distance euclidienne entre ces deux vecteurs, ici x^j et $w_{i,j}$:

$$t_i = \sqrt{\sum_{j=1}^m (x^j - w_{i,j})^2} \quad (\text{III. 2})$$

La valeur proche de zéro de t_i signifiera que les deux vecteurs sont relativement identiques. On applique cette opération à tous les neurones et le neurone qui a le plus grand produit scalaire ou la distance euclidienne minimale sera choisi en tant que gagnant.

Afin de maintenir la similitude des neurones les plus proches d'un point de vue topologique, les poids de tous les neurones dans un rayon choisi doivent être ajustés. Le neurone gagnant est le mieux ajusté par un processus modifiant ses poids de telle sorte qu'ils soient plus près des données initiales. Les poids des autres neurones sont ajustés en introduisant un gain inversement proportionnel au nombre d'itérations et au rayon de voisinage par rapport au neurone gagnant. À la fin de l'apprentissage, le vecteur poids de chaque neurone représente une référence, appelé vecteur référent.

4. Procédure de calcul dans les cartes auto-organisatrices

L'application des cartes auto-organisatrices de Kohonen pour la projection dans un espace réduit que celui des données initiales peut être résumé sous la forme de l'algorithme descriptif suivant :

1. **Initialisation de données :** la matrice de données doit être centrée et réduite en utilisant les équations (I.2), (I.3), (I.4), (I.5) et (I.6). Ceci empêchera tout signal de maîtriser le

reste de l'ensemble des signaux pendant les étapes de projection. Tous les échantillons des mesures dans l'état sein du système sont pris à chaque cycle d'itération.

2. **Initialisation des poids de la carte :** chaque neurone aura un vecteur de poids instantané incident d'un vecteur de la matrice d'observation. Des valeurs aléatoires sont affectées à chacun des poids et également au rayon de voisinage. Ce rayon sera important au début, en respectant la dimension de la carte choisie, pour couvrir le plus grand nombre de neurones adjacents. Au cours des itérations, le rayon est systématiquement réduit à une distance égale à l'unité. Un compteur indique le nombre de fois consécutives qu'un neurone déclaré gagnant est utilisé. Si un neurone 'gagne' excessivement, il permettra à d'autres neurones d'être déclarés gagnant.

3. **Détermination du neurone gagnant :** le produit scalaire ou la distance euclidienne, définies par les équations (II.1) et (II.2) respectivement, est calculé pour l'ensemble des données et les poids chaque neurone de la carte. Le neurone avec une distance minimale ou un produit scalaire maximal est déclaré gagnant. Ceci conduit à un neurone avec les poids les plus semblables aux données normalisées. Avant que le neurone gagnant ne soit identifié, son nombre de 'victoires' consécutives est évalué. Si la 'victoire' d'un neurone dépasse une limite prédéfinie, alors le nombre de neurones dont les poids sont ajustés, sont réduits proportionnellement pour permettre à d'autres neurones de 'gagner'. Le compteur de neurones perdants est remis à zéro et celui de neurones gagnants est incrémenté d'une unité.

4. **Ajustement des poids du neurone gagnant :** les poids de neurone gagnant sont modifiés en utilisant la formule de récurrence suivante :

$$W^{i+1}(j, k) = W^i(j, k) + \eta(i) \left((X^n(k))^T - W^i(j, k) \right) \quad (\text{III. 3})$$

$W^i(j, k)$ représente le vecteur poids à l'instant k du neurone gagnant j à la i -ème itération. Le pas du gradient $\eta(i)$ est une fonction décroissante du nombre d'itérations i . Elle peut être constante par morceaux, égale à $\frac{1}{\sqrt{i}}$ ou prendre d'autres formes.

5. Adaptation des poids des neurones voisins : la modification des poids des neurones adjacent au neurone gagnant peut être facilement calculée par une fonction décroissante du nombre d'itérations et le rayon de voisinage. La version adaptative de l'algorithme aux neurones de voisinage est donnée par :

$$W^{i+1}(r, k) = W^i(r, k) + d_j(r, i) \eta(i) \left((X^n(k))^T - W^i(r, k) \right) \quad (\text{III. 4})$$

avec $r \neq j$.

$d_j(r, i)$ est une fonction du nombre d'itérations i et de la distance entre le neurone gagnant j et son neurone voisin r . Tous les autres neurones en dehors du rayon associé au neurone gagnant ne subiront aucune correction.

6. Critère d'arrêt : une caractéristique principale des cartes de Kohonen est la détermination des neurones avec des poids associés. Par conséquent, une bonne projection est quantifiée par la distance minimale prédéfinie entre les poids et les données initiales. La procédure itérative est arrêtée lorsque la distance euclidienne entre les données et les poids associés est en dessous d'une certaine limite définie à l'avance. Si cette condition n'est pas remplie, une série d'itérations additionnelle sera nécessaire. Les poids trouvés sont des fonctions mémorisées qui seront employés pour la détermination des composantes principales.

5. Réseau de neurones à fonction de base radiale

La plupart des applications courantes utilisent une architecture du « réseau perceptron à couche multiples » dont la fonction d'activation à chaque nœud est de type sigmoïde [34,35,35,36].

Il existe un certain nombre de théorèmes garantissant les capacités d'approximation universelle dans certaines conditions (continuité de la fonction à approcher, nombre suffisant d'unités dans les couches cachées), même si le réseau ne comporte qu'une seule couche cachée. Cependant, ces théorèmes ne fournissent pas des méthodes systématiques constructives guidant le choix de la topologie adaptée pour une application particulière. En outre, ce type d'architecture entraîne des difficultés d'apprentissage [37] dont les plus importantes sont :

- la présence de minima locaux dans la fonction de coût,
- la possibilité de paralysie durant l'apprentissage à cause de la saturation de la fonction sigmoïde,
- la possibilité de désapprentissage dû au caractère « non local »,
- La difficulté d'introduire l'information a priori, ce qui est dû au caractère distribué et non local de la représentation.

On peut, dès lors, songer à adapter un autre type d'architecture pour éviter au mieux ces problèmes. Les réseaux de neurones à fonction de base radiale surmontent toutes les difficultés d'apprentissage dans une architecture neuronale de type sigmoïde [38]. Cette variante de réseaux a comme particularité la généralisation locale, c'est-à-dire qu'elle ne donne des réponses utiles que pour des valeurs relativement proches du centre de fonctions, appelées fonctions noyaux. Ces fonctions sont assemblées pour une bonne couverture de leurs domaines d'approximation de l'ensemble des signaux d'entrée. Les fonctions radiales gaussiennes peuvent être définies, de manière générale, par :

$$\rho(R) = \exp\left(-\frac{1}{2} R^2\right) \quad (\text{III.5})$$

Ces fonctions sont ensuite pondérées et leurs valeurs sont sommées pour produire des valeurs en sortie comme suit :

$$\hat{x} = w\rho(t) \quad (\text{III. 6})$$

où

$$\rho(\cdot) = [\rho_1(\cdot), \rho_2(\cdot), \rho_3(\cdot), \dots, \rho_{n_{cc}}(\cdot)]^T$$

tel que:

$$\rho_j = \exp\left(-\frac{1}{2}(t - c_j)^T Q_j^{-1}(t - c_j)\right) \quad (\text{III. 7})$$

L'équation (III.7) représente la sortie du neurone j de la couche cachée (figure (III.3)) où c_j représente le vecteur des centres, Q_j est la matrice des rayons gaussiennes, exprimée par l'équation suivante :

$$Q_j = (\text{diag}[\sigma_{1,j}, \sigma_{2,j}, \sigma_{3,j}, \dots, \sigma_{n_e,j}])^2 \quad (\text{III. 8})$$

et t est le vecteur d'entrée. n_{cc} et n_e représentent respectivement le nombre de neurones dans la couche cachée et le nombre des entrées. La matrice de poids définie par w est de dimension $n_s \times n_{cc}$ où n_s est le nombre de neurones dans la couche de sortie.

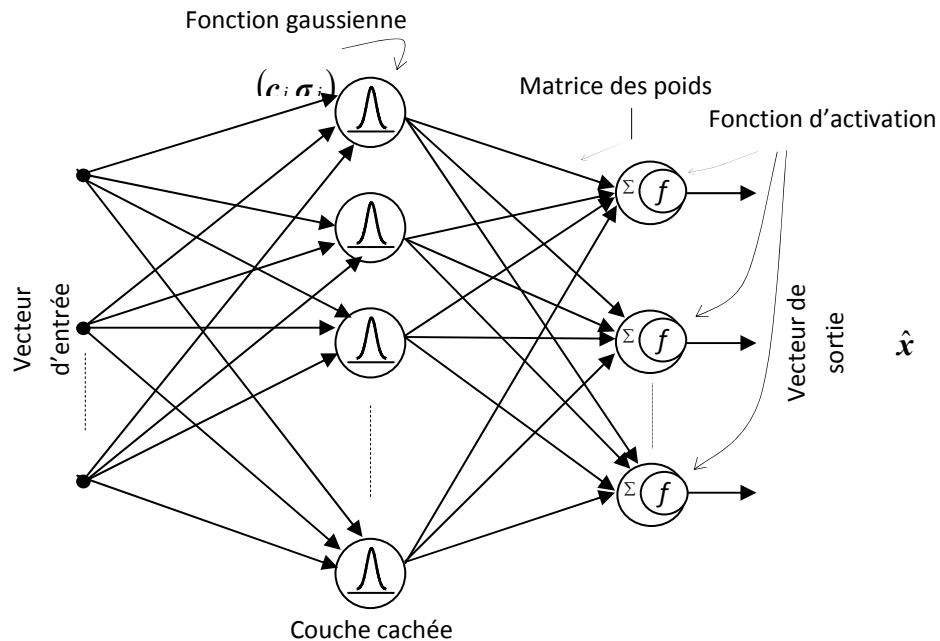


Fig. III.3 : Réseau de neurones à fonction de base radiale

6. Apprentissage du réseau de fonction à base radiale

Les réseaux de neurones réalisent des fonctions non linéaires paramétrées. Ils peuvent approcher toute fonction afin de trouver un réseau de neurones qui rend minimal la fonction définie par la somme des carrés des différences calculées par le réseau de neurones et les valeurs mesurées des données originelles. Cette fonction objective notée J est donnée par :

$$J(\mathbf{w}, \mathbf{c}, \sigma) = \frac{1}{2\xi} \sum_{i=1}^{\xi} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 \quad (\text{III. 9})$$

où ξ est le nombre d'échantillons.

La mise en œuvre d'un réseau de neurones de fonction à base radiale nécessite donc

- de déterminer les entrées représentées par les composantes principales non linéaires obtenues à partir de la projection des données initiales en utilisant la carte de Kohonen ;
- de collecter les données nécessaires à l'apprentissage et à l'évaluation des performances du réseau de neurones à la généralisation ;
- de trouver le nombre de neurones dans la couche cachée nécessaires pour obtenir une approximation satisfaisante ;
- d'effectuer un apprentissage de type supervisé.

Détermination du nombre de composantes non linéaires

Webb [39] propose d'utiliser un indice classique pour la détermination du nombre ε de composantes principales à retenir:

$$\varepsilon = \left(\frac{\|\hat{X} - X\|^2}{\|X - M\|^2} \right)^{\frac{1}{2}} \quad (\text{III. 10})$$

où \hat{X} est la matrice des données estimées par le modèle neuronale. Cet indice dépend fortement de la qualité de l'approximation du second réseau qui est utilisé pour la

reconstruction des données initiales. Dans ce cas, on augmente progressivement le nombre de composantes principales à retenir et on surveille la valeur de ε .

Pour des raisons liées à la qualité de l'apprentissage, nous proposons d'étendre les critères de choix du nombre de composantes principales à retenir, illustrée dans le chapitre 1, dans le cas non linéaire.

Collecte des données

Il faut disposer des données qui doivent être en nombre suffisant pour constituer trois bases à la fois suffisamment représentatives et distinctes, et qui seront présentées au modèle lors de son utilisation. La première base de données est utilisée pour l'apprentissage du réseau, par contre les deux autres sont collectées pour la généralisation et la validation du modèle neuronal obtenu. Dans les réseaux de neurones de fonction à base radiale, Il est possible de réaliser un pavage régulier dans tout le domaine de variation des entrées lorsque le modèle possède plusieurs entrées. Ceci se fait par fixation des centres et les écart-types des fonctions gaussiennes.

