

**In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.**

**While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.**

UNIVERSITÉ DE MONTRÉAL

DÉVELOPPEMENT D'UN NEZ ÉLECTRONIQUE
APPLIQUÉ À L'ODEUR DE BIOGAZ

PHILIPPE GALARNEAU-MICONE
DÉPARTEMENT DE GÉNIE CHIMIQUE
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DÎPLOME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE CHIMIQUE)
JUILLET 2003



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisitions et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-86399-9

Our file Notre référence

ISBN: 0-612-86399-9

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :
DÉVELOPPEMENT D'UN NEZ ÉLECTRONIQUE
APPLIQUÉ À L'ODEUR DE BIOGAZ

présenté par : GALARNEAU-MICONE Philippe

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. KLIVANA Danilo, Ph. D., Président

M. GUY Christophe, Ph. D., membre et directeur de recherche

M. PERRIER Michel, Ph. D., membre et codirecteur de recherche

M. BERTRAND François, Ph. D., membre

À mes parents

RÉSUMÉ

L'instauration de normes régissant la concentration des odeurs aux abords de sites industriels nécessite l'utilisation de techniques qui permettent la quantification de celles-ci. Dans le même ordre d'idée, l'évaluation de la performance des techniques d'abattement des odeurs doit pouvoir être mesurée objectivement. L'odeur à l'étude dans le présent mémoire est celle du biogaz, l'émission principale et la première cause de préoccupation des riverains des sites d'enfouissement sanitaire.

La technique actuelle de mesure de l'odeur est l'olfactométrie, une méthode longue et onéreuse reposant sur l'expertise d'un jury humain. Des techniques de mesure des odeurs ont été tentées avec des instruments analytiques, comme des GC/MS ou des GC/FID, sans grand succès, la relation entre la concentration de l'odeur et celles des composés mesurés par les instruments étant difficiles à réaliser.

Une nouvelle technique employée dans la quantification des odeurs est celle du nez électronique. Le terme *nez électronique* est utilisé lorsqu'un réseau de capteurs chimiques non spécifiques est jumelé à une unité de traitement informatique dans le but de reconnaître ou de mesurer la concentration d'un gaz ou d'une odeur. Le nez électronique permet de s'affranchir de la subjectivité humaine de l'olfactométrie et des limitations des instruments analytiques.

La relation entre les signaux provenant des capteurs chimiques d'un nez électronique et la concentration-odeur du mélange gazeux mesuré est hautement non-linéaire. Cette non-linéarité implique l'utilisation de techniques de régression pouvant contourner la malédiction de la dimensionalité inhérente à l'approximation de ce type de fonctions. Les machines neuronales, utilisées comme approximateurs universels, couplées à un algorithme de minimisation du risque structurel, permettent d'estimer efficacement les relations non-linéaires à plusieurs variables et de minimiser l'erreur de généralisation du modèle testé.

L'ajout de l'algorithme de minimisation du risque structurel permet de déterminer avec certitude le modèle optimal de machine neuronale devant être utilisé avec l'odeur étudiée et le nez électronique employé. Pour faire l'estimation de la relation entre les concentrations-odeurs et les signaux des capteurs, les machines neuronales utilisées sont les perceptrons multicouches (*multilayer perceptrons*) ou réseaux de neurones artificiels (*artificial neural networks*), les réseaux de fonctions à base radiale (*radial basis functions*) et les machines à supports vectoriels (*support vector machines*).

Pour le nombre de données disponibles, la machine neuronale minimisant l'erreur de généralisation pour l'ensemble de la plage de concentration-odeur à approximer, soit de 0 u.o./m³ (unité-odeur par mètre cube) à 200 u.o./m³, est le réseau de neurones artificiel à deux couches cachées, 4 neurones pour la première couche et 24 pour la seconde, et à sortie linéaire unique. Pour les concentrations-odeurs inférieures à 30 u.o./m³, la

machine à support vectoriel a été de loin la plus performante. Les caractéristiques de celle-ci sont un noyau de fonctions à bases radiales avec un écart-type de 0.2 et une constante de coût C de 1.

La faisabilité de l'utilisation d'un nez électronique comme outil de mesure de la concentration de l'odeur de biogaz a été prouvée. Il a aussi été démontré que pour une même odeur provenant de deux sources différentes, le modèle de machine neuronale minimisant l'erreur de généralisation était le même.

ABSTRACT

The respect of the standards governing the odor concentration at the perimeter of industrial sites requires the use of techniques which allow the quantification of odors. Furthermore, the evaluation of the performance of the odor abatement techniques must be measured objectively. In the present project, the quantification of the odor of biogas is being studied. Biogas is the principal release and the first cause of concern of the residents near wastetreatment works.

Olfactometry is currently the technique used worldwide for odor measurement. It is a time-consuming and expensive method relying on human expertise. Measurement of odors were tried with analytical instruments, like GC/MS or GC/FID, without much success. The relationships between the odor concentration and those of the compounds measured by the instruments are difficult to realize.

A new technique employed in the quantification of odors is the electronic nose. The term *electronic nose* is used when a network of nonspecific chemical sensors is twinned with a data-processing treatment unit with an aim of recognizing or measuring the concentration of a gas or an odor. The electronic nose makes it possible to get rid of the human subjectivity of the olfactometry and the limitations of the analytical instruments.

The relationship between the signals coming from the chemical sensors of an electronic nose and the odor concentration of the measured gas mixture is highly non-linear. This non-linearity implies the use of regression techniques able to circumvent the curse of dimensionality inherent in the approximation of this type of functions. The neural machines, used as universal approximators, coupled to an algorithm of structural risk minimization, make it possible to effectively construct non-linear relations at several variables and to minimize the generalization error of the model tested.

The structural risk minimization algorithm permits the determination of the optimal model of neural machine to be used with the odor studied and the electronic nose employed. To make the estimation of the relationship between the odor concentrations and the sensors signals, the neural machines tested are multilayer perceptrons (or artificial neural networks), the radial basis functions networks and the support vector machines.

For the amount of data available, the neural machine minimizing the generalization error on the total interval of interest, 0 o.u./m³ (odor unit per cubic meter) to 200 o.u./m³, is the artificial network with two hidden layers, 4 neurons for the first layer and 24 for the second, and a single linear exit. For the odor concentrations lower than 30 u.o./m³, the support vector machine was by far most powerful. The characteristics of the most powerful one was a kernel of radial basis functions, with standard-deviation of 0.2 and a constant of cost C of 1.

The feasibility of using an electronic nose as an odor concentration measurement instrument for biogas was proven. It has been shown also that for the same odor coming from two different sources, the model of neural machine minimizing the generalization error is the same for both sources.

TABLE DES MATIÈRES

Dédicace	iv
Résumé	v
Abstract	viii
Table des matières	xi
Liste des tableaux	xiv
Liste des figures	xv
 Introduction	 1
 CHAPITRE 1 – Théorie et revue de la littérature scientifique	 6
La théorie de l'apprentissage statistique	6
<i>Le principe de minimisation du risque empirique</i>	9
<i>La dimension VC</i>	11
<i>La minimisation du risque structurel</i>	14
<i>Conclusion</i>	19
Les réseaux de neurones artificiels	20
<i>Introduction et révision</i>	20
<i>Particularités sur la capacité de généralisation</i>	24
Les réseaux de fonctions à base radiale	28
<i>Introduction et révision</i>	28
<i>Particularités sur la capacité de généralisation</i>	30
Les machines à supports vectoriels	33
<i>L'hyperplan optimal</i>	33
<i>Construction de l'hyperplan optimal</i>	36
<i>Généralisation pour le cas non-séparable</i>	39
<i>Généralisation pour un grand espace de projection</i>	42

<i>Machines à supports vectoriels pour la régression</i>	46
La quantification des odeurs	51
<i>L'olfactométrie</i>	54
<i>Les instruments analytiques classiques</i>	55
<i>Le nez électronique</i>	57
<i>La quantification des odeurs environnementales par nez électroniques</i>	63
<i>dans la littérature scientifique</i>	63
CHAPITRE 2 - Objectifs	74
CHAPITRE 3 - Méthodologie	77
Échantillonnage et olfactométrie	77
Le nez électronique	80
Le pré-traitement	83
Entraînement et validation des machines neuronales	86
<i>Réseaux de neurones artificiels</i>	87
<i>Les réseaux de fonctions à bases radiales</i>	89
<i>Les machines à supports vectoriels</i>	90
CHAPITRE 4 - Résultats et discussion	91
Pré-traitement	91
Le respect de la théorie de l'apprentissage statistique	97
Les machines optimales pour l'ensemble des sources de biogaz	103
Machine neuronale	109
Le nombre de données d'entraînement	110
Un modèle optimal unique fonction du nombre de données d'entraînement	113
Comparaison des performances	120
Conclusion	121
Conclusion	122
Recommandations	125

Bibliographie	126
ANNEXE 1 – Calcul de la concentration-odeur	132
ANNEXE 2 – Liste des fichiers électroniques fournis sur la disquette	139

LISTE DES TABLEAUX

Tableau 1.1 Principaux composés constituant le biogaz (Barlaz <i>et al.</i> , 1989).....	52
Tableau 1.2 Principaux composés traces constituant l'odeur de biogaz.....	53
Tableau 1.3 Tableau comparatif des principales caractéristiques des capteurs (Haugen et Kvaal, 1998).....	62
Tableau 4.1 Modèles optimaux de machines neuronales.....	104
Tableau 4.2 Erreur de généralisation des machines neuronales optimales pour l'approximation des faibles concentrations-odeurs.....	109
Tableau 4.3 Erreur de généralisation selon la source des données d'entraînement	120

LISTE DES FIGURES

Figure 1.1 Modélisation de l'erreur	6
Figure 1.2 Vecteurs séparables et non-séparables linéairement dans le plan	12
Figure 1.3 Schématisation de la structure imbriquée d'ensembles de fonctions	16
Figure 1.4 Borne du risque exprimée comme la sommation du risque empirique	17
Figure 1.5 Schématisation du réseau de neurone à deux couches cachées	22
Figure 1.6 Schématisation de l'arrêt prématuré	25
Figure 1.7 Réseau de fonctions à base radiale	29
Figure 1.8 Approximation locale par fonctions gaussiennes	31
Figure 1.9 Hyperplan optimal avec la marge Δ de chacun des côtés.....	35
Figure 1.10 Hyperplan optimal pour le cas de données non-séparables linéairement	39
Figure 1.11 Projection du vecteur d'entrée dans un espace de plus grande dimension ...	42
Figure 1.12 Marge et fonction de perte ε -Insensitive.....	47
Figure 1.13 Schématisation de la régression avec des machines à supports vectoriels ...	49
Figure 1.14 Comparaison entre le nez électronique et le système olfactif humain.....	58
Figure 1.15 Explication du fonctionnement d'un capteur SOM	60
Figure 1.16 Graphique de la concentration en H_2S des mélanges odorants en fonction de deux plages de concentration-odeur, A) 0 – 400000 u.o/m ³ et B) 0 – 4000 u.o/m ³ (Stuetz <i>et al.</i> , 1998).....	66
Figure 1.17 Graphique de la sortie du nez électronique sous forme de variable canonique en fonction de la concentration-odeur (Stuetz <i>et al.</i> , 1998)	67
Figure 1.18 Graphique de la sortie du nez électronique sous forme de variable canonique en fonction de la concentration-odeur (Stuetz <i>et al.</i> , 1998)	68
Figure 1.19 Graphique de la sortie du nez électronique brute en fonction de la concentration-odeur (Stuetz <i>et al.</i> , 1998).....	69

Figure 1.20 Graphique de la concentration-odeur en fonction de la de la quantité totale de carbone organique (TOC) (Bockreis et Jager, 1999)	71
Figure 3.1 Pouvoir odorant des composés odorants des deux champs à l'étude.....	77
Figure 3.2 Échantillonnage du biogaz à la source.....	78
Figure 3.3 Fût et opérateur de l'olfactomètre.....	79
Figure 3.4 Jurée d'olfactométrie.....	80
Figure 3.5 Photographie du nez électronique.....	82
Figure 3.6 Réponse d'un capteur.....	84
Figure 4.1 Rapports R_S/R_0 des capteurs en fonction de la concentration-odeur pour l'ensemble des résultats expérimentaux.....	92
Figure 4.2 Rapports R_S/R_0 du capteur 03 en fonction de la concentration-odeur	95
Figure 4.3 Rapports R_S/R_0 du capteur 09 en fonction de la concentration-odeur	96
Figure 4.4 Erreur de généralisation (MSE) en fonction du nombre de neurones de la 1 ^{ère} couche cachée (la seconde est fixée à 12 neurones)	99
Figure 4.5 Erreur empirique en fonction de l'erreur de généralisation pour des réseaux de neurones entraînés et validés avec les mêmes points expérimentaux	100
Figure 4.6 Erreur de généralisation (MSE) en fonction du nombre de neurones (centres) de la couche cachée du réseau de fonctions à bases radiales	102
Figure 4.7 Comparaison des valeurs calculées par le réseau de neurones artificiel optimal et les valeurs attendues.....	105
Figure 4.8 Comparaison des valeurs calculées par un réseau de neurones artificiel non optimal et les valeurs attendues.....	106
Figure 4.9 Comparaison des valeurs calculées par le réseau de fonctions à bases radiales optimal et les valeurs attendues.....	107
Figure 4.10 Comparaison des valeurs calculées par la machine à supports vectoriels optimale et les valeurs attendues.....	108
Figure 4.11 Erreur de généralisation en fonction du nombre de données d'entraînement	111

Figure 4.12 Erreur de généralisation en fonction du nombre de paramètres pour la source de biogaz A modélisée par réseaux de neurones artificiels.....	114
Figure 4.13 Erreur de généralisation en fonction du nombre de paramètres pour la source de biogaz B modélisée par réseaux de neurones artificiels.....	114
Figure 4.14 Erreur de généralisation en fonction du nombre de centres pour la source de biogaz A modélisée par réseaux de fonctions à bases radiales	115
Figure 4.15 Erreur de généralisation en fonction du nombre de centres pour la source de biogaz B modélisée par réseaux de fonctions à bases radiales	116
Figure 4.16 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau polynomial pour la source de biogaz A	117
Figure 4.17 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau polynomial pour la source de biogaz B.....	117
Figure 4.18 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau de fonction à bases radiales pour la source de biogaz A.....	118
Figure 4.19 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau de fonction à bases radiales pour la source de biogaz B.....	119

INTRODUCTION

L'instauration de normes gouvernementales propres à la nuisance olfactive a débuté dans les années quatre-vingt, lorsque le concept général de pollution a été élargi à la détérioration de la qualité de vie, englobant également la pollution sonore et visuelle. Au cours des années qui ont suivi, la majorité des pays développés ont établi des règles claires quant au désagrément que pouvaient entraîner les émissions odorantes d'un site industriel pour les résidants jouxtant celui-ci. Ces règles impliquent l'utilisation de techniques et d'instruments permettant la quantification des odeurs afin de respecter une limite précise en ce qui a trait à la concentration des odeurs.

Les techniques développées pour faire le suivi et la surveillance des odeurs et s'assurer du respect des limites permises de concentration-odeur reposent essentiellement sur l'expertise humaine (*inspecteur-odeur*, jury olfactométrique, programme de suivi avec les citoyens). Ceci entraîne une demande importante en ressources logistiques et financières. Les résultats obtenus sont aussi en partie tributaires de la qualité du personnel employé. Le problème inhérent à cette situation est l'impossibilité de faire une campagne de caractérisation et de modélisation de l'impact de la nuisance olfactive d'un site industriel problématique sans un investissement significatif d'argent et de temps.

Pour plusieurs types de sites industriels, les émissions odorantes sont la source principale de pollution affectant les résidants avoisinants. Le taux des émissions peut

varier dans le temps et rendre circonstanciées les analyses olfactives ponctuelles. Les sites d'enfouissement sanitaire, objet de la présente recherche, sont des exemples d'activités industrielles malodorantes n'émettant pas de façon continue leurs odeurs. Les émissions odorantes sont liées à la production de biogaz, résultat du processus de dégradation biologique de la matière organique qui se met en place rapidement lors d'un stockage de produits fermentescibles et qui fluctue sur une longue période. Un suivi en continu des odeurs est donc nécessaire pour s'assurer du respect des normes en tout temps et, dans une optique d'amélioration de la gestion du site, pour identifier les causes du dépassement de celles-ci. En effet, il est impossible de mener une campagne de réduction des émissions si celles-ci ne peuvent être quantifiées.

Plusieurs essais de surveillance d'odeurs ont été tentés avec des instruments analytiques classiques, comme des FID (*flame ionization detector*) ou des GC/MS (*gas chromatography/mass spectrometry*). Malheureusement, la construction d'une relation mathématique valable entre la concentration des composés détectés et la concentration-odeur s'est butée à la complexité des effets du mélange des molécules. Les nombreuses synergies et inhibitions odorifères entre les molécules ne sont pas toutes décrites si seulement une dizaine sont mesurées alors que jusqu'à une centaine peuvent être présentes. De plus, le temps de réponse ou de traitement des données de ce type de montage ne permet souvent pas de le qualifier d'instrument de mesure en continu.

L'utilisation d'un nez électronique comme outil de mesure a la particularité de permettre de s'affranchir des problèmes de temps de réponse et de mono-détection des molécules. Les capteurs chimiques, utilisés comme interface de relation entre le mélange odorant gazeux et le système logiciel effectuant la quantification, ont une dynamique de réponse beaucoup plus rapide que les instruments analytiques classiques; de plus, ils ont la particularité de réagir à des familles chimiques ou à des groupes caractéristiques et non pas à un seul composé. Cette non-spécificité des capteurs permet de tirer une grande quantité d'information générale sur le mélange odorant dont peuvent être ensuite extraites des caractéristiques spécifiques.

La relation entre les signaux provenant des capteurs d'un nez électronique et la concentration-odeur du mélange gazeux ainsi mesuré est hautement non-linéaire. Cette non-linéarité implique l'utilisation de techniques de régression pouvant contourner la malédiction de la dimensionalité inhérente à l'approximation de ce type de fonctions. Les machines neuronales (au nombre desquelles se trouvent les réseaux de neurones), utilisées comme approximateurs universels et couplées à un algorithme de minimisation du risque structurel, permettent à la fois d'estimer efficacement les relations non-linéaires à plusieurs variables et de minimiser l'erreur de généralisation du modèle testé. L'ajout de l'algorithme de minimisation du risque structurel permet de déterminer avec certitude le modèle optimal de machine neuronale devant être utilisé avec l'odeur étudiée et le nez électronique employé selon le nombre de données d'entraînement disponibles ou souhaitées.

Le mémoire qui suit est divisé en quatre sections, soit une revue de la littérature scientifique, la présentation des objectifs et de la méthodologie, la présentation et l'analyse des résultats et, finalement, la conclusion et les recommandations.

La revue de la littérature débute par la présentation des travaux effectués à ce jour dans le domaine de la quantification des odeurs à l'aide d'instruments analytiques et de nez électroniques. Ensuite, une brève présentation de la théorie de l'apprentissage statistique précède la présentation des machines neuronales utilisées pour l'approximation de la relation existant entre les lectures des capteurs du nez électronique et la concentration-odeur mesurée, soit les perceptrons multicouches (*multilayer perceptrons*), les réseaux de fonctions à base radiale (*radial basis functions*) et les machines à support vectoriel (*support vector machines*).

La seconde section s'intéresse aux objectifs de recherche découlant des informations révélées par la revue de la littérature scientifique et à la méthodologie permettant de les atteindre.

La troisième partie de ce mémoire présente les résultats intéressants et signale les informations pertinentes permettant de remplir les objectifs de la recherche.

Finalement, la conclusion rappelle les objectifs fixés ainsi que les résultats probants et les recommandations présentent les pistes de continuation du projet de recherche.

CHAPITRE 1 – THÉORIE ET REVUE DE LA LITTÉRATURE SCIENTIFIQUE

LA THÉORIE DE L'APPRENTISSAGE STATISTIQUE

La théorie statistique de l'apprentissage fournit des outils d'estimation du risque et des méthodologies permettant la recherche de modèles optimaux affranchis de la nécessité de connaître la distribution des données et, dans certain cas, d'avoir des données en grande quantité. De ces outils et méthodologies, seuls ceux applicables à l'approximation de fonctions sont présentés.

L'objectif de la modélisation est d'arriver à identifier une fonction d'approximation réelle mais inconnue dans un espace d'hypothèse qui approxime le plus justement la vraie fonction de l'espace cible. Les erreurs possibles sont de deux types, soit les erreurs d'approximation et les erreurs d'estimation.

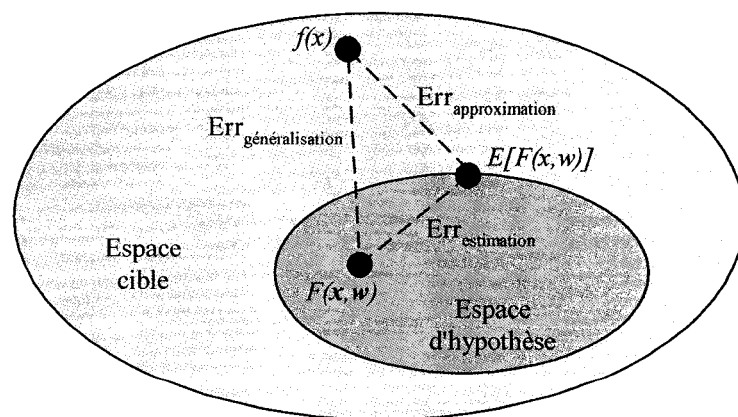


Figure 1.1 Modélisation de l'erreur

Les erreurs d'approximation surviennent lorsque l'espace d'hypothèse est trop petit par rapport à l'espace cible, entraînant la possibilité que la fonction réelle se trouve à l'extérieur de l'espace d'hypothèse. Les erreurs d'estimation résultent du choix d'un modèle non-optimal dans l'espace d'hypothèse suite à des lacunes dans le processus d'apprentissage. La somme de ces erreurs constitue l'erreur de généralisation, c'est-à-dire l'erreur visible lorsque la fonction est utilisée avec de nouvelles données n'ayant pas servies à son élaboration (son *entraînement*). Le développement théorique qui suit présente la théorie et la méthodologie, telles que construites et exposées par Vapnik (Vapnik, 1995), permettant de réduire le risque d'utilisation de la fonction, soit la minimisation de l'erreur de généralisation.

Soit une fonction f qui associe un scalaire d au vecteur d'entrées connu $\mathbf{x}$

$$d = f(\mathbf{x}) \quad (1.1)$$

où d représente la réponse scalaire mesurée du système étudié au vecteur connu $\mathbf{x}$ des entrées. Seul la relation entre les entrées et la sortie est inconnue. Soit un ensemble de N paires de données expérimentales

$$\Gamma = \{(\mathbf{x}_i, d_i)\}_{i=1}^N \quad (1.2)$$

indépendantes et identiquement distribuées tirées de cet environnement expérimental. Le processus d'apprentissage consiste à identifier la fonction $F(\mathbf{x}, \boldsymbol{\alpha})$, où $\boldsymbol{\alpha}$ représente les paramètres de la fonction, permettant l'approximation la plus juste de la fonction réelle inconnue $f(\mathbf{x})$. La fonction $F(\mathbf{x}, \boldsymbol{\alpha})$ peut être une machine neuronale si cette dernière se qualifie à titre d'approximateur universel (modèle générique permettant d'approcher une

fonction non-linéaire suffisamment régulière inconnue avec une certaine précision). Toutes les machines neuronales mentionnées dans ce mémoire, soit les réseaux de neurones artificiels, les réseaux de fonctions à bases radiales et les machines à support vectoriels, ont la propriété de se comporter comme des approximateurs universels.

Dans le cas idéal, la fonction de distribution des données d'entrées $F_{x,D}(\mathbf{x})$ est connue et le nombre de données est infini ($N \rightarrow \infty$). Le processus d'apprentissage consiste alors à minimiser la fonctionnelle du risque, $R[F]$, mesurant, pour la fonction $F(\mathbf{x}, \boldsymbol{\alpha})$, l'amplitude et la propension à faire des erreurs de généralisation,

$$R[F] = \int Q(\mathbf{x}, \boldsymbol{\alpha}) dF_{x,D}(\mathbf{x}) \quad (1.3)$$

où $Q(\mathbf{x}, \boldsymbol{\alpha})$ est la fonction de perte, ou d'erreur, mesurant l'écart entre la valeur désirée et celle obtenue de la fonction d'approximation. Une forme habituelle pour la fonction de perte est le carré de la différence entre la réponse désirée et celle obtenue de la fonction d'approximation.

$$Q(\mathbf{x}, \boldsymbol{\alpha}) = (d - F(\mathbf{x}, \boldsymbol{\alpha}))^2 \quad (1.4)$$

Malheureusement, la connaissance de la fonction de distribution des données est extrêmement rare dans la pratique, tout comme des données en quantité presque infinie. L'évaluation du risque, donc de l'habilité de généralisation, d'une fonction devient impossible si la formulation de la fonctionnelle du risque n'est pas adaptée.

Le principe de minimisation du risque empirique

La méthode habituellement employée pour minimiser la fonctionnelle du risque, lorsque la fonction de distribution $F_{X,D}(\mathbf{x})$ est inconnue, est résumée dans le principe inductif suivant (Vapnik, 1995) :

1. La fonctionnelle du risque $R[F]$ est approchée à la fonctionnelle du risque empirique $R_{emp}[F]$

$$R_{emp}[F] = \frac{1}{N} \sum_{i=1}^N Q(\mathbf{x}_i, \mathbf{a}) \quad (1.5)$$

construite sur la base de l'ensemble Γ des données expérimentales.

2. La fonction $Q(\mathbf{x}, \mathbf{a}_0)$ qui minimise le risque est approximée par la fonction $Q(\mathbf{x}, \mathbf{a}_N)$ minimisant le risque empirique

Appliquée avec la fonction de perte présentée ci-dessus, le principe de minimisation du risque empirique consiste à minimiser la fonction

$$R_{emp}[F] = \frac{1}{N} \sum_{i=1}^N (d_i - F(\mathbf{x}_i, \mathbf{a}))^2 \quad (1.6)$$

et la méthode résultante est tout simplement celle des moindres-carrés.

Par contre, une faille énorme existe dans l'utilisation du principe de minimisation du risque empirique comme estimation du risque de généralisation. La minimisation du risque empirique n'est valable que si

$$\lim_{N \rightarrow \infty} R_{emp}[F] = R[F] \quad (1.7)$$

ce qui est vrai si le nombre de données est en quantité suffisante, et que celles-ci sont bien distribuées, pour que soit respectée la loi des grands nombres. Une seconde condition doit être remplie, soit que

$$\lim_{N \rightarrow \infty} \min_{f \in H^n} R_{emp}[f] = \min_{f \in H^n} R[f] \quad (1.8)$$

qui n'est valide que lorsque l'espace H^n , l'espace dans lequel est défini la fonction, est assez petit. Cette condition est moins intuitive et requiert que les minima convergent aussi. Ceci signifie donc que pour que l'application du principe de minimisation du risque empirique soit acceptable comme estimation du risque de généralisation, il faut absolument utiliser un très grand nombre de données d'entraînement et connaître la dimension de l'espace définissant le modèle optimal. En conséquence, même si le nombre de données expérimentales consacrées à l'entraînement est en nombre suffisant pour croire que la première condition est respectée, il pourra toujours subsister un doute quant à la qualité du modèle utilisé pour réaliser l'approximation de la fonction.

Cette situation entraîne la nécessité d'évaluer le risque non pas seulement par rapport à l'erreur empirique mais aussi par rapport au nombre de données disponibles et à l'habilité du modèle à fournir une généralisation adéquate selon sa forme. Ces nouvelles exigences constituent l'objectif du principe de la minimisation du risque structurel.

Sans nécessairement quantifier le risque exactement, il est possible de définir une borne supérieure sur le risque qui soit fonction de ces nouveaux éléments et qui permet de déterminer l'architecture du modèle de fonctions fournissant la généralisation la plus performante.

Pour comprendre comment il est possible d'évaluer le risque de généralisation selon les nouveaux éléments soulevés, il est nécessaire d'introduire le concept de dimension VC (VC pour Vapnik-Chervonenkis, les deux chercheurs ayant défini celle-ci).

La dimension VC

La dimension VC est une valeur scalaire mesurant la capacité d'un ensemble de fonctions. Il est plus aisé de comprendre le concept de dimension VC en commençant par l'explication de la dimension VC pour les fonctions discriminantes à deux classes et ensuite le généraliser aux fonctions réelles.

La dimension VC d'un ensemble de fonctions discriminantes $S(\mathbf{x}, \alpha)$, $\alpha \in \Lambda$, est le nombre maximal h de vecteurs $x_1, \dots, x_h$ linéairement indépendants positionnés selon les 2^h façons possibles pouvant être séparés parfaitement en deux classes (Vapnik et Chervonenkis, 1971). Le graphique qui suit illustre le cas de vecteurs à deux dimensions positionnées dans le plan.

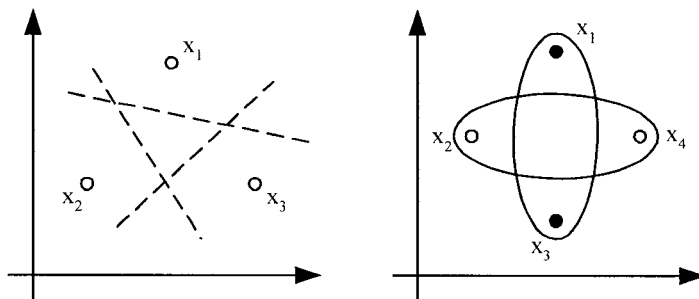


Figure 1.2 Vecteurs séparables et non-séparables linéairement dans le plan

La dimension VC des lignes séparatrices dans le plan est égale à trois. Comme le montre la partie de gauche de la figure, trois vecteurs à deux dimensions peuvent être en tout temps séparés par une ligne en deux classes distinctes. L'ajout d'un vecteur supplémentaire rend la séparation impossible, tel que montré dans la partie droite de la figure, les vecteurs x_2 et x_4 ne pouvant être séparés par une ligne des vecteurs x_1 et x_3 .

