

CLASSIFICATION DE DEFAUTS PAR RESEAUX DE NEURONES.  
APPLICATION AU CONTROLE NON DESTRUCTIF  
PAR COURANTS DE FOUCAULT PULSES.

J-C. BOUR, I. STANGHELLINI\*, A. BILLAT

Laboratoire d'Applications de la Microélectronique  
Faculté des sciences, Moulin de la Housse, BP 347  
51062 Reims Cedex

\*Université de Bologne, Faculté d'Ingénierie  
DEIS, via Risorgimento 2, 40100 Bologne, Italie

RÉSUMÉ

Après l'étude de différents critères de sélection de caractères pertinents, nous proposons deux techniques pour effectuer le choix définitif parmi les sous espaces possibles.

Nous étudions ensuite les capacités de deux classifieurs neuronaux sur le sous espace de données retenu. Le premier classifieur est basé sur l'algorithme de la rétropropagation du gradient, le second étant à fonctions de transferts gaussiennes (R.B.F.).

Ce travail s'inscrit dans un projet général de contrôle non destructif par courants de Foucault pulsés, de plaques d'acier austénitique.

ABSTRACT

After the study of different selection criterions for relevant features, two techniques are proposed which make possible a choice among the allowed subspaces. Then the classifying ability of the two neural networks on the better data subspaces are also examined. The first classifier is a backpropagation neural network whereas the second is a R.B.F one the basis functions of which are gaussians.

This work is a part of a more broad project of non-destructive testing by pulsed eddy currents.

1. Introduction

Le but de cet article est de présenter une méthode de détection et de classification de défauts artificiels introduits dans des plaques d'acier austénitique. La technique utilisée pour détecter les défauts est celle des courants de Foucault pulsés, dans laquelle le champ électromagnétique est généré par une impulsion de courant émise dans un capteur inductif plat [1]. Cette technique présente l'avantage, par rapport à celle des courants de Foucault sinusoïdaux, d'une pénétration plus profonde dans la matière.

Après l'examen de différents critères de sélection de caractères pertinents, nous étudierons deux classifieurs neuronaux et évaluerons leurs capacités.

2. Enoncé du problème

Des défauts artificiels ont été usinés par perçage ou électroérosion dans des plaques d'acier austénitique. Nous travaillons sur 44 échantillons différents. La détection d'un défaut et l'extraction de sa signature (Figure 1) sont issus d'un traitement particulier des signaux différentiels de réception, que nous avons déjà décrits par ailleurs [2]. Le jeu de caractères retenu est formé de 7 éléments dont 6 sont extraits directement de cette signature. Le septième étant la valeur maximale de la puissance moyenne du signal de réception pendant la détection.

{T1, T2, T3, E1, E2, E3, Max}

L'ensemble de ces mesures est divisé en 5 classes dont la description et le cardinal sont reportés dans le tableau 1. Le fait de fixer le nombre de classe a priori est une méthode classique (user-supervised classification).

Cet étiquetage permet de fournir une information sur le type de défaut détecté, ainsi que sur sa position relativement à la surface de contrôle.

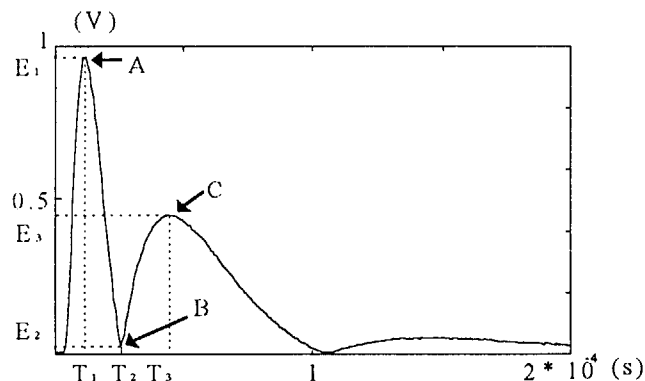


Figure 1. Signature d'un défaut.

Classe	Description	# d'éléments
1	Fentes débouchantes	5
2	Trous débouchants Ø < 2.5 mm	8
3	Trous débouchants Ø > 2.5 mm	6
4	Perçages (// surface contrôlée) (non débouchants)	20
5	Trous et fentes non débouchantes	5

Tableau 1. Description et cardinal des 5 classes.