Les centres sont choisis pour représenter les zones de grande concentration des données ou de les disposer en treillis réguliers pour couvrir uniformément la partie utile de l'espace d'entrée et respecter ainsi les contraintes de couverture et de généralisation localisée. La matrice de «rayons» est choisie le plus souvent diagonale. Dans le cas où les centres sont disposés en treillis réguliers, c'est-à-dire que la distance entre deux centres voisins est constante et les valeurs de rayons sont fixes, dans le but d'assurer un certain degré de recouvrement entre les différentes gaussiennes exprimé par :

$$\tau = \exp\left(-\frac{1}{2} \frac{\left(\frac{\delta}{2}\right)^2}{\sigma^2}\right) \tag{III. 11}$$

avec δ la distance entre deux centres adjacents.

Le degré de recouvrement τ illustré par la figure (III.4) résulte d'un compromis entre les propriétés de localisation et de généralisation, l'étendu de la généralisation est donc contrôlable.

Nombre de neurones dans la couche cachée

Barron [40] a proposé un nombre de neurones cachés qui est inversement proportionnel à l'écart entre l'approximation neuronale et la fonction à approcher. Cependant, ce résultat, comme pour d'autres considérations théoriques telles que la dimension de Vapnik-Cervonenkis [41] n'est pas constructif, en ce sens qu'il ne peut que donner des estimations du nombre de neurones nécessaire dans la couche cachée. Il n'existe pas de résultat théorique permettant de prévoir le nombre de neurones cachés requis pour obtenir une performance spécifiée du modèle, compte tenu des données disponibles. Nous tentons dans cette thèse de considérer dans la partie reconstruction des données initiales, un réseau à treillis régulier dont les centres sont uniformément disposés dans les domaines d'approximation. Le nombre n des centres et le nombre n_e des entrées du second réseau définissent son architecture qui est de n^{n_e} neurones cachés.

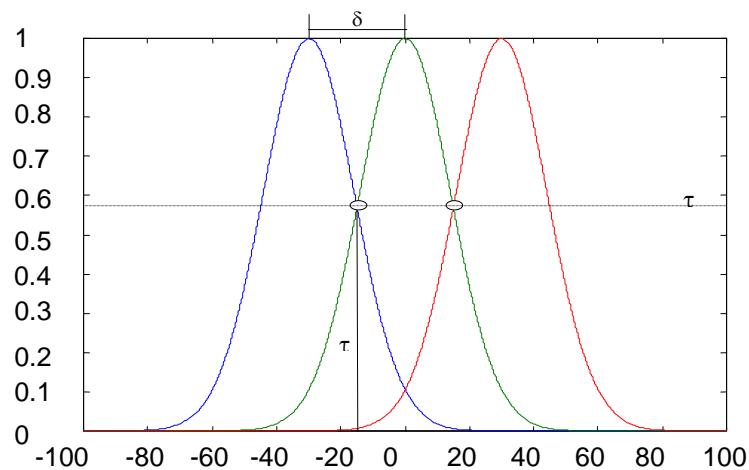


Fig. III.4 : Taux de recouvrement pour un treillis régulier unidimensionnel.

Algorithme d'apprentissage

Une fois le nombre de neurones cachés déterminé, on doit procéder à l'apprentissage. C'est l'estimation des paramètres de réseau qui permet d'obtenir une valeur minimale de la fonctionnelle définie par l'équation (III. 9). Elle représente une fonction de tous les paramètres ajustables w , c , et σ de tous les neurones et de toutes les connexions du second réseau.

La plupart des applications utilisant les fonctions à base radiale disposent de centres et de rayons fixés une fois pour toutes. Dans ce cas où la sortie du réseau est linéaire par rapport à ces paramètres, il est tout à fait concevable d'utiliser les techniques d'optimisation linéaire classique pour obtenir un apprentissage rapide et précis. La matrice des poids est obtenue en utilisant la technique de régression linéaire par

$$w = X \rho \text{inv}(\rho^T \rho) \quad (\text{III. 12})$$

En outre, la disposition des centres a priori s'avère quelques fois non adaptée. En effet, le nombre de centre requis est parfois trop élevé [42,43,44,45].

Le recours aux réseaux à fonction de base radiale modifiés appelés aussi réseaux à base radiale généralisés s'avère nécessaire. Ils permettent non seulement l'apprentissage des poids mais aussi celui des centres et des rayons. Par conséquent, la sortie n'est plus linéaire par rapport aux paramètres, cette caractéristique rend l'utilisation des algorithmes d'optimisation linéaire inutiles. L'ajustement effectué par une descente du gradient sur les positions des centres ainsi que les valeurs des rayons associés, risque de tomber dans des minima locaux. L'introduction des connaissances a priori sur la distribution initiale des centres, rayons et des poids permet de disposer des valeurs trop éloignées de celles optimales.

Comme alternative, nous proposons d'utiliser une méthode de recherche directe basée sur la division continue de l'espace de recherche dont le centre est le candidat du minimum de la fonction coût. Cette technique appelée DIRECT est appliquée sur les centres et les rayons des fonctions de base radiales gaussienne du second réseau.

7. Principe d'optimisation par application de l'algorithme DIRECT

L'algorithme d'optimisation globale DIRECT est une méthode déterministe d'échantillonnage a été introduit pour la première fois par [46]. Par fractionnement continu de l'espace de recherche et l'évaluation de la fonction objectif, est exprimé le processus itératif de l'optimisation de la fonction coût (III. 9) ; aucune information secondaire n'est nécessaire. L'algorithme DIRECT est une méthode de recherche par motifs (au sens de Torczon [47]) qui équilibre la recherche locale et globale dans l'intention de trouver efficacement un optimiseur global. D'autres méthodes déterministes d'échantillonnage, telles que le Filtrage Implicite [48], la Recherche Multidirectionnelle (MDS), les méthodes de Hooke-Jeeves [49], ou Nelder-Mead [50], conduisent le gradient à tendre vers zéro et (ou) ne sont pas destinées pour une recherche globale. La technique DIRECT est destinée à explorer complètement l'espace des variables, même après qu'un ou plusieurs minima locaux soient été identifiés.

La caractéristique remarquable de cette approche est en ce sens qu'il utilise la constante de Lipschitz. À chaque itération, les régions de l'espace des solutions doivent être explorées. Il n'y a cependant pas de test de convergence automatique proposé pour DIRECT. Plutôt il est suggéré que l'algorithme sera simplement exécuté pour un nombre d'itérations fixé à l'avance.

Un algorithme qui met trop l'accent sur la recherche locale sera facilement pris au piège des optima locaux. Inversement, un algorithme qui passe trop de temps à effectuer une recherche globale convergent très lentement. L'ajustement des paramètres que cet équilibre contrôle lui-même est souvent un problème d'optimisation non linéaire. Par conséquent, une technique qui ne dépend pas de tels paramètres serait souhaitable. Dans cette section, nous présentons une méthode lipschitzienne modifiée appelé DIRECT, qui peut trouver un optimum global d'une fonction à valeurs réelles de n'importe quelle dimension. Il s'agit d'une méthode de recherche directe qui est extrêmement robuste comparée aux autres méthodes déterministes, car elle fonctionne avec les évaluations de fonction et elle n'a pas besoin d'autres informations sur la fonction objectif. Elle est également 'parallélisée' dans le sens où

un grand nombre d'appels à la fonction objectif ne peut être distribué simultanément à de nombreux transformateurs. Toutefois, DIRECT exige parfois un grand nombre d'évaluations de fonctions pour trouver un minimum global, en particulier si un minimum se trouve aux frontières. Cela conduit à ralentir la convergence. DIRECT, combiné avec d'autres algorithmes aidera à surmonter cette situation.

L'algorithme DIRECT est initialisé en normalisant l'espace de recherche du minimum de la fonction coût (III. 9) à un hyper-rectangle unité. La fonction objective est évaluée au point central de cet hyper-rectangle, qui est identifié comme le premier hyper rectangle potentiellement optimale. L'algorithme DIRECT entame le processus de recherche en évaluant la fonction coût dans toutes les directions en des points qui sont définis comme égaux à $m \pm \delta e_i$ et qui sont équidistants du centre de coordonnées m . où δ représente le tiers de la longueur de l'hyper-rectangle dans la direction de i -ème vecteur unité e_i . Après avoir sélectionné l'hyper-rectangle potentiellement optimal, l'algorithme divise de la même manière ce nouvel espace à l'itération qui suit. La procédure de partage est effectuée par trisection dans toutes les directions. La trisection est basée sur les sens avec la plus petite valeur de la fonction objective. Une interprétation des trois premières itérations de l'algorithme est représentée dans la figure (III.5) pour un problème d'optimisation bidimensionnel.

Les points d'évaluations sont indiqués en rouge et noir, tandis que les points qui ont été préalablement évalués sont représentés par des cercles. Les zones ombrées sont les rectangles potentiellement optimaux.

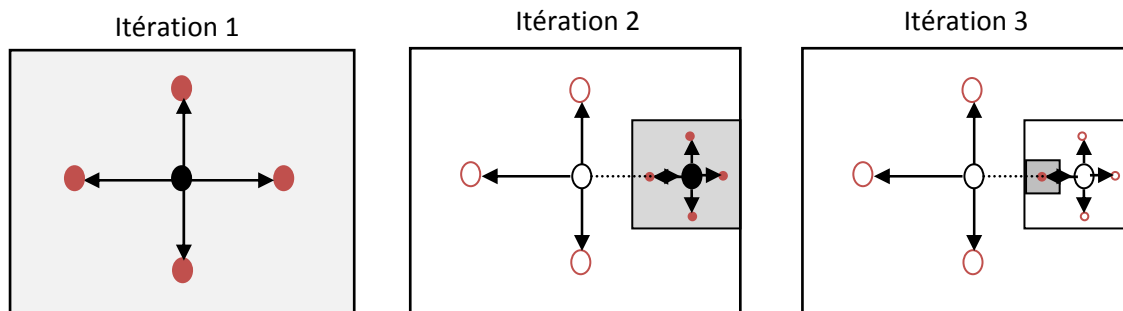


Fig. III.5 : Illustration des trois premières itérations de l'algorithme DIRECT.

Définition : En supposant que l'hyper rectangle unité de centre \mathbf{m}_i divisé en k hyper-rectangles, l'hyper-rectangle j est dit potentiellement optimal, s'il existe un certain taux de changement constant \tilde{K} , telle que

$$J(\mathbf{m}_j) - \tilde{K}d_j \leq J(\mathbf{m}_i) - \tilde{K}d_i, \quad \text{pour } i = 1, 2, 3, \dots, k \quad (\text{III. 12})$$

$$J(\mathbf{m}_j) - \tilde{K}d_j \leq J_{\min} - \varepsilon |J_{\min}| \quad (\text{III. 13})$$

où J_{\min} est la valeur minimale, de la fonction coût, située dans le bassin d'attraction du minimum global, d_j est la distance entre le centre et les frontières des hyper-rectangles, et le paramètre ε est utilisé ici pour protéger l'algorithme contre les 'préjugés' locaux excessifs dans la recherche. Les hyper-rectangles potentiellement optimaux sont ceux définissant le fond de la coque convexe d'un nuage de points de diamètre $J(\mathbf{m}_i)$ pour les centres \mathbf{m}_i de tous les hyper-rectangles, (figure (III.6)). La condition (III.12) exprime que l'intervalle sélectionné est situé sur l'enveloppe convexe inférieure droite des points (figure (III.6)), alors que la condition (III.13) force la borne inférieure de l'intervalle, à ne pas dépasser la meilleure solution courante par une quantité non triviale, ou en d'autres termes, assure la possibilité d'avoir une diminution suffisante dans l'espace de recherche. Le paramètre ε est positif et d'équilibre entre les recherches locale et globale. Dans certaines applications ε est compris entre 10^{-3} et 10^{-7} , [51,52,53]. Les espaces potentiellement optimaux sont situés sur l'enveloppe convexe inférieure droite. Chaque point dans le graphe de la figure (III.6) représente la valeur de la fonction coût au centre d'un ensemble d'intervalles ayant la même longueur.

L'hyper-rectangle i sera sélectionné si les pentes des courbes sur la droite de P sont supérieures à la ligne passant par P et J_{\min} . Cela permet à la convergence prématurée de ne produire. Une définition analogue pour mieux comprendre comment sélectionner les intervalles pour une éventuelle subdivision a été proposée dans les travaux de Gablonsky, [54].