Le cas précédent se généralise pour tout hyperplan à n dimensions. La dimension VC d'un hyperplan défini dans l'espace $\mathbf{R}^n$ est égale à $n + 1$, puisque $n + 1$ points pourront toujours être positionnés de façon à ce que si l'un d'eux est choisi comme origine, les autres seront toujours linéairement indépendants, mais jamais $n + 2$ points (sachant que $n + 1$ vecteurs dans $\mathbf{R}^n$ ne sauraient être linéairement indépendants). Ce résultat est intéressant puisqu'il servira de base dans la compréhension de l'avancée que représentent les machines à supports vectoriels dans l'apprentissage statistique (Burges, 1998).

La définition de la dimension VC d'un ensemble de fonctions réelles est beaucoup moins intuitive (Vapnik, 1982).

Soit

$$A \leq S(\mathbf{x}, \alpha) \leq B, \alpha \in \Lambda \quad (1.9)$$

un ensemble, en faisant varier α , de fonctions réelles bornées par les constantes A et B (A peut tendre vers $-\infty$ et B vers ∞). Supposons un ensemble de fonctions discriminantes

$$I(x, \alpha, \beta) = \theta[S(\mathbf{x}, \alpha) - B], \alpha \in \Lambda, \beta \in (A, B) \quad (1.10)$$

où $\theta(x)$ est la fonction échelon

$$\theta(z) = \begin{cases} 0 & \text{si } z < 0 \\ 1 & \text{si } z \geq 0 \end{cases} \quad (1.11)$$

La dimension VC de l'ensemble de fonctions réelles $S(\mathbf{x}, \alpha)$, $\alpha \in \Lambda$ est définie comme la dimension VC de l'ensemble de fonctions discriminantes défini ci-haut avec les paramètres $\alpha \in \Lambda$, $\beta \in (A, B)$.

Il arrive souvent que la dimension VC d'un ensemble de fonctions ne puissent pas être calculé. Ceci est particulièrement vrai dans le cas des machines neuronales. Les ordres de grandeur des dimensions VC peuvent être déterminés, mais pas leurs valeurs exactes. Il est alors habituellement possible de trouver une borne sur la dimension VC de l'ensemble de fonctions ou, plus précisément, de la machine neuronale.

La minimisation du risque structurel

La théorie de l'apprentissage statistique telle que développée par Vapnik a démontré la supériorité du principe de minimisation du risque structurel (risque associé au modèle) sur celui de la minimisation du risque empirique (risque associé uniquement à l'erreur commise lors de l'entraînement) pour la construction de machine neuronale utilisée comme approximateur universel (Gunn, 1998). Le principe de la minimisation du risque structurel permet d'obtenir le modèle de machine neuronale garantissant la meilleure généralisation.

Vapnik (Vapnik, 1995) a introduit la borne supérieure suivante sur le risque de généralisation d'une fonction $F(\mathbf{x}, \alpha)$, avec une probabilité $1 - \eta$,

$$R[F] \leq R_{emp}[F] + \sqrt{\frac{h \ln\left(\frac{2N}{h} + 1\right) - \ln\left(\frac{\eta}{4}\right)}{N}} \quad (1.12)$$

où le premier terme du membre de droite est le risque empirique et le second un intervalle de confiance sur l'habilité de généralisation de la fonction $F(\mathbf{x}, \alpha)$, évalué à l'aide de la capacité de la fonction en fonction de la dimension VC h et du nombre de données d'entraînement N .

Lorsque le rapport h/N est petit, soit $h/N < 0.05$, le second terme du membre de droite l'est aussi par rapport au risque empirique. Dans ce cas, une petite valeur du risque tel que déterminé par le principe de la minimisation du risque empirique garantit un faible risque de généralisation. Lorsque le rapport h/N est grand, ce qui est habituellement le cas, une faible valeur du risque empirique ne garantit pas un faible risque de généralisation. Dans ce cas, il devient nécessaire de minimiser simultanément les deux éléments du membre de droite de l'expression du risque, ce qui constitue l'objectif du principe de minimisation du risque structurel. Il est important de noter que le premier terme du membre de droite dépend uniquement d'une fonction de l'ensemble de fonctions alors que le second dépend de tout l'ensemble de fonctions. Pour minimiser le côté droit de la fonctionnelle du risque il devient alors nécessaire de faire de la dimension VC une variable de contrôle. Les explications qui suivent présentent le principe inductif de la minimisation du risque structurel permettant la minimisation simultanée du risque empirique et de l'intervalle de confiance (Vapnik, 1995).

Soit l'ensemble Φ de fonctions $S(\mathbf{x}, \alpha)$, $\alpha \in \Lambda$, imbriqué dans une structure de sous-ensembles de fonctions

$$\Phi_i = \left\{ S(\mathbf{x}, \alpha), \alpha \in \Lambda_i \right\}_{i=1}^p \quad (1.13)$$

tel que

$$\Phi_1 \subset \Phi_2 \subset \dots \subset \Phi_n \quad (1.14)$$

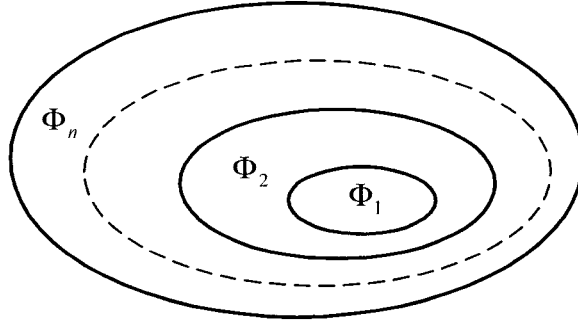


Figure 1.3 Schématisation de la structure imbriquée d'ensembles de fonctions

Où les éléments de la structure satisfont les deux propriétés

1. La dimension VC h_i de chacun des ensembles Φ_i de fonctions est finie, donc que

$$h_1 \leq h_2 \leq \dots \leq h_n \quad (1.15)$$

2. N'importe quel élément Φ_i de la structure contient un ensemble de fonction totalement bornée

$$0 \leq S(\mathbf{x}, \alpha) \leq B_i, \alpha \in \Lambda_i, \forall i \quad (1.16)$$

Cette structure est alors appelée une *structure admissible*. Pour un certain ensemble de données expérimentales, le principe de minimisation du risque structurel choisit la fonction $S(\mathbf{x}, \alpha)$ qui minimise le risque empirique dans le sous-ensemble Φ_i pour lequel le risque garanti de généralisation, déterminé par le membre de droite de la fonctionnelle du risque, est minimal.

Le principe de minimisation du risque structurel permet de trouver le juste milieu entre la qualité de l'approximation pouvant être tirée des données fournies et la complexité du modèle de la fonction d'approximation. La figure qui suit représente graphiquement ce principe.

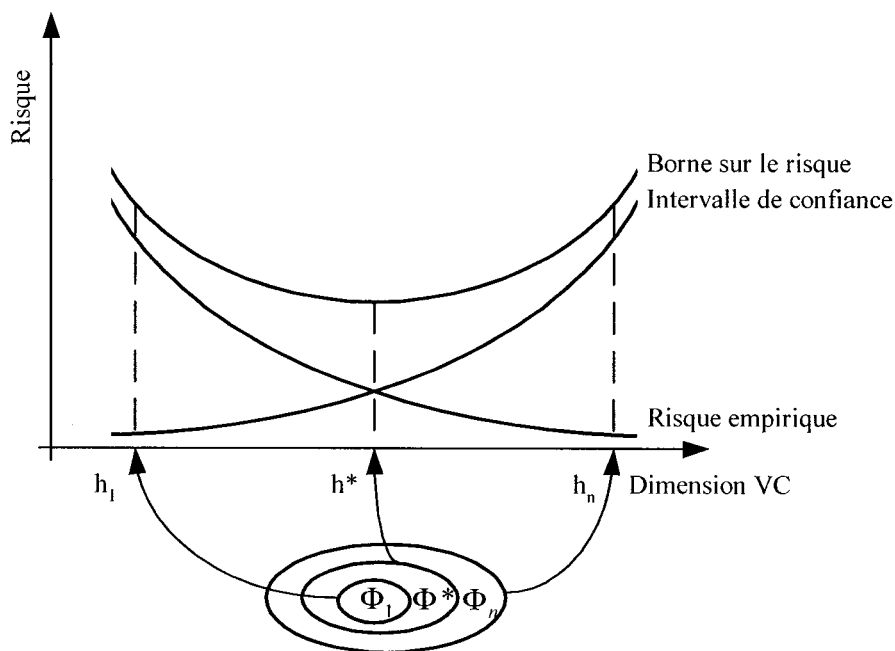


Figure 1.4 Borne du risque exprimée comme la sommation du risque empirique et l'intervalle de confiance

Il existe deux approches constructives permettant l'incorporation de ce principe dans des algorithmes d'apprentissage pour machines neuronales agissant comme approximateurs universels. Les explications qui suivent sont tirées des commentaires faits par Vapnik (Vapnik, 1995) par rapport à la théorie de l'apprentissage statistique appliquée aux machines neuronales.

La première approche consiste à fixer l'intervalle de confiance en utilisant l'ensemble de fonctions présentant la capacité idéale pour le problème d'approximation et de minimiser le risque empirique. La recherche du modèle de fonctions approprié peut se faire en fixant à une valeur N le nombre de données d'entraînement et par la construction

d'une structure de sous-ensemble de fonctions ayant chacune leur dimension VC h^* correspondante. La valeur de la dimension VC fixe l'intervalle de confiance pour chacun des ensembles de fonctions. Le processus d'apprentissage consiste donc à choisir l'élément approprié de la structure admissible, pour une quantité fixe de données expérimentales, qui minimise l'erreur empirique au maximum.

Le piège inhérent à cette approche est l'utilisation d'une fonction trop complexe pour le nombre de données d'entraînement. La conséquence directe est une sur-précision (*overfitting*) de la fonction résultante. Le risque empirique est réduit à zéro par la grande capacité de la fonction à modéliser les données d'entraînement, mais l'intervalle de confiance devient très grand. Pour éviter la sur-précision, la capacité, ou dimension VC, de la fonction doit être relativement petite. La difficulté réside donc à identifier un modèle dont la capacité n'est ni trop grande, ni trop faible, ce qui est difficile à faire sans connaissances *a priori* du problème d'approximation.

Les réseaux de neurones et de fonctions à bases radiales sont des types de machines neuronales utilisées comme approximateurs universels implantant cette première approche. La prochaine section traite des réseaux neuronaux et explique plus spécifiquement cette méthode.

La seconde approche consiste à maintenir le risque empirique à une valeur fixe, idéalement zéro, et de minimiser l'intervalle de confiance.

Les machines à supports vectoriels implantent cette deuxième approche. La construction de machines à supports vectoriels étant l'implantation directe de cette approche, les explications relatives à cette seconde approche seront données dans la section présentant les machines à supports vectoriels utilisées comme approximateurs universels. Cette section suit celles explicitant les réseaux neuronaux.

Conclusion

La conclusion importante de cette section est que l'utilisation du principe de minimisation du risque structurel est supérieure à celle du principe de minimisation du risque empirique pour la détermination d'une fonction d'approximation, ou plus précisément, dans le cas qui nous concerne, de machines neuronales jouant le rôle d'approximateurs universels. De plus, l'utilisation du principe de minimisation du risque structurel garantit la construction du modèle présentant la capacité de généralisation la plus élevée.

LES RÉSEAUX DE NEURONES ARTIFICIELS

Introduction et révision

Les réseaux de neurones, aussi appelés perceptrons multicouches, sont les premiers approximateurs universels de fonctions rencontrés dans cette section. Ceux-ci implantent la première stratégie rencontrée dans la section précédente sur la minimisation du risque structurel, soit garder l'intervalle de confiance sur le risque de généralisation fixe et minimiser le risque empirique.

Un réseau de neurones artificiel à deux couches cachées, à vecteur d'entrées réelles x et à sortie linéaire réelle y prend la forme schématique présentée à la page suivante.

L'exemple donné considère un vecteur d'entrées à m dimensions, suivi de deux couches cachées comportant chacune n et r neurones. Les fonctions non-linéaires d'activation de des neurones des couches cachées sont des tangentes hyperboliques de la forme :

$$\varphi(v) = \frac{2}{1 + \exp(-2v)} - 1 \quad (1.17)$$

où v est la sommation des entrées aux neurones, chacune étant la multiplication de la sortie du neurone de la couche précédente multipliée par le poids synaptique w_{ij} contenant l'information acquise par le réseau.

Dans le cas de l'approximation d'une fonction réelle, la dernière couche est composée d'un seul neurone, identifié par s , dont la fonction d'activation linéaire ne fait que la somme des entrées et l'additionne au biais b^3 . Chaque couche comprend un biais b qui s'additionne aux autres entrées des neurones de la couche inférieure. La forme analytique générale du réseau présenté prend la forme suivante :

$$F(\mathbf{x}, \mathbf{w}) = \varphi_{lin} \left(\sum_r w_{sr} \varphi_{\tanh} \left(\sum_n w_{rn} \varphi_{\tanh} \left(\sum_m w_{nm} x_m \right) \right) \right) \quad (1.18)$$

où les biais sont tout simplement considérés comme l'élément 0 des sommations.

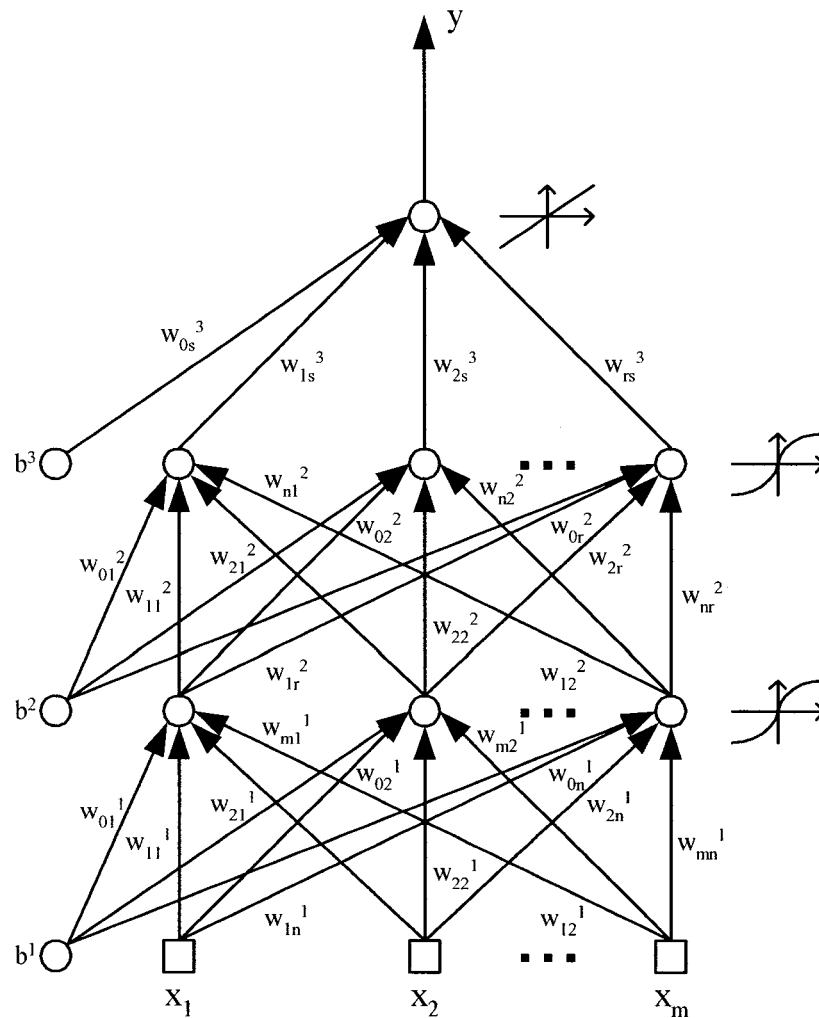


Figure 1.5 Schématisation du réseau de neurone à deux couches cachées

Le concept général d'entraînement de ce type de machine est d'estimer la valeur des poids du réseau en utilisant l'algorithme de rétro-propagation, basé sur le calcul du gradient du risque empirique. L'erreur entre la valeur attendue et celle calculée par le réseau est rétro-propagée dans celui-ci et en modifie la valeur des poids selon une des

nombreuses procédures possibles. Le processus est itératif puisque les données d'entraînement sont présentées plusieurs fois au réseau en construction.

La formule récursive est la clé de l'algorithme de rétro-propagation. Elle permet à l'erreur sur le signal d'une couche inférieure d'être calculée comme une combinaison linéaire des erreurs sur les signaux de la couche supérieure. De cette manière, les erreurs sur les signaux sont propagées vers l'arrière (d'où le terme *rétro-propagation*) sur toutes les couches, de la dernière vers la première. Ceci implique aussi que l'influence d'une couche supérieure sur une couche inférieure (et vice-versa) peut seulement être effectuée via les erreurs sur les signaux d'une couche intermédiaire.

Selon la procédure choisie, le calcul du gradient et la rétropropagation de l'erreur se font soit après présentation de l'ensemble des données (mode *batch*), soit après présentation de chaque donnée (mode séquentiel). Le mode *batch* est habituellement préféré car il donne un estimé plus juste du gradient, n'étant pas autant affecté par les données aberrantes que le mode séquentiel. Chaque cycle de calcul du gradient et de rétro-propagation de l'erreur est appelé *epoch*.

L'objectif de ce mémoire n'est pas de décortiquer l'algorithme de rétro-propagation et il ne sera donc pas présenté en détail. Les lecteurs intéressés à comprendre en détail celui-ci sont invités à se procurer la référence réalisée par Haykin (1999). Dans la même

référence, sont disponibles les explications des heuristiques et des procédures utilisées dans le présent travail et qui seront mentionnées dans la section sur la méthodologie.

Il est beaucoup plus intéressant de présenter dans cette section les capacités et limitations des réseaux de neurones par rapport aux autres machines neuronales utilisées ainsi que les informations touchant plus spécialement la recherche présentée.

Particularités sur la capacité de généralisation

La qualité de la généralisation d'un réseau de neurones entraîné par rétro-propagation de l'erreur repose essentiellement sur trois facteurs, soit la grandeur du bassin de données d'entraînement, l'architecture du réseau et la complexité physique du problème à modéliser (Haykin, 1999).

La quantité de données N nécessaires à la construction d'un réseau de neurones se généralisant bien peut être dérivée de la règle du pouce de Widrow pour l'algorithme LMS (*least-mean squares*) (Haykin, 1999)

$$N = O\left(\frac{W}{\varepsilon}\right) \quad (1.19)$$

où W est le nombre de poids synaptiques du réseau et ε l'erreur empirique acceptable. Par contre, lorsque cette égalité n'est pas respectée (le nombre de données n'est pas assez grand), ce qui peut souvent être le cas dans le type d'étude dont ce mémoire fait partie, le risque de sur-précision (*overfitting*) est très élevé et il est conseillé de pratiquer

la méthode d'entraînement de l'*arrêt prématuré*. Cette méthode consiste à arrêter l'entraînement de la machine avant que le gradient ait atteint un minimum. La figure qui suit présente schématiquement la raison de l'utilisation de cette technique.

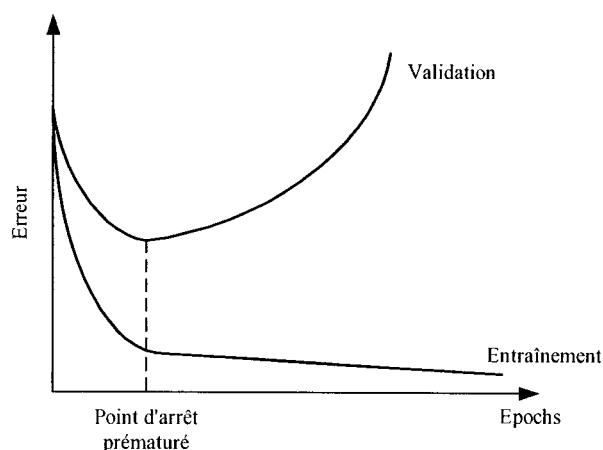


Figure 1.6 Schématisation de l'arrêt prématuré

La courbe d'apprentissage de l'entraînement décroît invariablement avec l'augmentation du nombre d'*epochs*. Par contre, la courbe de validation atteint un minimum et commence par la suite à augmenter. L'objectif est d'arrêter le processus d'entraînement lorsque ce minimum est atteint, garantissant le risque de généralisation le plus faible.

Le concept de malédiction de la dimensionalité, introduit par Bellman en 1961 (Bellman, 1961), donne une explication de l'influence de l'architecture du réseau de neurones et de la complexité physique du problème à modéliser sur la qualité de

généralisation. La malédiction de la dimensionalité, telle que définie par Friedman, est donnée ici (Friedman, 1995):

Une fonction définie dans un espace à haute dimension est habituellement beaucoup plus complexe qu'une fonction définie dans un espace de plus petite dimension, et les complications résultantes sont plus difficiles à discerner.

Une augmentation linéaire dans la dimensionalité d'un réseau de neurones se traduit en une augmentation exponentielle de sa complexité. Ceci se traduit par une détérioration de l'interpolation dans les régions non définies de l'espace. Ainsi, pour un problème hautement non-linéaire à plusieurs facteurs, le réseau devra contenir un nombre élevé de neurones pour permettre à l'espace cible d'être contenu dans l'espace d'hypothèse. Par contre, l'entraînement de celui-ci et sa généralisation seront beaucoup plus ardues que pour un réseau plus simple et nécessiteront un nombre élevé de données et une plus grande vigilance de la part de l'utilisateur. L'unique façon de combattre la malédiction de la dimensionalité est d'incorporer des connaissances *a priori* dans la construction du réseau. Par contre, dans le cas présent, ceci est évidemment impensable, sauf en ce qui à trait aux nombres de couches cachées que doit contenir le réseau de neurones.

Ainsi, Sontag (1992) a prouvé que lorsque le réseau de neurones à utiliser pour modéliser l'inverse d'une fonction dynamique, même à plusieurs paramètres, est un réseau à deux couches cachées est nécessaire. Des réseaux à une seule couche cachée

peuvent souvent faire l'affaire, mais ils ne peuvent pas garantir la convergence vers la solution optimale de tous les problèmes d'inversion dynamique.

Un des problèmes majeurs des réseaux de neurones réside dans l'algorithme de rétro-propagation de l'erreur qui ne garantit pas la convergence vers la solution optimale. La fonctionnelle du risque a plusieurs minima locaux et la rétro-propagation ne fait que converger vers l'un d'entre eux (Vapnik, 1995; Haykin, 1999). La qualité de la solution obtenue est tributaire de plusieurs facteurs, dont celui des valeurs d'initialisation des poids synaptiques et de l'heuristique utilisée pour déterminer la qualité du minimum atteint.

Un autre problème est celui de l'utilisation des fonctions sigmoïdes qui contiennent un facteur d'échelle qui affectent la qualité de l'approximation (Vapnik, 1995). Le choix de ce facteur d'échelle est un compromis entre la qualité de l'approximation et le taux de convergence, qui peut être très faible dans le cas de méthodes basées sur le gradient.

Comme les techniques de détermination des paramètres du modèle sont basées sur la minimisation de la différence des moindres-carrés de la valeur calculée et de la valeur prédite, les réseaux de neurones artificiels présentés peuvent tenir compte de l'incertitude associée à la valeur de sortie mais pas de celle associée aux valeurs d'entrées.

LES RÉSEAUX DE FONCTIONS À BASE RADIALE

Introduction et révision

Le théorème d'approximation universelle des réseaux de fonctions à base radiales (RBF) tel que présenté dans (Park et Sandberg, 1991) s'énonce comme suit :

Pour toute application entrée-sortie décrite par la fonction continue $f(x)$ il existe un réseau de fonctions à base radiale avec un ensemble de centres $\{t_i\}_{i=1}^{m_1}$ et un écart-type $\sigma > 0$ commun qui assure que la fonction $F(x)$ réalisée par le réseau de fonctions à base radiale est proche de $f(x)$ dans la norme L_p pour $p \in [1, \infty[$.

Ce théorème donne l'assurance théorique de l'application pratique des réseaux RBF pour l'approximation de fonctions non-linéaires (Haykin, 1999).

L'équation analytique d'un réseau RBF à entrées multiples et sortie unique est présentée ici.

$$F(x) = \sum_{i=1}^{m_1} w_i \exp \left(-\frac{\|x - t_i\|^2}{2\sigma^2} \right) \quad (1.20)$$

La distance entre le vecteur d'entrée x et les centres t des fonctions à base radiale est divisée par deux fois l'écart-type σ fixé par l'utilisateur au carré. L'écart-type est le même pour tous les neurones. Les poids synaptiques w_i déterminent l'importance

accordée à chacun des neurones. Sous forme schématique, un réseau à m entrées et réponse unique prend la forme suivante.

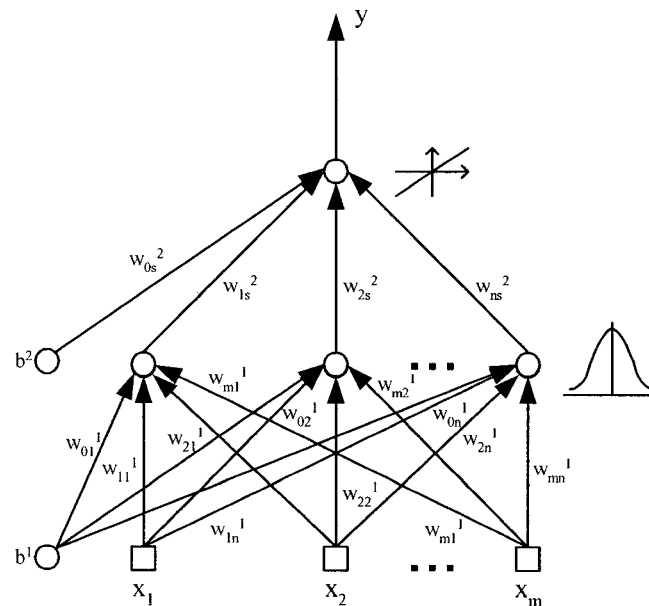


Figure 1.7 Réseau de fonctions à base radiale

L'entraînement du réseau se fait par l'ajustement des poids synaptiques w_i et sa morphologie est déterminée par le nombre de centres t_i .

Il existe plusieurs stratégies pour entraîner les réseaux de fonctions à base radiale. Celles-ci dépendent de la manière dont sont déterminés les centres. La liste suivante énumère les quatre principales stratégies utilisées (Vapnik, 1995) :

1. Centres fixes sélectionnés aléatoirement
2. Sélection non supervisée des centres
3. Sélection supervisée des centres

4. Interpolation stricte avec régularisation

Les lecteurs intéressés à comprendre en détail celles-ci sont invités à se procurer la même référence que pour les réseaux de neurones (Haykin, 1999). Dans la même référence sont disponibles les explications des heuristiques et des procédures utilisées dans le présent travail et mentionnées dans la section présentant la méthodologie.

Particularités sur la capacité de généralisation

Une borne supérieure sur le risque de généralisation, comparable au modèle général présenté dans la section sur la théorie de l'apprentissage statistique, a été dérivée par Niyogi et Girosi (1996). Cette borne prend la forme suivante

$$R[F] \leq O\left(\frac{1}{n}\right) + O\left(\frac{mn}{N} \log(nN) + \frac{1}{N} \log\left(\frac{1}{\eta}\right)^{1/2}\right) \quad (1.21)$$

où m et n sont respectivement le nombre d'entrées et de neurones de la couche cachée, et N et $1 - \eta$, le nombre de données d'entraînement et la probabilité $1 - \eta$ de confiance d'application de la borne. Certaines déductions peuvent être tirées de cette borne. Ainsi, l'erreur de généralisation converge vers zéro seulement si le nombre de neurones n de la couche cachée augmente plus lentement que le nombre de données d'entraînement N . Aussi, pour un certain nombre de données d'entraînement N , le nombre optimal de neurones n de la couche cachée se comporte selon la relation $n \propto N^{1/3}$.

Les réseaux de fonctions à base radiale sont des approximateurs universels dans le même sens que les réseaux de neurones. Par contre, si les réseaux de neurones construisent des fonctions d'approximation globales non-linéaires, les réseaux de fonction à base radiale construisent des approximations locales (Haykin, 1999). La figure qui suit présente ce phénomène d'approximation locale pour un problème en deux dimensions.

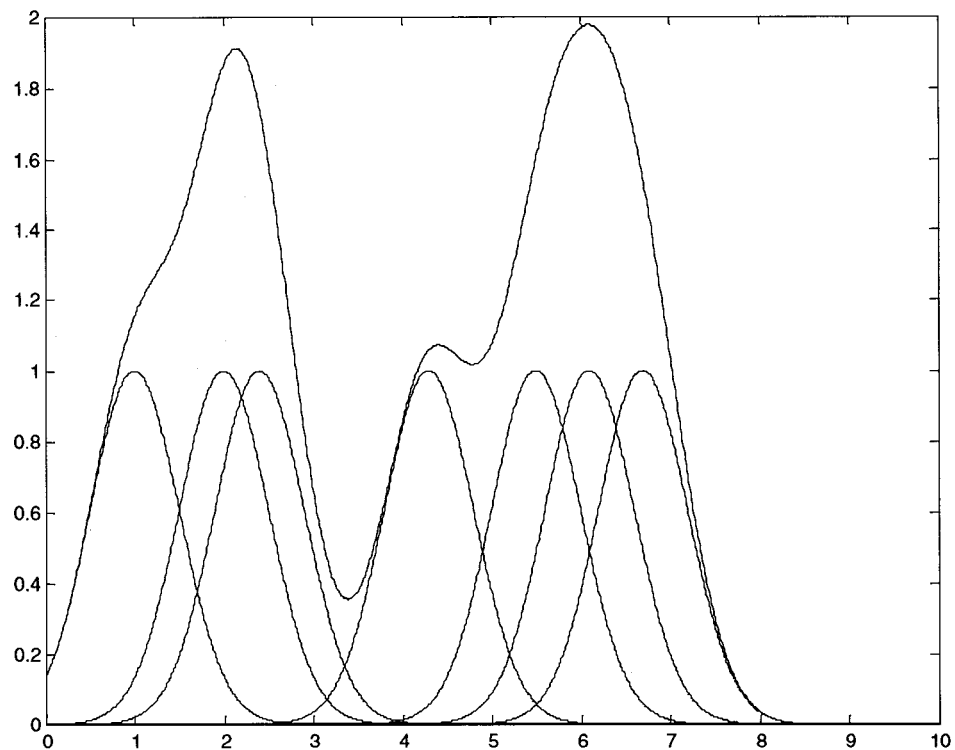


Figure 1.8 Approximation locale par fonctions gaussiennes

La sommation des différentes fonctions gaussiennes (en bleu sur le graphique) donnent la fonction d'approximation (en noir).

