

UNIVERSITÉ DU QUÉBEC

**MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN INGÉNIERIE**

par

ERIC LAROUCHE

**EXPLORATION DE DIFFÉRENTES ARCHITECTURES DE
RÉSEAUX DE NEURONES POUR LA PRÉDICTION DE
LA GLACE ATMOSPHÉRIQUE SUR LES CONDUCTEURS
DES RÉSEAUX ÉLECTRIQUES**

Hiver 2002



Mise en garde/Advice

Afin de rendre accessible au plus grand nombre le résultat des travaux de recherche menés par ses étudiants gradués et dans l'esprit des règles qui régissent le dépôt et la diffusion des mémoires et thèses produits dans cette Institution, **l'Université du Québec à Chicoutimi (UQAC)** est fière de rendre accessible une version complète et gratuite de cette œuvre.

Motivated by a desire to make the results of its graduate students' research accessible to all, and in accordance with the rules governing the acceptance and diffusion of dissertations and theses in this Institution, the **Université du Québec à Chicoutimi (UQAC)** is proud to make a complete version of this work available at no cost to the reader.

L'auteur conserve néanmoins la propriété du droit d'auteur qui protège ce mémoire ou cette thèse. Ni le mémoire ou la thèse ni des extraits substantiels de ceux-ci ne peuvent être imprimés ou autrement reproduits sans son autorisation.

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

RÉSUMÉ

Dans le but d'établir un modèle informatique empirique visant à prédire l'évolution temporelle des charges mécaniques dues à l'accumulation de la glace atmosphérique sur les conducteurs de transport d'énergie électrique, cinq architectures de réseaux de neurones artificiels ont été étudiées et comparées. Deux réseaux de nature statique, soit le Perceptron multicouches (MLP) et le réseau à fonctions de base radiales (RBF), ainsi que deux réseaux orientés vers le traitement temporel, soit le réseau à réponse impulsionnelle finie (FIR) et le réseau récurrent Elman, ont été comparés à la régression linéaire multiple (ADALINE).

Les données utilisées pour faire l'entraînement des modèles basés sur les réseaux de neurones proviennent du site instrumenté du Mont-Bélair qui fait partie du système de surveillance en temps réel SYGIVRE d'Hydro-Québec. On retrouve sur ce site une ligne de 315 kV dotée d'un capteur de force pour la mesure des forces mécaniques, plusieurs instruments météorologiques standards ainsi qu'un givromètre permettant d'évaluer l'intensité des conditions givrantes environnantes. Les modèles créés dans le cadre de cette recherche utilisèrent la température ambiante, la vitesse

normale du vent et le signal du givromètre comme paramètres d'entrée et le signal de charge (ou son taux de variation) comme paramètre de sortie.

Les résultats indiquent que les réseaux de neurones (le FIR par exemple) constituent un outil prédictif potentiellement avantageux ayant une puissance de représentation supérieure à des techniques statistiques simples comme la régression linéaire multiple. Les réseaux de neurones de nature statique donnent généralement de moins bons résultats que les réseaux orientés vers le traitement temporel et semblent donc moins adéquats. Toutefois, les performances affichées par les modèles montrent bien que la problématique de la prédiction des charges mécaniques de glace atmosphérique est plutôt difficile et qu'il ne faut pas considérer les réseaux de neurones comme une solution magique. L'utilisation de l'historique passé du signal de sortie comme variable d'entrée supplémentaire améliore notablement la situation cependant dans le cadre d'une prédiction en avance. Dans ce cas, les modèles fournissent une prédiction raisonnablement précise.

REMERCIEMENTS

Cette étude a été effectuée dans le cadre de la Chaire industrielle CRSNG/Hydro-Québec/UQAC sur le givrage atmosphérique des équipements des réseaux électriques (CIGELE) en collaboration avec le professeur Jean Rouat du groupe ERMETIS de l'UQAC. Je tiens à remercier le titulaire de la CIGELE, le professeur Masoud Farzaneh, pour sa contribution à la définition de mon projet, ses conseils judicieux et le support financier qu'il m'a accordé durant mes études.