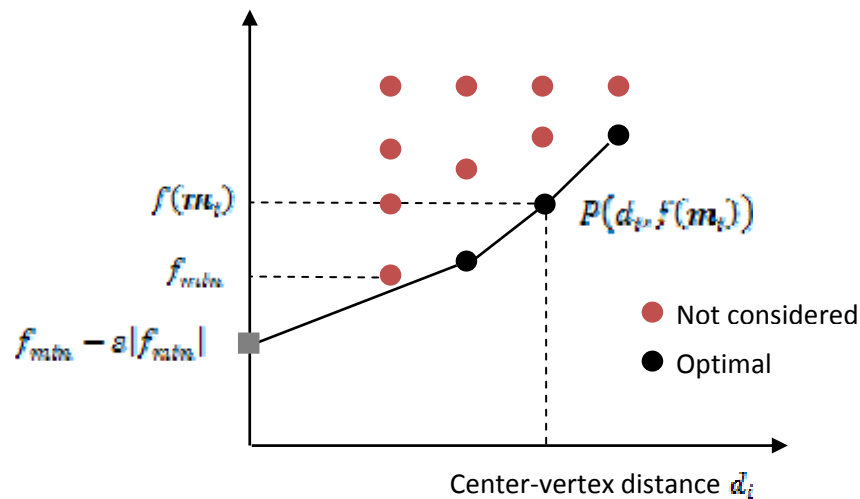


Fig. III.6 : Evaluations de $J(m_i)$ en fonction de la taille des hyper rectangles d_i .

Pour une meilleure compréhension de cette technique d'optimisation, nous illustrons l'algorithme multidimensionnel standard de DIRECT.

Algorithme du DIRECT standard

1. Normaliser l'espace de recherche à un hyper-rectangle unité de centre m_1 ,
2. Calculer $J(m_1)$.
3. Identifier l'ensemble S des hyper-rectangles potentiellement optimaux,
4. Choisir un hyper-rectangle quelconque $hr \in S$.
5. Pour l'hyper-rectangle hr :
 - 5a. Identifier l'ensemble I des dimensions dont la longueur du côté est maximale en utilisant le compteur i . Soit δ égale au tiers de la longueur maximale du côté.
 - 5b. Echantillonner l'hyper-rectangle contenant m aux points $m \pm \delta e_i$ pour tout $i \in I$ et diviser en tiers le long des dimensions dans I , où m est le centre d'hyper-rectangle hr et e_i représente le i -ème vecteur unité.
6. Mettre à jour S . Poser $S = S - \{hr\}$.
Si S est non vide, aller à l'étape 4.

Sinon aller à l'étape 7.

7. Itérer. Reporter les résultats de cette itération, et revenir à l'étape 3. Si l'itération se limite, aller à l'étape 8.
8. Si l'optimisation est complète, reporter les paramètres recherchés en valeurs réelles et la valeur de la fonctionnelle J_{\min} .

9. Conclusion

Dans ce chapitre, une nouvelle structure pour le modèle ACPNL a été introduite. Une carte auto-organisatrice de Kohonen représente une application de l'espace d'observation vers un espace de dimension réduit. Chaque vecteur de l'espace d'observation est un codage particulier des mesures issues d'une matrice de données. Ce codage permet de définir les vecteurs poids référents. La connaissance des poids référents détermine une quantification vectorielle de la projection. La détermination de ces poids projecteurs se fait par minimisation d'une fonction de coût. Cette fonction coût est l'objet de la maximisation du produit scalaire ou minimisation d'une distance euclidienne entre les observations et les poids référents. Un bon optimum s'obtient après un nombre d'itérations suffisamment grand de l'algorithme du gradient à pas variable pour la modification des poids de la carte de Kohonen. Les poids référents obtenus seront employés pour la détermination des composantes principales.

Le second réseau définit la transformation inverse permettant de calculer l'estimation des données initiales à partir des composantes principales obtenues par la carte de Kohonen. Nous avons exposé les éléments essentiels qui permettent de comprendre pourquoi et dans quel cas il est avantageux de mettre en œuvre un réseau de neurones à fonction de base radiale. La spécificité de cet approximateur universel réside précisément dans le fait qu'il nécessite moins de paramètres ajustables par rapport aux réseaux perceptrons multicouches. Il est sans doute utile de rappeler que les réseaux de neurones sont utilisés comme outils de modélisation par apprentissage. L'apprentissage est réalisé par des algorithmes d'optimisation dans le but est de minimiser la fonction de coût qui définit l'écart entre les sorties du réseau et les sorties réelles, en modifiant les paramètres d'une manière itérative.

Un algorithme d'optimisation global a été introduit. Cet algorithme appelé DIRECT est une méthode déterministe d'échantillonnage par fractionnement continu de l'espace de recherche. Elle est basée sur l'évaluation de la fonction coût dont aucune information supplémentaire n'est nécessaire.

Sur le plan théorique, nous avons présenté une nouvelle structure du modèle ACPNL et l'algorithme standard de DIRECT dans un cadre aussi général que possible. Une étude consacrée à la détection et à la localisation de défauts dans un système à trois réservoirs

interconnectés avec une application des concepts et des techniques introduits dans cette partie, fera l'objet du chapitre 4.

1. Présentation du système

La maquette de la figure (IV.1) est constituée de trois cuves cylindriques, de hauteur maximale h_{\max} , de section S , couplées par des vannes de transfert et qui peuvent être vidangées dans une cuve de rétention par des vannes de fuites. Ces dernières ont des sections efficaces identiques S_p . Les cuves 1 et 2 peuvent être alimentées en fluide par l'intermédiaire de deux pompes de débit maximale q_{\max} . Trois capteurs piézo-résistifs de la pression différentielle permettent de mesurer la hauteur du liquide présent dans chacune des cuves et deux autres capteurs de débit sont utilisés pour mesurer le débit de refoulement des deux pompes.

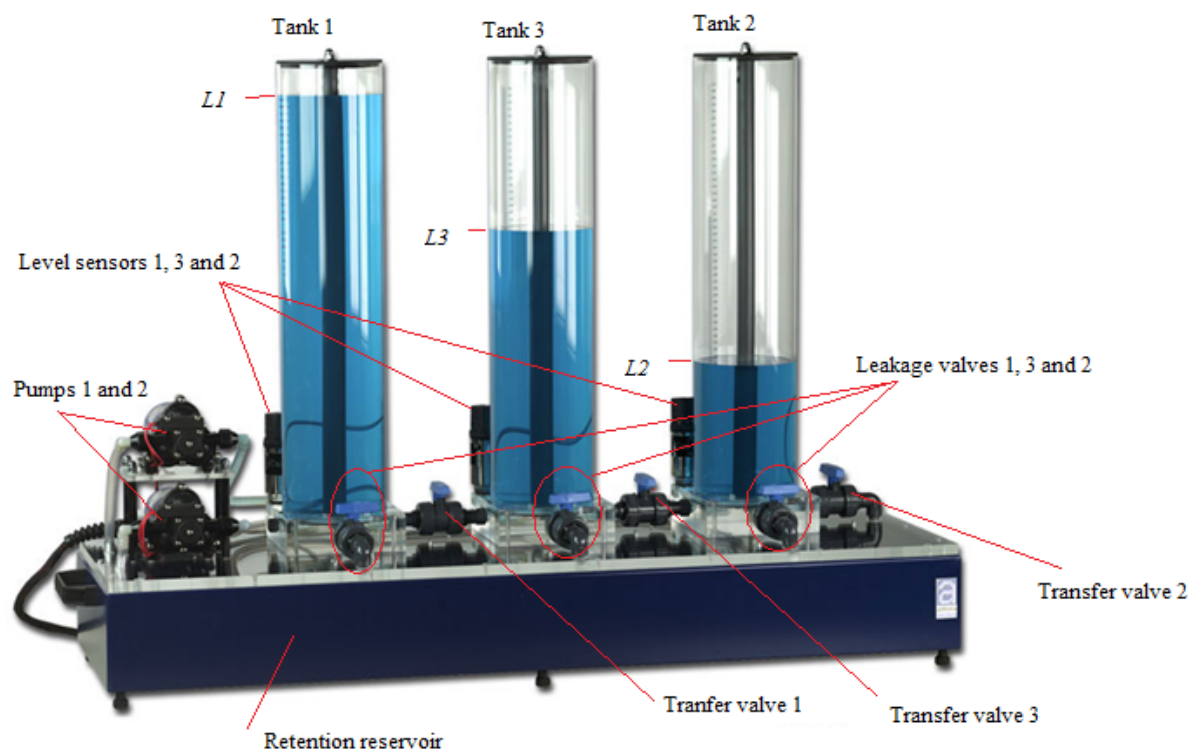


Fig. IV.1 : Stand expérimental de trois réservoirs interconnectés

Deux correcteurs décentralisés de type proportionnel et intégrale (PI) sont mis en œuvre sous l'environnement de programmation Matlab à partir d'une carte d'acquisition NIDAQ 6025E afin de maintenir les niveaux h_1 et h_2 dans les réservoirs 1 et 2 à leurs points de consigne au moyen de deux pompes à débit notées q_1 et q_2 . Les correcteurs sont également introduits pour fournir une performance stable et acceptable et pour rejeter les perturbations. Le niveau h_3 dans le réservoir 3 est une variable non commandable. Plus de détails sur le système de trois réservoirs interconnectés DTS-200 comprenant le montage, la mise en œuvre, et les données techniques peuvent être trouvés dans le manuel [56]

Les mesures de niveau et de débit sont exprimées en volts. Le temps de simulation considéré dans le présent travail est de 400 secondes. La période d'échantillonnage est choisie égale au cinquième de la seconde pour produire en fin d'expérience un ensemble de données de 2000 échantillons. Le tableau (IV.1) regroupe les valeurs de tous les paramètres du système ainsi que les valeurs nominales du mode de fonctionnement à l'état sain.

q_{10}	Débit de refoulement de la pompe 1	Valeur nominale	-1.25 (V)
q_{20}	Débit de refoulement de la pompe 2	Valeur nominale	-1.06 (V)
h_{10}	Hauteur du fluide dans le réservoir 1	Valeur nominale	0.42 (V)
h_{20}	Hauteur du fluide dans le réservoir 2	Valeur nominale	2.65 (V)
h_{30}	Hauteur du fluide dans le réservoir 3	Valeur nominale	1.6 (V)
S	Section du réservoir 1, 2 et 3	Constante	0.0149 (m ²)
S_p	Section effective des vannes de transfert et de fuite	Constante	5.10 ⁻⁵ (m ²)

Tableau IV.1 : Différentes paramètres techniques du système à trois réservoirs DTS-200

2. Extraction des composantes principales

Les données de 2000 échantillons de chaque signal sont organisées dans une matrice X . Celle-ci contient donc cinq signaux, h_1, h_2, h_3, q_1 et q_2 correspondant, respectivement, aux niveaux dans les trois réservoirs et les débit des deux pompes d'alimentation en eau.

Un prétraitement a tout d'abord été effectué, par le calcul de la matrice centrée et réduite $X^{(n)}$, en utilisant les équations (I.2), (I.3), (I.4), (I.5) et (I.6).

L'algorithme proposé et utilisé dans cette étude pour l'extraction des composantes principales est l'algorithme des cartes auto-organisatrices de Kohonen. Cet algorithme projette l'espace de données de dimension 5 sur un espace de plus faible dimension. Cet espace, qu'on appelle carte, est constitué de 5 neurones numérotés de 1 à 5 et disposés topologiquement en une seule dimension. La structure graphique de cette carte est illustrée sur la figure (IV.2).

L'algorithme procède d'une manière itérative pour déterminer l'ensemble des vecteurs référents, en optimisant l'une des fonctions coût qui sont décrites par les équations (III.1) et (III.2). L'optimisation s'obtient en affectant chaque vecteur d'observation $x_i^{(n)}$ au référent $w_j(i)$ qui représente le vecteur poids correspondant au j -ème neurone de la carte de Kohonen associée.