Par conséquent, il est souvent nécessaire d'utiliser un réseau de fonctions à base radiale contenant plus de paramètres à déterminer que pour le réseau de neurones présentant une généralisation équivalente.

Comme dans le cas des réseaux de neurones artificiels, les techniques de détermination des paramètres du modèle sont basées sur la minimisation de la différence des moindres-carrés de la valeur calculée et de la valeur prédite, les réseaux de fonctions à bases radiales présentés peuvent tenir compte de l'incertitude associée à la valeur de sortie mais pas de celle associée aux valeurs d'entrées.

LES MACHINES À SUPPORTS VECTORIELS

Les machines à support vectoriels implantent la seconde stratégie rencontrée dans la section de la minimisation du risque structurel, soit garder le risque empirique fixe et minimiser l'intervalle de confiance sur le risque de généralisation.

Il est beaucoup plus facile de comprendre les machines à supports vectoriels en commençant par la présentation de leur utilisation pour la construction de fonctions discriminantes. Cette sous-section abordera d'abord l'hyperplan optimal, pour ensuite présenter la construction de celui-ci dans le cas de données séparables et par la suite non-séparables. Les noyaux pouvant être utilisés pour la projection de l'hyperplan dans un espace de grande dimension seront présentés subséquemment. Finalement, la généralisation pour l'approximation de fonction sera exposée.

Les explications et preuves qui suivent sont toutes tirées des références suivantes (Vapnik, 1995; Burges, 1998; Gunn, 1998; Haykin, 1999).

L'hyperplan optimal

Soit un ensemble de données d'entraînement

$$\Gamma = \{(\mathbf{x}_i, d_i)\}_{i=1}^N, \mathbf{x}_i \in \mathbf{R}^n, d_i \in \{-1, 1\} \quad (1.22)$$

pouvant être séparées par l'hyperplan

$$(\mathbf{w} \cdot \mathbf{x}) - b = 0 \quad (1.23)$$

L'ensemble de vecteurs est séparé par l'hyperplan optimal si la séparation se fait sans erreur et que la distance entre l'hyperplan et le vecteur le plus proche de celui-ci est maximale.

Pour décrire l'hyperplan, il est possible de le présenter sous la forme :

$$d_i[(\mathbf{w} \cdot \mathbf{x}_i) - b] \geq 1, \text{ pour } i = 1, 2, \dots, N \quad (1.24)$$

La distance entre un point $\mathbf{x}_i$ et l'hyperplan $(\mathbf{w}, b)$ est

$$d(\mathbf{w}, b, \mathbf{x}_i) = \frac{|(\mathbf{w} \cdot \mathbf{x}_i) - b|}{\|\mathbf{w}\|} \quad (1.25)$$

L'hyperplan optimal est trouvé en maximisant la marge Δ , sujet aux contraintes d'inégalité ci-haut. La marge est donnée par

$$\Delta = \min_{\mathbf{x}_i: d_i = -1} d(\mathbf{w}, b; \mathbf{x}_i) + \min_{\mathbf{x}_i: d_i = 1} d(\mathbf{w}, b; \mathbf{x}_i) \quad (1.26)$$

$$\Delta = \min_{\mathbf{x}_i: d_i = -1} \frac{|(\mathbf{w}, \mathbf{x}_i) - b|}{\|\mathbf{w}\|} + \min_{\mathbf{x}_i: d_i = 1} \frac{|(\mathbf{w}, \mathbf{x}_i) - b|}{\|\mathbf{w}\|} \quad (1.27)$$

$$\Delta = \frac{1}{\|\mathbf{w}\|} \left(\min_{\mathbf{x}_i: d_i = -1} |(\mathbf{w} \cdot \mathbf{x}_i) - b| + \min_{\mathbf{x}_i: d_i = 1} |(\mathbf{w} \cdot \mathbf{x}_i) - b| \right) \quad (1.27)$$

$$\Delta = \frac{2}{\|\mathbf{w}\|} \quad (1.28)$$

Donc, l'hyperplan qui sépare de façon optimale les données est celui qui minimise, par rapport à $\mathbf{w}$ et b la fonctionnelle

$$\Phi(\mathbf{w}) = \|\mathbf{w}\|^2 \quad (1.29)$$

La figure qui suit présente cet hyperplan optimal.

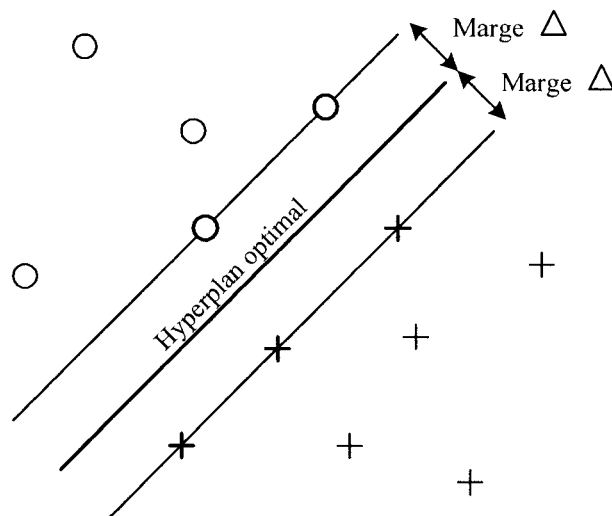


Figure 1.9 Hyperplan optimal avec la marge Δ de chacun des côtés

Le théorème suivant, présenté par Vapnik, vient faire le lien entre la construction de l'hyperplan optimal et la minimisation du risque structurel.

Soit l'ensemble de vecteurs x compris dans une sphère de rayon R , alors l'ensemble des hyperplans séparateurs de marge Δ a la dimension VC h bornée par l'inégalité

$$h \leq \min \left(\left\lceil \frac{R^2}{\Delta^2} \right\rceil, n \right) + 1 \quad (1.30)$$

La dimension VC d'un hyperplan peut donc être inférieure à $n + 1$, tel que présenté précédemment. Elle est minimisée par la maximisation de la marge Δ . Donc, maximiser

la marge revient à diminuer l'intervalle sur le risque de généralisation, ce qui est une implantation directe du principe de minimisation du risque structurel.

La construction des machines à supports vectoriels, qui sont des hyperplans séparateurs, maximise cette marge par un procédé d'optimisation, présenté dans la prochaine sous-section.

Construction de l'hyperplan optimal

Il est important de souligner que le développement qui suit s'applique uniquement dans le cas où les données utilisées pour l'entraînement sont linéairement séparables.

La construction de l'hyperplan optimal est un problème de programmation quadratique et se fait par la minimisation de la fonctionnelle

$$\phi(\mathbf{w}) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) \quad (1.31)$$

sous les contraintes

$$d_i[(\mathbf{w} \cdot \mathbf{x}_i) - b] \geq 1, \text{ pour } i = 1, 2, \dots, N \quad (1.32)$$

La solution de ce problème d'optimisation est donnée par le point de selle de la fonctionnelle de Lagrange suivante

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) - \sum_{i=1}^N \alpha_i \{[(\mathbf{x}_i \cdot \mathbf{w}) - b]d_i - 1\} \quad (1.33)$$

où les α_i sont les multiplicateurs de Lagrange. Le Lagrangien est minimisé par rapport à $\mathbf{w}$ et b et est maximisé en respectant $\alpha_i > 0$. Au point de selle, les solutions $\mathbf{w}^*$, b^* et $\boldsymbol{\alpha}^*$ respectent les conditions

$$\frac{\partial L(\mathbf{w}^*, b^*, \boldsymbol{\alpha}^*)}{\partial b} = 0 \quad (1.34)$$

$$\frac{\partial L(\mathbf{w}^*, b^*, \boldsymbol{\alpha}^*)}{\partial \mathbf{w}} = 0 \quad (1.35)$$

Par contre, il est certain qu'une forme permettant un calcul plus direct est nécessaire. Le développement qui suit permet d'atteindre cette forme.

La retranscription des équations précédentes sous formes explicites donne les trois équations suivantes, qui définissent totalement l'hyperplan optimal :

1. Les coefficients α_i^* doivent satisfaire les contraintes

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^N \alpha_i^* d_i = 0, \alpha_i^* \geq 0, i = 1, 2, \dots, N \quad (1.36)$$

2. L'hyperplan optimal est une combinaison linéaire de vecteurs provenant des données d'entraînement

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w}^* = \sum_{i=1}^N \mathbf{x}_i \alpha_i^* d_i, \alpha_i^* \geq 0, i = 1, 2, \dots, N \quad (1.37)$$

3. Seulement les *vecteurs de support* peuvent avoir des coefficients α_i^* non nul dans l'expansion de $\mathbf{w}^*$. Les *vecteurs de support* sont ceux pour lesquels, dans l'inégalité $d[(\mathbf{w} \cdot \mathbf{x}_i) - b] \geq 1$, pour $i = 1, 2, \dots, N$, l'égalité est atteinte.

$$\mathbf{w}^* = \sum_{\text{vecteurs de support}} \mathbf{x}_i \alpha_i^* d_i, \alpha_i^* \geq 0 \quad (1.38)$$

Cette réponse fait suite à l'utilisation des conditions du théorème de Kuhn-Tucker qui prévoit que les conditions nécessaires et suffisantes pour l'hyperplan optimal sont que les hyperplan séparateurs satisfassent les conditions

$$\alpha_i^* \left[(\mathbf{x}_i \cdot \mathbf{w}^* - b^*) d_i - 1 \right] = 0, i = 1, 2, \dots, N \quad (1.39)$$

En remplaçant ces expressions dans le Lagrangien et en tenant compte des conditions de Kuhn-Tucker, la fonctionnelle résultante suivante est obtenue

$$W(\mathbf{a}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j d_i d_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (1.40)$$

C'est cette fonctionnelle qui est maximisée selon les contraintes

$$\sum_{i=1}^N \alpha_i d_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, N \quad (1.41)$$

pour obtenir la fonction discriminante finale, maximisant la marge,

$$F(\mathbf{x}^*, \mathbf{a}^*) = \text{sign} \left(\sum_{\text{vecteurs de support}} d_i \alpha_i^* (\mathbf{x}^* \cdot \mathbf{x}) - b^* \right) \quad (1.42)$$

où $\mathbf{x}^*$ sont les vecteurs de support optimaux, $\mathbf{a}^*$ les coefficients de Lagrange optimaux correspondants et b^* la constante correspondante pouvant être déterminée utilisant un vecteur de chaque classe.

Généralisation pour le cas non-séparable

La construction d'un hyperplan optimal dans le cas où les vecteurs d'entraînement ne sont pas linéairement séparable nécessite l'introduction de variables de coût $\xi \geq 0$, mesure des erreurs de classification, et d'une fonction de pénalité

$$F_{\alpha}(\xi) = \sum_{i=1}^N \xi_i^{\alpha}, \alpha > 0 \quad (1.43)$$

Le problème d'optimisation est maintenant posé comme devant minimiser l'erreur de classification et la borne sur la dimension VC de l'hyperplan. Les contraintes sont alors modifiées par l'introduction de ces variables et prennent alors la forme

$$d_i[(\mathbf{w} \cdot \mathbf{x}_i) - b] \geq 1 - \xi_i, \text{ pour } i = 1, 2, \dots, N, \xi \geq 0 \quad (1.44)$$

L'introduction des variables de coût peuvent être visualisé sur la figure suivante.

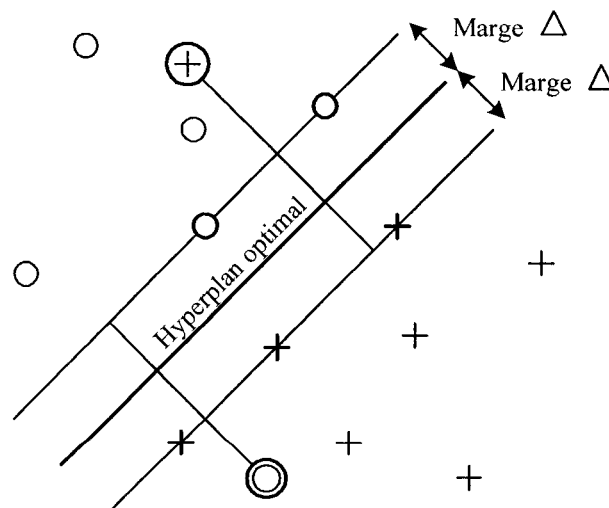


Figure 1.10 Hyperplan optimal pour le cas de données non-séparables linéairement

La construction de l'hyperplan optimal est un problème de programmation quadratique et se fait par la minimisation de la fonctionnelle

$$\phi(\mathbf{w}, \xi) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + C \sum_{i=1}^N \xi_i \quad (1.45)$$

où C est une constante fixée par l'utilisateur (la valeur à donner à celle-ci sera discuté plus loin). La solution est alors donnée par le point de selle de la fonctionnelle de Lagrange

$$L(\mathbf{w}, b, \alpha, \xi, \beta) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) - \sum_{i=1}^N \alpha_i \{[(\mathbf{x}_i \cdot \mathbf{w}) - b]d_i - 1 + \xi_i\} - \sum_{i=1}^N \beta_i \xi_i \quad (1.46)$$

où α_i et β_i sont les multiplicateurs de Lagrange. Le Lagrangien est minimisé dans le respect de $\mathbf{w}$ et b et est maximisé en respectant $\alpha_i > 0$, $\beta_i > 0$. Au point de selle, les solutions $\mathbf{w}^*$, b^* , α^* et β^* respectent les conditions

$$\frac{\partial L(\mathbf{w}^*, b^*, \alpha^*)}{\partial b} = 0 \quad (1.47)$$

$$\frac{\partial L(\mathbf{w}^*, b^*, \alpha^*)}{\partial \mathbf{w}} = 0 \quad (1.48)$$

$$\frac{\partial L(\mathbf{w}^*, b^*, \alpha^*)}{\partial \xi} = 0 \quad (1.49)$$

La dernière condition est nouvelle et implique la contrainte additionnelle

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \alpha_i + \beta_i = C \quad (1.50)$$

La fonctionnelle devant être maximisé est toujours la même, soit

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j d_i d_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (1.51)$$

et la fonction discriminante finale obtenue, maximisant la marge, reste de la même forme, soit,

$$F(\mathbf{x}^*, \boldsymbol{\alpha}^*) = \text{sign} \left(\sum_{\text{vecteurs de support}} d_i \alpha_i^* (\mathbf{x}^* \cdot \mathbf{x}) - b^* \right) \quad (1.52)$$

Par contre, une nouvelle contrainte, contenant la variable C , s'ajoute à celle existante pour donner la paire de contraintes

$$\sum_{i=1}^N \alpha_i d_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, N \quad (1.53)$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \dots, N \quad (1.54)$$

La solution est donc identique à celle obtenue pour le cas des données d'entraînement linéairement séparables, sauf à l'exception de l'ajout d'une borne supérieure sur les multiplicateurs de Lagrange. Cette variable C demeure la seule inconnue demandant à être déterminée. Elle ajoute une capacité de contrôle additionnelle à la fonction discriminante et peut, par le fait même, être directement reliée à un paramètre de régularisation. Il est important de noter que plus la valeur de C est petite, plus la marge de l'hyperplan est élevée. Par contre, le taux d'erreur peut devenir beaucoup plus élevé dans ce cas. Il est parfois suggéré de l'utiliser comme outil d'intégration d'une certaine connaissance du bruit des données, mais il reste qu'il faut idéalement tester plusieurs valeurs de C pour savoir laquelle abaisse le plus le taux d'erreur de validation.

Par contre, le cas non séparable dans cette forme n'est que rarement rencontré puisque toute la puissance des machines à supports vectoriels réside dans leur capacité à projeter les données d'entrée dans un espace à grande dimension, où elles y sont linéairement séparables, et d'y construire l'hyperplan optimal, processus expliqué à la section suivante.

Généralisation pour un grand espace de projection

Dans le cas où la séparation des données d'entraînement ne peut être faite linéairement dans l'espace de celles-ci, la machine à support vectoriel peut projeter le vecteur d'entrées, $\mathbf{x}$, dans un espace à haute dimension, $\mathbf{z}$. La machine à support vectoriel construit alors l'hyperplan optimal dans ce nouvel espace, tel présenté ici.

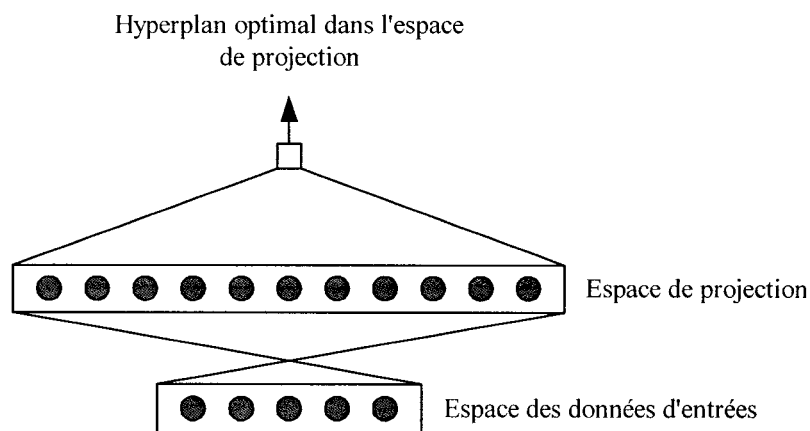


Figure 1.11 Projection du vecteur d'entrée dans un espace de plus grande dimension

Il existe plusieurs restrictions quant aux projections non-linéaires possibles, mais la plupart des fonctions connues, telles les fonctions à base radiale, polynomiales et sigmoïdes, sont utilisables.

La fonctionnelle à maximiser devient alors

$$W(\mathbf{a}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j d_i d_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (1.55)$$

où $K(\mathbf{x}_i, \mathbf{x}_j)$ est le noyau de fonctions réalisant la projection dans l'espace à grande dimension. Les contraintes restent inchangées, soit

$$\sum_{i=1}^N \alpha_i d_i = 0, \alpha_i \geq 0, i = 1, 2, \dots, N \quad (1.56)$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \dots, N \quad (1.57)$$

L'idée d'utiliser un noyau de fonctions est de permettre la réalisation des opérations dans l'espace d'entrée plutôt que dans l'espace de projection. En effet, le produit scalaire n'est pas fait dans l'espace de grande dimension, mais dans celui de petite dimension. La preuve suivante se base sur la théorie de Reproducing Kernel Hilbert Spaces. Un produit scalaire dans l'espace de projection a toujours un noyau équivalent dans l'espace d'entrée.

$$K(\mathbf{x}, \mathbf{x}') = (\phi(\mathbf{x}) \cdot \phi(\mathbf{x}')) = (\mathbf{z} \cdot \mathbf{z}') \quad (1.58)$$

Cette expression est valable si K est une fonction symétrique définie positive satisfaisant les Conditions de Mercer,

$$K(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\infty} a_i \varphi_i(\mathbf{x}) \varphi_i(\mathbf{x}') , \quad a_m \geq 0 \quad (1.59)$$

$$\iint K(\mathbf{x}, \mathbf{x}') g(\mathbf{x}) g(\mathbf{x}') d\mathbf{x} d\mathbf{x}' > 0 , \quad g \in L_2 \quad (1.60)$$

Cette situation permet d'outrepasser la malédiction de la dimensionalité qui causait problème avec les deux autres types de machines neuronales.

La preuve de la validité de l'utilisation des noyaux qui suivent peut être trouvée dans la référence de Gunn (1998). Il existe une grande quantité de noyaux dont certains sont présentés ici.

1. Polynomial

La projection polynomiale est une méthode populaire de modélisation non-linéaire. Le noyau prend une des deux formes suivantes

$$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}')^d \quad (1.61)$$

$$K(\mathbf{x}, \mathbf{x}') = [(\mathbf{x} \cdot \mathbf{x}') + 1]^d \quad (1.62)$$

La seconde forme est préférable dans la mesure où le problème du Hessian valant zéro est évité lors de la résolution du problème quadratique.

2. Fonctions à base radiale gaussienne

L'attrait de l'utilisation des noyaux de fonctions à base radiale gaussienne est que la sélection des centres de chacune des gaussiennes devient implicite avec l'utilisation des machines à support vectoriel. Chacun des vecteurs de support devient le centre d'une fonction gaussienne.

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right) \quad (1.63)$$

Le seul paramètre devant être fixé est l'écart-type σ de la fonction gaussienne.

3. Splines

La flexibilité des splines les rend très intéressantes pour la modélisation. Une spline finie à N nœuds situés à τ_s et d'ordre κ est donnée par

$$K(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\kappa} \mathbf{x}^i \cdot \mathbf{x}'^i + \sum_{s=1}^N (\mathbf{x} - \tau_s)_+^{\kappa} (\mathbf{x}' - \tau_s)_+^{\kappa} \quad (1.64)$$

Dans le cas d'une spline infinie définie sur l'intervalle $[0,1]$ le noyau donne

$$K(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\kappa} \mathbf{x}^i \cdot \mathbf{x}'^i + \int_0^1 (\mathbf{x} - \tau_s)_+^{\kappa} (\mathbf{x}' - \tau_s)_+^{\kappa} d\tau \quad (1.65)$$

Dans le cas où $\kappa = 1$, (S_1^∞) , le noyau devient

$$K(\mathbf{x}, \mathbf{x}') = 1 + (\mathbf{x}' \cdot \mathbf{x}') + \frac{1}{2}(\mathbf{x} \cdot \mathbf{x}') \min(\mathbf{x}, \mathbf{x}') - \frac{1}{6} \min(\mathbf{x}, \mathbf{x}')^3 \quad (1.66)$$

qui n'est autre que la solution cubique segmentée.

4. B-splines

Une autre formulation de la splines est celle communément appelée la B-splines.

Le noyau formé est très simple et défini sur la plage $[-1,1]$.

$$K(\mathbf{x}, \mathbf{x}') = B_{2N+1}(\mathbf{x} - \mathbf{x}') \quad (1.67)$$

Toute la difficulté réside dans le choix du noyau approprié. La borne supérieure sur la dimension VC peut permettre de comparer les noyaux entre eux, mais cela nécessite le calcul du rayon de la sphère contenant les données dans l'espace de grande dimension, ce qui peut être hasardeux. Idéalement, il faut tester les différents noyaux par validation croisée ou par la méthode du *bootstrap* et identifier celui qui s'applique le mieux à l'approximation non-linéaire tentée.

Machines à supports vectoriels pour la régression

La formulation du problème initial change légèrement. Soit un ensemble de données d'entraînement

$$\Gamma = \{(\mathbf{x}_i, d_i)\}_{i=1}^N, \mathbf{x} \in \mathbf{R}^n, d \in \mathbf{R} \quad (1.68)$$

approximées par la fonction

$$F(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b \quad (1.69)$$

La fonction optimale de régression est donnée par le minimum de la fonctionnelle

$$\phi(\mathbf{w}, \xi) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + C \sum_{i=1}^N (\xi_i^- + \xi_i^+) \quad (1.70)$$

où C est une valeur fixée par l'utilisateur et ξ^- et ξ^+ sont des variables de coût représentant les contraintes supérieures et inférieures sur les réponses du système.

Pour pouvoir appliquer la régression aux machines à support vectoriel, une fonction de coût alternative doit être élaborée afin d'inclure une mesure de distance. Plusieurs types de fonctions peuvent être abordées comme la fonction de perte quadratique, de Laplace ou de Huber ($\epsilon = 0$). Cependant, la fonction de perte ϵ -Insentive, telle que présentée à la figure qui suit, s'avère la plus performante (Smola et Scholkopf, 1998).

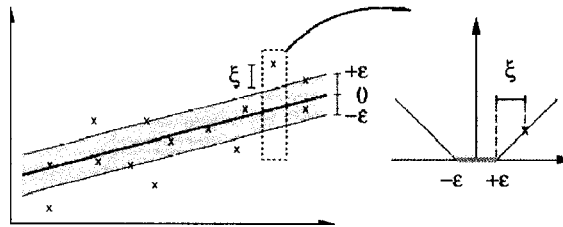


Figure 1.12 Marge et fonction de perte ϵ -Insentive

La valeur du paramètre ϵ influence directement le nombre de supports vectoriels. Plus elle est élevée, moins important sera le nombre de vecteurs de support sélectionnés. Il est important de noter que l'utilisation de ce type de fonction implique que les erreurs d'entraînement plus petites que ϵ ne sont pas considérées. Des travaux récents menés par Cherkassky et Ma (2003) présentent des méthodes pratiques pour la sélection des

paramètres et pour l'estimation du bruit dans le cadre de régression par machines à supports vectoriels.

Le remplacement de la fonction ε -Insensitive dans la fonctionnelle 1.70 donne la fonctionnelle suivante à maximiser

$$W(\mathbf{a}, \mathbf{a}^*) = \sum_{i=1}^N (\alpha_i^*(d_i - \varepsilon) - \alpha_i(d_i - \varepsilon)) - \frac{1}{2} \sum_{i,j} (\alpha_i^* - \alpha_j)(\alpha_j^* - \alpha_i) K(\mathbf{x}_i, \mathbf{x}_j) \quad (1.71)$$

avec les contraintes

$$\sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0, i = 1, 2, \dots, N \quad (1.72)$$

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, N \quad (1.73)$$

La résolution de ces équations permet de déterminer les multiplicateurs de Lagrange $\mathbf{a}^*$ et $\mathbf{a}$ ainsi que la fonction de régression

$$F(x) = \sum_{\text{vecteurs de support}} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (1.74)$$

La fonction ε -Insensitive est intéressante parce que, contrairement aux autres fonctions de perte, elle n'utilise pas tous les points comme vecteurs de supports, ce qui peut être dangereux dans le cas de données d'entraînement éparses ou aberrantes.

La figure qui suit présente schématiquement la fonction d'approximation obtenue à l'aide des machines à supports vectoriels.

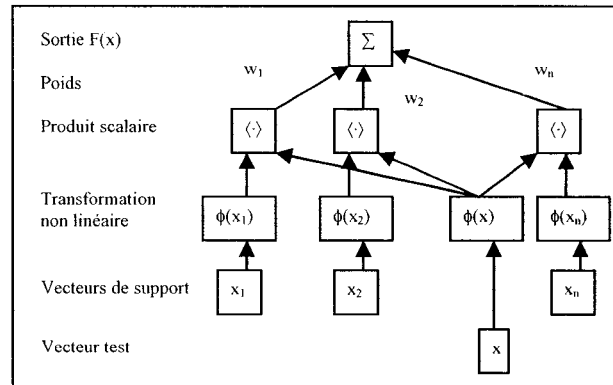


Figure 1.13 Schématisation de la régression avec des machines à supports vectoriels

Le procédé de construction des machines à support vectoriel peut être résumé comme suit :

1. Choisir un noyau de fonctions pour faire la projection des données d'entrées dans un hyper-espace
2. Fixer les valeurs des paramètres du noyau choisi, de la constante C et du paramètre ϵ de la fonction ϵ -Insensitive
3. Construire un problème d'optimisation quadratique dans lequel l'objectif est d'utiliser les variables d'entraînement comme supports d'un hyperplan optimal
4. Résoudre le problème d'optimisation quadratique pour trouver les valeurs des coefficients de Lagrange régissant l'hyperplan optimal selon l'équation 1.74
5. Trouver la valeur de b dans l'équation d'approximation générale 1.74 en utilisant un des points d'entraînement ou une moyenne des valeurs obtenues à partir de plusieurs points d'entraînement

Comme dans les deux cas des machines neuronales précédentes, les techniques de détermination des paramètres du modèle sont basées sur la minimisation de la différence des moindres-carrés de la valeur calculée et de la valeur prédite, les machines à supports vectoriels présentées peuvent tenir compte de l'incertitude associée à la valeur de sortie mais pas de celle associée aux valeurs d'entrées.

LA QUANTIFICATION DES ODEURS

La quantification des odeurs est presque exclusivement une problématique environnementale. La revue de la littérature scientifique qui suit s'intéresse donc aux techniques et résultats publiés dans les revues scientifiques ayant une relation directe avec la quantification d'odeurs nuisibles émanant de sites industriels.

La littérature scientifique est riche de techniques et de résultats sur la reconnaissance des odeurs, sujet qui occupe la majorité des chercheurs du domaine de l'olfaction artificielle. La quantification n'intéresse qu'un groupe très restreint d'entre eux. Ainsi, comme le présent mémoire fait une totale abstraction de la reconnaissance des odeurs, il est normal de ne pas y rencontrer des références habituelles.

La section se divise en quatre parties intimement liées entre elles, soit la quantification par olfactométrie, la quantification des odeurs par instruments analytiques classiques, GC/MS, GC/FID, détecteur H₂S, la quantification des odeurs par nez électronique et la présentation des recherches ayant porté sur la quantification des odeurs environnementales. Le lien direct est que la plupart des recherches menées jusqu'à maintenant ont comparé les diverses techniques entre elles.