3. Sélection des caractères

L'étape de la sélection de caractères (diminution de la dimension de l'espace de représentation) est très importante. Conserver les d caractères du vecteur forme d'origine (d=7) risque non seulement d'alourdir les calculs (caractères non informatifs ou redondants) mais également de rendre le problème trivial dans le cas d'un faible nombre d'exemples. En effet, plus la dimension de l'espace est grande, plus il est



aisé de trouver les hyperplans séparateurs des différentes classes. Ces derniers sont alors peu représentatifs des classes réelles et leur orientation peut varier de multiple façon tout en séparant toujours les classes étudiées.

La répartition des défauts dans les différents plans formés par les caractères pris deux à deux permet de fixer *a priori* l'ordre du sous espace à 3. Ce choix a été validé *a posteriori* par une comparaison des performances obtenues avec un réseau à rétropropagation admettant en entrée d'une part le sous espace retenu, et d'autre part les 7 caractères du jeu initial.

Parmi les critères disponibles pour mener à bien cette étape, nous en avons retenu 4.

Le premier (critère de Fisher) est défini par l'équation 1.

$$F_{(Ci)} = \frac{[\hat{m}_1(Ci) - \hat{m}_2(Ci)]^2}{\sum_{y \in X_1} [y(Ci) - \hat{m}_1(Ci)]^2 + \sum_{y \in X_2} [y(Ci) - \hat{m}_2(Ci)]^2} \quad (1)$$

Où  $\hat{m}_m(Ci)$  est l'estimée de la moyenne du caractère  $Ci$  de la classe  $X_m$  et  $y(Ci)$  est le caractère  $Ci$  d'un exemple particulier appartenant à la classe  $X_m$ .

Il ne peut être appliqué qu'à deux classes simultanément. Employé sur les deux classes formées respectivement de l'ensemble des défauts débouchants d'une part, et des défauts non débouchants d'autre part, il fournit 2 caractères principaux (T1, E1) qui contiennent 54.12% de la quantité d'information du jeu complet. L'étude graphique des défauts dans le plan obtenu montre une bonne séparabilité des classes 1, 2 et 3 entre elles ainsi que de leur ensemble par rapport à la totalité des défauts non débouchants. La séparabilité des classes 4 et 5 entre elles étant jugée insuffisante, ce critère a été appliqué de nouveau sur ces deux classes. Les 2 paramètres principaux sont T1 et E2 (84.26% de la quantité d'information totale). Ce premier critère nous fournit donc comme sous espace :  $S1 = \{T1, E1, E2\}$

Le second critère est une extension du précédent à un nombre quelconque de classes [3]. Il est défini par l'équation 2.

$$F_{mc(Ci)} = \frac{2}{k \cdot (k-1)} \sum_{m=1}^{k-1} \sum_{n=m+1}^k \frac{[\hat{m}_m(Ci) - \hat{m}_n(Ci)]^2}{N_m \cdot \hat{S}_m^2(Ci) + N_n \cdot \hat{S}_n^2(Ci)} \quad (2)$$

Où  $\hat{S}_m^2(Ci)$  est l'estimée de la variance du caractère  $Ci$  de la classe  $m$ ,  $N_m$  est le nombre d'éléments de la classe  $m$  et  $k$  est le nombre de classes.

Les deux derniers critères [4] sont basés sur les matrices de covariance (équations 3 et 4).

$$\text{tr}(W^{-1} B) \quad (3) \qquad \frac{\det(V)}{\det(W)} \quad (4)$$

Où  $W$  est la matrice de dispersion intraclasse,  $B$  la matrice de dispersion interclasse et  $V$  la matrice de dispersion totale.

Ces critères de sélection sont tous basés sur le principe suivant : les classes seront d'autant mieux séparables qu'elles seront plus éloignées les unes des autres et plus regroupées sur elles mêmes. Ils fournissent chacun des sous espaces différents qui sont :

- S1={T1, E1, E2}
- S2={E1, E3, Max}
- S3={E1, E2, Max}
- S4={T2, T3, E1}

Le choix final du sous espace est obtenu après l'application de deux algorithmes différents.