Mes remerciements s'adressent tout particulièrement à mon directeur et mon co-directeur de recherche, les professeurs Gilles Bouchard de la CIGELE et Jean Rouat, pour leur encadrement et le support scientifique dont ils m'ont fait part tout au long de ma recherche. Ce travail n'aurait pu se faire sans leur soutien constant.

Je remercie tout le personnel de la CIGELE qui m'a soutenu et a contribué à ce que mon travail se passe dans de très bonnes conditions. En particulier, j'aimerais remercier le Dr Konstantin Savadjiev pour ses conseils et pour avoir mis à ma disposition des données provenant de sites naturels de givrage d'Hydro-Québec.

Je remercie également ma famille et mes amis pour tous les encouragements et le support qu'ils m'ont donnés durant mes études.

TABLE DES MATIÈRES

Résumé	iv
Remerciements	vi
Liste des figures	x
Liste des tableaux	xiv
Glossaire	xv
CHAPITRE 1	
INTRODUCTION	1
1.1 PROBLÉMATIQUE	2
1.2 OBJECTIFS	8
1.3 MÉTHODOLOGIE	9
1.4 SURVOL DU MÉMOIRE	10
CHAPITRE 2	
REVUE DE LA LITTÉRATURE	11
2.1 INTRODUCTION	12
2.2 DÉTAILS DES APPROCHES	13
2.2.1 APPROCHE PHYSIQUE	13
2.2.2 APPROCHE STOCHASTIQUE	25

2.2.3 APPROCHE EMPIRIQUE PAR RÉSEAUX DE NEURONES	28
2.3 INSERTION DU PRÉSENT TRAVAIL	33
CHAPITRE 3	
LES RÉSEAUX DE NEURONES ARTIFICIELS	36
3.1 INTRODUCTION	37
3.2 MODÈLE D'UN NEURONE ARTIFICIEL	41
3.3 TYPES DE TOPOLOGIE	47
3.4 ALGORITHME STANDARD D'APPRENTISSAGE	51
3.5 STRUCTURE PRÉDICTIVE TEMPORELLE	59
CHAPITRE 4	
DONNÉES D'ENTRAÎNEMENT ET ARCHITECTURES	61
4.1 SOURCE DE DONNÉES	62
4.2 PRÉ-TRAITEMENT	67
4.3 ANALYSE STATISTIQUE	70
4.4 ARCHITECTURES EXPÉRIMENTÉES	81
4.4.1 ARCHITECTURES STATIQUES	82

4.4.1.1 PERCEPTRON MULTICOUCHES (MLP)	82
4.4.1.2 RÉSEAU À FONCTIONS DE BASES RADIALES (RBF)	83
4.4.2 ARCHITECTURES DÉPENDANTES DU TEMPS	85
4.4.2.1 RÉSEAU À RÉPONSE IMPULSIONNELLE FINIE (FIR)	85
4.4.2.2 RÉSEAU RÉCURRENT D'ELMAN	87
CHAPITRE 5	
RÉSULTATS EXPÉRIMENTAUX	89
5.1 INTRODUCTION	90
5.2 CRITÈRE D'ERREUR	91
5.3 VARIATION DES PARAMÈTRES	92
5.4 COURBES DE PRÉDICTION	105
5.5 COURBES DE PRÉDICTION (MODE DE PRÉDICTION FUTURE)	109
CHAPITRE 6	
CONCLUSION	115
BIBLIOGRAPHIE	119