Pour bien comprendre ce dont il est question lorsque l'on parle de mesure des odeurs, il est important de faire la distinction entre l'odeur et les odorants qui la constituent. Cette distinction est importante puisque les deux grandes classes de techniques de mesure des

odeurs s'appuient sur l'un ou l'autre de ces concepts. Ainsi, la méthode traditionnelle de l'olfactométrie, dite de mesure sensorielle, repose exclusivement sur la perception de l'odeur par un jury de sujets humains. À l'opposé, les techniques de mesure analytique s'appuient sur la reconstruction de l'odeur à partir de la reconnaissance de ses constituants odorants par des instruments de mesure. Les composés constituant en majeure partie l'odeur de biogaz sont présentés dans le tableau qui suit. Les odeurs proviennent presque exclusivement des produits de la dégradation biologique due à l'activité bactérienne dans les déchets.

Tableau 1.1 Principaux composés constituant le biogaz (Barlaz *et al.*, 1989)

Composé	Formule chimique	% volume	Caractère de l'odeur
Méthane	CH ₄	47.4	Sans odeur
Dioxyde du carbone	CO ₂	47.0	Sans odeur
Eau (vapeur) à 25°C	-	3.0	Sans odeur
Hydrocarbures (Paraffine)	-	0.1	----
Hydrocarbures aromatiques	-	0.2	----
Hydrogène	H ₂	0.1	Sans odeur
Sulfure d'hydrogène	H ₂ S	0.01	Odeur d'œufs pourris
Monoxyde de carbone	CO	0.1	Sans odeur

Une grande part de l'odeur de biogaz vient des composés traces. Le tableau suivant en présente une partie. Il est important de savoir que même les molécules ne présentant pas

d'odeur influence l'odeur résultante du mélange (Stuetz *et al.*, 1998). L'humidité du mélange est d'ailleurs une variable importante de l'intensité de celle-ci.

Tableau 1.2 Principaux composés traces constituant l'odeur de biogaz
(Barlaz *et al.*, 1989)

Composantes traces du biogaz				
Groupe	Composé	Formule chimique	Concentration mg/m ³	Caractère de l'odeur
Alcane	Éthane	C ₂ H ₆	0.8 ~48	sans odeur
	Propane	C ₃ H ₈	0.04 ~10	sans odeur
	Butane	C ₄ H ₁₀	0.3 ~23	N.D.
	Pentane	C ₅ H ₁₂	0~1	odeur d'essence
	Hexane	C ₆ H ₁₄	3~18	odeur douce d'essence
	Cyclohexane	C ₆ H ₁₂	2~6	odeur sucrée douce
	Heptane	C ₇ H ₁₆	3~8	odeur douce d'essence
	Octane	C ₈ H ₁₈	0.05~75	odeur d'essence
	Nonane	C ₉ H ₂₀	0.005~400	N.D.
	Décane	C ₁₀ H ₂₂	0.2~137	N.D.
	Undécane	C ₁₁ H ₂₄	7~48	N.D.
	Dodécane	C ₁₂ H ₂₆	2~4	N.D.
	Tridécanne	C ₁₃ H ₂₈	0.2~1	N.D.
Alcène	Éthylène	C ₂ H ₄	0.7~31	odeur sucrée
	Butylène	C ₄ H ₈	1~2	N.D.
Autres	Cumène	C ₉ H ₁₂	0~32	odeur benzénique pénétrante
	Benzène	C ₆ H ₆	0.03~7	odeur de diluant à peinture
	Toluène	C ₇ H ₈	0.2~615	odeur benzénique

D'autres composés sont en présence encore plus faible mais ont une certaine incidence sur l'odeur résultante. Une étude portant sur ces composés est parue en 2001 et fait l'inventaire d'une quarantaine d'autres produits (Davoli *et al.*, 2003).

Les auteurs s'entendent habituellement sur quatre dimensions de l'odeur, soit la concentration, l'intensité, le caractère et le ton hédonique (Gostelow *et al.*, 2000). La concentration de l'odeur est une mesure physique de l'odeur, faisant l'objet d'unité de mesure (unité-odeur/m³) et dont la dispersion peut être modélisée. L'intensité fait référence à la magnitude de la sensation que procure la perception de l'odeur. Le caractère de l'odeur est ce qui la rend reconnaissable par rapport aux autres odeurs, telle la saveur pour les aliments. Le ton hédonique est le niveau d'appréciation, donc du plaisir, du sujet par rapport à l'odeur perçue.

L'olfactométrie

L'analyse olfactométrique consiste à déterminer le seuil de perception olfactif d'un échantillon gazeux à l'aide d'un jury humain. Le seuil de perception olfactif est défini comme le nombre de fois que l'échantillon odorant se doit d'être dilué avec de l'air pur pour que 50% des membres du jury perçoivent l'odeur et que l'autre 50 % ne la perçoivent pas. Le seuil de perception olfactif correspond à une concentration-odeur d'une unité-odeur par mètre cubique (1 u.o./m³). La relation entre les dilutions et la concentration-odeur est linéaire. La concentration-odeur de l'échantillon est donc égale au nombre de dilutions nécessaires à l'atteinte du seuil de perception olfactif. Il est important de faire la différence entre le seuil de perception olfactif et la reconnaissance de l'odeur, qui a habituellement lieu à des concentrations de 2 u.o./m³ ou 3 u.o./m³.

Dans le cadre de la présente recherche, la technique d'olfactométrie utilisée est celle exigée dans la méthode de référence de la CUM en matière de mesure des odeurs et dans le projet de modification du règlement sur la qualité de l'atmosphère (Q-2 r.20) du Ministère de l'Environnement du Québec (MENV), soit l'olfactométrie à dilution dynamique. La méthode consiste à présenter à un jury de six personnes différentes dilutions de l'échantillon gazeux, réalisées par un olfactomètre mesurant précisément les débits de l'air propre et de l'odeur, dans des cornets de *flairage*. Les réponses du jury sont ensuite analysées selon une des nombreuses méthodes de détermination du seuil olfactif. L'Annexe I décrit ces méthodes. La précision atteinte est de loin supérieure à l'olfactométrie statique (Gostelow *et al.*, 2000).

L'olfactométrie est la clé de voûte de toutes les techniques subséquentes de quantification des odeurs. Comme il n'existe pas de modèle de la concentration-odeur des mélanges gazeux fonction des concentrations des odorants, sauf dans certains cas très particuliers, l'olfactométrie est toujours nécessaire à la construction d'une relation entre les mesures fournies par les instruments.

Les instruments analytiques classiques

Toutes les techniques de quantification reposant sur l'utilisation d'instruments analytiques cherchent à détecter et à quantifier certains odorants clés ainsi qu'à développer une corrélation entre leur concentration dans l'échantillon et la concentration-odeur correspondante. La qualité des relations et la réussite de ces

techniques divisent par contre les chercheurs, comme on le verra dans la dernière partie de cette section.

Les instruments de mesure analytiques pouvant être utilisés sont présentés dans la liste suivante (Gostelow *et al.*) :

- GC/MS (*gas chromatography, mass spectrometry*)
- GC/FID (*gas chromatography, flame ionization detector*)
- FID (*flame ionization detector*)
- MFD (*microwave plasma detector*)
- GC/FPD (*gas chromatography, flame photometric detector*)
- GC/O (*gas chromatography, odor*)
- Détecteur H₂S

L'intérêt des instruments analytiques réside essentiellement dans leur grande répétabilité, leur précision et leur faible coût d'utilisation par rapport à la quantification de l'odeur à l'aide de l'olfactométrie.

Dans le cas de la quantification d'odeurs environnementales, certains désavantages surgissent. Comme ce type d'odeur est très complexe, l'analyse des données peut devenir beaucoup plus ardue que la seule reconnaissance par un jury. Le nombre de molécules odorantes est élevé et la construction d'un modèle basé uniquement sur certains composés peut ne pas faire l'affaire. De plus, les nombreuses interactions entre les odorants peuvent rendre la construction d'un modèle difficile. Finalement, certaines

molécules sont présentes en infimes quantités, souvent sous le seuil de détection de l'instrument, et malgré leur effet sur l'odeur, ne pourront jamais être insérées dans le modèle.

Le nez électronique

Le terme *nez électronique* est utilisé lorsqu'un réseau de capteurs chimiques non spécifiques est jumelé à une unité de traitement informatique dans le but de reconnaître ou de mesurer la concentration d'un gaz ou d'une odeur (Stuetz *et al.*, 1998). Les nez électroniques tirent leur nom de leur similitude avec le système olfactif humain, dont ils sont fortement inspirés.

Le nez humain contient environ cinquante millions de neuro-récepteurs, reliés à dix milles neurones primaires. Ces dernières sont à leur tour en contact avec une deuxième couche de neurones qui vont se greffer au bulbe olfactif, dans le cortex cérébral, où l'odeur est reconnue (Persaud et Dodd, 1982). Dans le système du nez électronique, les neuro-récepteurs sont remplacés par un réseau de capteurs chimiques non-spécifiques, dont le temps de réponse est beaucoup plus rapide que celui des instruments analytiques. Les interactions entre les différentes molécules gazeuses et les capteurs altèrent la conductivité de ces derniers dont les lectures représentent un vecteur de données propre à une certaine odeur, gaz ou concentration. Les deux couches de neurones et le cortex cérébral sont suppléés par un élément logiciel de reconnaissance ou de quantification sur

ordinateur, soit une fonction non-linéaire calculant le facteur désiré en utilisant les réponses des capteurs comme variables d'entrées.

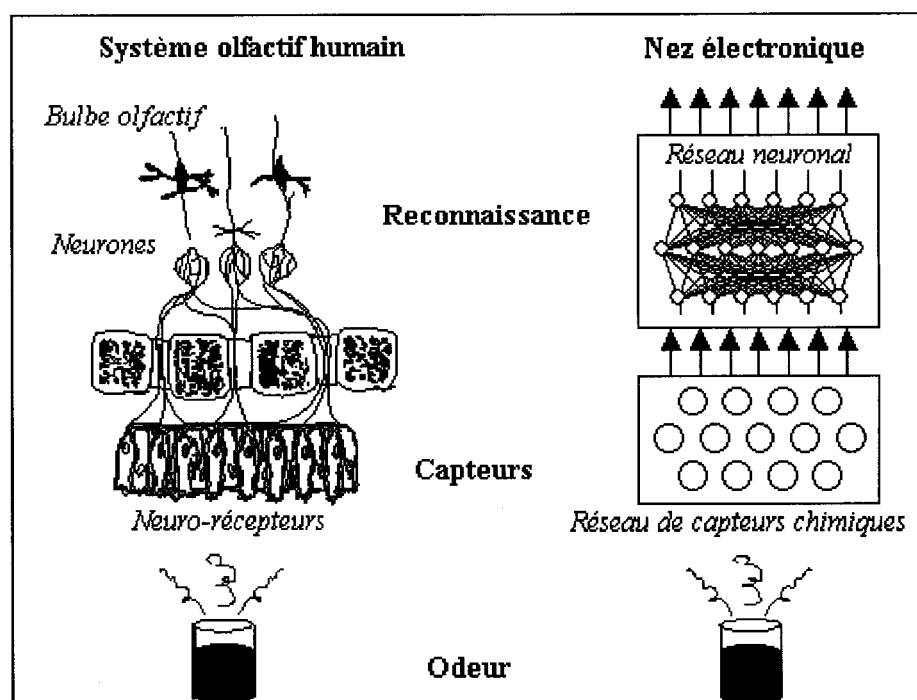


Figure 1.14 Comparaison entre le nez électronique et le système olfactif humain

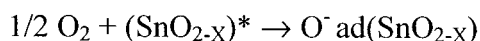
Les capteurs chimiques sont un domaine de recherche en soi et ne sont donc pas spécialement créés pour les applications qu'en font les nez électroniques. Par contre, au cours des dernières années, le développement de capteurs spécialisés s'est fait sous l'impulsion des entreprises offrant des nez électroniques commerciaux. Le premier nez électronique développé à l'École Polytechnique, et dont le présent projet est la continuité, permettait de quantifier des mélanges odorants en terme d'intensité (Hudon *et al.*, 2000). Cette section présentera les différents types de capteurs, spécialisés ou non, pouvant être utilisés dans la fabrication d'un nez électronique.

Les principes de fonctionnement des capteurs chimiques reposent sur l'adsorption et la désorption chimique ou physique, l'absorption optique ou les réactions chimiques entre les molécules gazeuses et la surface et/ou la matrice des matériaux du capteur. Ces interactions causent des changements physiques dans les capteurs qui peuvent être détectés au niveau de la température, de la conductivité, de l'activité électrochimique ou des propriétés optique, diélectrique ou magnétique (Haugen et Kvaal, 1998).

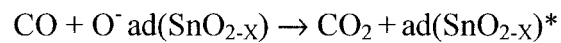
Les capteurs chimiques sont habituellement séparés en deux grandes familles, soit les capteurs chauds et les capteurs froids.

Les capteurs chauds sont les semi-conducteurs à base d'oxyde métalliques (SOM) et les transistors d'oxyde métallique à effet de champ (SOMTEC), qui fonctionnent à de hautes températures.

Les capteurs SOM (TGS Sensor, 1996), de loin les plus répandus, opèrent à des températures oscillant entre 200 °C et 500 °C, ce qui a pour but d'adsorber les atomes d'oxygène chargés négativement présents dans l'air. Cette adsorption d'une molécule chargée négativement crée un transfert de charge de l'atome du métal à l'oxygène et provoque une barrière de potentiel. Cette barrière empêche les électrons de se déplacer et augmente la résistivité du matériau. L'équation de cette réaction est donné ici.



Lorsque le capteur se trouve dans un milieu contenant des gaz réducteurs, les atomes d'oxygène à la surface du capteur sont récupérés par les molécules réductrices, l'intensité de la barrière de potentiel diminue et la résistivité du matériau varie, tel que présenté dans l'équation suivante.



La figure suivante représente ce phénomène.

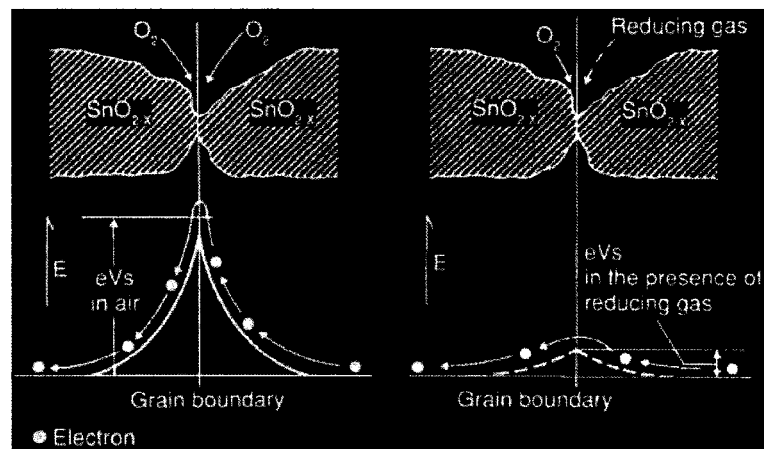


Figure 1.15 Explication du fonctionnement d'un capteur SOM

Les capteurs SOMTEC (Lundström *et al.*, 1990) sont chauffés dans la plage de température de 100-200 °C. La matrice semi-conductrice d'oxyde d'étain est dopée de métaux nobles et est recouverte d'une couche métallique catalytique. Le signal de sortie varie en fonction du changement de potentiel dans le capteur lorsqu'il y a modification de la polarisation électrique de celui-ci par la réaction des molécules avec la surface catalytique (Haugen et Kvaal, 1998).

Les capteurs froids fonctionnent à température ambiante, soit entre - 30 °C et 40 °C. Il va s'en dire qu'ils sont donc grandement influencés par la température du milieu dans lequel ils se trouvent. Les capteurs polymériques conducteurs (PC), les capteurs à oscillation, les capteurs optiques et les cellules électrochimiques forment la famille des capteurs froids.

Les capteurs PC (Bartlett *et al.*, 1989), les plus utilisés avec les capteurs SOM, sont habituellement composés de pyrroles ou d'anilines. L'adsorption en surface et dans la matrice des molécules gazeuses altère la conductivité du capteur et la différence de potentiel mesuré est utilisée comme ensemble de données. En changeant les groupements ou la structure du polymère et en utilisant différents ions dopants, il est possible d'en changer la sélectivité et la sensibilité.

Les capteurs à oscillation fonctionnent selon le principe que l'adsorption de molécules sur la couche de contact du capteur résulte en une diminution de la fréquence dû à l'augmentation de la masse et parfois dans un changement de la viscosité de la couche de contact.

C'est le cas des capteurs à quartz recouvert de polymère ou d'un type comparable de revêtement (QMB, *quartz micro balance*) (Nakamoto *et al.*, 1995; Garrigues, 2002) dans lequel un courant est passé et où est calculé la fréquence retournée par l'élément piézoélectrique. Les capteurs acoustiques (SAW, *surface acoustic wave*) (Liron *et al.*,

1997) fonctionnent de façon comparable. L'élément piézoélectrique recouvert d'un revêtement sert de base à deux paires d'électrodes et le courant est induit aux électrodes plutôt qu'au cristal.

Dans les capteurs optiques, des sources lumineuses excitent les molécules gazeuses et le signal de sortie varie en fonction des propriétés telles l'absorbance, le facteur de réflexion, la fluorescence et la chimiluminescence (Haugen et Kvaal, 1998).

Les cellules électrochimiques consistent en un bain électrolytique dans lequel baigne plusieurs paires d'électrodes. Avec l'exposition aux molécules gazeuses, des réactions d'oxydoréduction prennent place sur les différentes électrodes et le courant généré entre chaque paire est mesuré.

Le tableau suivant présente les différentes caractéristiques des principaux capteurs utilisés dans la fabrication des nez électroniques.

Tableau 1.3 Tableau comparatif des principales caractéristiques des capteurs
(Haugen et Kvaal, 1998)

Caractéristiques	Type de capteur				
	SOM	SOMTEC	PC	QMB	SAW
Sélectivité	Faible	Moyenne	Moyenne	Élevée	Élevée
Sensibilité	> 0.1 ppm	> 0.1 ppm	> 0.01 ppm	> 0.1 ppm	Ppb
Reproductibilité	Faible	Bonne	Bonne	Moyenne	Moyenne
Thermo-dépendance	Faible	Faible	Élevée	Moyenne	Élevée
Gaz vecteur	O ₂	O ₂	Inerte / O ₂	Inerte / O ₂	Inerte / O ₂
Hygro-dépendance	Faible	Moyenne	Élevée	Faible	Faible
Temp. d'opération (°C)	200 - 500	100 - 200	Ambiante	Ambiante	Ambiante
Temps de réponse (s)	0.5 - 5	0.5 - 5	20 - 50	20 - 50	20 - 50

Temps de récupération	Rapide	Rapide	Lente	Lente	Lente
Temps de vie (année)	3 - 5	1 - 4	1 - 2	< 2	< 2

Les premiers travaux sérieux sur les nez électroniques se penchèrent surtout sur la détection des composés volatils toxiques ou inflammables (Zarcomb et Stetter, 1984) et sur la discrimination des odeurs. Depuis, l'utilisation des nez électroniques s'est généralisée sous l'influence de la recherche au niveau des capteurs, des méthodologies d'entraînement et des algorithmes de reconnaissance.

Les travaux utilisant le nez électronique dans la détermination des concentrations-odeurs sont, par contre, au stade de balbutiements. La plupart des travaux réalisés jusqu'à maintenant dans le domaine des odeurs se sont concentrés sur la discrimination des odeurs, soit entre des odeurs provenant de sources distinctes (Romain *et al.*, 2000), soit entre les variantes d'odeurs d'une seule et même source (par exemple les différentes essences de café). Aussi étonnant que cela puisse être, une partie de la recherche sur la quantification des odeurs par un nez électronique se concentre sur la fabrication de systèmes contournant la problématique de la variation de la concentration-odeur pour les problèmes de reconnaissance (Maekawa *et al.*, 2001).

La quantification des odeurs environnementales par nez électroniques dans la littérature scientifique

Tel que signalé précédemment, la littérature scientifique est plutôt mince en ce qui a trait aux comptes-rendus de recherche portant sur la quantification d'odeur à l'aide de nez

électroniques. La section qui suit présente les plus importants d'entre eux relativement à la recherche du présent mémoire.

Les premières études sur la quantification des odeurs à l'aide de nez électroniques traitaient essentiellement de la démonstration de la faisabilité de ce type de système.

Une étude portant sur l'odeur de lisier de porc (Persaud *et al.*, 1996) révélait que le nez électronique arrivait à discriminer différents odorants et fournissait une réponse proportionnelle à la concentration de ceux-ci. L'objectif était de suppléer aux instruments analytiques classiques un nez électronique composé de vingt capteurs de polymères conducteurs. Par contre, bien que ce ne soit resté qu'à l'état de projet, les informations provenant des capteurs auraient été utilisées d'abord pour la détermination de la concentration des composés odorants. Les réponses des capteurs n'auraient donc pas été employées directement pour la quantification de l'odeur. L'approximation aurait été faite en deux étapes, soit la détermination de la concentration des composés et, ensuite, le calcul de la concentration-odeur.

Une autre étude (Misselbrook *et al.*, 1997), étudiant les odeurs de lisier de bovin, s'attaqua pour la première fois à l'utilisation directe des réponses des capteurs pour la détermination de la concentration-odeur. Deux nez électroniques furent comparés, chacun composé d'un nombre différent de capteurs de polymères conducteurs, soit vingt et trente-deux. La moyenne des réponses des capteurs était calculée pour différentes dilutions dans l'optique de bâtir une relation entre cette moyenne et la concentration-

odeur. Les concentrations-odeur, déterminées par olfactométrie, présentées aux capteurs étaient comprises entre 100 u.o./m³ et 1000 u.o./m³. Il fut déterminé que les moyennes des réponses des capteurs exprimaient environ 60 % de la variance des concentrations-odeurs.

Un projet de grande envergure sur les odeurs émanant de dix stations d'épuration d'eaux usées fut mené en Angleterre à la fin des années quatre-vingt-dix (Stuetz *et al.*, 1998). La recherche consistait à comparer la valeur d'un détecteur de H₂S (Jerome 631-X) et d'un nez électronique commercial (NOSE Neotronics modèle D) dans la détermination de la concentration-odeur des différents sites et éléments des procédés de traitement des eaux usées. La valeur réelle de concentration-odeur était déterminée par olfactométrie dynamique. Le premier résultat intéressant dégagé par la recherche est que, comparativement à une idée souvent préconçue chez certains chercheurs, il n'y avait pas de relation directe entre la concentration en H₂S du mélange odorant et sa concentration-odeur (TON sur le graphique). La figure qui suit présente ces résultats.

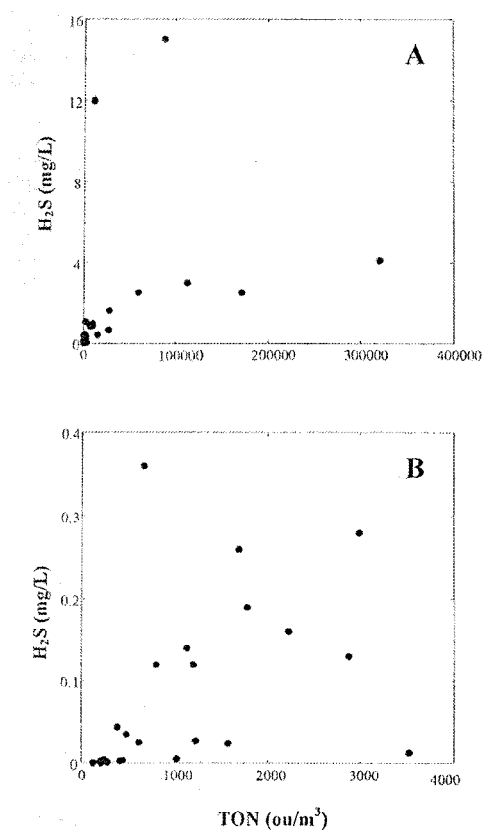


Figure 1.16 Graphique de la concentration en H₂S des mélanges odorants en fonction de deux plages de concentration-odeur, A) 0 – 400000 u.o/m³ et B) 0 – 4000 u.o/m³ (Stuetz *et al.*, 1998)

Les auteurs expliquent la différence entre leur résultats aux antipodes de ceux présentés par d'autres (Hobson et Walsh, 1990; Walsh, 1995) par la nature même de l'odeur, qui n'était pas majoritairement composée de H₂S.

Selon les auteurs, le nez électronique est plus performant que le détecteur de H₂S. Les informations utilisées comme variables d'entrées de la fonction d'approximation de la concentration-odeur sont la différence de voltage, en pourcentage, subie par les capteurs

du nez électronique entre leur ligne de base et lorsque soumis pendant une minute à une dilution de l'odeur.

Les données sont traitées par analyse canonique discriminante et corrélation canonique. Ce type de traitement statistique permet de réduire la dimension des entrées à une ou deux dimensions, plutôt que les douze de départ (puisque'il y avait douze capteurs). La corrélation canonique forme une combinaison linéaire des variables qui expriment la majeure partie de la variation dans les données. Comparativement à l'analyse par composante principale qui maximise la variance, l'analyse de corrélation canonique maximise la corrélation. Les résultats présentés sont impressionnants. Pour un élément du procédé, la relation entre la sortie du nez électronique, sous forme de variable canonique, et la concentration-odeur est linéaire. La première figure présentée montre la relation obtenue sur l'odeur provenant d'un biofiltre d'une des stations d'épuration.

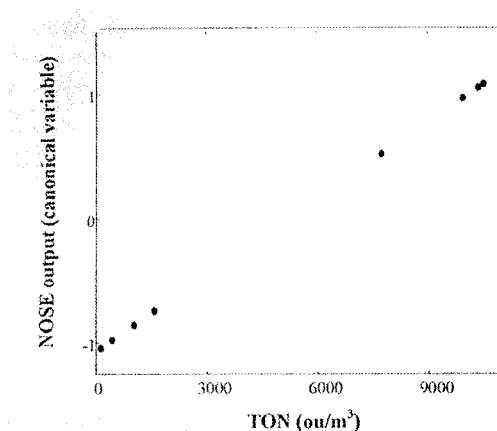


Figure 1.17 Graphique de la sortie du nez électronique sous forme de variable canonique en fonction de la concentration-odeur (Stuetz *et al.*, 1998)

La seconde figure présente le même type de relation, mais avec l'odeur ambiante de l'usine de traitement.

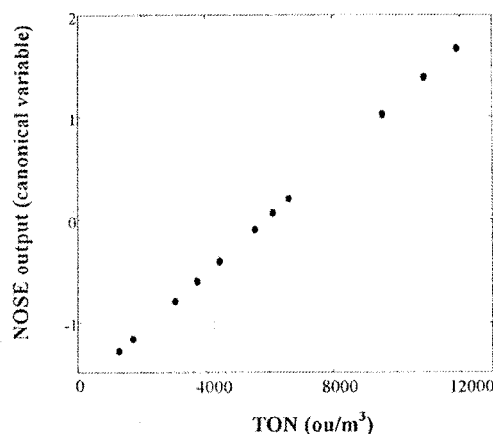


Figure 1.18 Graphique de la sortie du nez électronique sous forme de variable canonique en fonction de la concentration-odeur (Stuetz *et al.*, 1998)

Il est important de signaler que chaque relation, ainsi que toutes celles présentées dans l'article, est construite avec une analyse de corrélation canonique indépendante et n'est valable donc que pour une seule source d'odeur, sur un site particulier. Les relations produites ne sont valables que pour les faibles concentrations-odeurs, soit au maximum 12000 u.o/m³. Les auteurs n'ont pas réussi à modéliser la concentration-odeur des émissions provenant de certaines sources semblables, par exemple deux biofiltres de deux sites différents ou deux bassins du même site. Le profil chimique des différentes sources n'était pas précisé. Le graphique qui suit présente l'ensemble des points expérimentaux, pour les dix sites de traitement, avant le traitement statistique pour deux des capteurs.

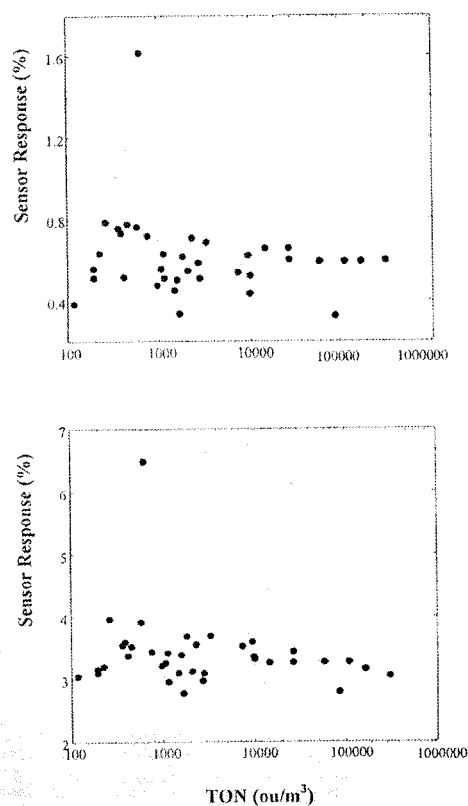


Figure 1.19 Graphique de la sortie du nez électronique brute en fonction de la concentration-odeur (Stuetz *et al.*, 1998)

Il appert en effet qu'aucune relation intuitive ne peut en être dégagée. Il est souligné par les auteurs que le succès de l'analyse par corrélation canonique pour des odeurs de site et de source spécifiques est une preuve qu'une relation existe.

Une faille énorme dans cette recherche est l'absence de validation des modèles proposés. On ne mentionne nulle part la réussite de prédiction des modèles sur des données nouvelles qui n'ont pas été employées pour l'entraînement. De ce fait, il est impossible

de dire si ceux-ci sont utilisables par la suite pour quantifier des odeurs de concentrations inconnues.

Un autre problème est l'utilisation de l'analyse de corrélation canonique. Cette dernière cherche à linéariser au maximum les données dépendantes (la concentration-odeur) en fonction des données indépendantes (les sorties du nez électronique). Les résultats d'entraînement sont donc très bons, mais il n'est pas aussi certain qu'une nouvelle donnée se positionnerait aussi bien sur la droite. En effet, cette technique minimise le risque empirique mais laisse béant la minimisation du risque structurel. La sur-précision du modèle devient donc très élevée, ce qui peut être déduit de l'alignement quasi-parfait des points sur la droite. Les données provenant de l'olfactométrie contiennent une erreur qui aurait dû faire dévier au moins certains points hors de la droite de régression, comme dans n'importe quelle régression linéaire à partir de données expérimentales.