Le premier est à base d'un réseau neuronal linéaire (Madaline) dont nous comparons les performances dans chaque sous espace. Pour éviter au réseau de placer ses frontières au plus près des classes les plus représentées, les vecteurs cibles (habituellement une ligne de la matrice identité (I) de taille  $M \times M$  ( $M$  est le nombre de classes) où chaque zéro est remplacé par la valeur -1) ont été modifiés de façon à rendre les résultats indépendants du cardinal de chaque classe. Ainsi, le vecteur codant la  $j$ -ième classe est obtenu en multipliant la  $j$ -ième ligne de I par  $n_{\max}/n_j$ , où  $n_{\max}$  est le nombre d'élément de la classe la plus représentée et  $n_j$  est le cardinal de la classe  $j$ . Cette modification précède le remplacement des 0 par -1. Le codage obtenu tend à privilégier le taux de classification moyen par classe.

Le second algorithme peut être scindé en deux parties. La première est constituée d'une méthode de coalescence (type centres mobiles) dont le but est de fournir les 5 groupes de défauts les plus regroupés, dans le sous espace étudié. La seconde partie utilise la méthode hongroise (méthode des moindres regrets) [5] pour estimer la "similitude" entre les classes réelles et celles fournies par la méthode de coalescence. Le tableau 2 récapitule les résultats obtenus avec ces deux algorithmes appliqués aux 4 sous espaces présélectionnés.

	Madaline	Méthode hongroise
Sous espace	T*	"Similitude"
S1	80%	75%
S2	56%	63.6%
S3	56%	72.7%
S4	77.5%	72.7%

Tableau 2. Sélection du sous espace de représentation

T\* : Taux moyen de bonne classification par classe.

Ces résultats montrent que le sous espace séparant le mieux les 5 classes étudiées est S1.

#### 4. Classification par réseau à rétropropagation du gradient

Le premier classifieur utilisé est un réseau à une couche cachée, dont la règle d'apprentissage est celle de la rétropropagation du gradient de l'erreur, dans sa formulation *batch*. Le réseau est étagé, de façon stricte, en couche (aucune connexion pontée). La couche cachée possède une fonction d'activation du type "tangente hyperbolique" avec des valeurs comprises entre -1 et 1. La fonction associée aux sorties est une fonction logistique (valeurs entre 0 et 1). Pour permettre d'accélérer la convergence, on utilise un pas adaptatif et un filtrage sur les incréments d'adaptation (momentum) [6]. La normalisation des données consiste en une transformation affine sur chaque entrée, en fonction des valeurs extrémales qu'elles prennent dans l'ensemble du jeu de valeurs. Ceci afin de rapporter chaque caractère dans l'intervalle [-1;+1]. L'utilisation d'une telle architecture nécessite l'optimisation du nombre de cellules cachées, celui des entrées et des sorties étant fixé par la taille du sous espace retenu et le nombre de classes. La figure 2 présente les taux de généralisation obtenus pour différents nombres de cellules cachées, en fonction du nombre d'itérations. Dans cet exemple, 60% des données sont utilisées en apprentissage, le reste en test. Les résultats

présentés sont les résultats moyens obtenus avec 10 initialisations différentes. Ceci a été fait dans le but de minimiser l'influence des conditions initiales, notamment des poids des connexions. Le seuil de décision retenu est fixé à 0.1. Ainsi un défaut sera-t-il affecté à une classe si la sortie la plus élevée est supérieure à 0.9, toutes les autres étant inférieures à 0.1.