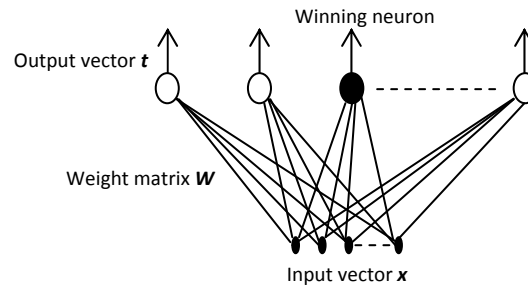


Fig. IV.2 : Architecture de la carte d'auto-organisation de Kohonen unidimensionnelle

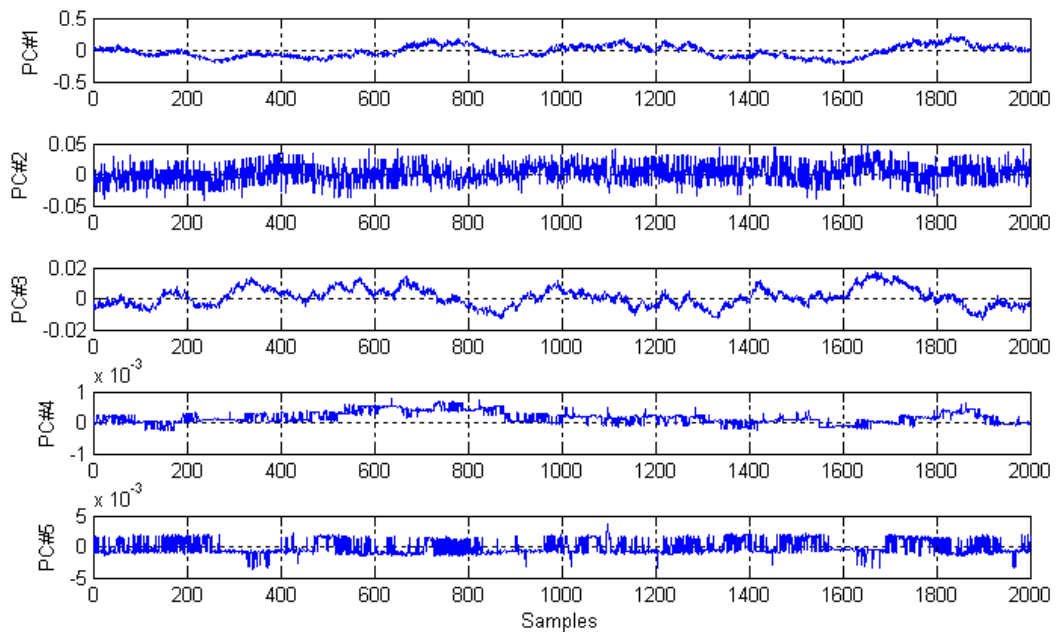


Fig. IV.3 : Différentes Composantes Principales

Cette technique peut être expliquée par analogie avec l'ACP linéaire. Cette dernière est basée sur le développement statistique des données, en formant une matrice de covariance et en calculant les composantes principales. Ceci fournit des valeurs et vecteurs propres classés des plus importants vers les moins significatifs. Le vecteur propre correspondant à la plus grande valeur propre représente la combinaison des signaux qui sont les plus significatifs dans l'ensemble des données. La prochaine valeur propre représentera la classe suivante la plus significative et ainsi de suite. Ceux-ci se traduisent dans la carte de Kohonen par la recherche d'un vecteur poids qui minimise la distance euclidienne avec le vecteur de données. Pour obtenir ce vecteur, Nous avons opté pour la méthode du gradient à pas optimal qui est une fonction du gradient instantané et évalué à chaque itération. Un nombre d'itérations de

1000 est fixé a priori pour permettre de placer les vecteurs référents du neurone gagnant au centre de gravité de l'ensemble des signaux du système à trois réservoirs interconnectés. Les prochains vecteurs poids référents sont déterminés à partir de l'introduction d'une contrainte supplémentaire liée à la conservation de la topologie et de l'imposition de la notion de voisinage. Deux vecteurs poids référents sont proches par rapport à la distance euclidienne sur l'espace de données. Il est possible, en choisissant une fonction noyau positive et symétrique, de considérer que la zone d'influence devienne progressivement décroissante avec la distance de voisinage et le nombre d'itérations.

Les composantes principales sont calculées à partir de l'équation (III.1), leur évolution est illustrée sur la figure (IV.3). On remarque que les composantes t_4 et t_5 ne représentent que du bruit alors que les trois premières composantes sont porteuses d'information et sont fortement corrélées avec les signaux originaux.

Cependant, pour l'estimation des signaux originaux, on ne doit garder que les composantes porteuses d'information significative permettant d'expliquer les différentes variables. Pour cette raison, Nous avons utilisé le critère de la variance cumulée pour définir la structure de modèle ACPNL. Le tableau (IV.2) représente la variance, la variance expliquée en pourcentage et celle cumulée des différentes composantes principales.

Ainsi, pour notre application, un modèle ACPNL à trois composantes principales a été retenu ce qui explique 95.29 % des corrélations entre les variables.

Composante principale	Variance	Variance expliquée (%)	Variance cumulée (%)
01	04.58	44.95	44.95
02	03.32	32.58	77.53
03	01.81	17.76	95.29
04	0.36	03.53	98.82
05	0.12	01.18	100

Tableau IV.2 : Différentes variances de l'ensemble des composantes principales

3. Reconstruction de données fondamentales

L'identification du modèle ACPNL se fait par la détermination de la structure de ce modèle d'une part et, l'estimation des données initiales d'autre part. Cette estimation est calculée à partir des composantes principales obtenues par la carte de Kohonen et retenu dans le modèle ACPNL. Ces composantes sont injectées dans un réseau de neurones de fonction à base radiale. La reconstruction des données de départ se fait par modification des centres, des rayons et des poids synaptiques du réseau de neurones tout en minimisant la fonction de coût décrite par l'équation (III.9). Deux bases de données ont été créées. Les premiers 1000 échantillons de l'ensemble de composantes principales retenues sont utilisés dans la phase d'apprentissage. La validation/généralisation du réseau de neurones de fonction à base radiale obtenu se fait par les derniers échantillons.

Avant de décrire la stratégie de l'apprentissage plus en détail, nous soulignons que l'algorithme DIRECT modifié, est utilisé pour la recherche des paramètres d'un réseau à topologie optimale fixe. Cette dernière est une description architecturale du réseau spécifiant le nombre de neurones dans la couche cachée; la technique essai/erreur a été alors employée

pour faire évoluer favorablement une population de topologie. Le modèle approché, que nous avons adopté dans cette application, consiste en un réseau de neurones à base radiale caractérisé par un nombre de centres égale à trois et de rayons vectoriels trois. Les poids du réseau qui ont un nombre 40 font aussi partie de l'ensemble de paramètres recherchés. Nous avons considéré dans cette partie identification du modèle ACPNL un réseau de 8 neurones dans la couche cachée. Toutes les combinaisons possibles de la disposition des centres dans le domaine d'approximation ainsi que le nombre des entrées du réseau définissent l'architecture neuronale. Les connexions de moindre influence sur la valeur de la fonction coût ont été éliminées.

Cependant, une modification majeure a été apportée dans le fonctionnement de l'algorithme DIRECT. Nous avons introduit un facteur de régularisation dans le but d'adapter la longueur des côtés du l'espace de recherche. La figure (IV.4) illustre un espace de recherche à deux dimensions et dont les propriétés de l'algorithme standard de la méthode DIRECT restent conservées. L'intérêt principal de cette modification est d'accélérer la convergence de l'algorithme d'optimisation. De plus, la division suivant une base aléatoire permet de surmonter le problème de balayage de l'espace de recherche, qui est partitionné sur une base euclidienne dans l'algorithme standard de DIRECT, dans des différentes directions sans augmentation du nombre de candidat dans l'algorithme. Soit m le centre de l'hyper rectangle, les points de la division proposée sont données par :

$$m \pm \alpha \delta e_i \tag{IV.1}$$

où α est choisi dans l'intervalle $[0 \ 1]$. δ équivaut au tiers de la longueur de l'hyper-rectangle correspond à l'espace de recherche normalisé et e_i est le i -ème vecteur dont les éléments sont choisis aléatoirement dans l'intervalle $[0 \ 1]$. A chaque itération, Nous obtenons la valeur minimale de la fonction coût en comparant l'évaluation de cette fonction à chaque point des partitions comme suit :

$$J_{\min,k} = \min\{J(\mathbf{m}_{1r,k}), J(\mathbf{m}_{1r,k} + \alpha\delta e_{1r,k}), J(\mathbf{m}_{1r,k} - \alpha\delta e_{1r,k})\} \quad (\text{IV.2})$$

où $J(\cdot)$ est la fonction objective définie par l'équation (III.9) à l'itération k . Elle est évaluée à tous les r partitions de l'espace de recherche le long de l'ensemble des directions définies par la matrice $e_{1r,k}$.

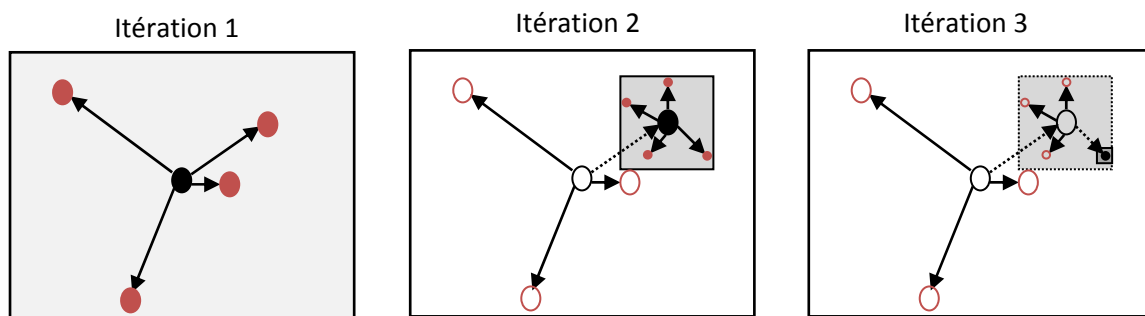


Fig. IV.4 : Illustration des trois premières itérations de l'algorithme DIRECT modifié.

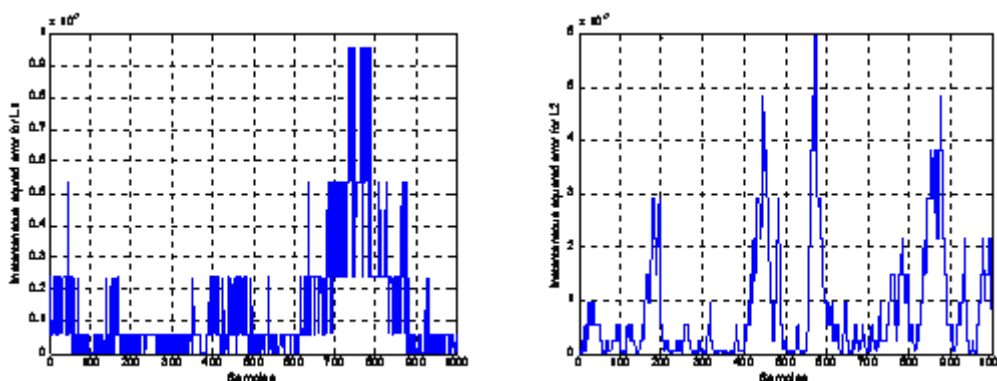
Figure (IV.5) montre l'erreur d'approximation quadratique instantanée en utilisant le modèle neuronal obtenu après apprentissage par l'algorithme DIRECT modifié. Le processus itératif s'est arrêté après 10 itérations dans la fonction coût (III.9) a été évaluée 176 fois. Cette fonction coût, qui exprime la différence quadratique entre les signaux réels et celui approchés par le réseau de neurones de fonction à base radiale, est égale à $2.3872e-005$. En examinant les courbes de la figure (IV.5), On peut constater que les niveaux prévus correspondent assez bien aux niveaux désirés et générés par le système à trois réservoirs ainsi que les débits des deux pompes. Les différences entre ces signaux sont, en général, inférieures à $3e-5$ Volts.

D'après les résultats obtenus, il apparaît clairement que les caractéristiques du système à trois réservoirs DTS-200 sont parfaitement reproduites par le modèle neuronal obtenu après apprentissage.

En fin d'apprentissage, on utilise le réseau de neurones trouvé pour estimer les sorties la matrice des données initiales à lequel la deuxième base de données, qui n'a pas été présentée lors de l'apprentissage, figure sur les entrées du réseau. Cette phase est utilisée pour examiner l'efficacité et l'aptitude du modèle neuronal à l'identification du modèle ACPNL.