LISTE DES FIGURES

Figure 3.1	<i>Modèle d'un neurone élémentaire</i>	42
Figure 3.2	<i>Fonction sigmoïdale</i>	44
Figure 3.3	<i>Fonction tangente hyperbolique</i>	44
Figure 3.4	<i>Fonction de transfert linéaire</i>	45
Figure 3.5	<i>Réseau à propagation avant</i>	48
Figure 3.6	<i>Réseau à propagation avant multicouches</i>	49
Figure 3.7	<i>Réseau récurrent avec neurones cachés</i>	50
Figure 3.8	<i>Apprentissage supervisé</i>	51
Figure 3.9	<i>Processus de rétropropagation</i>	58
Figure 3.10	<i>Détails du processus de rétropropagation</i>	58
Figure 3.11	<i>Structure temporelle prédictive</i>	60
Figure 4.1	<i>Schéma du site de mesure du Mont-Bélair</i>	63
Figure 4.2	<i>Schéma global du modèle</i>	66
Figure 4.3	<i>Graphique de la température en fonction du temps</i>	71
Figure 4.4	<i>Graphique de la vitesse normale du vent en fonction du temps</i>	71
Figure 4.5	<i>Graphique du signal de givromètre en fonction du temps</i>	72

Figure 4.6	<i>Graphique du signal du capteur de force en fonction du temps</i>	72
Figure 4.7	<i>Graphique du taux d'accrétion en fonction du temps</i>	73
Figure 4.8	<i>Nuage de points du signal de charge avec le signal de givromètre</i>	74
Figure 4.9	<i>Nuage de points du taux d'accrétion avec le signal de givromètre</i>	74
Figure 4.10	<i>Nuage de points du signal de charge avec la température</i>	75
Figure 4.11	<i>Nuage de points du taux d'accrétion avec la température</i>	76
Figure 4.12	<i>Nuage de points du signal de charge avec la vitesse normale du vent</i>	77
Figure 4.13	<i>Nuage de points du taux d'accrétion avec la vitesse normale du vent</i>	77
Figure 4.14	<i>Architecture du réseau FIR</i>	85
Figure 5.1	<i>Effet du nombre de neurones cachés pour le Perceptron Multicouches avec le signal du capteur de force</i>	92
Figure 5.2	<i>Effet du nombre de neurones cachés pour le réseau RBF avec le signal du capteur de force</i>	93
Figure 5.3	<i>Effet du nombre de neurones cachés pour le réseau FIR avec le signal du capteur de force</i>	94
Figure 5.4	<i>Effet du nombre de neurones cachés pour le réseau Elman avec le signal du capteur de force</i>	95

Figure 5.5	<i>Effet du nombre de neurones cachés pour le Perceptron Multicouches avec le taux d'accrétion</i>	96
Figure 5.6	<i>Effet du nombre de neurones cachés pour le réseau RBF avec le taux d'accrétion</i>	97
Figure 5.7	<i>Effet du nombre de neurones cachés pour le réseau FIR avec le taux d'accrétion</i>	98
Figure 5.8	<i>Effet du nombre de neurones cachés pour le réseau Elman avec le taux d'accrétion</i>	99
Figure 5.9	<i>Résumé des résultats de prédiction avec le signal du capteur de force</i>	101
Figure 5.10	<i>Résumé des résultats de prédiction avec le taux d'accrétion</i>	102
Figure 5.11	<i>Courbes de prédiction pour l'ADALINE avec le signal du capteur de force en phase d'accrétion</i>	105
Figure 5.12	<i>Courbes de prédiction pour le Perceptron Multicouches avec le signal du capteur de force en phase d'accrétion</i>	106
Figure 5.13	<i>Courbes de prédiction pour le réseau FIR avec le signal du capteur de force en phase d'accrétion</i>	107
Figure 5.14	<i>Prédiction d'un pas de temps d'avance pour le réseau FIR avec le capteur de force</i>	110
Figure 5.15	<i>Prédiction d'un pas de temps d'avance pour le réseau FIR avec le taux d'accrétion</i>	111