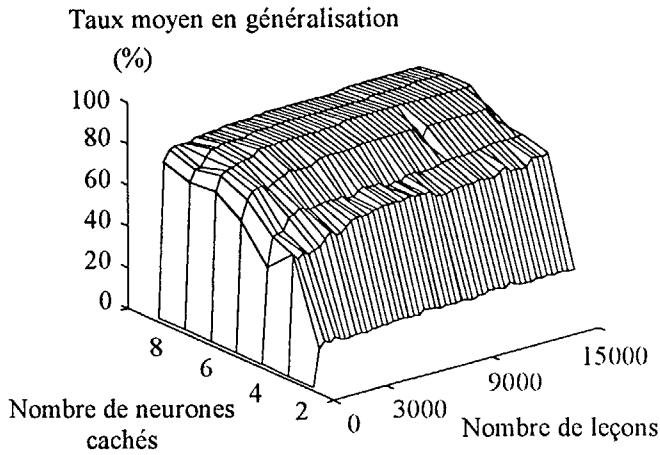


Figure 2. Définition du critère d'arrêt

Le taux de généralisation moyen, après une phase montante, se stabilise autour d'un certain niveau de performance comportant de faibles variations. Cette constatation permet de définir un critère d'arrêt pour l'apprentissage que nous avons fixé à 8000 leçons. La figure 2 permet également de montrer qu'un réseau trop petit n'atteint que de faibles taux de généralisation, même avec un apprentissage très long. Nous avons testé 4 réseaux (de 5 à 8 cellules cachées) qui sont non seulement ceux présentant les meilleures capacités de classifications, mais également la plus faible sensibilité aux conditions initiales (inférieure à 10%). Pour comparer les performances de ces réseaux, nous avons utilisé le test du leave-one-out (N-1 exemples en apprentissage et 1 en test, puis on calcule les performances moyennes sur toutes les partitions possibles).

Pour chacun des 4 réseaux, le seuil de décision a été optimisé de la façon suivante : on évalue pour chaque valeur du seuil, le pourcentage de rejet (défauts non classés) ainsi que le pourcentage de mal classés rapportés au nombre de défauts classés. La figure 3 montre l'évolution de ces deux indices en fonction du seuil, pour 6 cellules cachées (meilleure configuration). Le seuil retenu est fixé à 0.065. La règle de décision reste celle exposée précédemment. Les performances

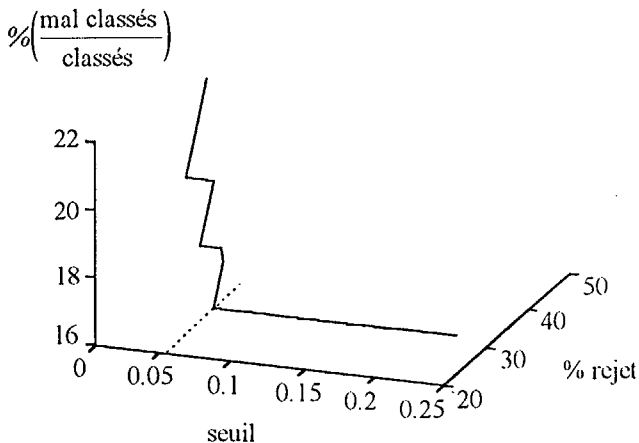


Figure 3. Optimisation du seuil de décision.

obtenues sont résumées dans les deux matrices M1 et M2. Les trois colonnes de la première matrice (M1) contiennent respectivement le pourcentage de Bien Classés (BC), de Mal Classés (MC) et de Rejet (R) pour les 5 classes étudiées. La matrice M2 (matrice de confusion) présente la distribution des différents exemples, en comparaison de leur classe réelle. Dans le cas d'une classification idéale, la première colonne de la matrice M1 serait identiquement égale à 100%, toute autre colonne étant nulle. La seconde matrice serait nulle, excepté la diagonale (100%).

$$\begin{matrix}
 & \text{BC} & \text{MC} & \text{R} \\
 \text{Classe 1} & \left( \begin{matrix} 20 & 0 & 80 \\ 87.5 & 0 & 12.5 \\ 100 & 0 & 0 \\ 70 & 5 & 25 \\ 40 & 60 & 0 \end{matrix} \right) & = & \text{M1}
 \end{matrix}$$

$$\begin{matrix}
 & \text{Classes réelles} \\
 \text{Classes prédites} & \left( \begin{matrix} 100 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 82.35 & 17.65 \\ 0 & 0 & 0 & 33.4 & 64.4 \end{matrix} \right) & = & \text{M2}
 \end{matrix}$$

Les performances atteintes sont relativement élevées, excepté dans les classes sous représentées, pour lesquelles les frontières de décision peuvent être très variables d'un défaut test à l'autre. Néanmoins cette étude montre les aptitudes d'un classifieur neuronal dans ce type d'applications. Un tel classifieur est cependant limité puisqu'il n'intègre pas la notion de rejet de distance [7]. Ceci est dû au type de fonction (sigmoïde) associée aux sorties du réseau, qui se trouve être très peu localisée autour des classes. Dans le cas d'une application industrielle, cette limitation peut être rédhibitoire, du fait de la probabilité élevée de l'apparition d'un défaut n'appartenant à aucune classe prédéfinie. L'utilisation de fonctions de transfert gaussiennes permet la prise en compte des deux notions de rejet : ambiguïté (doute entre plusieurs classes), et distance.