Les résultats de validation du modèle neuronal sont illustrés par la figure (IV.6). Cette figure montre l'erreur quadratique instantanée de validation. On peut constater que les niveaux et les débits prédits par le modèle du réseau de neurones de fonction à base radiale sont assez proches aux signaux réels correspondants. Il faut aussi souligner la performance de couverture de l'ensemble de la région d'approximation des données exprimée par une valeur maximale de l'erreur quadratique instantanée égale à $4e-5$ dans l'ensemble des signaux du système à trois réservoirs. Cependant, l'erreur quadratique moyenne calculée en utilisant l'équation (III.9) vaut $3.7893e-5$.

L'ensemble des résultats, ainsi obtenus, ne peut que confirmer l'exactitude relative du modèle ACPNL neuronal avec trois composantes principales retenues à simuler toutes les caractéristiques dynamiques du système à trois réservoirs DTS-200 dans un environnement complexe et variable. Cette précision dans l'approximation repose principalement sur la qualité d'apprentissage par la technique de DIRECT modifié.



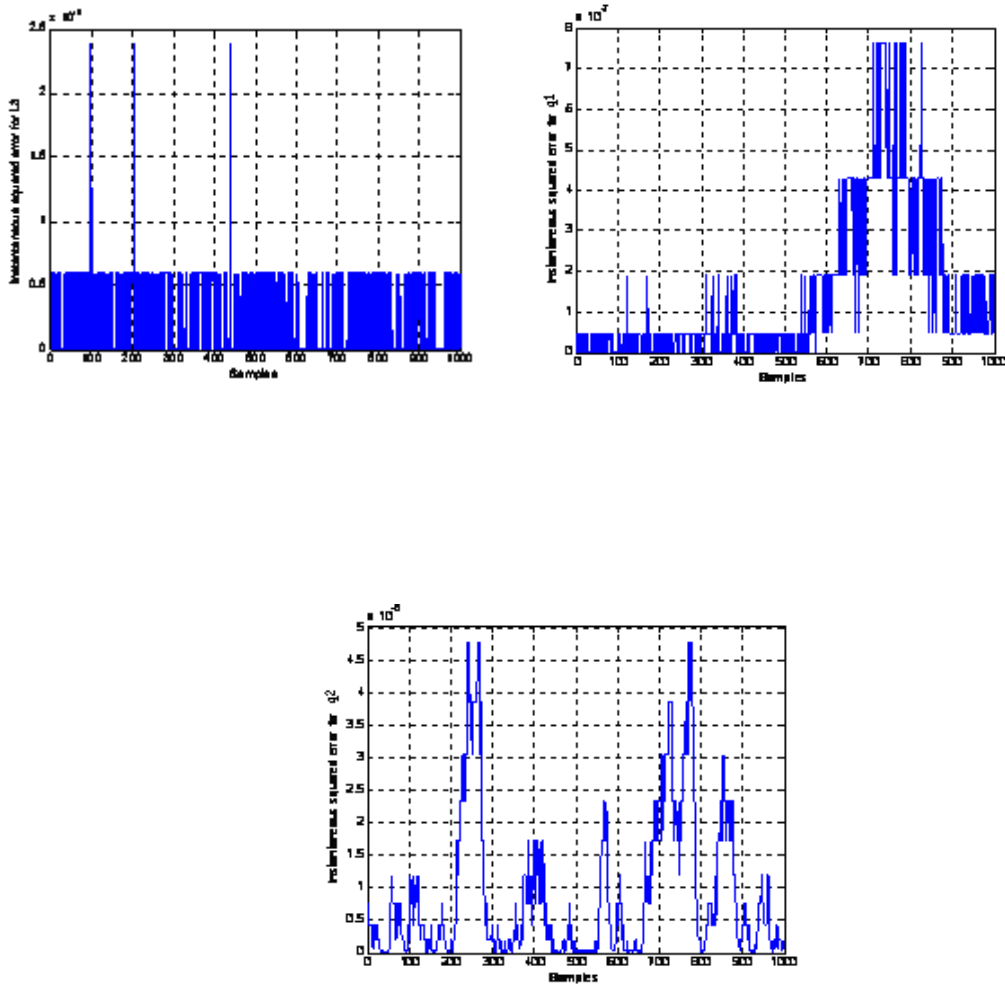
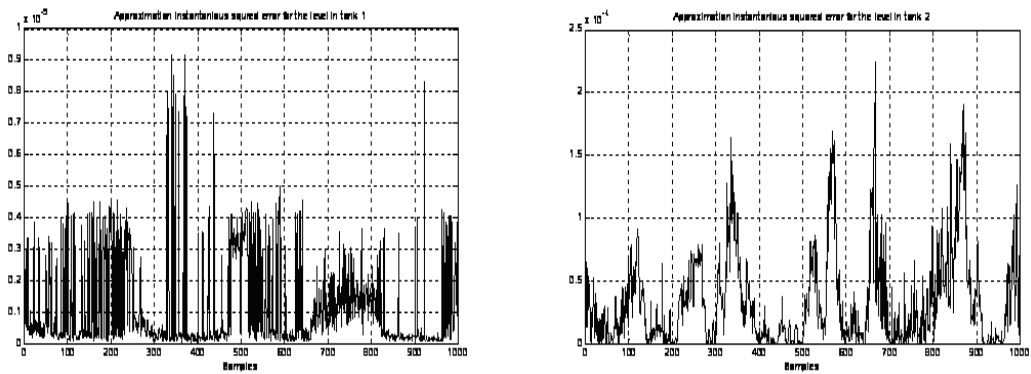


Fig. IV.5: Erreur quadratique instantanée d'estimation pour différentes variables du système dans la phase d'apprentissage



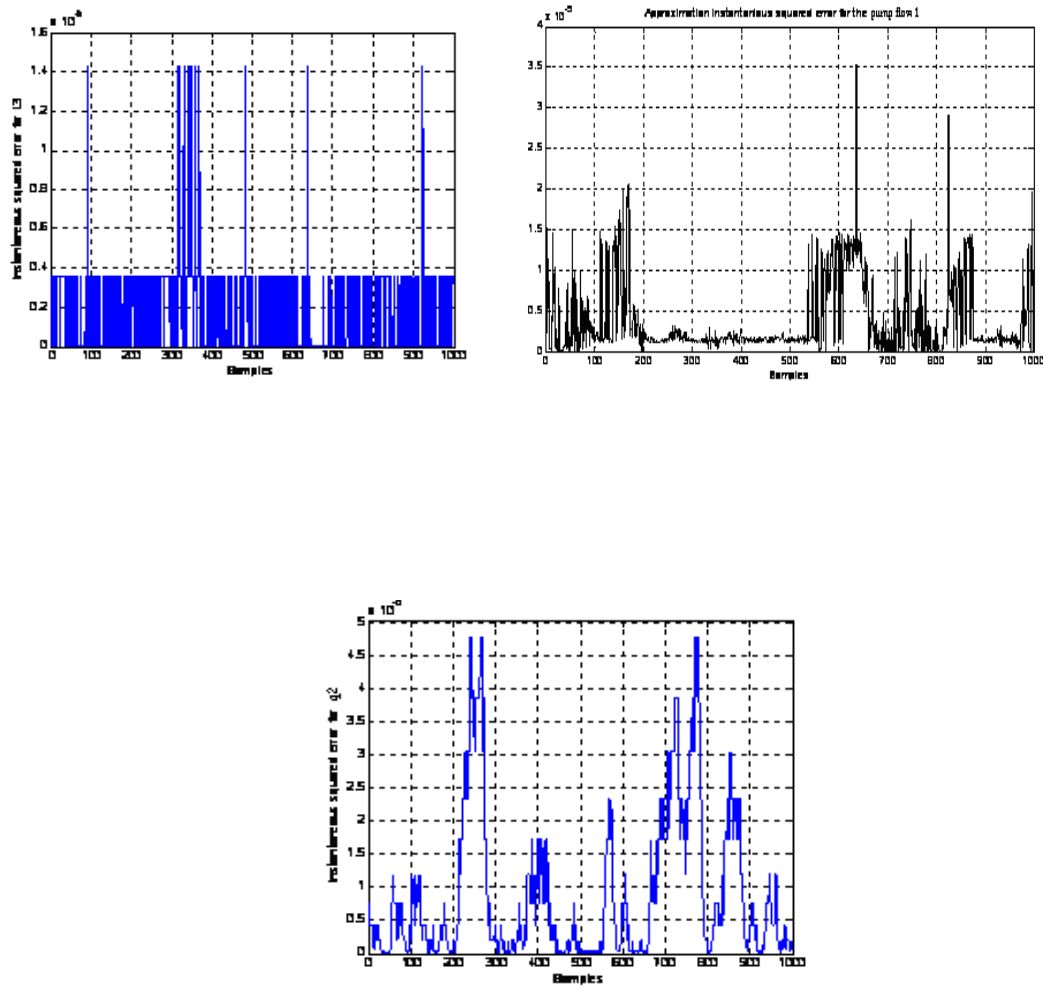


Fig. IV.6: Erreur quadratique instantanée d'estimation pour différentes variables du système dans la phase de validation

4. Détection de défauts

Avant de simuler les divers défauts, intéressons nous au seuil de détection. Le seuil calculé d'après la formule (I.19) est égale à $2.3958e-4$. Cependant, le seuil ainsi déterminé ne permet pas de procéder à une détection fiable et précise. Nous avons donc élaborés une technique assez robuste pour augmenter la sensibilité de l'indice de détection, qui est basé sur l'erreur quadratique d'estimation EQE donnée par l'équation (I.17), à détecter le défauts et d'annuler dans le plus possible les fausses alarmes. Cette technique repose sur une analyse de la distribution du signal EQE ; il s'agit donc d'estimer sa loi de distribution. Pour se faire, deux hypothèses peuvent être envisagées :

H_0 : EQE suit la loi normale $N(m_{EQE}, \sigma_{EQE})$;

H_1 : EQE ne suit pas la loi normale $N(m_{EQE}, \sigma_{EQE})$;

Un test statistique s'avère nécessaire qui permettra de trancher entre ces deux hypothèses au vu des résultats d' EQE dont une et une seule est vraie. La forme de l'histogramme de l' EQE sur la figure (IV.7) peut conduire à éliminer certains modèles. La symétrie constatée sur cet histogramme mène le plus souvent à poser l'hypothèse de normalité de l' EQE . À un niveau de probabilité $\alpha = 5\%$ l'hypothèse H_1 pourrait être acceptée si le résultat obtenu faisait partie d'une éventualité improbable qui n'avait que 5 chances sur 100 de se produire, nous avons utilisés dans ce travail le test non paramétrique de Kolmogorov-Smirnov. Ce test d'ajustement à une distribution entièrement spécifiée de fonction de répartition confirme l'acceptance de l'hypothèse de la normalité de l' EQE à un niveau de vérité égale à 95 %.

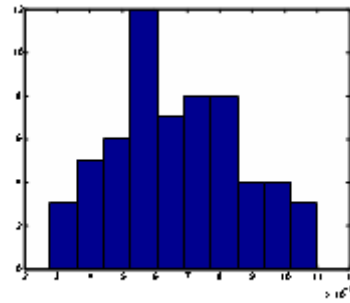


Fig. IV.7 : Histogramme du signal EQE

L'évaluation des propriétés statistiques du signal EQE est nécessaire pour la détermination du seuil de confiance qui sera utilisé pour le besoin de détection de défauts. Ce seuil est calculé dans notre cas à un niveau de signification égal à 0.95. En terme probabiliste, cela revient à déterminer ε tel que :

$$P(EQE < \varepsilon) \geq 0.95 \quad (\text{IV.3})$$

La loi de la variable réduite peut être approchée par la loi normale $N(0,1)$ où le seuil de confiance pour EQE à un niveau de signification $(1 - \alpha)$ où α représente le risque d'erreur encouru en affirmant que EQE est bien situé en dessous du seuil considéré ; ceci vérifie la condition de confiance de l'équation (IV.3) est alors donné par :

$$SC_{\alpha} = m_{EQE} + \varepsilon \quad (\text{IV.4})$$

avec

$$\varepsilon = 1.96\sigma_{EQE} \quad (\text{IV.5})$$

Un défaut quelconque serait mis en évidence si le signal EQE dépasse le seuil de confiance adopté.