Cette recherche permet tout de même de croire que la quantification d'odeur à l'aide des sorties d'un nez électronique peut s'avérer supérieure à la simple utilisation d'un détecteur de H_2S lorsque l'odeur comporte une grande quantité d'odorants et que le H_2S n'en est pas le constituant majoritaire.

Une seconde étude, portant sur la quantification d'odeur de site de traitement des déchets, avait comme objectif l'utilisation d'un nez électronique dans la détermination de la concentration-odeur et la comparaison de celui-ci avec certains instruments

analytiques (Bockreis et Jager, 1999). C'est, à ce jour, l'étude la plus complète dans le domaine. La détermination des concentrations-odeurs étalons ont été faites à l'aide d'un olfactomètre à dilution dynamique et d'un jury, en conformité avec la norme européenne sur l'olfactométrie (CEN/TC264/WG2Odours).

Le premier instrument analytique testé fut un GC/MS. Par contre, la méthode fut déterminée rapidement trop difficile et onéreuse. Les auteurs ont invoqué qu'un trop grand nombre d'odorants était présent dans l'odeur d'un site de traitements des déchets et qu'aucune relation valable ne pouvait être remarquée avec les concentrations de certains de ces odorants.

Le second instrument analytique employé fut un GC/FID. La méthode consiste à établir une relation entre la quantité totale de carbone organique et la concentration-odeur du mélange. La figure qui suit présente les résultats obtenus par les auteurs au moyen de cette méthode.

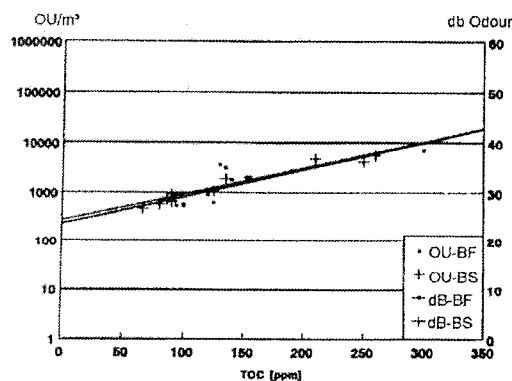


Figure 1.20 Graphique de la concentration-odeur en fonction de la de la quantité totale de carbone organique (TOC) (Bockreis et Jager, 1999)

Les auteurs précisent que cette relation est valable uniquement pour un site et un élément précis. Une corrélation générale est impossible car les données d'un site de traitement à un autre sont trop dissemblables. Ils soulignent que l'absence du calcul des concentrations de H_2S et de NH_3 , deux odorants importants, peut être une explication de la piètre performance de ce système.

Le dernier système éprouvé par les deux chercheurs est un nez électronique commercial (FOX3000) couplé à un réseau de neurones artificiel. Ils précisent que l'entraînement de la machine neuronale avec des données provenant de l'olfactométrie ne peut être meilleur que le sens de l'odorat du jury olfactométrique ayant produit les résultats. Plusieurs réseaux ont été construits et comparés. Quatre méthodes d'entraînement ont été utilisées, soit la rétro-propagation par mode *batch*, la rétropropagation par mode séquentielle, la propagation rapide (*quickpropagation*) et l'algorithme appelé *Weigen Weight Eliminator*. L'entraînement s'est fait avec 27 données expérimentales couvrant une plage de concentration-odeur entre 40 u.o./m³ et 1000000 u.o./m³. L'architecture optimale des réseaux pour chacun des algorithmes d'entraînement a été déterminée par essais et erreurs par les chercheurs, mais n'est malheureusement pas donnée.

Les résultats obtenus vont dans le même sens que l'étude précédente (Stuetz *et al.*, 1998). Les réseaux ne sont valables que pour un élément d'un site spécifique. L'erreur moyenne d'entraînement, oscillant entre 12 % et 19 %, pour les réseaux quantifiant l'ensemble des odeurs des différents sites, semble trop importante pour les auteurs

comparativement aux succès obtenus par des réseaux déterminant la concentration-odeur d'un élément d'un site particulier, dont la moyenne des erreurs lors des entraînements est de 1 % pour trois des quatre réseaux. Le quatrième réseau obtient une moyenne d'erreur de 10 %.

Les auteurs font l'erreur de considérer que la minimisation du risque empirique entraîne une bonne généralisation. Ils arrivent mal à s'expliquer par la suite la mauvaise performance de leurs réseaux face à des données de validation, dont ils ne donnent pas le nombre. Ils font alors l'hypothèse que la performance de généralisation est fonction de l'algorithme utilisé. Ils taisent leurs résultats de validation et finissent par conclure que leur nombre de données expérimentales disponibles pour l'entraînement est trop petit.

Le nombre de données disponibles pour l'entraînement n'est pas suffisant pour l'utilisation d'une minimisation maximale du risque empirique. Les auteurs auraient sûrement obtenu de meilleurs résultats de généralisation lors de l'étape de validation en employant la technique de l'arrêt prématuré.

La conclusion générale pouvant tout de même être tirée de cette recherche, et de la littérature scientifique en général, est que le nez électronique présente une option intéressante de suivi en continu des odeurs, dont la technique se doit d'être peaufinée.

CHAPITRE 2 - OBJECTIFS

Dans cette section, on trouvera d'abord les objectifs de recherche dont ce mémoire de maîtrise veut rendre compte, puis les explications relatives à leur choix.

L'objectif général de la recherche est de déterminer la faisabilité de l'utilisation de trois types de machines neuronales comme approximateurs de la concentration-odeur de l'odeur de biogaz habituellement rencontrée en air ambiant. Les trois types de machines sont les réseaux de neurones artificiels, les réseaux de fonctions à bases radiales ainsi que les machines à supports vectoriels. Le corollaire direct de celui-ci est d'arriver à déterminer quels modèles de machines neuronales minimisent le risque de généralisation pour un nombre fixe de données d'entraînement.

À cet objectif général et à son corollaire se greffe un ensemble de sous-objectifs présentés dans la liste suivante :

1. Vérifier le respect de la théorie de l'apprentissage statistique prévoyant que la minimisation du risque structurel est supérieure à la minimisation du risque empirique pour la détermination de la machine la plus performante.
2. Déterminer s'il existe une différence notable entre la performance des machines neuronales entraînées pour effectuer l'approximation de la concentration-odeur de deux sources distinctes d'odeur de biogaz et celles entraînées pour une seule.

3. Déterminer si pour un type de machine neuronale et un nombre fixe de données disponibles pour l'entraînement, l'architecture optimale est la même pour deux sources d'odeur de biogaz distinctes modélisées séparément
 - a. Pour les réseaux de neurones, ceci consiste à vérifier si le nombre de couches cachées et de neurones par couches cachées est identique pour les deux réseaux optimaux
 - b. Pour les réseaux de fonctions à bases radiales, ceci consiste à déterminer si le nombre de centres et l'écart-type des fonctions gaussiennes sont les mêmes pour les deux réseaux optimaux
 - c. Pour les machines à supports vectoriels, ceci consiste à vérifier si le noyau utilisé et les paramètres de celui-ci sont identiques pour les deux régressions optimales
4. Déterminer s'il existe un nombre de données d'entraînement permettant un compromis acceptable entre une bonne généralisation et une expérimentation minimale.

Ce dernier sous-objectif est important dans la mesure où la récolte de données pour l'entraînement est une étape fastidieuse et onéreuse devant être réduite au maximum. Ceci s'inscrit dans une volonté de développer une méthodologie d'entraînement permettant l'implantation de nez électroniques dans un laps de temps le plus court possible.

Le troisième sous-objectif continue dans le même sens en donnant l'assurance, s'il est vérifié, que pour une odeur de biogaz, le modèle optimal de machine neuronale permettant la meilleure généralisation, pour un nombre de données d'entraînement fixe, serait déjà connu. Seul l'entraînement des paramètres de ce modèle particulier devrait être fait, sans tester un nombre d'architectures important, dans le cas des réseaux, ou de noyaux différents, dans le cas des machines à supports vectoriels. Pour faire suite aux propos du paragraphe précédent, la vérification de ce troisième sous-objectif permettrait de s'assurer que pour un nombre de données minimisant le coût d'implantation du système, le modèle de machine neuronale à utiliser serait déjà fixé. La détermination du modèle optimal serait éliminé du processus d'entraînement et un gain temporel substantiel serait obtenu.

Le second sous-objectif tente de vérifier si les machines neuronales et les algorithmes d'entraînement utilisés dans le cadre de cette recherche permettent une approximation plus générale de la concentration-odeur que celles présentées dans la revue de la littérature scientifique de la section précédente. Dans une moindre mesure, ce sous-objectif vérifie si la modélisation de l'odeur de biogaz est reproductible ou si chaque source nécessite une machine neuronales qui leur est propre.

Le dernier sous-objectif évalue la validité de l'utilisation de la minimisation du risque structurel dans le cadre de cette recherche et de sa supériorité sur la minimisation du risque empirique.

CHAPITRE 3 - METHODOLOGIE

Cette section s'intéresse en grande partie au traitement de l'information provenant des nez électroniques. L'échantillonnage, l'olfactométrie, le nez électronique ainsi que ses capteurs seront d'abord présentés. Le pré-traitement de l'information fera l'objet d'une courte sous-section suivie de la présentation des algorithmes d'entraînement.

ÉCHANTILLONNAGE ET OLFAC TOMÉTRIE

Les échantillons de biogaz proviennent de deux sites distincts d'enfouissements, soit les champs 1 et 3 d'un site d'enfouissement sanitaire de la grande région de Montréal. Dans la présentation et l'analyse des résultats qui suivra, le champ 1 sera identifié sous le vocable de *Source A* et le champ 3 sous celui de *Source B*. Les deux champs émettent des odeurs de composition chimique identique mais les odorants sont en proportions différentes, tel que prouvé par la figure qui suit.

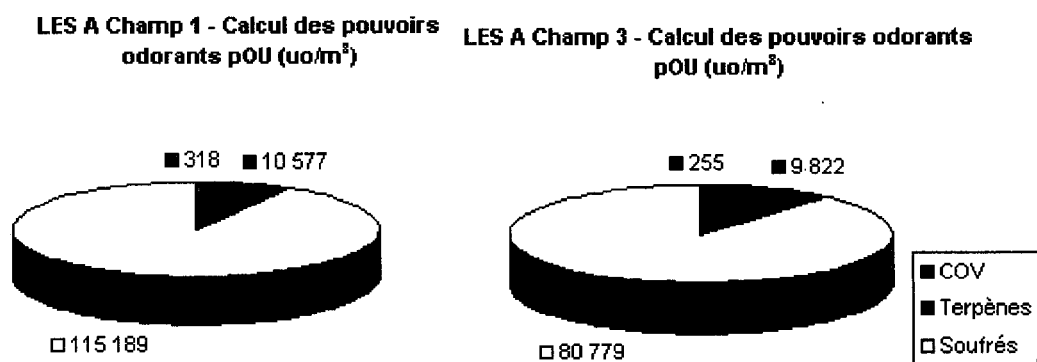


Figure 3.1 Pouvoir odorant des composés odorants des deux champs à l'étude

L'analyse complète des compositions chimiques ainsi que les résultats de la détermination des concentrations-odeurs par olfactométrie des échantillons sont fournis à l'Annexe II. À la source, le biogaz est composé de 58 % de méthane pour la source A et de 59 % pour la source B.

Le nombre d'échantillons de biogaz pur prélevé dans des sacs de Tedlar est respectivement de quatre pour la *Source A* et de 6 pour la *Source B*. La figure qui suit présente l'échantillonnage du biogaz directement au puits de captage.

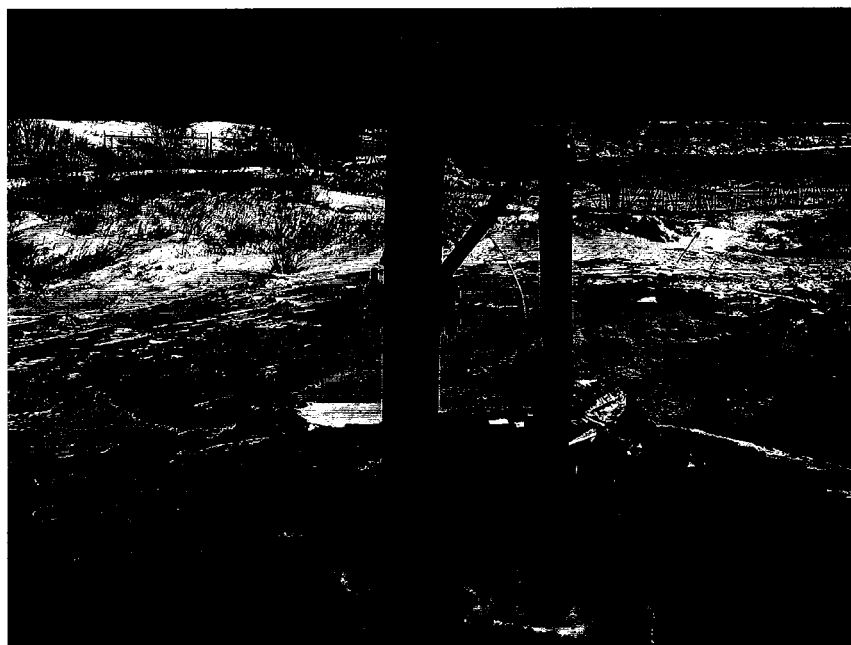


Figure 3.2 Échantillonnage du biogaz à la source

Les concentrations-odeurs des échantillons oscillent autour de 1 200 000 u.o./m³. Ces échantillons sont dilués dans d'autres sacs de Tedlar avec de l'air pur et sec pour atteindre une concentration-odeur d'environ 5000 u.o./m³. Les échantillons purs sont utilisés pour produire plusieurs échantillons dilués. Par contre, chaque nouvel

échantillon créé par dilution est quantifié par olfactométrie. Ce sont ces mêmes échantillons dilués qui serviront ensuite à l'entraînement du nez électronique.

La détermination des concentrations-odeurs a été effectuée par le même jury de six personnes pour l'ensemble des échantillons. Chaque échantillon dilué de biogaz est quantifié par olfactométrie à dilution dynamique, selon la méthode présentée dans la section de la revue de la littérature scientifique. La figure qui suit présente le fût où sont installés les sacs de Tedlar contenant l'odeur diluée ainsi que l'opérateur de l'olfactomètre.



Figure 3.3 Fût et opérateur de l'olfactomètre

La figure suivante présente une jurée lors du processus de quantification de l'odeur de biogaz.

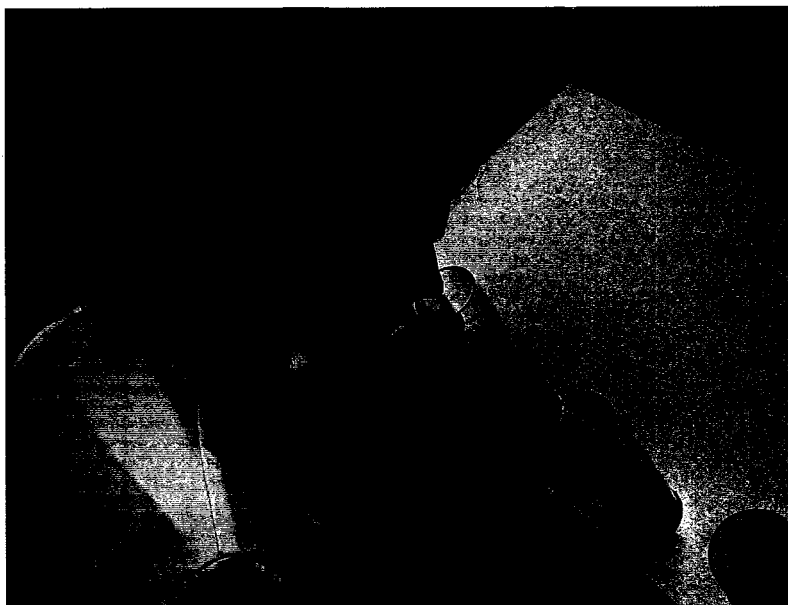


Figure 3.4 Jurée d'olfactométrie

LE NEZ ÉLECTRONIQUE

Le nez électronique employé dans le cadre de cette recherche est commercialisé par la firme Odotech inc. Il est composé de 16 capteurs commerciaux d'oxyde d'étain non-spécifiques contenus dans une chambre en acier inoxydable de 1 litre. La liste de ces capteurs est présentée au tableau qui suit.

Les différentes dilutions nécessaires à l'obtention des concentrations-odeurs sont effectuées par le même olfactomètre à dilution dynamique que pour la quantification des échantillons dilués de biogaz. Suite à la quantification des échantillons, le nez électronique est relié directement à la sortie de l'olfactomètre et soumis aux dilutions désirées de l'odeur. Tous le processus d'échantillonnage, de dilution, de quantification

des odeurs diluées et de la caractérisation des odeurs par le nez électronique est exécuté en moins de 24 heures, pour assurer que la composition du mélange gazeux ne subisse pas de variations notables.

La prise de données par le nez électronique fonctionne selon la procédure suivante:

1. Purge en continu de la chambre des capteurs avec de l'air pur (air zéro), débit de 20 L/min, jusqu'à l'atteinte d'une ligne de base stable.
2. Alimentation de l'odeur en continu, au débit de 20 L/min, à la chambre des capteurs pour une période de 90 secondes.
3. Purge en continu de la chambre des capteurs avec de l'air pur, débit de 20 L/min, pour une période de 5 minutes.
4. Répéter les étapes 2 et 3

La variable de sortie mesurée des capteurs est la tension, en volts, fourni à celui-ci pour maintenir le courant constant. La source d'énergie se comporte donc en source de courant. La résistance offerte par les capteurs est calculée par le système d'acquisition des données et sauvegardée dans un fichier. La fréquence de lecture est de une par seconde pour chacun des capteurs. La figure qui suit présente le nez électronique (le boîtier métallique) ainsi que l'ordinateur effectuant la récupération des signaux.

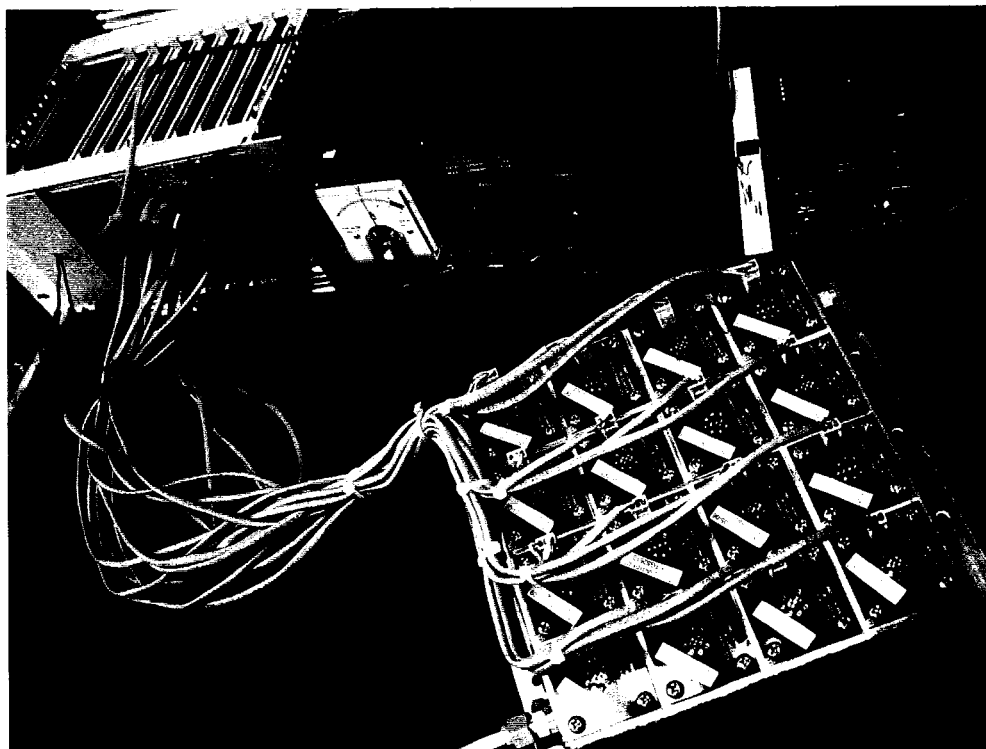


Figure 3.5 Photographie du nez électronique

Le nombre de données expérimentales produites au total et disponibles pour l'entraînement et la validation est de 143 couples vecteur-réponse/concentration-odeur. Les groupes de données d'entraînement et de validation sont donnés à l'Annexe II, avec leur résultats correspondants. Les concentrations-odeurs mesurées par le nez électronique sont comprises entre 0 u.o./m³ et 200 u.o./m³.

Toute cette partie a été effectuée en collaboration entre l'auteur de ce mémoire et des employés de l'entreprise Odotech inc. L'auteur du présent mémoire a donc traité les données brutes de résistance des capteurs.

LE PRÉ-TRAITEMENT

La figure de la page suivante présente la réponse d'un capteur d'oxyde d'étain soumis à une odeur de biogaz ainsi que sa modélisation.

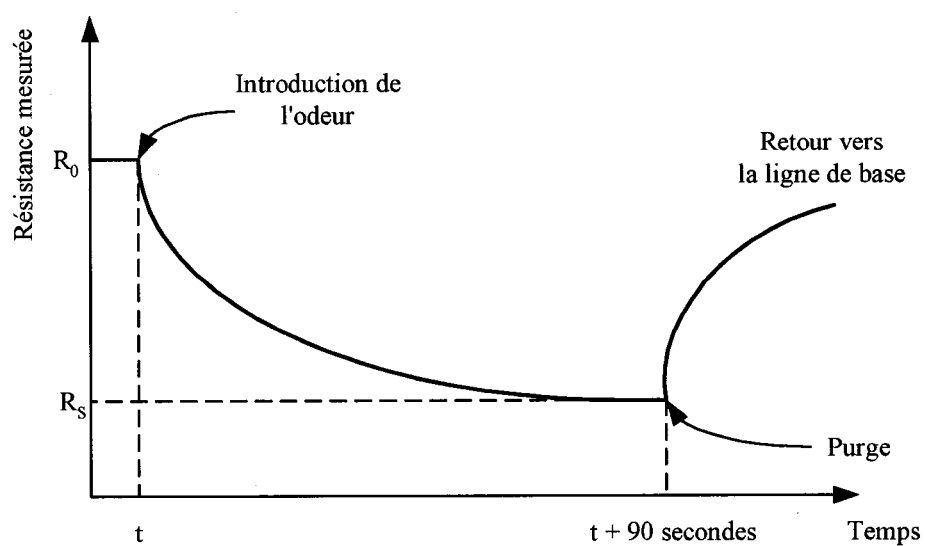
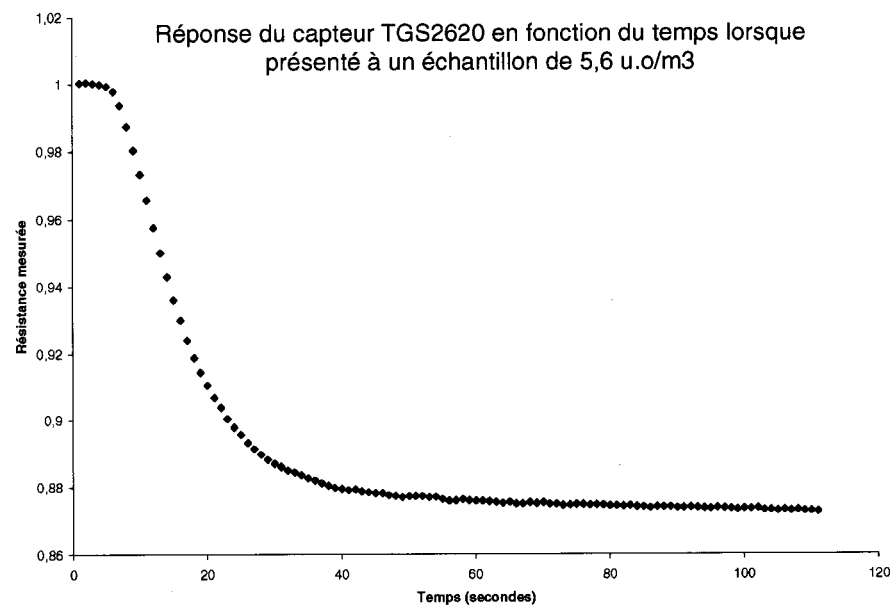


Figure 3.6 Réponse d'un capteur

La valeur de la résistance du capteur lorsque celui-ci est soumis à de l'air pur est dénoté R_0 . Lorsque soumis à l'odeur, la résistance du capteur diminue habituellement jusqu'à l'atteinte d'un plateau. La valeur de la résistance après 90 secondes d'exposition à l'odeur est nommé R_s . Pour une certaine concentration-odeur, une certaine résistance R_s

est atteinte. Pour chaque concentration-odeur, la résistance R_S de chacun des capteurs est divisée par leur ligne de base respective R_0 prise juste avant l'introduction de l'odeur dans la chambre. Cette transformation permet d'obtenir un vecteur à 16 dimensions, une dimension par capteur, représentatif de la dilution. La matrice des données du nez électronique prend alors la forme

$$\begin{aligned}
 X = [& \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur1}}^{\text{Conc-Odeur1}} & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur2}}^{\text{Conc-Odeur1}} & \dots & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur16}}^{\text{Conc-Odeur1}} \\
 & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur1}}^{\text{Conc-Odeur2}} & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur2}}^{\text{Conc-Odeur2}} & \dots & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur16}}^{\text{Conc-Odeur2}} \\
 & \dots & \dots & \dots & \dots \\
 & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur1}}^{\text{Conc-OdeurN}} & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur2}}^{\text{Conc-OdeurN}} & \dots & \left\{ \frac{R_S}{R_0} \right\}_{\text{Capteur16}}^{\text{Conc-OdeurN}}]
 \end{aligned}$$

pour un nombre de 16 capteurs et de N concentrations-odeurs mesurées.

Ce sont ces données qui seront ensuite utilisées comme entrées pour l'entraînement et la validation des machines neuronales, en association avec la concentration-odeur correspondante comme variable désirée de sortie.

ENTRAÎNEMENT ET VALIDATION DES MACHINES NEURONALES

L'objectif de l'entraînement est de trouver la fonction d'approximation de la concentration-odeur, avec comme intrants les quotients des résistances R_S/R_0 , permettant la meilleure généralisation pour chacun des types de machine neuronale. La généralisation est évaluée avec un groupe de données de validation non utilisées lors de l'entraînement des machines neuronales.

Les données recueillies sont bien distribuées et indépendantes, conditions nécessaires pour l'analyse.

La validation est évaluée pour chaque machine neuronale avec la moyenne des erreurs au carré entre la valeur de concentration-odeur attendue et celle calculée par la machine neuronale.

Toute la méthodologie qui suit a été implantée à l'aide du logiciel *MatLab*. Les fonctions utilisées propriété de *MatLab* se trouvent dans le *Neural Network Toolbox*, le *Statistics Toolbox* et le *Optimization Toolbox*. Les autres fonctions sont libres d'accès, ce qui est le cas du *SVM Toolbox*, ou ont été développées par l'auteur du mémoire. L'ensemble des fonctions non propriétaires de *MatLab* se trouvent à l'Annexe II.

Réseaux de neurones artificiels

La méthodologie d'entraînement et de validation des réseaux de neurones artificiels permettant l'identification du modèle de réseau diminuant au maximum le risque de généralisation doit inclure une composante de minimisation du risque structurel à celle de minimisation du risque empirique déjà incluse dans l'algorithme de rétro-propagation.

Cette minimisation du risque structurel s'effectue selon le principe inductif présenté dans la section sur la théorie de l'apprentissage statistique. Pour minimiser le côté droit de la fonctionnelle du risque, il devient alors nécessaire de faire de la dimension VC une variable de contrôle. Le principe inductif de la minimisation du risque structurel permettant la minimisation simultanée du risque empirique et de l'intervalle de confiance (Vapnik, 1995).

Soit l'ensemble Φ de fonction $S(\mathbf{x}, \alpha)$, imbriqué dans une structure de sous-ensembles de fonctions

$$\Phi_i = \left\{ S(\mathbf{x}, \alpha), \alpha \in \Lambda_i \right\}_{i=1}^p \quad (3.1)$$

tel que

$$\Phi_1 \subset \Phi_2 \subset \dots \subset \Phi_n \quad (3.2)$$

Où les éléments de la structure satisfont que la dimension VC h_i de chacun des ensembles Φ_i de fonctions est finie, donc que

$$h_1 \leq h_2 \leq \dots \leq h_n \quad (3.3)$$

Il est possible de traduire ce principe en processus à deux étapes :

1. Pour une structure de sous-ensembles d'architectures de réseaux de neurones artificiels, l'erreur empirique est réduite au maximum pour chacune des architectures
2. L'architecture avec le risque total (risque empirique + intervalle de confiance) est identifiée grâce au processus de validation.

L'entraînement, soit la minimisation du risque empirique, des différents réseaux de neurones testés s'est fait avec quatre algorithmes modifiés de rétro-propagation :

1. *Fletcher-Powell conjugate gradient backpropagation*
2. *Polak-Ribiere conjugate gradient backpropagation*
3. *Scaled conjugated gradient backpropagation*
4. *BFGS quasi-Newton backpropagation*

Les architectures de réseaux testés sont à une ou deux couches cachées, chacune oscillant entre 1 et 30 neurones, à fonctions d'activation de forme tangente hyperbolique et d'une couche de sortie à neurone unique à fonction d'activation linéaire.

Les réseaux de fonctions à bases radiales

La méthodologie d'entraînement et de validation des réseaux de fonctions à bases radiales permettant l'identification de l'architecture de réseau diminuant au maximum le risque de généralisation doit inclure une composante de minimisation du risque structurel à celle de minimisation du risque empirique déjà incluse dans l'algorithme de rétro-propagation.