### 5. Classification par réseau R.B.F.

La figure 4 présente l'architecture que nous avons retenue. Elle est constituée d'une couche de cellules à fonction de transfert gaussienne et d'une couche de sortie à fonction de transfert linéaire. Le vecteur d'entrée est connecté aux cellules gaussiennes par des poids fixes (unitaires), les sorties par des poids adaptatifs.

Dans une telle configuration, chacune des cellules gaussiennes est connectée à une sortie unique à laquelle elle est dédiée. Le rôle de cette dernière est d'estimer l'appartenance d'un vecteur d'entrée à la classe qu'elle représente. Pendant l'apprentissage, seuls les coefficients liant les cellules gaussiennes aux sorties, ainsi que l'écart type des gaussiennes suivant chacune de leurs composantes sont modifiés. Les poids des connexions adaptatives sont initialisés de façon arbitraire à 1. La durée de l'apprentissage sera suffisamment longue pour fournir un résultat indépendant de ces initialisations.

L'utilisation d'un réseau R.B.F. nécessite l'initialisation des centres dans l'espace de représentation des données. Dans notre application, ces derniers sont figés à leur position



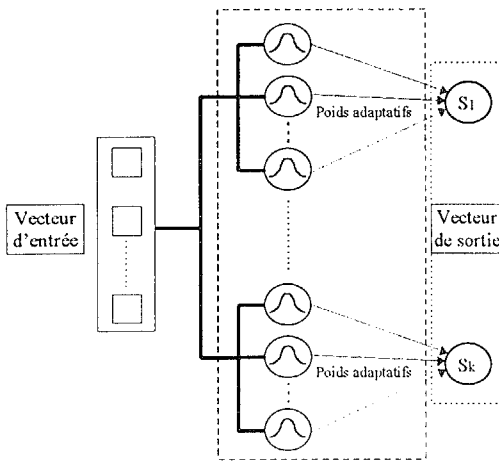


Figure 4. Architecture du réseau R.B.F.

initiale. Plusieurs procédures peuvent être utilisées pour ce faire, tel le placement aléatoire dans le sous espace de l'ensemble des données [8] ou encore celle proposée dans [9] dont l'avantage est, malgré une complexité élevée, d'initialiser à la fois les centres et les poids des connexions adaptatives.

La méthode que nous utilisons est basée sur un algorithme de coalescence (type centres mobiles) appliqué sur la base d'apprentissage. Le but est de structurer chaque classe en groupes, dont les centres de gravité sont utilisés pour initialiser les centres des cellules, l'écart type de chaque classe étant employé pour initialiser celui des centres qui lui sont associés. La partition retenue sera celle minimisant la dispersion intra-classe.

Aucune étude n'a rendu possible l'émergence de règles permettant de déterminer le nombre de centres. Cependant, on constate généralement que l'augmentation de ce dernier diminue l'erreur globale sur les vecteurs d'apprentissages. Ce phénomène conduit souvent à un apprentissage "par cocur" qui réduit le pouvoir de généralisation du réseau. Cette méthode appliquée à notre étude nous a permis d'obtenir un réseau comportant au total 18 centres. Les résultats obtenus montrent, comme pour le classifieur précédent, de faibles capacités à discerner les deux dernières classes. Pour cette raison, nous avons augmenté le nombre de centre dans ces deux classes. Le tableau 3 récapitule le nombre de centres retenus pour notre étude.

Classe	1	2	3	4	5
Nombre de centres	2	3	2	12	2

Tableau 3. Nombre de centres retenus

L'optimisation du seuil de décision a été menée de façon identique à celle de l'étude précédente. Seule la règle de décision est modifiée pour intégrer la notion de rejet de distance. Elle est définie comme suit :

- Rejet de distance (RD) si :  $S_M < (1\text{-seuil})$
- Rejet d'ambiguïté (RA) si :  $S_M > (1\text{-seuil})$  et  $S_{M-1} > (\text{seuil})$
- Affectation à la classe correspondant à  $S_M$  sinon.