Pour mettre en évidence l'intérêt du formalisme proposé dans la détection de défauts, nous avons simulé trois cas de défauts sur le système expérimental. Plus précisément, il s'agit d'un défaut d'actionneur et de deux défauts de capteurs, comme mentionné ci-après :

- **Cas 1** : Un défaut de 10% sur le capteur de niveau 1
- **Cas 2** : Un défaut de 10% sur la pompe de refoulement 1
- **Cas 3** : Un défaut de 10% sur le capteur de niveau 2

La figure (IV.8) illustre la variation du niveau d'eau h_1 dans le réservoir 1 lorsqu'un défaut de 10% est supposé survenir sur le capteur de niveau 1 à l'échantillon $k = 650$ (cas 1). Une telle situation fait agir le contrôleur de niveau qui intervient sur la pompe 1 pour rattraper le point de consigne, ce qui revient en fait à compenser l'effet du défaut sur le système ; chose qui rend difficile sa mise en évidence. En analysant le signal EQE indiqué sur la figure (IV.9), il a été possible de remarquer le changement dans l'allure de ce dernier et il se trouve bien en dessus de seuil de confiance établi en fonctionnement normal, ce qui traduit la présence du défaut. Cette variation du signal EQE est due principalement au changement remarquable dans les signaux du niveau h_1 et du débit q_1 . De plus, les signaux h_3 et q_2 sont aussi affectés en raison des interactions bien connues dans ce type de procédé.

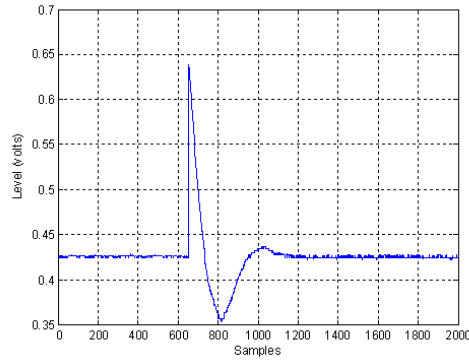


Fig. IV.8 : Evolution de niveau dans le réservoir obtenue avant et après l'occurrence du défaut (cas 1).

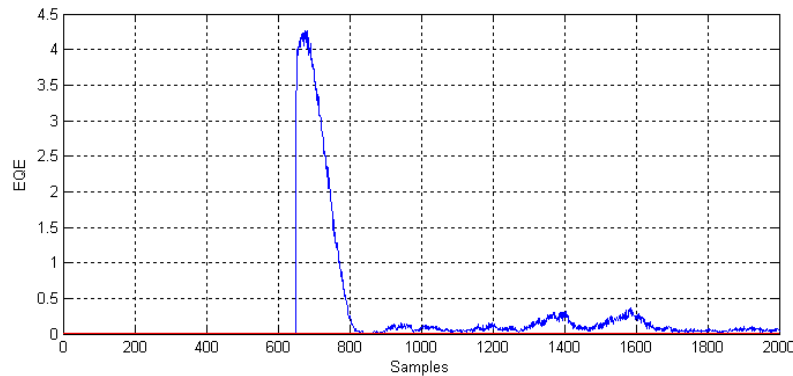


Fig. IV.9 : Evolution de l'erreur d'estimation quadratique avant et après l'occurrence du défaut (cas 1).

Les résultats obtenus pour les cas 2 et 3 sont aussi significatifs et montrent clairement que les défauts survenant séparément sur la pompe 1 et le capteur de niveau 2 sont bien détectés à travers le changement dans la courbe du signal *EQE* et le franchissement du seuil de confiance établi comme le montrent les figures (IV.10) et (IV.11).

On constate également que le fait de diminuer le débit d'alimentation de la deuxième cuve comme il est indiqué sur la figure (IV.12), le niveau d'eau dans cette cuve commence à chuter, et le débit inter-réservoirs 3 et 2 augmente. Par conséquent, le niveau 3 diminue

progressivement ainsi que le niveau 1 est influencé par l'augmentation de débit inter-réservoir 1 et 3. La pompe 1 commence à refouler plus pour rendre le niveau 1 à sa position demandée.

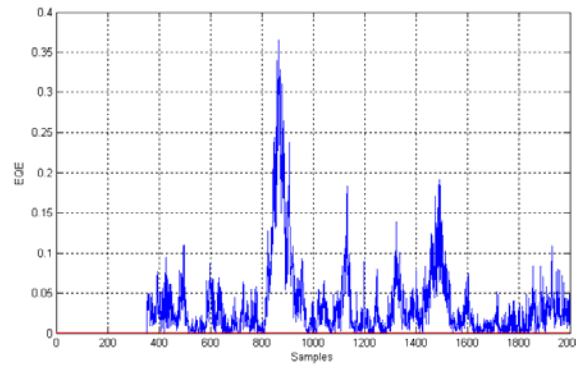


Fig. IV.10 : Evolution de l'erreur d'estimation quadratique avant et après l'occurrence du défaut (cas 2).

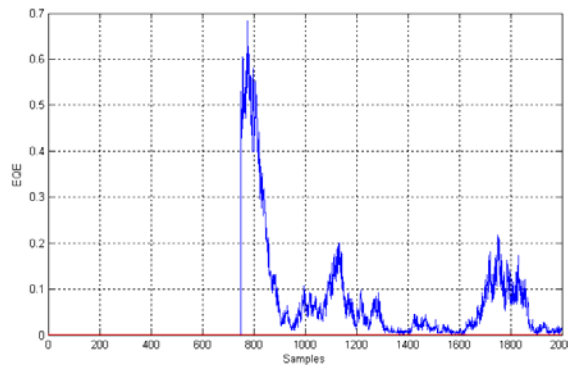


Fig. IV.11 : Evolution de l'erreur d'estimation quadratique avant et après l'occurrence du défaut (cas 3).

Dans le cas 3, Il s'agit d'un défaut affectant le capteur de niveau 2. Celui-ci ne délivre que 90% de la valeur normale comme le montre la figure (IV.13). Dans ce cas, la pompe 2

commence à refouler plus de débit dans le réservoir 2 pour ramener le niveau dans ce réservoir à la hauteur demandée. Ce remplissage diminue le débit inter-réservoir 3 et 2, ce qui permet l'élévation du niveau d'eau dans le réservoir 3 afin de se stabiliser. Ce changement doux dans le signal h_3 influe pour sa part considérablement sur le débit inter-réservoir 1 et 3. Par conséquent, le débit q_1 diminue brusquement en se répercutant sur le niveau h_1 .

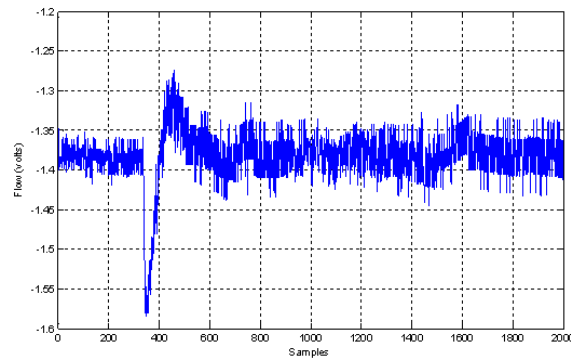


Fig. IV.12 : Evolution du débit de refoulement de la pompe 1 obtenue avant et après l'occurrence du défaut (cas 2).

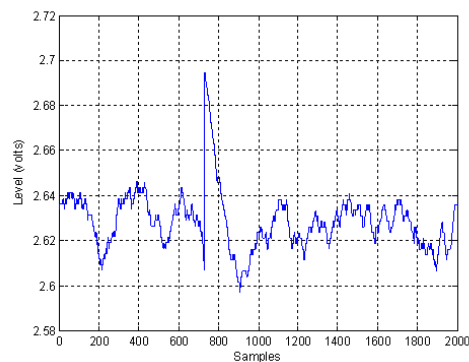


Fig. IV.13 : Evolution de niveau dans le réservoir 2 obtenue avant et après l'occurrence du défaut (cas 3).

Notons cependant que pour l'ensemble des cas envisagés les changements observés sur les allures des signaux affectés sont rapidement compensés. Il ne s'agit absolument pas de la

disparition du défaut, mais plutôt c'est la boucle de commande qui intervient pour compenser l'erreur de réglage. Il est alors nécessaire d'évaluer pour chaque cas l'indice de détection qui nous permet de donner une conclusion quant à la l'occurrence ou non du défaut. L'indice de détection adopté permet bien de déceler les anomalies de fonctionnement du capteur de niveau 1 et 2 ainsi que la pompe 1. Aussi, il convient de constater que le taux de dispersion de certains signaux devient relativement élevé en présence de défaut, ce qui permet d'expliquer l'influence du défaut sur le signal considéré.

5. Localisation de défauts

Après avoir détecté un défaut, il faudrait ensuite le localiser. Pour ce faire, Nous avons calculé les contributions de chaque signal du système à trois réservoirs interconnectés à l'instant de détection. Le signal ayant la plus forte contribution est considéré comme la variable en défaut. La contribution de la i -ème variable à l'instant k est définie par l'équation (I.26).

A partir de cette formule, on trace les différentes contributions sur lesquelles les défauts affectant, respectivement, les signaux h_1 , h_2 et q_1 ont été détectés par le signal EQE . Les résultats de l'application de l'analyse des contributions aux instants où l'indice de détection EQE franchit le seuil de confiance établi, sont illustrés sur les trois figures (IV.14), (IV.15) et (IV.16). Il est tout à fait clair que, le signal causant le défaut est bien localisé dans les trois cas envisagés. La simplicité, l'efficacité et la robustesse d'identification de la cause de défauts reposent essentiellement sur la qualité de l'approche neuronale proposée. La robustesse ne peut être en aucun cas remplie si l'on n'adopte pas une stratégie de diagnostic basée fondamentalement sur un modèle neuronal du processus qui n'offre plus « le droit à l'erreur ».

En tenant compte de la nature du procédé et de son environnement de fonctionnement, les résultats obtenus sont très satisfaisants avec le modèle ACPNL obtenu.

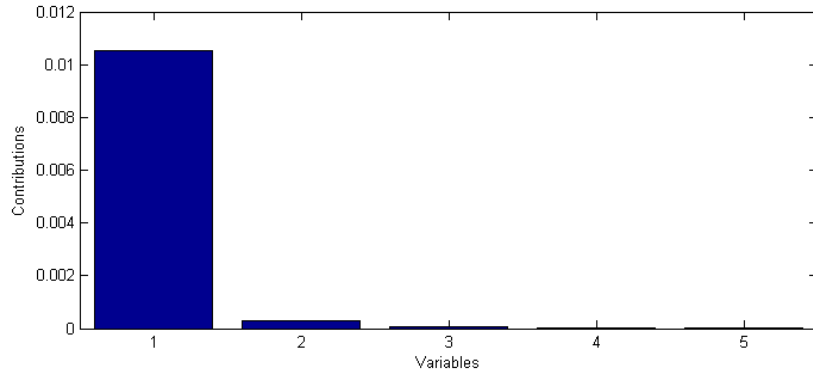


Fig. IV.14 : Contributions des différentes variables à l'indice *EQE* à l'instant 655 (Cas 1).

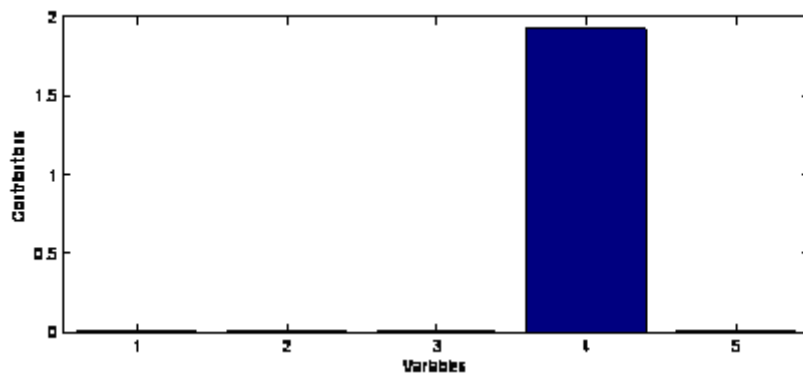


Fig. IV.15 : Contributions des différentes variables à l'indice *EQE* à l'instant 355 (Cas 2).