- Figure 5.16 *Prédiction de deux pas de temps d'avance pour le réseau FIR
avec le signal du capteur de force mais sans le givromètre* 112
- Figure 5.17 *Prédiction de deux pas de temps d'avance pour le réseau FIR
avec le givromètre mais sans le signal du capteur de force* 113
- Figure 5.18 *Prédiction de trois pas de temps d'avance pour le réseau FIR
avec le capteur de force* 114

LISTE DES TABLEAUX

Tableau 2.1	<i>Comparaison des modèles</i>	24
Tableau 4.1	<i>Matrice de corrélation normalisée des paramètres pour l'ensemble avec la phase d'accrétion et le signal du capteur de force</i>	78
Tableau 4.2	<i>Matrice de corrélation normalisée des paramètres pour l'ensemble avec l'événement complet et le signal du capteur de force</i>	79
Tableau 4.3	<i>Matrice de corrélation normalisée des paramètres pour l'ensemble avec la phase d'accrétion et le taux d'accrétion</i>	80
Tableau 4.4	<i>Matrice de corrélation normalisée des paramètres pour l'ensemble avec l'événement complet et le taux d'accrétion</i>	80

GLOSSAIRE

Neurone formel: élément calculatoire simple qui effectue un traitement élémentaire (sommation pondérée suivie d'une transformation non-linéaire), unité de base des réseaux neuronaux.

Réseau de neurones: ensemble de plusieurs neurones interconnectés sous différentes configurations.

Réseau à propagation avant: réseau de neurones dans lequel le signal se propage dans une seule direction spatiale sans rétroaction.

Réseau récurrent: réseau de neurones dans lequel une partie de la sortie peut revenir à l'entrée; caractérisé par la présence de rétroaction.

Apprentissage supervisé: processus par lequel un réseau de neurones est adapté pour émuler une certaine relation entrée-sortie sur la base d'exemples fournis.

Rétropropagation: algorithme d'apprentissage standard des réseaux de neurones à propagation avant basé sur la propagation de signaux d'erreur.

Poids synaptique: coefficient associé à une entrée afférente d'un neurone; paramètre librement ajustable.

Givromètre: instrument composé d'une sonde vibrante qui estime l'intensité de présence de conditions givrantes.

Capteur de force: instrument qui mesure les forces mécaniques dans les lignes de transport d'énergie électrique.

Série temporelle: séquence temporellement ordonnée d'une grandeur physique.

Identification de système: processus par lequel la fonction de transfert entrée-sortie d'un système complexe est simulée ou estimée.

CHAPITRE 1

INTRODUCTION

1.1 PROBLÉMATIQUE

L'accumulation de glace atmosphérique sur les structures civiles créées par l'homme est un phénomène météorologique qui survient fréquemment dans les régions nordiques telles que le Canada, la Norvège, le Royaume-Uni, etc. Il en résulte souvent des conséquences technologiques et économiques graves dans différents domaines d'activité pour les collectivités humaines touchées par ce phénomène. L'effet perturbateur de la glace atmosphérique se manifeste de diverses manières et à différents degrés. Son impact peut causer dans le meilleur des cas un ralentissement ou un arrêt temporaire des réseaux de transport, de communication ou d'énergie mais peut aussi provoquer une paralysie majeure prolongée des services lors de bris matériels importants. Les dommages observés au cours des années, qu'on pense à la tempête de janvier 1998 au Québec, montrent bien l'importance que peut prendre ce phénomène dans les environnements climatiques froids: réseaux de transport et de distribution de l'énergie électrique, réseaux de télécommunication, les routes, les aéroports, les avions, les chemins de fer, les tours d'émission de micro-onde, les instruments de météorologie, les systèmes mécaniques automatisés, etc.