Cette minimisation du risque structurel s'effectue selon le même principe inductif que celui présenté dans la sous-section précédente sur la méthodologie d'entraînement et de validation des réseaux de neurones artificiels.

Les deux types d'entraînement varient la dimension VC en modifiant la structure des réseaux. Dans le cas des réseaux de fonctions à bases radiales, le nombre de neurones de l'unique couche cachée varie entre 2 et le nombre maximal possible de neurones, équivalent au nombre de données d'entraînement. L'écart-type des fonctions gaussiennes a aussi été varié, entre des valeurs de 0,01 et 1,01.

Deux méthodes de détermination des centres des fonctions gaussiennes sont utilisées pour s'assurer de trouver la solution minimisant le risque empirique au maximum, soit la sélection supervisée des centres et la solution utilisant tous les points comme centres.

Les machines à supports vectoriels

Les machines à support vectoriels testées ont utilisé deux différents types de noyaux. Ainsi, pour le noyau polynomial le degré varie de 1 à 4. Pour le noyau de fonctions à bases radiale, l'écart-type varie de 0,01 à 3,01. Quant au paramètre C , il varie de 0,001 jusqu'à 1000 par saut de puissance.

CHAPITRE 4 - RESULTATS ET DISCUSSION

Au total, 143 points expérimentaux ont été recueillis, couvrant les réponses des 16 capteurs du nez électronique sur une plage de concentrations-odeurs allant de 0 u.o./m³ à 200 u.o./m³. Les mêmes concentrations-odeurs ont été présentées à plusieurs reprises, soit deux, trois ou quatre fois, au nez électronique. Les odeurs de biogaz de deux zones d'enfouissement distinctes ont été utilisées. La plage des concentrations mesurées à l'aide du nez électronique ainsi que le choix des répétitions de ces mesures ont été faits par Odotech inc.

La section qui suit présente les résultats obtenus des données brutes fournies par Odotech inc. La première sous-section présentera la transformation des données brutes par le pré-traitement. La seconde sous-section présentera les résultats généraux de la construction des machines neuronales. Les sous-sections qui suivront répondront chacune à un des sous-objectifs présentés précédemment.

PRÉ-TRAITEMENT

Les intrants de la fonction d'approximation générée par la machine neuronale sont les rapports R_s/R_0 , introduits et expliqués dans la section de la méthodologie. Les rapports de chacun des 16 capteurs sont calculés pour chacun des 143 points expérimentaux. La figure qui suit présente le graphique des rapports pour le capteur 03 en fonction des concentrations-odeurs soumises au nez électronique.

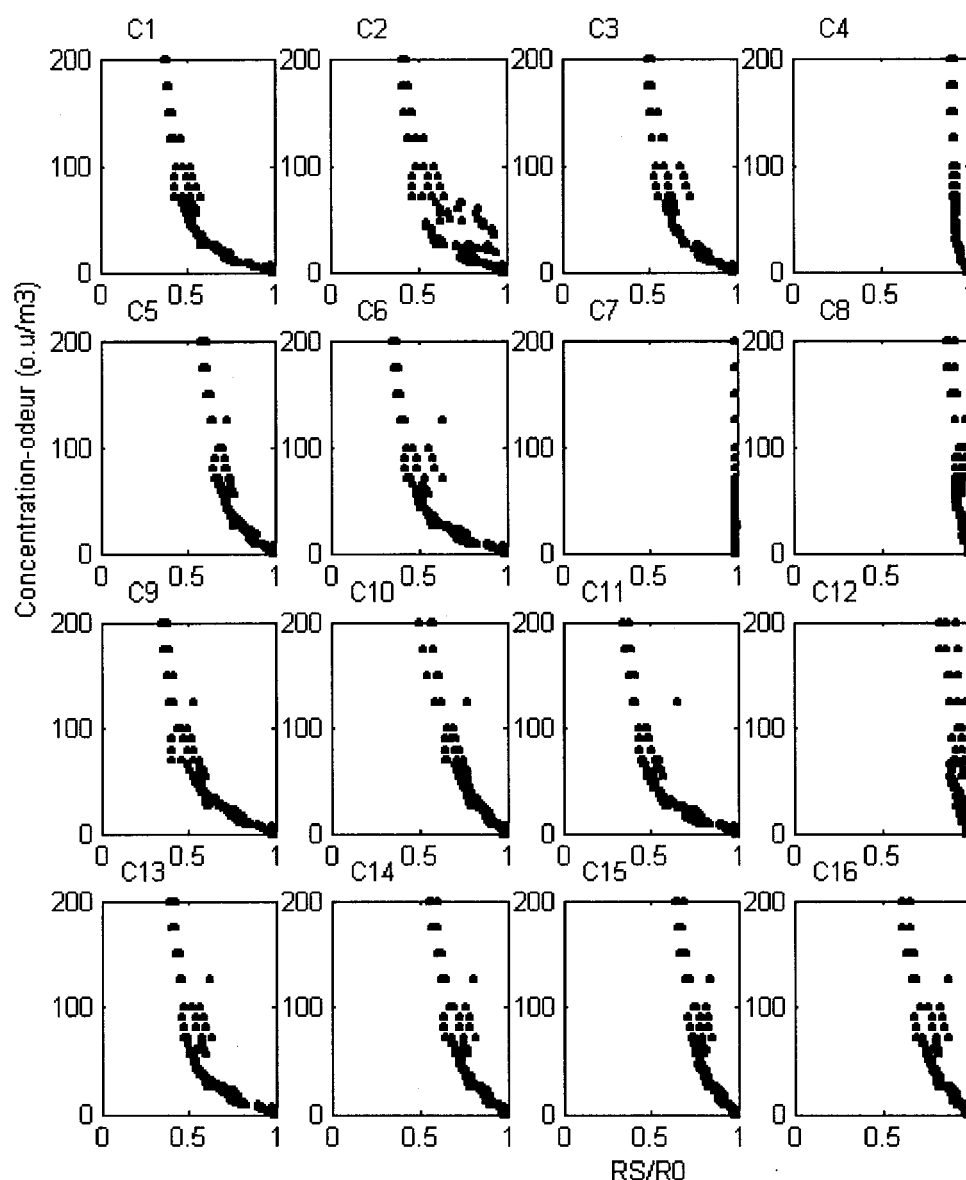


Figure 4.1 Rapports R_S/R_0 des capteurs en fonction de la concentration-odeur pour l'ensemble des résultats expérimentaux

Ce type de relation est caractéristique des capteurs utilisés et des résultats obtenus dans le cadre de cette recherche. La valeur du rapport R_S/R_0 diminue avec l'augmentation de la concentration-odeur jusqu'à l'atteinte d'un plateau dans les concentrations-odeurs les plus élevées où les capteurs saturent. Les valeurs de R_S/R_0 sont toujours comprises entre

0 et 1, sauf de rares exceptions. Ces exceptions peuvent survenir lorsqu'un des composés de l'odeur a une action oxydante majeure sur la surface du capteur, inhibant les effets réducteurs des autres molécules, et en augmente la résistance. Une autre explication peut être que, dans les très faibles concentrations-odeurs où le rapport R_S/R_0 est encore très proche de 1, le bruit du signal représente une proportion non négligeable de la différence entre la résistance R_S et la ligne de base R_0 . Une action additive du bruit sur la valeur mesurée peut donc entraîner un dépassement de la valeur unitaire du rapport. C'est cette dernière situation qui occasionne habituellement les dépassements de la valeur unitaire du rapport R_S/R_0 . Il n'a d'ailleurs pas été observé dans les résultats utilisés des manifestations de la première explication.

Une autre caractéristique intéressante à constater est la difficulté pour le capteur à renvoyer une réponse précise et reproductible sur une certaine plage de concentrations-odeurs. Pour le capteur 03, cette plage se situe entre 70 u.o./m³ et un peu plus de 100 u.o./m³. Cette situation peut être expliquée par l'une des deux raisons suivantes :

1. Les compositions chimiques des échantillons diffèrent et entraîne donc des réponses différentes pour les capteurs malgré une même concentration-odeur
2. Les dilutions effectuées par l'olfactomètre à dilution dynamique sont critiques dans l'intervalle utilisé pour atteindre ces concentrations-odeurs

Cette dernière explication peut être éliminée dans la mesure où différentes pré-dilutions de l'odeur ont été effectuées. L'olfactomètre à dilution dynamique n'enregistre pas une erreur constante, ou même linéaire, en fonction de la dilution qui lui est demandée. Cette erreur varie selon son degré d'utilisation des différentes vannes réalisant la dilution de l'odeur avec de l'air propre. Ainsi, pour prévenir l'emploi unique de certaines plages de dilution, et donc d'erreur, pour certaines concentrations-odeurs, différentes pré-dilutions de l'odeur ont été effectuées pour s'assurer que pour une même concentration-odeur l'olfactomètre travaille dans différentes zones de dilutions et donc d'erreur.

Avec les données sur les compositions chimiques des deux champs présentées dans la section de la méthodologie, la seconde explication prend tout son sens. Les effets synergiques et inhibiteurs des odorants peuvent entraîner une concentration-odeur ultime identique, bien que la composition chimique de l'odeur soit différente. Comme les capteurs réagissent selon la composition chimique (la concentration des molécules), il est normal que les valeurs des capteurs soient différentes pour une même concentration-odeur provenant d'échantillon différents.

L'utilisation de machines neuronales - plutôt que de simples régressions non-linéaires - et d'un ensemble de capteurs réagissant différemment aux différents composés pouvant être présent dans l'air ambiant - plutôt que d'un seul - pour l'approximation des concentrations-odeurs à partir des réponses des capteurs, provient essentiellement de la

nécessité ci-haut mentionnée. La figure qui suit illustre ce qui adviendrait de l'utilisation d'un seul capteur et d'une relation non-linéaire simpliste.

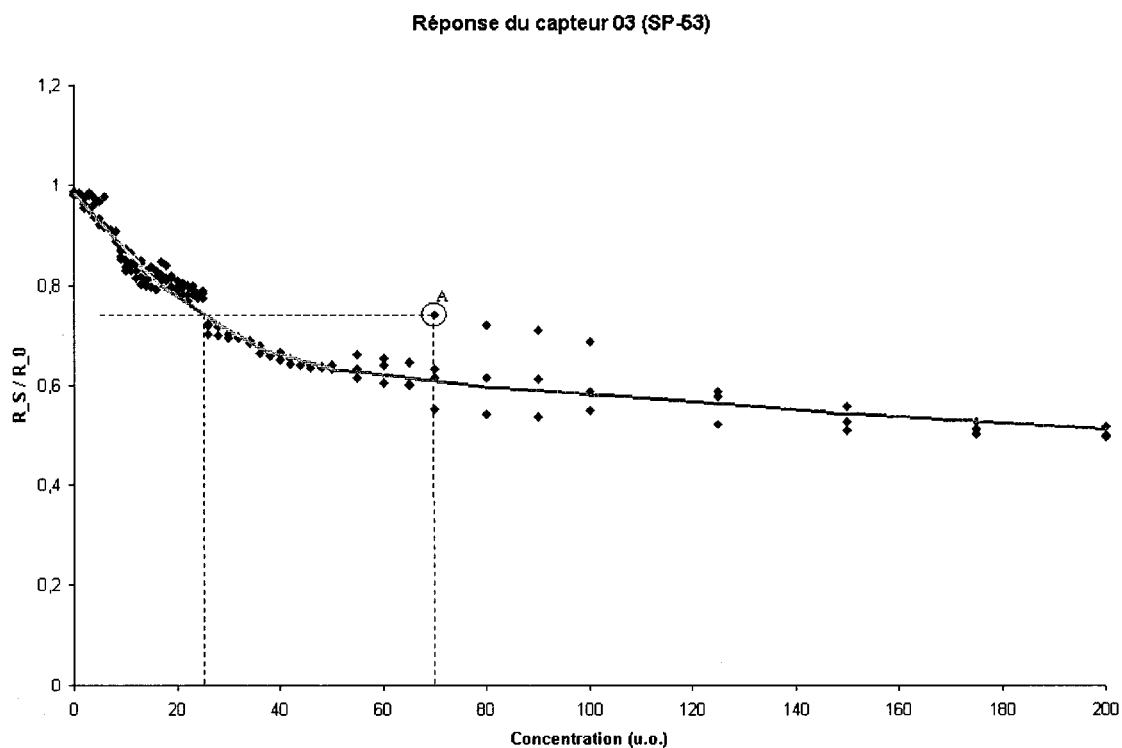


Figure 4.2 Rapports R_S/R_0 du capteur 03 en fonction de la concentration-odeur

Le point expérimental A est une des lectures obtenues du capteur 03 lorsque soumis à une concentration-odeur de 70 u.o./m³. Si cette lecture avait été utilisée avec une corrélation ayant une allure comparable à celle présentée par la courbe bleu clair sur la figure ci-haut, le résultat aurait tourné autour de 25 u.o./m³, soit une erreur de prédiction de plus de 60 %.

En utilisant, ne serait-ce qu'un autre capteur, il est possible d'approximer plus justement la valeur de la concentration-odeur en répartissant l'importance des deux lectures sur la valeur d'approximation finale. La figure qui suit montre, pour le capteur 09, la valeur mesurée pour l'échantillon A présenté dans la figure plus haut.

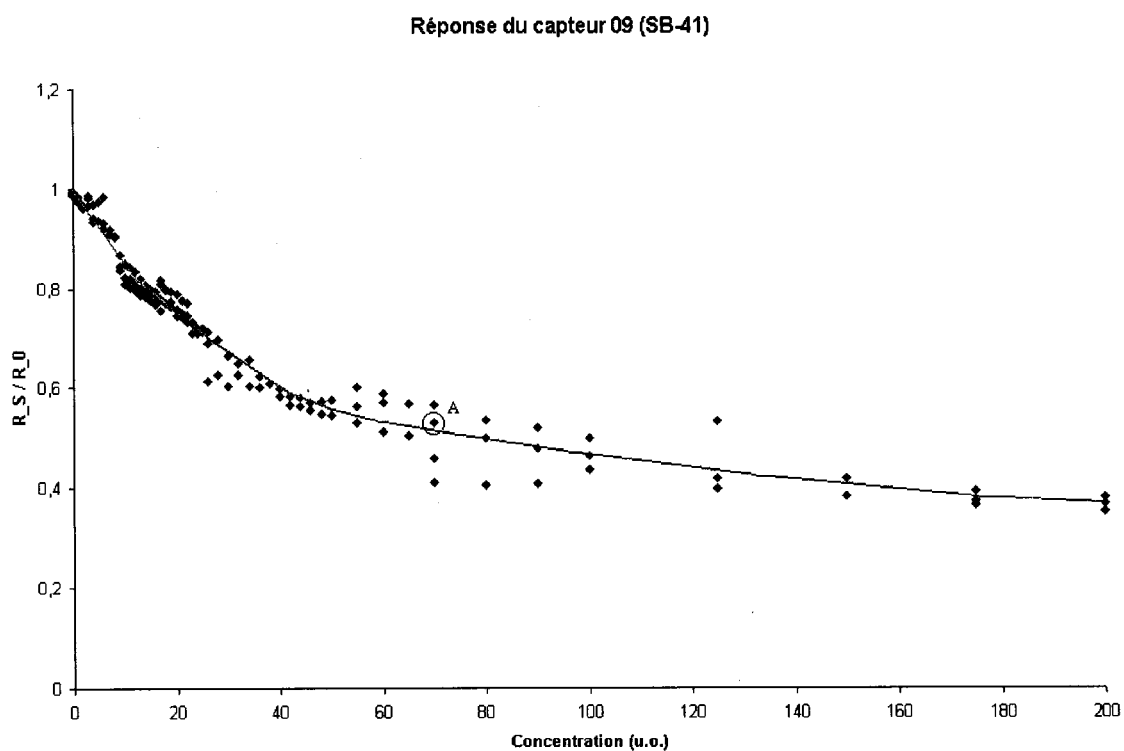


Figure 4.3 Rapports R_s/R_0 du capteur 09 en fonction de la concentration-odeur

La valeur est maintenant plus proche de ce qui pourrait être la relation entre les lectures du capteur 09 et la concentration-odeur réelle. La valeur retournée serait donc plus juste. Dans ce cas, il serait donc important d'accorder plus de valeur au capteur 09 qu'au capteur 03, si l'on veut réduire l'erreur sur la plage critique. Par contre, il serait malvenu

d'éliminer complètement le capteur 03, puisqu'il répond mieux que le capteur 09 pour des valeurs de concentrations-odeurs se trouvant entre 25 u.o./m³ et 50 u.o./m³.

Le choix de travailler avec 16 capteurs s'inscrit dans cette nécessité de réduire au maximum l'influence des zones critiques et du bruit sur la qualité de l'approximation et de la réponse finale obtenue. La combinaison des signaux des 16 capteurs permet d'extraire plus facilement l'information nécessaire à la prédiction de la concentration-odeur.

Toute la difficulté de l'approximation réside dans la recherche de l'importance optimale à accorder à chacun des capteurs et aux informations pouvant être tirées des comparaisons croisées entre les valeurs enregistrées par ceux-ci. Les machines neuronales, couplées à des algorithmes de régression, sont des outils permettant ce type de travail.

LE RESPECT DE LA THÉORIE DE L'APPRENTISSAGE STATISTIQUE

Cette première étape consiste à vérifier que les algorithmes d'entraînement créés par l'auteur de ce mémoire pour implanter la minimisation du risque structurel génèrent des résultats en accord avec la théorie de l'apprentissage statistique. Cette preuve est nécessaire dans la mesure où elle garantit que le réseau optimal atteint est bien celui minimisant le risque de généralisation.

Des trois types de machines testées, seulement les machines à supports vectoriels sont une implantation directe de la minimisation du risque structurel. Par contre, tel qu'exposé dans la section sur la méthodologie, il est possible de travailler avec les réseaux de neurones artificiels et les réseaux de fonctions à bases radiales dans un contexte de minimisation du risque structurel en créant une structure de sous-ensembles de fonctions. La recherche du modèle optimal pour chacun des deux types de réseaux se fait selon ce principe. Pour les machines à supports vectoriels, il n'y a que les paramètres définissant les noyaux utilisés qui se doivent d'être variés.

Dans le cas des réseaux de neurones artificiels, l'augmentation du nombre de neurones d'une des couches cachées entraîne une augmentation de la dimension VC du réseau. Ainsi, selon la théorie de l'apprentissage statistique, la variation du nombre de neurones d'une des couches cachées et le calcul de l'erreur de généralisation pour chacun des modèles élaborés et entraînés permettent de désigner celui minimisant le risque de généralisation. Le risque de généralisation propre à chaque modèle est évalué avec l'erreur de généralisation, soit la moyenne des erreurs de validation au carré. Il est important de noter que la théorie, qui prévoit que l'erreur de généralisation se présente sous la forme d'une cloche inversée en fonction de la dimension VC, est respectée dans la très grande majorité des vérifications ayant fait partie de cette étude. La figure qui suit présente un exemple de ce résultat.

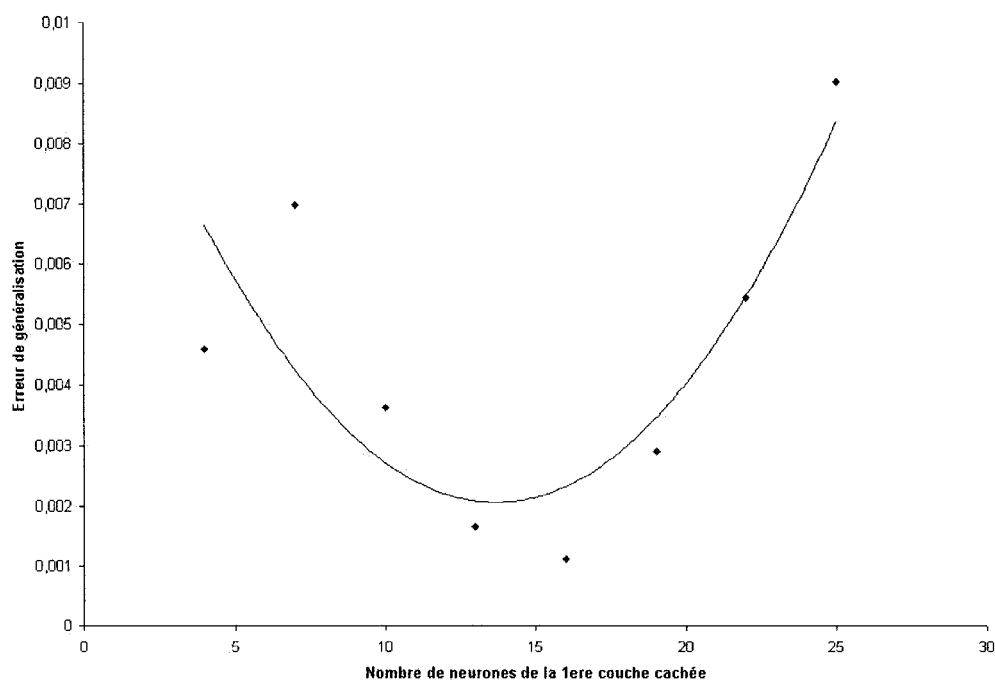


Figure 4.4 Erreur de généralisation (MSE) en fonction du nombre de neurones de la 1^{ère} couche cachée (la seconde est fixée à 12 neurones)

Dans le cas précis de la figure 4.4, le nombre de neurones idéal pour la première couche cachée est de seize si la suivante en compte douze. Le risque le plus faible sur la réponse du réseau à des données inconnues est atteint lorsque cette architecture est privilégiée. Il est certain que ce résultat n'est valable que si le nombre de douze neurones de la deuxième couche cachée est imposé. Par contre, il est possible d'étendre le travail de recherche du modèle optimal en modifiant aussi le nombre de neurones de la seconde couche. C'est ce travail de recherche du modèle optimal, en modifiant le nombre de neurones sur chacune des couches cachées et en calculant l'erreur de généralisation pour chacun des modèles élaborés et entraînés, qui constitue l'implantation de la recherche du modèle minimisant le risque de généralisation.

La supériorité du principe de minimisation du risque structurel sur celui de minimisation du risque empirique est démontrée dans la figure qui suit. Ces résultats proviennent de l'entraînement et de la validation de tous les réseaux de neurones artificiels dont la première couche varie entre 1 et 30 neurones et la deuxième entre 0 et 30. Le groupe de données d'entraînement est composé de 95 points expérimentaux et celui de validation de 48. Le réseau de neurones artificiel offrant l'erreur de généralisation la plus faible est un de ceux ayant obtenu une des erreurs empiriques les plus élevées lors de l'entraînement.

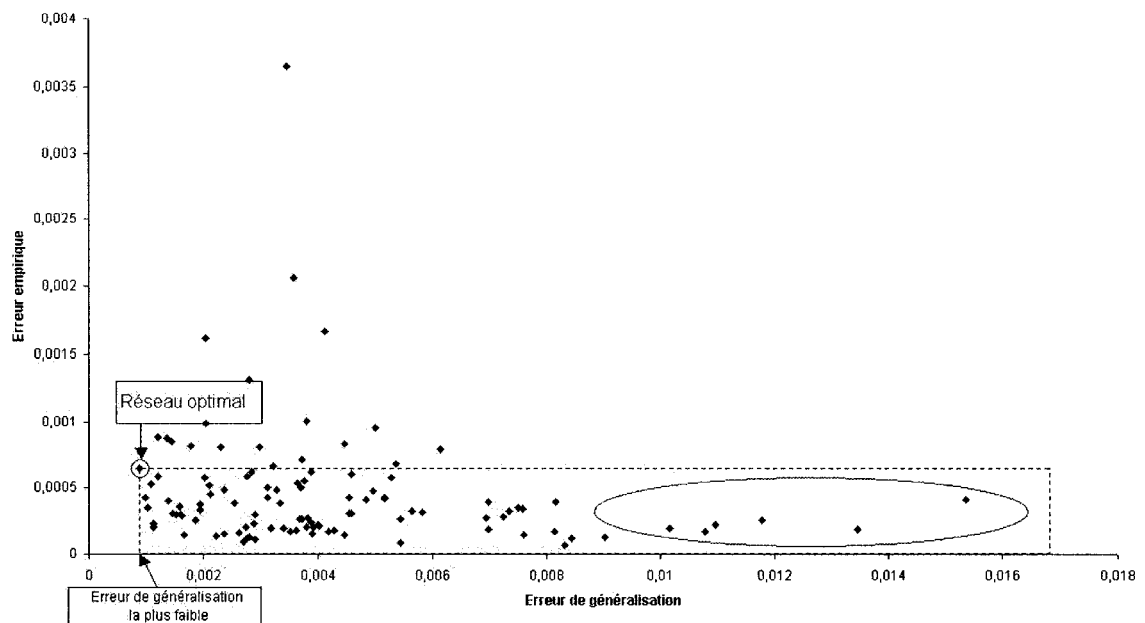


Figure 4.5 Erreur empirique en fonction de l'erreur de généralisation pour des réseaux de neurones entraînés et validés avec les mêmes points expérimentaux

Pour l'instant, la présentation de l'architecture du réseau optimal n'est pas l'objectif. Ceci sera fait dans une sous-section subséquente avec ces mêmes résultats. L'importance de cette figure est qu'elle démontre qu'une minimisation du risque empirique n'est pas un gage de minimisation du risque de généralisation. Tous les points se trouvant à l'intérieur du rectangle en pointillé représente des réseaux ayant obtenu une erreur plus faible, lors de l'entraînement, que le réseau optimal; mais n'ayant pas réussi aussi bien lorsque confrontés à des données inconnues. Cette situation est particulièrement vraie pour les réseaux dont les performances sont encadrées et qui présentent des erreurs de généralisation d'un ordre de grandeur plus élevées que celle du réseau optimal.

L'implantation de la minimisation du risque structurel pour les réseaux de fonctions à bases radiales donnent des résultats tout aussi probants. Cette minimisation se fait selon les mêmes principes que pour les réseaux de neurones artificiels, c'est-à-dire que l'augmentation du nombre de neurones entraîne une augmentation de la dimension VC du réseau et que le modèle optimal minimisant le risque de généralisation se trouve au creux de la courbe modélisant l'erreur de généralisation en fonction de la dimension VC.

La figure qui suit présente cette courbe qui prend encore la forme d'une cloche inversée, tel que le veut la théorie de l'apprentissage statistique.

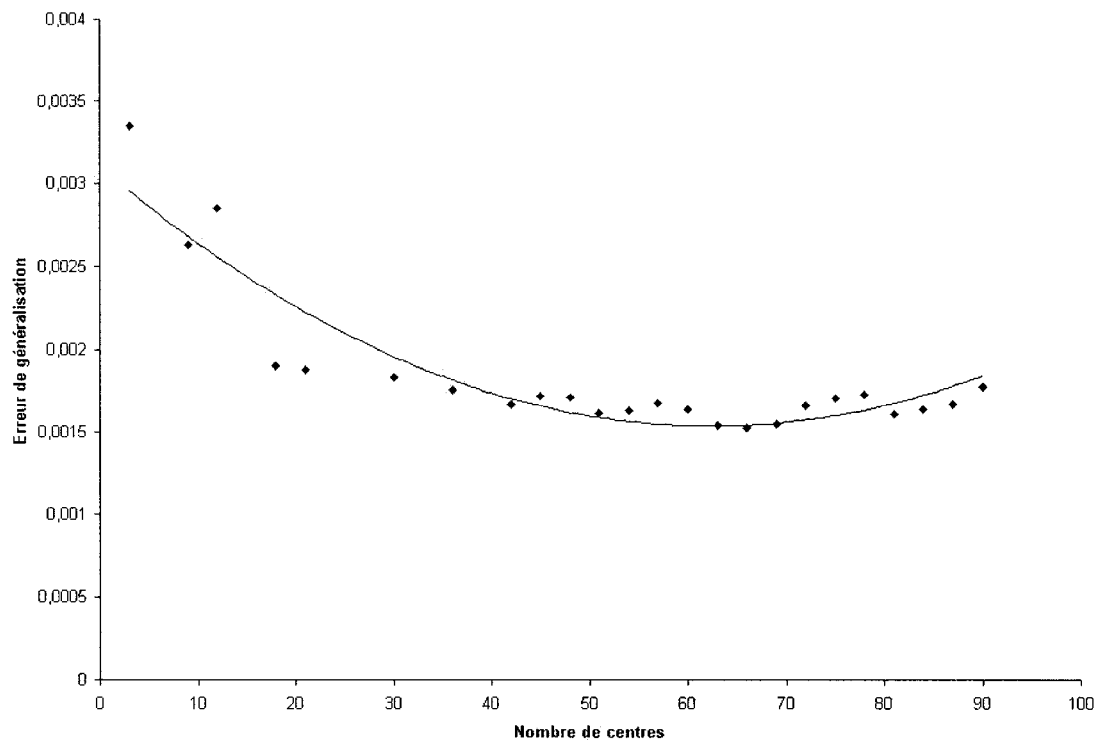


Figure 4.6 Erreur de généralisation (MSE) en fonction du nombre de neurones (centres) de la couche cachée du réseau de fonctions à bases radiales

Le nombre optimal de neurones pour ce réseau de fonctions à bases radiales oscille autour de 65. Ces résultats ont été obtenus à partir des mêmes ensembles de données d'entraînement et de validation que ceux employés pour les réseaux de neurones artificiels.

Les machines neuronales étant une implantation directe de la minimisation du risque structurel, les lecteurs sont encouragés à se référer à la section de la revue de la littérature scientifique pour avoir une preuve de leur capacité à minimiser le risque de généralisation.

Cette sous-section a montré que les algorithmes développés par l'auteur pour effectuer la minimisation du risque structurel de réseaux de neurones artificiels et à fonctions à bases radiales engendrent des résultats en accord avec la théorie statistique de l'apprentissage. De plus, il a été démontré que l'unique minimisation de l'erreur empirique sur une machine neuronale ne constitue pas une garantie de sa performance face à des données inconnues.