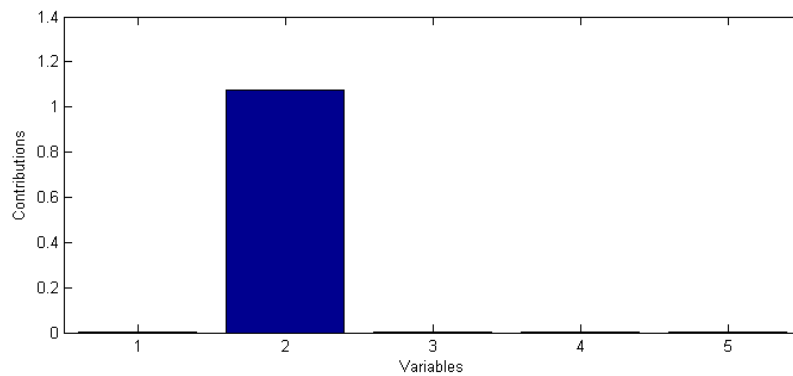


Fig. IV.16 : Contributions des différentes variables à l'indice *EQE* à l'instant 772 (Cas 3).

6. Conclusion

La procédure de détection et l'identification de défauts présentée dans ce chapitre se fonde sur l'utilisation conjointe de deux types de réseaux de neurones confortée par un algorithme d'apprentissage convenablement adéquat, en terme de convergence et rapidité, et une analyse statistique appropriée. Le principe de l'approche étant d'ordre probabiliste car son application requiert la réalisation d'un ensemble d'observations issues d'une expérience réalisée dans les mêmes conditions. La carte d'auto-organisation de Kohonen permet d'extraire les composantes principales dont le choix du nombre de ces dernières à retenir est basé sur sa variance cumulée explicative en pourcentage.

Après avoir déterminé les composantes principales porteuses des informations significatives, dans un espace compacté, sur les variables initiales, un deuxième réseau de neurones de fonction à base radiale est introduit. L'utilité de ce réseau est d'estimer les données initiales. L'apprentissage d'un tel réseau est parfaitement lié au temps de calcul et en plus au problème de convergence. Pour cette raison, nous avons proposé une nouvelle méthode directe d'optimisation non linéaire et globale dite 'DIRECT'. Cette technique, dans sa forme standard, utilise pour la recherche de l'optimum les centres de l'espace de recherche après sa division uniforme selon des vecteurs d'unité orthogonaux. Malgré les avantages de cet algorithme comparé à d'autres méthodes, il présente néanmoins quelques inconvénients. Parmi lesquels, celui où un minimum global se trouve aux voisinages des bords. En revanche, Nous avons introduit quelques modifications sur cet algorithme pour surmonter cet inconvénient. Des vecteurs directifs aléatoires ont été introduits dans le fonctionnement du DIRECT tout en gardant son principe de base par un paramètre de régularisation adaptatif. Il ne laisse la recherche de l'optimum que dans l'espace prédéfini. Comme la convergence est relativement garantie, l'algorithme du DIRECT modifié augmente le pouvoir d'approximation du réseau de neurones par la réduction énorme de nombre d'évaluation de la fonction coût. Ce qui se répercute sur le temps de calcul, et rend possible, d'un point de vue pratique, leur application en temps réel. Nous avons toutefois montré que le modèle neuronal de l'ACPNL illustre clairement dans les deux différentes situations (phase d'apprentissage et celle de validation) son aptitude de reproduction les données initiales du système à trois réservoirs.

Cette précision, simplicité, efficacité et robustesse du modèle neuronal repose essentiellement sur la qualité de l'algorithme d'apprentissage proposé.

La qualité de la reconstruction, définie par la somme successive des carrés de la différence entre les données initiales et celles prédites par le réseau de neurones à fonction de base radiale notée *EQE*, est utilisée comme indice de détection. Analyse des principales caractéristiques statistiques de ce signal mesuré, alors que l'évaluation statistique de ce signal vise à déceler d'éventuelles anomalies afin de trouver est fixer le seuil de détection de défauts approprié. Ceci exprime la précision de détection et qui rend l'indice de détection de défauts plus sensibles aux défauts qu'aux erreurs de modélisations. L'emploi du test de Kolmogorov-Smirnov a donné lieu à une performance appréciable du seuil de détection de défauts que celui basé sur la technique empirique classique, que nous avons bien démontré, à travers les résultats des différents cas de présence de défauts envisagées.

Les contributions de chaque variable du système ont été évaluées à l'instant de détection de défauts. La base de l'identification de défauts en cause repose principalement sur la plus grande valeur en contribution de la variable synonyme de défaut. Par conséquent, la précision de cette technique à la localisation de défauts est nettement appréciée dans la mesure où le système est composé des incertitudes aléatoires.

Conclusion générale

Dans les plans d'instrumentation des systèmes de production modernes, la plupart des procédés sont équipés d'organes de mesure et de contrôle sensibles et très sophistiqués, nécessitant la mise en place de systèmes de surveillance performants réalisant des tâches bien particulières liées aux aspects sûreté de fonctionnement et supervision. La conception d'un système de surveillance capable de prévoir et de détecter efficacement les anomalies de fonctionnement d'un procédé physique continue à faire l'objet de plusieurs études qui s'attachent à rechercher les outils adéquats pour une meilleure caractérisation des situations de pannes.

Durant ces dernières années, plusieurs travaux se sont axés sur les problèmes de diagnostic de défauts. Des approches de surveillance à base de modèle et sans modèle ont été établis et appliquées dans des domaines variés. Les méthodes qui se basent sur une modélisation de l'équipement sont naturellement tributaires de l'existence ainsi que de la qualité d'une modélisation physique de l'équipement. Ce modèle servira de référence pour un fonctionnement nominal et tout écart par rapport au point de fonctionnement nominal sera synonyme de défaut. Cela est tout à fait envisageable s'il s'agit d'un système de petite dimension. En revanche, la modélisation devient progressivement compliquer avec le nombre de variables du système à surveiller, ce qui rend son élaboration irréalisable. Comme alternative, l'Analyse en Composantes Principales (ACP) se trouve très intéressante pour identifier la structure de dépendance linéaire entre les observations du processus sans formuler de façon explicite le modèle du système, afin d'obtenir une représentation compacte de celles-ci.

Pour des raisons sur la précision exigible dans la plupart des applications industrielles, et la présence des observations à caractère non linéaire, qui sont issues des capteurs opérant dans des environnements complexes, incertains ou variants dans le temps, il vient s'ajouter les calculs importants, dans le cas où la matrice de corrélation est de taille importante, il s'avère

plus que nécessaire d'introduire une approche non linéaire de l'ACP. Un de nos objectifs a donc été une approche non linéaire de l'ACP pour le diagnostic de défauts.

En premier lieu, nous avons présenté le principe de l'ACP. Cette technique permet de privilégier les directions d'un espace de données porteuses du maximum d'informations au sens de la maximisation des variances de projection. L'amplitude des valeurs propres de la matrice de covariance des données quantifie pour chacune de ces directions la quantité d'information encodée. Les variables initiales sont ainsi transformées en de nouvelles variables, appelées composantes principales. Elles sont obtenues par combinaison linéaire des composantes originales, ordonnées et non corrélées. L'ACP cherche à identifier les vecteurs propres et les valeurs propres de la matrice de corrélation des variables de départ. Cependant, pour l'obtention de la structure du modèle, il faut déterminer le nombre de composantes principales à retenir dans celui-ci. Plusieurs critères ont été cités, notre choix s'est porté sur le critère de la variance cumulée.

Une fois le modèle ACP identifié, La méthode de détection fondée sur ce dernier utilise l'espace résiduel comme indice de détection. Cet indice révélateur de la présence de défauts, qui exprime l'erreur quadratique d'estimation, est généré à partir d'un modèle ACP de bon fonctionnement du processus et des mesures disponibles. Un défaut quelconque serait mis en évidence si EQE dépasse le seuil de confiance établi sous certaines considérations et propriétés statistiques du signal EQE dans le but de définir au mieux les phénomènes aléatoires qui lui sont associés. Lorsqu'un défaut abrupt est détecté, le calcul des contributions de chaque variable à la valeur de EQE permet d'identifier la ou les variables qui sont en cause. La variable ayant la plus grande contribution est considérée comme la variable incriminée.

Par la suite, nous avons présenté le principe de l'Analyse en Composantes Principales Non linéaire (ACPNL). Dans un premier temps, nous avons illustré le principe de l'Analyse en Composantes Curvilignes (ACC). Cette approche utilise la méthode de gradient pour la détermination des composantes curvilignes qui souffre de la présence des minima locaux qui

se répercute sur la qualité de la projection. Il vient de s'ajouter la difficulté de choix des composantes initiales de la règle d'adaptation qui pose un problème major sur le temps de calcul lors de l'utilisation en ligne de l'ACC. A cet effet, cette approche reste non recommandée pour des applications dans la surveillance des processus industriels. Nous avons présenté aussi la méthode des courbes principales. Cette technique est une généralisation non paramétrique de l'ACP. Cependant, elle ne permet pas d'avoir un modèle de représentation ACP. Les approches neuronales semblent plus intéressantes pour l'approximation des données à caractère non linéaire. Ainsi, plusieurs approches sont présentées utilisant différentes architectures des réseaux de neurones. De plus d'autres alternatives ont été proposé pour surmonter le problème d'apprentissage en combinant les courbes principales et les réseaux de neurones.

Ainsi, nous avons présenté une extension de l'ACP dans le cas non linéaire par l'utilisation de deux types de réseaux de neurones. Le premier réseau, représenté par la carte d'auto-organisation de Kohonen, sert à extraire les composantes principales, tandis que le deuxième de fonction à base radiale est utilisé pour la reconstruction des données initiales. La connaissance des poids référents de la carte de Kohonen détermine une quantification vectorielle de la projection dans l'espace de dimension réduit. La détermination de ces poids projecteurs se fait par minimisation d'une fonction de coût. Cette fonction coût est l'objet de la maximisation du produit scalaire ou minimisation d'une distance euclidienne entre les observations et les poids référents.

La transformation inverse permettant de calculer l'estimation des données initiales à partir des composantes principales obtenues par la carte de Kohonen se fait par un réseau de neurones de fonction à base radiale. Ce réseau de neurones est utilisé comme outil de modélisation par apprentissage. L'apprentissage est réalisé par des algorithmes d'optimisation dans le but est de minimiser la fonction de coût qui définit l'écart entre les sorties du réseau et les sorties réelles, en modifiant les paramètres d'une manière itérative. Nous avons présenté un algorithme d'optimisation direct et globale, appelé 'DIRECT' dans sa forme standard. Cet algorithme est basé sur la division continue de l'espace de recherche de la fonction coût.

En dernier lieu, nous avons présenté l'application de l'approche neuronale proposée au diagnostic de défauts dans un système à trois réservoirs DTS-200. La carte d'auto-organisation de Kohonen permet d'extraire les composantes principales dont le choix du nombre de ces dernières à retenir est basé sur sa variance cumulée explicative en pourcentage. Ces composantes principales retenues représentent l'entrée du deuxième réseau de fonction à base radiale. L'utilité de ce réseau est d'estimer les données initiales en modifiant leurs paramètres par apprentissage. L'apprentissage d'un tel réseau est parfaitement lié au temps de calcul et en plus au problème de convergence. Pour ces raisons, nous avons introduit quelques modifications dans l'algorithme standard de DIRECT. Des vecteurs directifs aléatoires ont été introduits dans le fonctionnement du DIRECT tout en gardant son principe de base par un paramètre de régularisation adaptatif. Nous avons montré que le modèle neuronal de l'ACPNL illustre clairement son aptitude de reproduire les données initiales du système à trois réservoirs. Cette précision de reconstruction revient à la qualité de l'algorithme d'apprentissage proposé.

La somme successive des carrés de la différence entre les données initiales et celles prédites par le réseau de neurones à fonction de base radiale notée EQE , est utilisée comme indice de détection. L'analyse des principales propriétés statistiques de ce signal mesuré vise à déceler d'éventuelles anomalies afin de trouver est fixer le seuil de détection de défauts approprié. Ceci exprime la précision de détection à travers les résultats des différents cas de présence de défauts envisagés, ce qui rend l'indice de détection de défauts plus sensibles aux défauts qu'aux erreurs de modélisations. Nous avons évalué après chaque détection la contribution de chaque variable afin de fixer celle synonyme de défaut.