Où  $S_M$  et  $S_{M-1}$  sont, par ordre décroissant, les deux sorties les plus élevées.

Le seuil optimisé est égal à 0.265. Les résultats obtenus sont consignés dans les deux matrices ci-dessous. Le critère d'arrêt a été fixé à 1000 itérations.

$$\begin{matrix} & \text{BC} & \text{MC} & \text{RA} & \text{RD} \\ \text{Classe 1} & \left( \begin{array}{cccc} 80 & 0 & 0 & 20 \\ 75 & 0 & 12.5 & 12.5 \\ 83.33 & 0 & 0 & 16.67 \\ 70 & 0 & 5 & 25 \\ 20 & 20 & 0 & 60 \end{array} \right) \\ \vdots & & & & \\ \text{Classe 5} & & & & \end{matrix} = M1$$

$$\begin{matrix} & \text{Classes réelles} \\ \text{Classes prédites} & \left( \begin{array}{ccccc} 100 & 0 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 & 0 \\ 0 & 0 & 100 & 0 & 0 \\ 0 & 0 & 0 & 93.33 & 6.67 \\ 0 & 0 & 0 & 0 & 100 \end{array} \right) \\ & \end{matrix} = M2$$

Les résultats obtenus montrent une nette diminution du nombre de défauts mal classés par rapport au premier classifieur. Ceci est dû principalement à l'intégration du rejet de distance dans le second classifieur. Cette amélioration permet d'augmenter la confiance que l'on peut accorder à la décision du classifieur, notamment pour les deux dernières classes (cf. matrice M2).

### Conclusion

Les classifieurs neuronaux ont été choisis de préférence à d'autres méthodes, en raison de leur rapidité en phase de test et de leur tolérance au bruit.

La taille limitée de la base d'exemples, due au coût prohibitif de la fabrication des défauts par électroérosion, ne permet pas d'arriver à des conclusions définitives, relativement à la validité générale des résultats obtenus. Néanmoins, ces derniers sont encourageants et démontrent la faisabilité de l'étude entreprise.

Les résultats obtenus montrent l'intérêt du rejet de distance (même en *user-supervised classification*), qui permet de diminuer le taux de défauts mal classés, et donc d'augmenter la confiance à accorder à la réponse du réseau.

### Références

- [1] J-C. Bour, M. Meddour, P. Vasseur, A. Billat, "Using flat coils for the detection of defects by eddy-current techniques", Proceeding of the International Conference on Signals and Systems ICSS'94, pp.11.20-11.24, Alger, 24-26 Septembre 1994.
- [2] J-C. Bour, M. Meddour, P. Vasseur, A. Billat, "Détection et reconnaissance de forme de défauts par courants de Foucault pulsés : sélection de paramètres pertinents", Actes de la 2<sup>ème</sup> Conférence Internationale sur l'Automatisation Industrielle, pp.723-728, Nancy, 7-9 juin 1995.
- [3] P.G. Doctor, T.P. Harrington, T.J. Davis, C.J. Morris, "Pattern-Recognition Methods for Classifying and Sizing Flaws Using Eddy-Current Data", *Eddy-current characterization of materials and structures*, eds. Birbaum G., Free G., ASTM STP 722, American Society for Testing and Materials, pp. 464-483, 1981.
- [4] B. Dubuisson, Diagnostic et reconnaissance de forme, Traité des nouvelles technologies, Hermès, Paris, 1990
- [5] A. Fron, Programmation par contraintes (Addison-Wesley, France 1994).
- [6] D.E. Rummelhart, G.E. Hinton, R.J. Williams, "Parallel Distributed Processing : Explorations in the Microstructure of Cognition Vol.1, Foundations", MIT Press, Cambridge, 1986.
- [7] M-H. Masson, "Contribution à l'élaboration d'une méthode de décision avec rejet par réseaux de neurones. Applications à la surveillance de systèmes", Thèse de doctorat, U.T.C. (France), Décembre 1992.
- [8] M. Bishop, Neural networks and their applications, *rev. Sci. Instrum.*, 65 (6), (1994) 1803-1832.
- [9] S. Chen, S. F. N. Cowan, P. M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, *IEEE Trans. Neural Networks*, 2, (1991) 302-309.