LES MACHINES OPTIMALES POUR L'ENSEMBLE DES SOURCES DE BIOGAZ

L'objectif principal de ce travail consiste à déterminer la validité de l'utilisation de machines neuronales comme approximateurs de la concentration de l'odeur de biogaz sans discrimination des sources. L'ensemble des données expérimentales étant majoritairement des triplicatas, il a été déterminé que les deux tiers de l'ensemble (95 données) seraient utilisées comme données d'entraînement et un tiers (48 données) comme données de validation. La répartition a été faite aléatoirement pour assurer que les deux ensembles, entraînement et validation, soient constitués de données provenant des deux sources de biogaz. Les 95 données d'entraînement sont considérées comme le maximum pouvant être utilisées et sont les mêmes pour toutes les machines entraînées. Dans les sous-sections subséquentes, le nombre de données d'entraînement sera réduit pour évaluer l'impact de celui-ci sur la qualité des machines neuronales obtenues.

La sous-section précédente a fait la preuve de la validité de l'algorithme de détermination du modèle optimal, pour chacun des types de machines neuronales à l'étude. Les résultats de cette détermination sont présentés au tableau suivant.

Tableau 4.1 Modèles optimaux de machines neuronales

Machine neuronale	Erreur de généralisation (MSE)	Erreur moyenne sur la prédiction	Architecture optimale
Réseau de neurones artificiel	0,0008963	3 %	16-4-24-1
Réseau de fonctions à bases radiales	0,0012308	3,5 %	16-65-1 $\sigma = 0.41$
Machines à support vectoriels	0,0009849	3,1 %	Noyau FBR $\sigma = 0.2, C = 1$

Les valeurs des erreurs de généralisation sont faibles si elles sont comparées aux valeurs présentées dans la revue de la littérature bibliographique de ce mémoire qui oscillaient entre 10 % et 20 %. Le succès de l'opération est sans équivoque. La discussion qui suit présente chacune des machines neuronales optimales.

Dans le cas des réseaux de neurones artificiels, le qualité du résultat était prévisible dans la mesure où pour la première fois une méthodologie garantissant le meilleur résultat était utilisée dans l'approximation de la concentration-odeur. Les méthodologies jusqu'à maintenant utilisées par les chercheurs reposaient sur le principe d'essais/erreurs. Le graphique qui suit présente les concentrations-odeurs escomptées en fonction des concentrations-odeurs calculées par le réseau de neurones artificiel optimal.

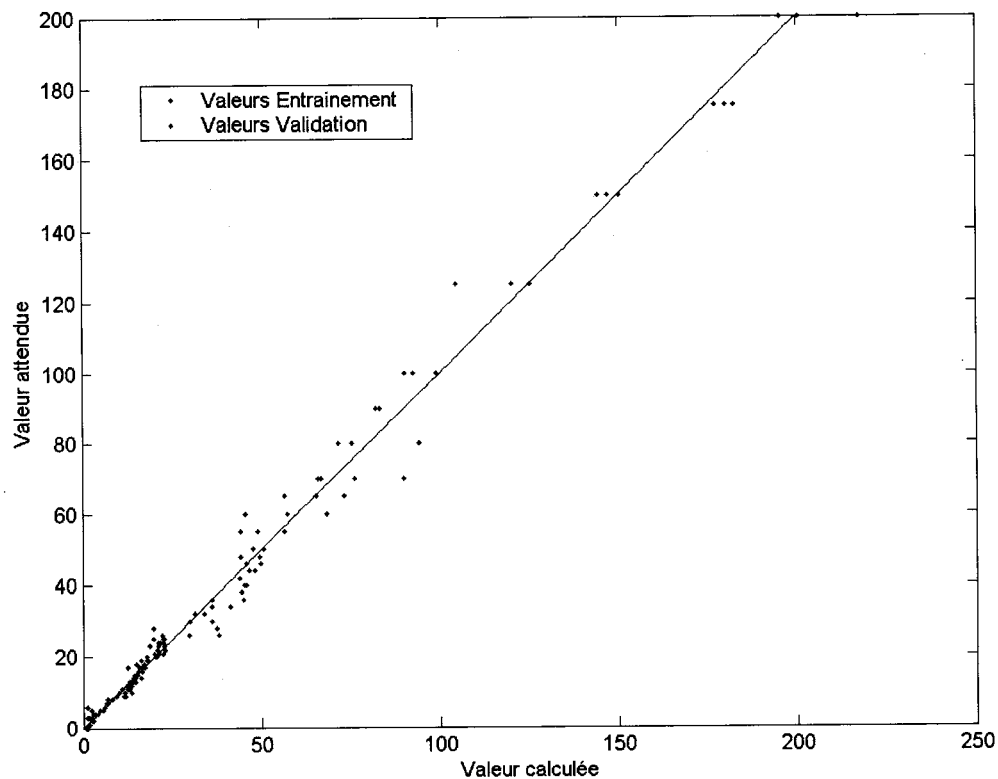


Figure 4.7 Comparaison des valeurs calculées par le réseau de neurones artificiel optimal et les valeurs attendues

Les points de validation sont presque tous dans les environs de la droite d'équivalence (*valeur attendue* = *valeur calculée*). Ce résultat prend toute sa mesure lorsqu'il est comparé à celui d'un réseau ayant une erreur de généralisation plus élevée. La figure qui suit présente les concentrations-odeurs attendues en fonction des concentrations-odeurs calculées pour un autre réseau de neurones artificiels à deux couches cachées (13 neurones pour la première, 6 pour la deuxième, $MSE = 0.0154$). Les points sont beaucoup plus éloignés de la droite d'équivalence que dans le cas du réseau optimal, et ce pour l'ensemble de la plage d'étude.

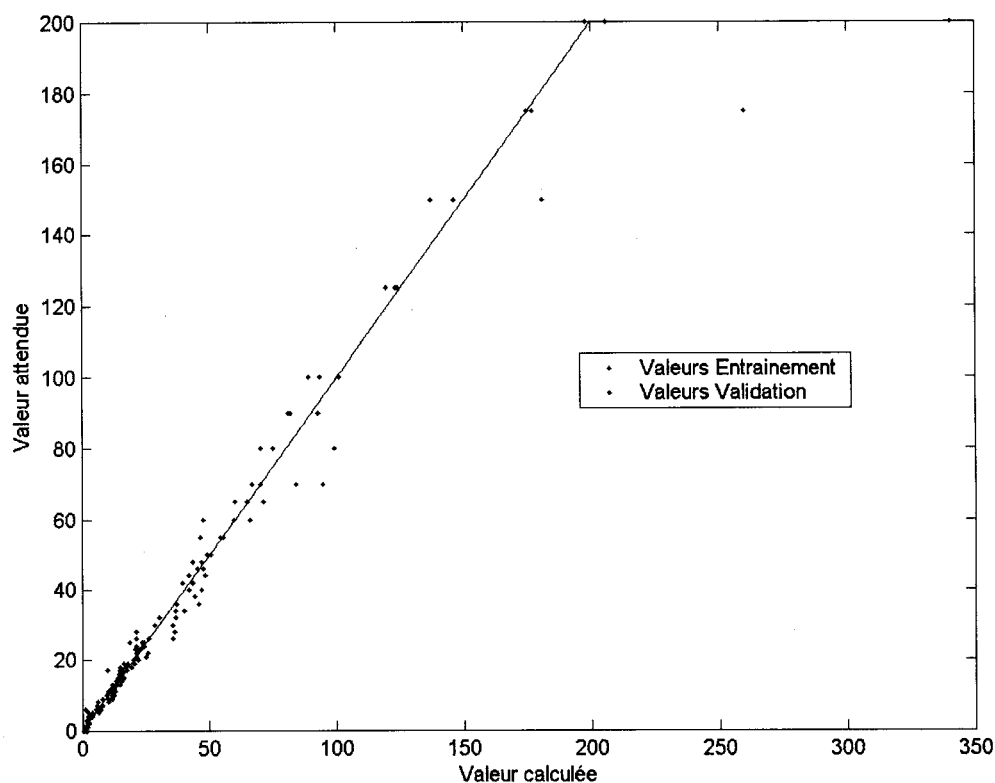


Figure 4.8 Comparaison des valeurs calculées par un réseau de neurones artificiel non optimal et les valeurs attendues

Dans la zone des basses concentrations (0 u.o./m^3 à 25 u.o./m^3), les points forment un nuage dispersé autour de la ligne d'équivalence. Dans la figure représentant le réseau optimal, ces points sont beaucoup plus rapprochés de cette ligne. Le problème devient majeur dans les plus hautes concentrations (150 u.o./m^3 et plus). Les prédictions faites par le réseau deviennent aberrantes et dangereusement erronées si ce réseau devait servir à la surveillance de l'odeur de biogaz dans un cadre législatif.

Le réseau de fonctions à bases radiales optimal n'est pas aussi performant que le meilleur réseau de neurones artificiels. Le réseau de fonctions à bases radiales minimisant le risque de généralisation contient 65 neurones. Le même type de figure qu'auparavant est présenté ici.

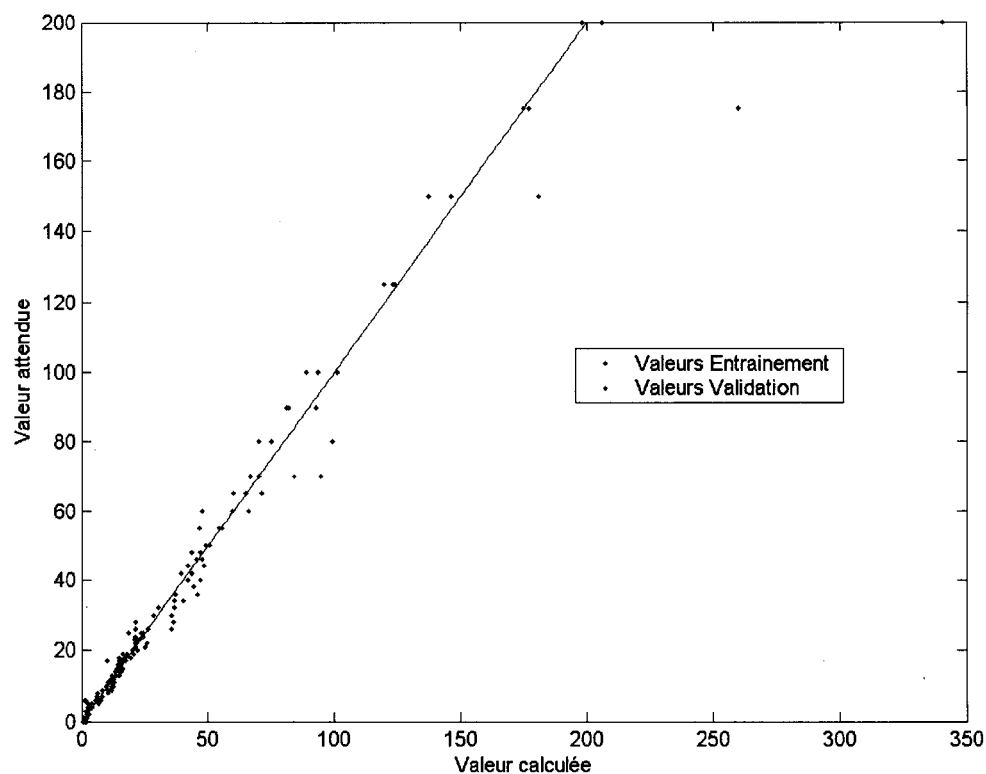


Figure 4.9 Comparaison des valeurs calculées par le réseau de fonctions à bases radiales optimal et les valeurs attendues

Bien que l'erreur moyenne des prédictions soit petite, les valeurs de concentrations-odeurs calculées par le réseau ne semblent pas s'approcher autant des valeurs attendues que dans le cas du réseau de neurones artificiel optimal. Il est intéressant de noter par contre que la majorité des points éloignés de la courbe sont des données d'entraînement.

Le résultat le plus étonnant est par contre celui de la performance des machines à supports vectoriels. La qualité de la régression se remarque par la qualité du positionnement des points sur le graphique des concentrations-odeurs attendues en fonction des concentrations-odeurs calculées par la machine à supports vectoriels optimale.

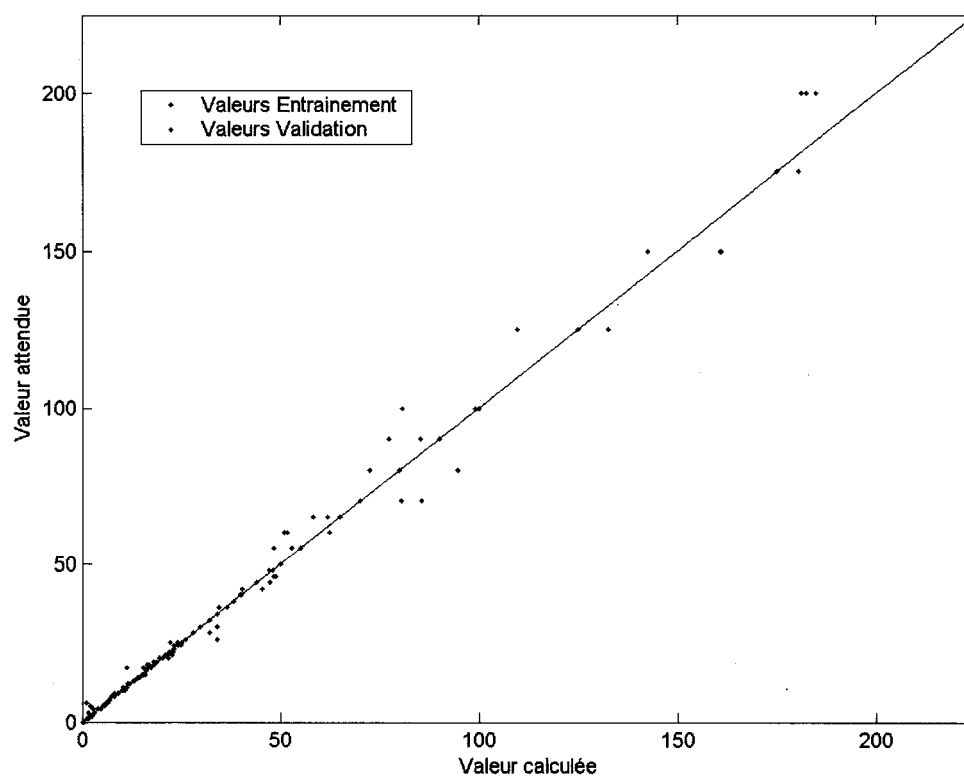


Figure 4.10 Comparaison des valeurs calculées par la machine à supports vectoriels optimale et les valeurs attendues

L'approximation des faibles concentrations (moins de 30 u.o./m³) est quasi-parfaite. En effet, la majorité des points de validation se retrouve presque exactement sur la droite d'équivalence.

Ainsi, si les trois machines neuronales optimales sont comparées d'après leur performance dans les faibles concentrations-odeurs, les machines à supports vectoriels ont une erreur de généralisation d'un ordre de grandeur moins élevé que les deux types de réseaux. Les résultats de la validation sur la plage de 0 u.o./m³ à 30 u.o./m³ sont donnés ici.

Tableau 4.2 Erreur de généralisation des machines neuronales optimales pour l'approximation des faibles concentrations-odeurs

Machine neuronale	Erreur de généralisation
Réseau de neurones artificiel	0,000325
Réseau de fonctions à bases radiales	0,000756
Machine à supports vectoriels	0,000054

Cette plage de 0 u.o./m³ à 30 u.o./m³ est la mieux définie par les données expérimentales. Chacune des trente unités de concentrations-odeurs a été caractérisée par au moins deux échantillons différents. Au-dessus de 30 u.o./m³, les données expérimentales commencent à s'espacer sur l'échelle des concentrations-odeurs et la machine à supports vectoriels optimale performe moins bien. D'ailleurs, il est impératif de souligner que la machine à supports vectoriels a obtenu une erreur de généralisation

équivalent à celles des deux autres machines neuronale sur la plage 0 u.o./m³ à 200 u.o./m³. Ceci signifie donc que la qualité de son approximation sur les concentrations-odeurs supérieures à 30 u.o./m³ est bien en deçà de celles des deux autres machines neuronales.

Face à ces résultats, il est possible de conclure que, si la plage de concentrations-odeurs à approximer est bien caractérisée par les données expérimentales, les machines à supports vectoriels sont celles des trois types de machines neuronales testées offrant le risque le plus faible d'utilisation avec des données inconnues.

Si certaines régions de la fonction à approximer sont mal définies, les trois machines neuronales se valent dans l'absolu. Par contre, selon les résultats obtenus et présentés, les réseaux de neurones artificiels semblent être les moins portés à retourner des valeurs aberrantes ou très distantes de la valeur attendue.

LE NOMBRE DE DONNÉES D'ENTRAÎNEMENT

Le nombre de données utilisées pour l'entraînement est une composante importante de la performance de généralisation des machines neurales. La figure qui suit présente le graphique de l'erreur de généralisation des machines neuronales optimales pour trois ensembles de données d'entraînement de grandeurs différentes.

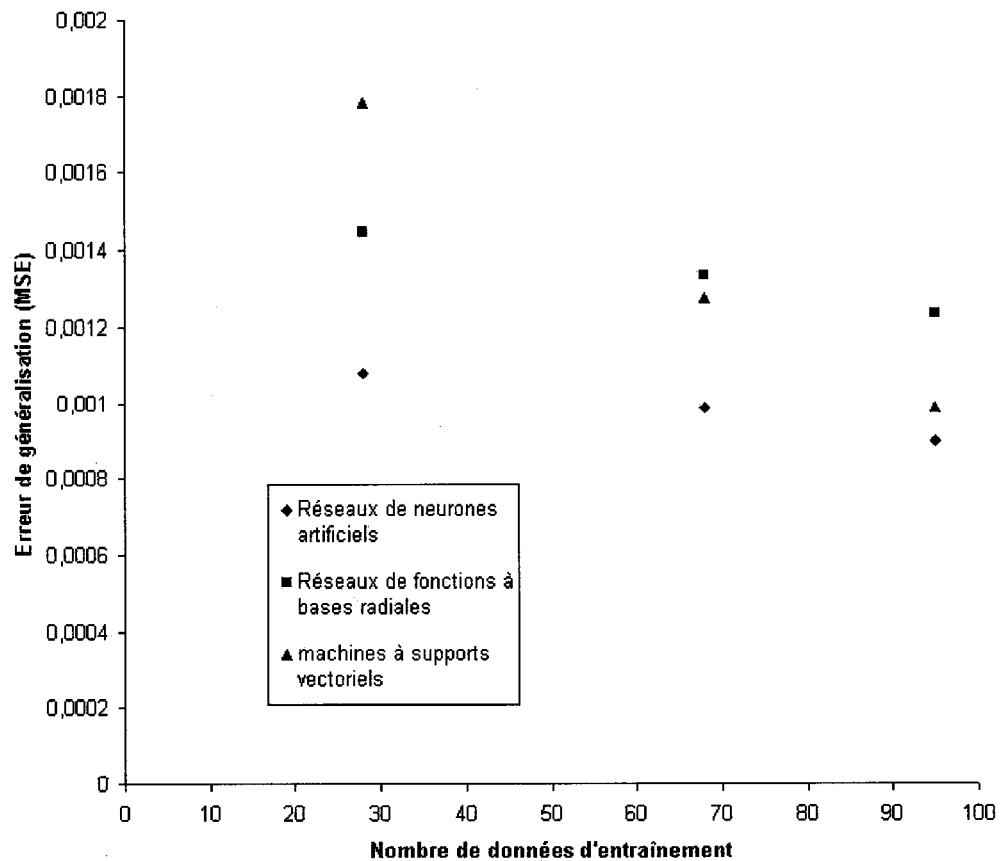


Figure 4.11 Erreur de généralisation en fonction du nombre de données d'entraînement

La théorie de l'apprentissage statistique prévoyait la diminution du risque de généralisation du modèle optimal de chacune des machines neuronales.

Un résultat intéressant est par contre la confirmation de l'importance de la définition de la région à approximer par les données d'entraînement pour que la machine à supports vectoriels offre une bonne performance. Pour la machine à supports vectoriels optimale entraînée uniquement avec 28 données d'entraînement, l'erreur est beaucoup plus grande

que pour les deux autres types de machines neuronales testées. Par contre, la diminution de l'erreur de généralisation est plus rapide dans leur cas lorsque le nombre de données augmente et caractérise davantage la plage de concentrations-odeurs à approximer. Ainsi, il peut être supposé que si le nombre de données avait été plus élevé et bien distribué sur la plage à l'étude, la machine à support vectoriels aurait offert la meilleure performance de généralisation. La preuve de cette affirmation a été faite dans la sous-section précédente. Lorsque testée uniquement sur la plage très bien caractérisée de 0 u.o./m³ à 30 u.o./m³, la machine à supports vectoriels a été beaucoup plus performante sur la minimisation du risque de généralisation.

Il n'a pas été possible de déterminer s'il existait un compromis acceptable entre le nombre de données d'entraînement et la minimisation de l'erreur de généralisation. Le nombre de données critique aurait été celui précédant une stabilisation de l'erreur de généralisation lorsque le nombre de données est augmenté.

Par contre, dans l'optique où ce nombre aurait été trouvé, il est intéressant de vérifier si le principe de minimisation du risque structurel, prévoyant qu'un nombre de données d'entraînement résulte en un seul type de modèle optimal, est vérifiée pour l'odeur de biogaz.

Dans ce cas, connaissant le niveau critique du nombre de données, le modèle optimal d'approximation de l'odeur de biogaz à utiliser minimisant le risque de généralisation

serait aussi connu et seul un ré-entraînement avec des données de la nouvelle source serait indispensable. Aucune nouvelle recherche du modèle optimal ne serait nécessaire.

UN MODÈLE OPTIMAL UNIQUE FONCTION DU NOMBRE DE DONNÉES D'ENTRAÎNEMENT

Deux ensembles de 48 données provenant de sources de biogaz distinctes (notées *Source A* et *Source B*) ont été utilisées séparément pour entraîner les trois types de machines neuronales jusqu'ici testées. Les concentrations des différents odorants ne sont pas identiques et influencent donc différemment l'odeur qui en résulte.

Les résultats de ces entraînements prouvent que le principe de minimisation du risque structurel est vérifié pour les odeurs de biogaz utilisées. Selon ce principe, pour un nombre fixe de données expérimentales, le même modèle de machine neuronal optimal s'applique à toutes les sources d'odeur de biogaz.

Pour les réseaux de neurones artificiels, la dimension VC du modèle peut être approximée par le nombre de paramètres engendrés par la structure choisie. Les deux figures qui suivent montrent que pour les deux groupes de données provenant de sources de biogaz distinctes, le nombre de paramètres idéal pour minimiser l'erreur de généralisation se trouve autour de 420 paramètres. Les deux erreurs minimales de généralisation obtenues sont aussi presque identiques. La forme est toujours caractéristique de celle attendue pour l'erreur de généralisation en fonction de la dimension VC, soit une cloche inversée.

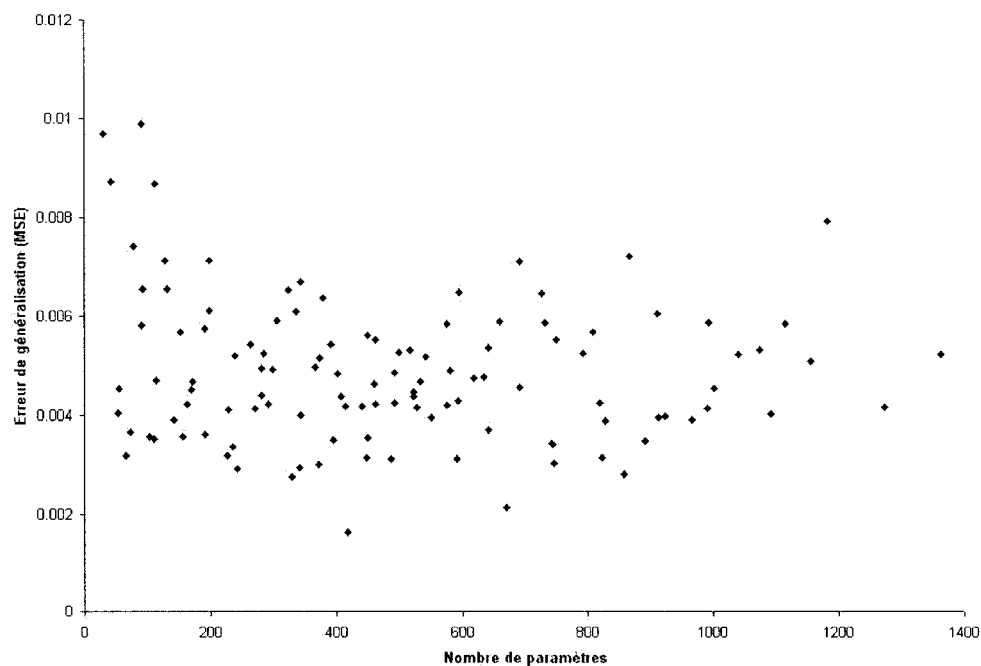


Figure 4.12 Erreur de généralisation en fonction du nombre de paramètres pour la source de biogaz A modélisée par réseaux de neurones artificiels



Figure 4.13 Erreur de généralisation en fonction du nombre de paramètres pour la source de biogaz B modélisée par réseaux de neurones artificiels

Dans le cas des réseaux de fonctions à bases radiales, la dimension VC du modèle peut être directement reliée à la quantité de neurones de la couche cachée formant les centres de la fonction. Les résultats sont encore en accord avec la théorie et la forme de l'erreur de généralisation est toujours caractéristique. Les deux réseaux de fonctions à bases radiales optimaux ont un nombre de neurones tournant autour de 6. Il est important de mentionner que l'entraînement des réseaux pour la source A a été faite sur toutes les structures possibles, alors que celui de la source B n'a été fait que pour certaines architectures, pour diminuer du temps de calcul.

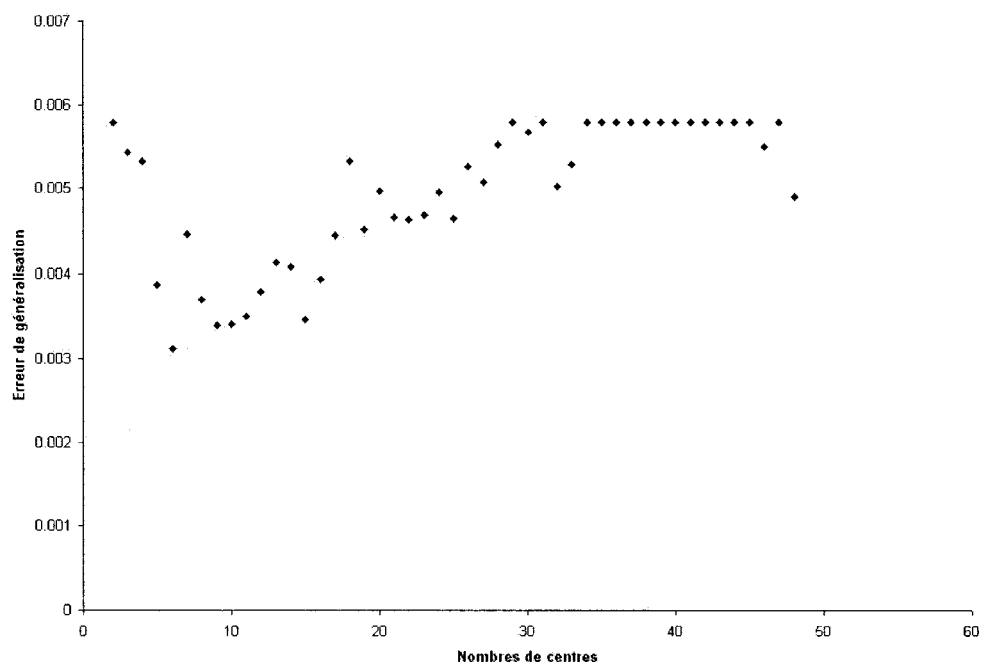


Figure 4.14 Erreur de généralisation en fonction du nombre de centres pour la source de biogaz A modélisée par réseaux de fonctions à bases radiales

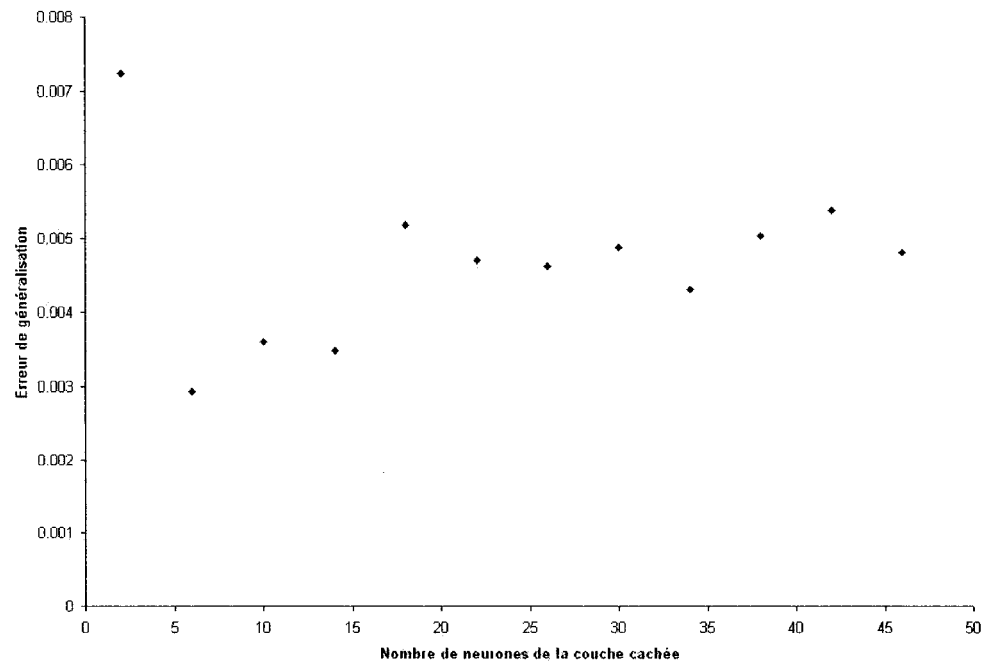


Figure 4.15 Erreur de généralisation en fonction du nombre de centres pour la source de biogaz B modélisée par réseaux de fonctions à bases radiales

Les erreurs minimales de généralisation des deux réseaux de fonctions à bases radiales optimaux obtenus pour les sources A et B sont très similaires.