Il est tout à fait clair que, d'un point de vue pratique, cette procédure est entièrement réalisable et peut servir au diagnostic de défauts dans des applications diverses. En effet, elle est simple à réaliser et peut s'appliquer sur l'historique des mesures que l'on prélève régulièrement sur les sites industriels. Sa mise en œuvre doit tenir compte des spécificités propres à chaque procédé. Néanmoins, il est impératif de traiter les résultats d'application avec rigueur et certitude car les systèmes de régulation peuvent jouer le rôle de compensateurs de défauts ou encore certaines entrées inconnues de fortes amplitudes peuvent être identifiées en tant que défauts. Ce dilemme est bien fréquent en diagnostic; une perception approfondie de la situation de défaillance courante permettrait d'aider à prendre les décisions appropriées.

Références bibliographies

- [1] Chen, J., and Patton, R., (1999), "Robust model-based fault diagnosis for dynamic systems", Kluwer, Boston, MA.
- [2] Frank, P., (1990), "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy a survey and some new results", *Automatica*, pp. 459-474.
- [3] Hammouri, H., Kinnaert, M., and El Yaqqoubi, E., (1999), "Observer-based approach to for detection and isolation for nonlinear systems", *IEEE, Trans. Autom. Control*, pp. 1879-1884.
- [4] Tao He, and *al.* (2006), "Process fault detection and diagnosis based on principal component analysis", *Proceedings of the 5th ICMLC, Dalian*.
- [5] Seongkyu, Yoon, MacGregor, J F., (2000), "Statistical and causal model-based approaches to fault detection and isolation", *AIChEJ*, pp. 1813-1824.
- [6] Song K., Wang H., and Li P., (2004), *PLS-based Optimal Quality Control Model for TE Process*, *IEEE International Conference on Systems, Man and Cybernetics*.
- [7] Debashis N., and Cory E. S., (1997), *Application of Multivariate Statistical Techniques for Monitoring Emulsion Batch Processes*, *Proceedings of the American Control Conference Albuquerque, New Mexico*.
- [8] Akaike H., (1974). *A new look at the statistical model identification*, *IEEE transactions on Automatic control*, 19, p. 716-723.
- [9] Onder Uluyol, Kyusung Kim, and Emmanuel O. Nwadiogbu, (2006), *Synergistic Use of Soft Computing Technologies for Fault Detection in Gas Turbine Engines*, *IEEE Trans. on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, Vol. 36, No. 4, p. 476-484.
- [10] F. Gustafsson, (2002), *Stochastic Fault Diagnosability in Parity Spaces*, *IFAC, 15th Triennial World Congress, Barcelona, Spain*.

- [11] Francesco Pierri, Gaetano Paviglianiti, Fabrizio Caccavale, and Massimiliano Mattei, (2008), Observer-based sensor fault detection and isolation for chemical batch reactors, *Engineering Applications of Artificial Intelligence*, 21, p. 1204–1216.
- [12] Shewhart, W.A., (1986), *Statistical Method from the Viewpoint of Quality Control*. The Graduate School, U.S. Department of Agriculture, Washington, 1939, 155 p. Réédité par Dover Publications, 1986.
- [13] Shewhart, W.A., (1980), *Economic Control of Quality of Manufactured Products*. New York: Van Nostrand, London: MacMillan, 1931, 501 p. Réédité en 1980 par l'American Society for Quality.
- [14] David J. H. Wilson, George W. Irwin, and Gordon Lightbody, (1999), RBF Principal Manifolds for Process Monitoring, *IEEE Trans. on Neural Network*, Vol. 10, No. 6, p. 1424-1434.
- [15] Jim Cao and Janos Gertler, (2001), Noise-Induced Bias in PCA Modeling of Linear System, *Proceedings of the American Control Conference*, Arlington, VA.
- [16] Sivan Lachman-Shalem, Nir Haimovitch, Eitan N. Shauly, and Daniel R. Lewin, (2002), MBPCA Application for Fault Detection in NMOS Fabrication, *IEEE Trans. On Semiconductor Manufacturing*, Vol. 15, No. 1, p. 60-70.
- [17] YANG Yinghua, LU Ningyun, WANG Fuli, MA Liling, (2002), A new fault detection and diagnosis method based on principal component analysis in multivariate continuous processes, *Proceedings of the 4th World Congress on Intelligent Control and Automation*, Shanghai, P.R.China, p. 3156-3160.
- [18] Jiangxin Wang, and Costas J. Spanos, (2002), Real-Time Furnace Modeling and Diagnostics, *IEEE Trans. on Semiconductor Manufacturing*, Vol. 15, No. 4, p. 393-403.
- [19] Zhaoj L., T.-Y. Chai, and G. Wang, (2003), A Nonlinear Modeling and Online Monitoring Method for the Batch Process Using Multiple Local PCA, *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, Xi'an, p. 1190-1194.
- [20] Ji-Hoon Cho, Jong-Min Lee, Sang Choi, Dongkwon Lee and In-Beum Lee, (2004), Sensor Fault Identification based on Kernel Principal Component Analysis, *Proceedings of the IEEE International Conference on Control Applications*, Taipei, Taiwan, p. 1223-1228.

- [21] H. Henry Yue, and Masayuki Tomoyasu, (2004), Weighted Principal Component Analysis and its Applications to Improve FDC Performance, 43rd IEEE Conference on Decision and Control, Atlantis, Paradise Island, Bahamas, p. 4262-4267.
- [22] L. Brito Palma, F. Vieira Coito, R. Neves da Silva, (2005), Diagnosis of Parametric Faults Based on Identification and Statistical Methods, Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, Seville, Spain, p. 3838-3843.
- [23] Gregory A. Cherry and S. Joe Qin, (2006), Multiblock Principal Component Analysis Based on a Combined Index for Semiconductor Fault Detection and Diagnosis, IEEE Trans. On Semiconductor Manufacturing, Vol. 19, No. 2, p. 159-172.
- [24] Demartines, P., (1995), Analyse de données par réseaux de neurones auto-organisés, thèse de l'Institut National Polytechnique de Grenoble.
- [25] Hérault, J. and Jutten, C., (1993), Réseaux de neurones et traitement du signal, Hermès.
- [26] Vigneron, V., (1997), Méthodes d'apprentissage statistiques et problème inverses- Application à la spectrographie, thèse Université d'Evry-Val-d'Essonne.
- [27] Pilato, V., (1998), Application des réseaux de neurones aux méthodes de mesure basées sur l'interaction rayonnement matière, thèse Université Paris-Sud.
- [28] Sung-Kwun Oh, Han-Jong Jang, and Witold Pedrycz, (2009), The design of a fuzzy cascade controller for ball and beam system: A study in optimization with the use of parallel genetic algorithms, Engineering Applications of Artificial Intelligence, 22, p. 261-271.
- [29] M. Kherallah, F. Bouri, A.M. Alimi, (2009), On-line Arabic handwriting recognition system based on visual encoding and genetic algorithm, Engineering Applications of Artificial Intelligence, 22, p. 153–170.
- [30] Jian-Da Wu, Yu-Hsuan Wang, Peng-Hsin Chiang, Mingsian R. Bai, (2009), A study of fault diagnosis in a scooter using adaptive order tracking technique and neural network, Expert Systems with Applications, 36, p. 49–56.

- [31] Jian-bo Yu, and Li-feng Xi, (2009), A neural network ensemble-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes, *Expert Systems with Applications*, 36, p. 909–921.
- [32] Chuan-Yu Chang, ChunHsi Li, Jia-Wei Chang, and MuDer Jeng, (2009), An unsupervised neural network approach for automatic semiconductor wafer defect inspection, *Expert Systems with Applications*, 36, p. 950–958.
- [33] Niranjan Roy, Ranjan Ganguli, (2009), Filter design using radial basis function neural network and genetic algorithm for improved operational health monitoring, *Applied Soft Computing*, 6, p. 154–169.
- [34] Y. Fang, T. W. S. Chow and X. D. Li, (1999), Use of Recurrent Neural Network in Discrete Sliding-Mode Control, *IEE. Proc.-Control Theory Appl.*, Vol. 146, No. 1, p. 84-90.
- [35] Asriel U. Leven and Kumpati S. Narendra, (1993), Control of Nonlinear Dynamical Systems using Neural Networks: Controllability and Stabilization, *IEEE Transactions on Neural Networks*, Vol. 4, No. 2, p. 194-206.
- [36] X. Luo, A. D. Patton and C. Singh, (2000), Real Power Transfer Capability Calculations using Multi-Layer Feedforward Neural Networks, *IEEE Transactions on Power Systems*, Vol. 15, No. 2, p. 903-908.
- [37] M. Zelmat and A. Kouadri, (2001), A Nonlinear Model for a Drum Boiler-Turbine System Using Neural Network Approach, *3^{ème} Séminaire National en Génie Electrique*, Biskra.
- [38] Jianbo Yu, Lifeng Xi, Xiaojun Zhou, (2009), Identifying source(s) of out-of-control signals in multivariate manufacturing processes using selective neural network ensemble, *Engineering Applications of Artificial Intelligence*, 22, p. 141–152.
- [39] Webb A. R., (1996), An approach to nonlinear principal component analysis using radially symmetric kernel functions. *Statist. Comput.*, Vol.6, p. 159-168.
- [40] Webb A. R., (1999), A loss function to model selection in nonlinear principal components. *Neural Networks*, Vol 12, p. 339-345.

- [41] N. Goldberg, T. G. Kolda, and A. S. Yoshimura, (2008), Concurrent Optimization with DUET: DIRECT Using External Trial Points. SANDIA report. Downloadable at: csmr.ca.sandia.gov/~tgkolda/pubs/bibtgkfiles/SAND2008-5844.pdf
- [42] J. D. Griffin and T. G. Kolda, (2008), Asynchronous parallel hybrid optimization combining DIRECT and GSS, SANDIA report, SAND2008-6553. Downloadable at: csmr.ca.sandia.gov/~tgkolda/pubs/bibtgkfiles/SAND2008-6553.pdf
- [43] Kaspar Riesen, Horst Bunke, (2009), Reducing the dimensionality of dissimilarity space embedding graph kernels, *Engineering Applications of Artificial Intelligence*, 22, p. 48– 56.
- [44] D. Aguadoa,_, T. Montoyaa, L. Borrassb, A. Secob, J. Ferrera, (2008), Using SOM and PCA for analysing and interpreting data from a P-removal SBR, *Engineering Applications of Artificial Intelligence*, 21, p. 919–930.
- [45] Varanon Uraikul, Christine W. Chan, Paitoon Tontiwachwuthikul, (2007), Artificial intelligence for monitoring and supervisory control of process systems, *Engineering Applications of Artificial Intelligence*, 20, p. 115–131.
- [46] Wenbo Cao _, Robert Haralick, (2009), Affine feature extraction: A generalization of the Fukunaga– Koontz transformation, *Engineering Applications of Artificial Intelligence*, 22, p. 40– 47.
- [47] D.R., Jones, C.D., Perttunen, and B.E., Stuckman, (1993), Lipschitzian optimization without the Lipschitz constant, *J. Optim. Th. Appl.*, 79, p. 157-181.
- [48] V. Torczon, Multi-directional search: a direct search algorithm for parallel machines, Ph.D. Thesis, Department of Mathematical Sciences, Rice University, Houston, Texas, 1989.
- [49] P., Gilmore and C.T., Kelley, (1995), An implicit filtering algorithm for optimization of functions with many local minima, *SIAM J. Optim.*, 5, p. 269-285.
- [50] R., Hooke, and T.A., Jeeves, (1961), Direct search solution of numerical and statistical problems, *Journal of the ACM*, 8, p. 212-229.
- [51] Nelder JA, Mead R. A simplex for function minimization. *Comput J* 1965; 7: 308-13.

[52] D. E. Finkel, C. T. Kelley, (2004), Convergence Analysis of the DIRECT algorithm, North Carolina State University, Center for Research in Scientific Computation, Raleigh, version of July 14, 2004.

[53] D. E. Finkel, (2003), DIRECT Optimization Algorithm User Guide, Center for Research in Scientific Computation, North Carolina State University, Raleigh.

[54] J.M. Gablonsky, (2001), DIRECT version 2.0 user guide, Technical Report, CRSC-TR01-08, Center for Research in Scientific Computation, North Carolina State University.

[55] J.M. Gablonsky, (2001), Modifications of the DIRECT Algorithm, PhD Thesis, North Carolina State University, Raleigh, North Carolina.

[56] Amira-DTS200. (1996). *Laboratory setup three tank system*. Amira GmbH, Duisburg-Germany.