Le cas des machines à supports vectoriels est quelque peu différent. Les paramètres de la machine avec un noyau polynomial sont identiques, soit un noyau du premier degré et une constante de coût C égale à 10. Les deux figures qui suivent montrent l'erreur de généralisation en fonction des paramètres.

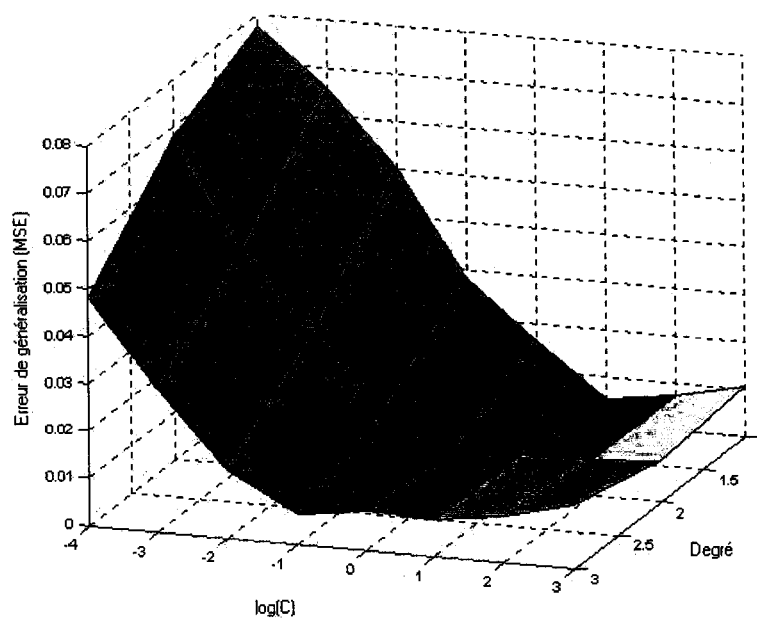


Figure 4.16 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau polynomial pour la source de biogaz A

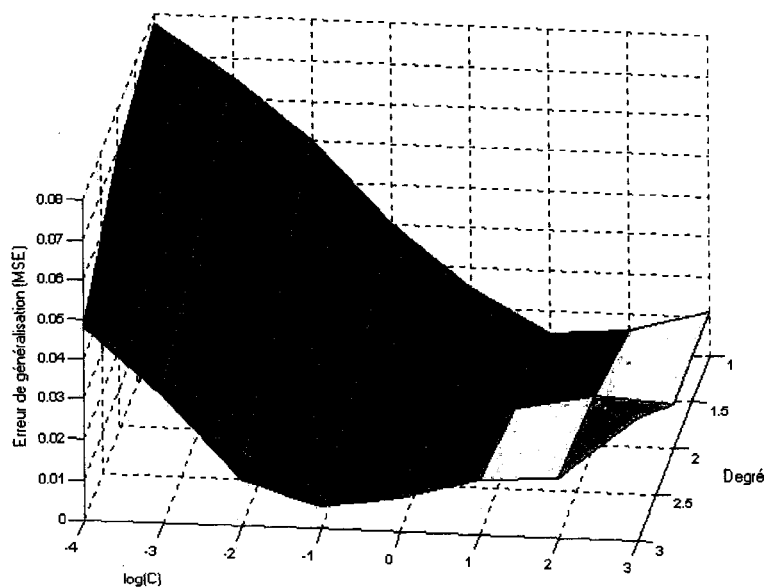


Figure 4.17 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau polynomial pour la source de biogaz B

Pour les machines à supports vectoriels à noyau de fonctions à bases radiales, la valeur de la constante de coût C est identique pour les deux régressions, soit 100. Par contre, la valeur de l'écart-type n'est pas la même pour les deux machines optimales. Ainsi, dans le cas de la source A de l'odeur de biogaz la valeur d l'écart-type est de 2,6. Pour la source B, l'écart-type est de 0,5.

Cette différence aurait sûrement pu être expliquée si plusieurs autres tests avaient été conduits sur des sources différentes. Par contre, il est impossible de dire avec seulement deux simulations si un des résultats est une exception ou si les noyaux de fonctions à bases radiales se comportent différemment. Les figures qui suivent présentent les résultats des simulations.

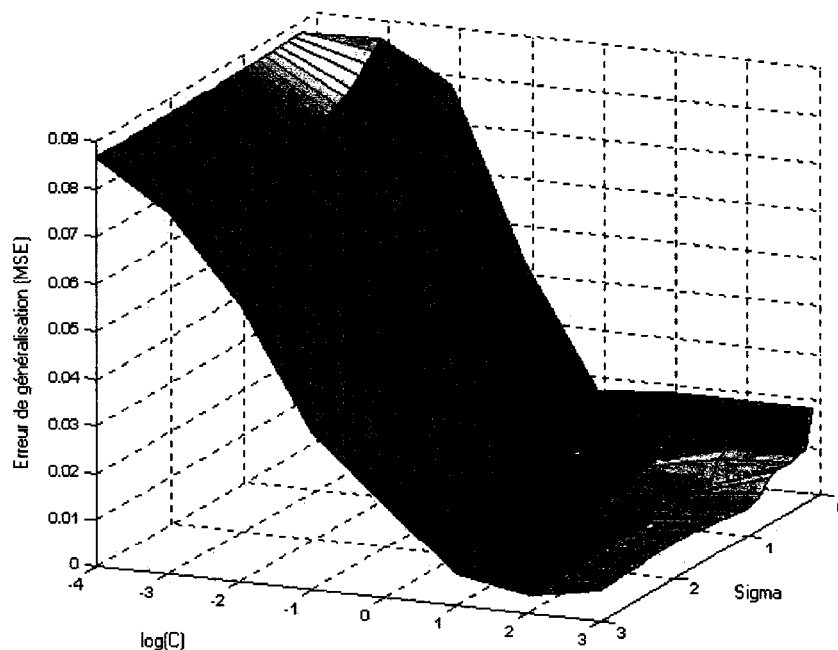


Figure 4.18 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau de fonction à bases radiales pour la source de biogaz A

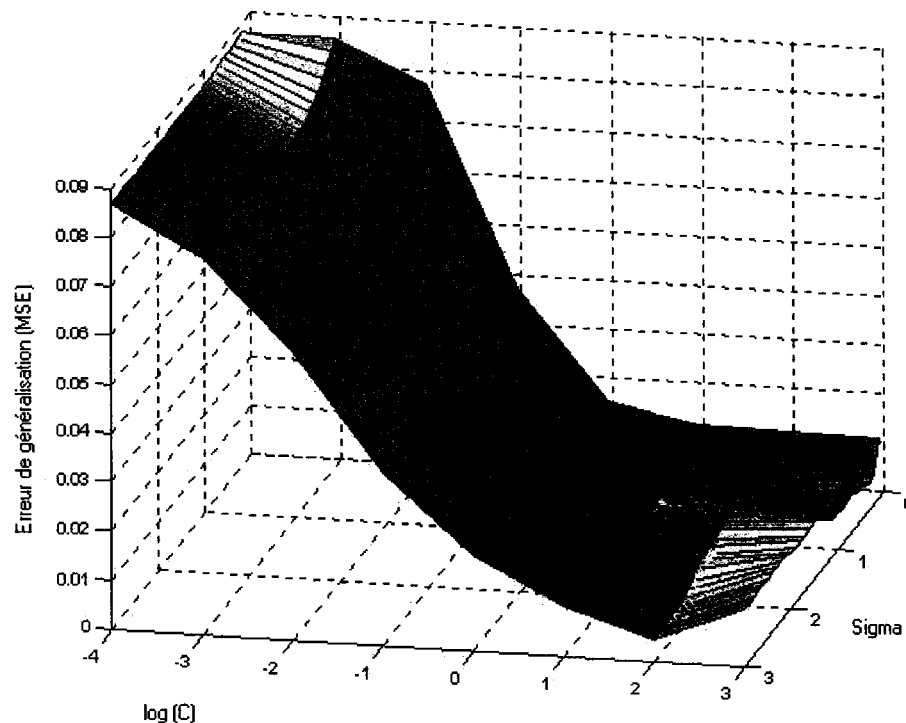


Figure 4.19 Erreur de généralisation en fonction des paramètres des machines à supports vectoriels à noyau de fonction à bases radiales pour la source de biogaz B

Il est intéressant de noter par contre que les surfaces des deux figures précédentes présentent des formes similaires. Trois minima sont visibles environ dans les mêmes régions. L'optimalité des machines n'était donc pas si éloignée dans la mesure où les deux groupes de données ont engendré des courbes comparables de l'erreur de généralisation, un peu comme dans le cas des deux types de réseaux.

Il peut être conclu que la preuve a été faite que l'approximation de la concentration-odeur de biogaz s'approxime selon le même modèle optimal de machine neuronal pour les deux sources à l'étude, sauf exception mineure. Par contre, des tests plus poussés

seraient nécessaires pour permettre la généralisation de ce principe à l'ensemble des sources de biogaz. Dans une certaine limite, ce principe implique que si des données expérimentales d'odeur de biogaz ont été récoltées à l'aide du même nez électronique, le modèle optimal de la machine neuronale à utiliser est connu.

COMPARAISON DES PERFORMANCES

Il est maintenant intéressant de comparer la performance des machines neuronales optimales approximant l'odeur de biogaz de l'ensemble des sources à celles en approximant une seule. Le tableau ci-dessous présente les résultats de l'erreur de généralisation minimale atteinte pour chacun des trois groupes de données. Il est important de rappeler que le nombre de données d'entraînement n'est pas identique pour chacun des groupes.

Tableau 4.3 Erreur de généralisation selon la source des données d'entraînement

	Erreur de généralisation		
	Réseau de neurones artificiel	Réseau de fonctions à bases radiales	Machine à support vectoriel
Ensemble des sources	0,0008963	0,0012308	0,0009849
Source A	0,0016265	0,0031164	0,0036352
Source B	0,0020184	0,002935	0,0038491

Les erreurs de généralisation se trouvent toutes dans le même ordre de grandeur. Il est intéressant de soulever le fait que l'impact du nombre de données d'entraînement, qui est près de deux fois plus grand pour les machines neuronales entraînées pour

l'ensemble des sources, est un facteur beaucoup plus important dans la minimisation du risque de généralisation que la provenance de l'odeur de biogaz. Ainsi, il est plus intéressant d'entraîner une seule machine neuronale pour les deux sources, que de créer des approximateurs pour chacune des sources, car l'erreur de généralisation qui en résulte est beaucoup plus petite.

CONCLUSION

La meilleure machine neuronale pour l'ensemble de la plage de concentration-odeur est le réseau de neurones artificiel à deux couches cachées, 4 neurones pour la première couche et 24 pour la seconde, et à sortie linéaire unique. Pour les concentrations-odeurs inférieures à 30 u.o./m^3 , la machine à support vectoriel a été de loin la plus performante. Les caractéristiques de celle-ci sont un noyau de fonctions à bases radiales avec un écart-type de 0.2 et une constante de coût de 1.

CONCLUSION

L'objectif général de la recherche consistait à déterminer la faisabilité de l'utilisation de trois types de machines neuronales, soit les réseaux de neurones artificiels, les réseaux de fonctions à bases radiales, ainsi que les machines à supports vectoriels, comme approximateurs de la concentration-odeur de l'odeur de biogaz habituellement présente en air ambiant. À cela s'ajoutait la détermination de la machine optimale par l'utilisation du principe de minimisation du risque structurel.

Il a été démontré que les machines neuronales employées dans le cadre de cette recherche peuvent approximer efficacement les concentrations de l'odeur de biogaz provenant de deux sources différentes sur une plage de 0 u.o./m³ à 200 u.o./m³, la plus performante dans ce cas étant le réseau de neurones artificiel à deux couches cachées, 4 neurones pour la première couche et 24 pour la seconde, et à sortie linéaire unique. Si la plage à approximer est bien définie par les données d'entraînement (i.e. un grand nombre de données bien distribuées est disponible sur l'intervalle de concentrations-odeurs à approximer), donc que l'interpolation entre les points d'entraînement est minime, les machines à supports vectoriels sont beaucoup plus performantes que les autres machines neuronales testées.

L'objectif de vérifier le respect de la théorie de l'apprentissage statistique a été atteint. La minimisation du risque structurel s'est avérée supérieure à la minimisation du risque empirique pour la détermination de la machine minimisant le risque de généralisation.

Il a été constaté qu'il n'existait pas de différence notable entre la performance des machines neuronales entraînées pour effectuer l'approximation de la concentration-odeur de deux sources distinctes d'odeur de biogaz et celles entraînées pour une seule. L'erreur de généralisation était comparable pour toutes les machines optimales entraînées avec les différents ensembles de données expérimentales.

Un résultat important a été de déterminer que, pour un type de machine neuronale et un nombre fixe de données d'entraînement, l'architecture, ou la dimension VC, optimale est la même pour deux sources d'odeur de biogaz distinctes modélisées séparément. Ceci permet de s'affranchir du processus de la recherche du modèle optimal si celui-ci a déjà été trouvé précédemment pour le nombre de données disponibles. La modélisation de l'odeur de biogaz est donc reproductible.

Malheureusement, le dernier sous-objectif consistant à déterminer un nombre de données d'entraînement permettant un compromis acceptable entre une bonne généralisation et une expérimentation minimale n'a été atteint qu'en partie. Le manque de données n'a pas permis d'atteindre un plateau sur l'erreur de généralisation en fonction du nombre de données d'entraînement. Par contre, il a été démontré qu'une

augmentation du nombre de données d'entraînement provoquait une baisse substantielle du risque de généralisation de la machine neuronale optimale.

Finalement, il est possible de conclure que les machines neuronales constituent des outils performants pour l'approximation des concentrations-odeurs de biogaz à l'aide des informations provenant des capteurs d'un nez électronique.

RECOMMANDATIONS

Les recommandations suivantes constituent des suggestions de recherches complémentaires à celle présentée dans le présent mémoire.

- Vérifier la validité des modèles optimaux de machines neuronales sur plusieurs autres sources de biogaz.
- Trouver les modèles optimaux pour plusieurs nombres de données d'entraînement et identifier les nombres minimaux pour chaque machines neuronales pour lequel l'erreur de généralisation se stabilise.
- Vérifier si les conclusions présentées peuvent être généralisées à d'autres types d'odeurs environnementales.
- Il serait intéressant de développer des méthodes de validation des réponses des capteurs. Une possibilité serait de prédire à partir des réponses des autres capteurs celle du capteur analysé et la comparer à la valeur réelle retournée par celui-ci.
- Une méthode qui tienne compte de l'incertitude sur les données d'entrée de la fonction d'approximation devrait être développée.
- La méthode du *Total Least-squares* pourrait être testée pour identifier les capteurs réellement utiles dans l'approximation de la concentration-odeur et minimiser le bruit sur la réponse de ceux-ci.

BIBLIOGRAPHIE

ANDRÉ, P., BOREUX, J.J., HANSON, A., NICOLAS, J., THOEN D. et VASSEL, J.L. (1995). Mesures olfactives dans Mesures et environnement – Chapitre 14 : Environnement, rédigé pour l'ouvrage collectif La mesure et l'instrumentation – État de l'art et perspectives, Masson.

BARLAZ, M.A., HAM, R.K. et SCHAEFER, D.M. (1989). Mass-balance analysis of anaerobically decomposed refuse, Journal of Environmental Engineering, 6, 1088-1102.

BARTLETT, P.N., ARCHER, P.B. et LING-CHUNG S.K. (1989). Conducting polymere gas sensor – 1. Fabrication and characterization, Sensor and Actuators, 19, 125-140.

BELLMAN, R. (1961). Adaptive Control Process: A Guided Tour, Princeton, New Jersey, Princeton University Press.

BOCKREIS, A. et JAGER, J. (1999). Odour Monitoring by the Combination of Sensors and Neural Networks, Environmental Modelling & Software, 14, 421-426.

BURGES, C.J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and KnowledgeDiscovery, 2, 121-167.

CHERKASSKY, V. et MA, Y. (2003), Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks, article en attente de publication.

DAVOLI, E., GANGAI, M.L., MORSELLI, L. et TONELLI, D. (2003). Characterisation of Odorants Emissions from Landfills by SPME and GC/MS, Chemosphere, 51, 357-368.

FRIEDMAN, J.H. (1995). An Overview of Prediction Learning and Function Approximation, dans V. Cherkassky, J. H. Friedman et H. Wechsler, eds, From Statistic to Neural Networks: Theory and Pattern recognition Applications, Springer-Verlag, New York.

GARRIGUES, S. (2002). Étude des odeurs de matériaux polymères de l'habitable automobile : Application des techniques de la métrologie sensorielle. Thèse de doctorat, Institut National Polytechnique de Toulouse, France.

GOSTELOW, P., PARSONS, S.A. et STUETZ, R.M. (2000). Odour measurement for sewage treatment works, Water Research, 35 (3), 579-597.

GUNN, S.R. (1998). Support Vector machines for Classification and Regression, Technical Report, University of Southampton, Faculty of Engineering and Applied Science.

HAUGEN, J.E. et KVAAL, K. (1998). Electronic nose and artificial neural network, Meat Science, 49, S273-S286.

HAYKIN, S. (1999). Neural networks – A Comprehensive Foundation, 2nd Edition, Prentice Hall, Toronto.

HOBSON, J.(1995). The odour potential : a new tool for odour management, Journal of Chartered Institution of Water And Environmental Management, 9, 458-463.

HUDON, G., GUY, C. et HERMIA, J. (2000) , Measurement of Odor Intensity by an Electronic Nose, Journal of the Air and Waste Management Association, 50 (10) , 1750-1759.

LIRON, Z., KAUSHANSKY, N., FRISHMAN, G., KAPLAN et GREENBLATT, J. (1997). The polymer-coated SAW sensor as a gravimetric sensor, Analytical Chemistry, 69, 2848-2854.

LUNDSTRÖM, I., SPETZ, A.,WINQUIST, F., ACKELID, U. et SUNDGREN, H. (1990). Catalytic metals and field-effect devices – a useful combination, Sensors and Actuators B, 1, 15-20.

MAEKAWA, T., CAI, K., SUZUKI, K., DOUGAMI, N., TAKAGA, T. et EGASHIRA, N. (2001) Compensatory methods for the odor concentration in a electronic nose system using software and hardware, Sensors and Actuators B, 76, 430-435.

MISSELBROOK, T.H., HOBBS, P.J., et PERSAUD, K.C. (1997). Use of an Electronic Nose to Measure Odour Concentration Following Application of Cattle Slurry to Grassland, Journal of Agricultural Engineering Research, 66, 213-220

NAKAMOTO, T., FUKUNISHI, N. et MORIZUMI, T. (1995). Identification capability of odor sensing using quartz-resonator array and neural network pattern recognition, Sensors and Actuators B, 18, 473-476.

NIYOGI, P. et GIROSI, F. (1996). On the Relationship Between Generalization Error, Hypothesis Complexity and Sample Complexity for Radial Basis Functions, Neural Computation, 8, 819-842.

PARK, J. et SANDBERG, I.W. (1991). Universal approximation using radial-basis function networks, Neural computation, 3, 246-257.

PERSAUD, K., DODD, G.H.(1982), Nature, 299, 352-355.

PERSAUD, K, PAYNE, P. A., KHAFFAF, S. M., DOWDESWELL, R. M., HOBBS, P. J., MISSELBROOK, T. H. et SNEATH, R. W. (1996), Application of conducting polymer odour sensing arrays to malodour monitoring, IEE Colloquium (Digest), 242, 6/1-6/4.

ROMAIN, A.C., NICOLAS, J., WIERTZ, V., MATERNOVA, J. et ANDRÉ, P. (2000). Use of a simple tin oxide sensor array to identify five malodours collected in the field, Sensors and Actuators B, 62, 73-79.

SMOLA, A. et SCHOLKOPF, B. (1998). A Tutorial on Support Vector Regression, NeuroColt2 Technical Report Series, Royal Holloway College, University of London, Royaume-Uni.

SONTAG, E.D. (1992). Feedback Stabilization Using Two-hidden-layer Nets, IEEE Transactions on Neural Networks, 3, 981-990.

STUETZ, R.M., FENNER, R.A. et ENGIN, G. (1998). Assesment of odours from sewage treatment works by an electronic nose, H₂S analysis and olfactometry, Water Research, 33 (2), 453-461.

TGS Sensor, General Information for TGS Sensor, Taguchi Sensor Co., 1996, 2-11

VAPNIK, V.N. et CHERVONENKIS, A.J. (1971). On the Uniform Convergence of Relative Frequencies of Events and their Probabilities, Theory Prob. Apl., 16, 264-280.

VAPNIK, V.N. (1982). Estimation of Dependencies based on Empirical Data, Springer, New York.

VAPNIK, V.N. (1995). The Nature Of Statistical Learning Theory, Springer, New York.

WALSH, M.E. (1995). Analytical method for white phosphorus in water, Bulletin of Environmental Contamination and Toxicology, 54 (3), 432.

ZARCOMB, S. et STETTER, J.R. (1984). Monitoring of Electrochemically Inactive Compounds by Amperometric Gas Sensors, Sensors and Actuators, 4, 225-243.

Odour concentration measurement by dynamic olfactometry, CENTC264/WG2, Comité Européen de Normalisation.

ANNEXE 1 – CALCUL DE LA CONCENTRATION-ODEUR

Ces explications proviennent d'un document préparé par Odotech inc., entreprise ayant fournie la technologie et les données bruts provenant du nez électronique.

1. Seuil de perception olfactif PROBIT

Cette technique d'analyse est basée sur une méthode statistique dose-réponse couramment utilisée en toxicologie. Elle consiste à déterminer, en utilisant une fonction de densité de probabilité suivant une distribution normale, la dilution à laquelle 50 % du jury perçoit l'odeur (EZ_{50}). Cette technique permet de minimiser l'effet des valeurs extrêmes et élimine l'erreur inhérente à l'estimation du seuil de perception individuel (EZ_{50p}) utilisée dans d'autres techniques.

Le modèle postulé suppose que la relation liant le $\log(\text{concentration})$ avec les réponses peut être modélisée par une droite, au travers la fonction de lien **probit**. Dans le cas des courbes concentration-réponse (ou dilution-réponse, ou encore dose-réponse) le choix du lien est lié à la forme de la distribution des seuils individuels (appelée distribution des seuils, ou distribution des tolérances, selon le contexte). La courbe dilution-réponse estimée pour une population (un jury) est le résultat des réponses individuelles. Chaque individu a son propre seuil de perception (de tolérance) au produit. On sait que dans le cadre d'expériences

olfactométriques, la distribution des seuils de tolérances est normale. Le lien **probit** est associé à une distribution normale.

$$probit(p) = \phi^{-1}(p) \text{ où } \phi \text{ est la fonction de répartition de la loi } N(0,1).$$

Les observations brutes sont binaires. En effet, les deux alternatives sont les suivantes : un individu détecte correctement l'odeur ou non. Ce qui amène à une loi de Bernouilli : on désigne une variable aléatoire : «détection de l'odeur»,

$$X \approx \mathcal{B}(1, p)$$

$$X = \begin{cases} 1 & \text{si bonne détection} \\ 0 & \text{sinon} \end{cases}$$

On considère $r \in \{0,1\}$, alors

$$P(X = r) = p^r (1 - p)^{1-r} \quad E(X) = p \text{ et } \text{VAR}(X) = p(1 - p)$$

avec p , la probabilité de détection.

Les observations binaires sont supposées se réaliser dans des conditions identiques et toutes indépendantes les unes des autres. En effet, le fait qu'un individu sente ou non n'a pas d'influence sur le juré voisin. Il n'y a pas de

phénomène de concurrence. Elles sont supposées toutes suivre la même loi de Bernoulli (même probabilité de détection dans les mêmes conditions de dilution). Comme on suppose également qu'aucun des juré n'est distinguable vis à vis d'un autre, on regroupe les observations binaires. Chaque groupe représente l'ensemble des réponses données par un jury à une dilution. Quand le nombre d'occurrences y d'un phénomène observé sur n individus est la somme de variables binaires indépendantes et de même probabilité p , ce nombre suit une loi binomiale $\mathcal{B}(n, p)$.

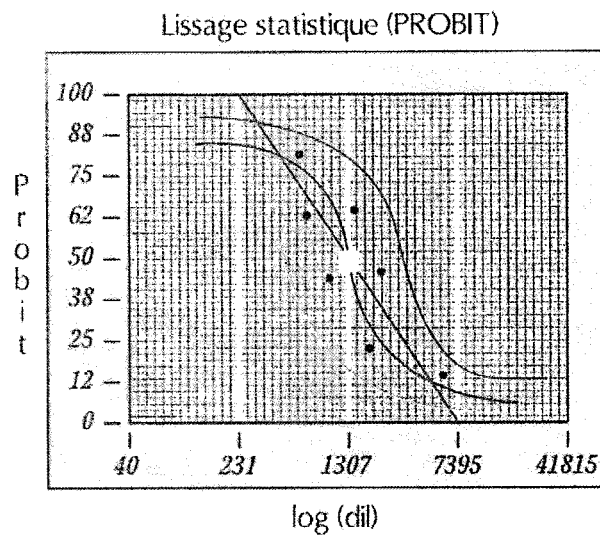
On peut formaliser de la façon suivante :

$X_i, i = 1, 2, \dots, n$ binaires, indépendantes, de même loi $\mathcal{B}(1, p)$, alors

$Y = \sum_{i=1}^n X_i$ prend la valeur $y \in [0, n]$ selon $P(Y = y) = C_n^y p^y (1 - p)^{n-y}$

$E(Y) = np$ et $VAR(Y) = np(1 - p)$

Dans le cas des proportions, on a $E\left(\frac{Y}{n}\right) = p$ et $VAR\left(\frac{Y}{n}\right) = \frac{p(1 - p)}{n}$



Donc le lien Probit est la loi normale des seuils de tolérance.

Finalement pour un individu, on a un seuil de détection, $x_i = \log$ de la dilution

U : variable aléatoire «Seuil de détection», $p_i = P(U \leq x_i) = \int_{-\infty}^{x_i} f(u) du$

$$U \approx N(\mu, \sigma^2), f(u) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_i} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du = \Phi\left(\frac{x_i - \mu}{\sigma}\right)$$

Φ : fonction de répartition de la loi normale centrée réduite

$$\text{On pose } \alpha = -\frac{\mu}{\sigma}, \beta = \frac{1}{\sigma} \quad p_i = \Phi(\alpha + \beta x_i) \quad \Phi^{-1}(p_i) = \alpha + \beta x_i$$

Ensuite on réalise une régression **probit** qui consiste à déterminer les paramètres

les plus probables de la loi normale des seuils de tolérance mesurés.

Une fois les paramètres de la loi normale déterminés, le seuil de perception du jury (EZ_{50}) en est extrait c'est à dire la dilution qui prévoit que 50% du jury sent l'odeur.

2. Seuil de perception olfactif ASTM E679-91

Cette technique utilise le protocole défini par la norme ASTM-E679-91¹. Elle est basée sur l'estimation des seuils de perception individuels de chaque juré et calcule la moyenne logarithmique des seuils personnels estimée (EZ_{50p}) pour estimer la valeur du seuil de perception du jury (EZ_{50}).

Le lecteur qui désire obtenir tous les détails de cette technique d'évaluation du seuil de perception olfactif est invité à lire le document ASTM E679-91.

3. Seuil de perception olfactif CEN

Cette technique utilise le protocole défini par la norme CEN prEN13725². Elle est basée sur la méthode ASTM E679-91 à l'exception qu'un post-traitement est effectué sur les valeurs des seuils individuels afin d'éviter de considérer dans le

¹ ASTM E679-91: Standard Practices for Determination of Odor Taste Thresholds by a Forced-choice Ascending Concentration Series Method of Limits.

² CEN prEN13725: Air quality – Determination of Odour Concentration by Dynamic Olfactometry, Comité Européen de Normalisation.

calcul du seuil moyen les jurés ayant eu des seuils individuels non représentatifs du groupe.

Le lecteur qui désire obtenir tous les détails de cette technique d'évaluation du seuil de perception olfactif est invité à lire le document CEN prEN13725.

4. Seuil de perception olfactif CUM

Cette technique utilise le protocole défini par la CUM. Elle est basée sur la méthode ASTM E679-91 pour évaluer les seuils de perception individuels. Le seuil moyen du jury est fixé à la valeur où 50 % des jurés ont atteint leur seuil de perception individuel.

Le lecteur qui désire obtenir tous les détails de cette technique d'évaluation du seuil de perception olfactif est invité à lire le document CUM Méthode de référence: Mesure du Nombre d'unité d'odeur³.

5. Régression log

³ CUM, Méthode de référence: Mesure du Nombre d'unités d'odeur (Olfactomètre dynamique) Révision 1994, Service de l'environnement, Direction de l'assainissement de l'air et de l'eau, Communauté Urbaine de Montréal.

Cette technique est basée sur une approche de type dose-réponse qui consiste à déterminer par la régression logarithmique du pourcentages de réponses positives en fonction du nombre de dilutions, la dilution à laquelle 50 % du jury perçoit l'odeur (EZ_{50}).

ANNEXE 2 – LISTE DES FICHIERS ELECTRONIQUES FOURNIS SUR LA DISQUETTE

La liste suivante présente les différents fichiers électroniques disponibles sur la disquette.

- Fichiers MatLab : Dossier contenant l'ensemble des fichiers non propriétaires implantés du plate-forme MatLab
- Résultats.zip : Ensembles des résultats bruts obtenus du nez électronique
- Résumé modifié.zip : Résultats présentés sur tableur et graphique EXCEL
- Analyses.zip : Analyses chimiques des champs de biogaz