Cependant, la présente étude porte uniquement sur les réseaux de transport de l'énergie électrique. Dans la province de Québec, les lignes de transport doivent traverser de vastes étendues (plus de 1000 km) où elles sont exposées à différentes

conditions de glace atmosphérique avant de se rendre au bassin de population, qui se trouve plus au sud du territoire, où se consomme l'électricité. Cette contrainte rend le réseau d'énergie électrique très susceptible aux dommages causés par l'accumulation de glace atmosphérique. L'impact de celle-ci peut se distinguer selon deux grands aspects physiques principaux: les effets mécaniques et les effets électriques.

La présence de givre et de verglas sur les câbles et les structures de lignes aériennes de transport d'électricité affecte leur comportement mécanique à plusieurs égards [20,24]. Tout d'abord, la glace atmosphérique constitue une surcharge directe qui provoque un changement d'état de nature statique dans la ligne. La traction dans les câbles augmente en conséquence, de même que les efforts sur les structures et les accessoires d'attache. Selon les longueurs des portées réelles, les charges longitudinales, transversales ou en torsion peuvent être déséquilibrées. L'effet du vent sur les structures gonflées par la glace provoque des charges horizontales additionnelles. La présence de glace atmosphérique affecte également grandement les propriétés dynamiques des lignes (fréquences naturelles de vibration, amortissement interne). Le mouvement de galop des câbles glacés est un phénomène d'instabilité aérodynamique complexe qui peut causer des dommages importants aux câbles et aux accessoires de ligne. La présence de glace atmosphérique sur les câbles peut éventuellement causer des efforts transitoires lors de délestages soudains.

La glace atmosphérique peut aussi causer des problèmes électriques majeurs sur les équipements de transport et de distribution d'énergie. L'accumulation de glace et/ou de neige sur les isolateurs externes (souvent combinée avec la pollution de l'air ambiant) peut modifier considérablement leur caractéristique d'isolation diminuant ainsi leur performance électrique - tel que cela a été rapporté dans plusieurs pays [4,6,12,13,25,26]. Si les conditions s'y prêtent bien et que la tension de tenue baisse suffisamment, il peut y avoir contournement électrique, c'est-à-dire un passage brusque du courant électrique à travers l'isolateur [15]. Il s'ensuit inévitablement une panne électrique. En effet, la couverture de glace perturbe considérablement la performance des isolateurs haute tension causant une diminution de la distance d'isolement.

Diverses stratégies peuvent être utilisées pour contrer les multiples effets nuisibles de la glace atmosphérique. Il est à noter que ces différentes stratégies ne doivent pas être considérées comme étant mutuellement exclusives mais peuvent se compléter de manière synergique. Un problème aussi complexe que le givrage atmosphérique des structures doit être attaqué sur plusieurs fronts à la fois. À un premier niveau, il y a la recherche fondamentale, soit l'exploration des principes théoriques et l'expérimentation en situation contrôlée de laboratoire ou sur des sites naturels instrumentés, qui vise à mieux comprendre les phénomènes physiques impliqués dans le processus de formation du givre et du verglas et des mécanismes par lesquels ils causent des dommages matériels aux équipements (bris mécaniques,

mouvement de galop des conducteurs, vibrations éoliennes sévères, développement d'arcs électriques et contournement des isolateurs, etc.) L'application de ces connaissances scientifiques à des situations concrètes pourra éventuellement aider à prévenir les bris matériels de manière plus efficace ou du moins à diminuer la fréquence et/ou la gravité de ceux-ci. Elle permettra de mieux gérer les risques encourus. Les études de type probabiliste ou statistique peuvent apporter à ce sujet un appui important pour la gestion de risque et la prédiction [8,9,14].

À un deuxième niveau, on retrouve les efforts de création de modèles mathématiques et/ou informatiques visant à simuler et à décrire les processus physiques impliqués pendant l'accumulation de la glace. Ils pourront éventuellement prédire avec suffisamment de précision les charges de glace atmosphérique sur les conducteurs électriques ou les structures de support. Le succès de ces modèles dépend fortement des informations accessibles aux chercheurs; ces modèles sont en grande partie tributaires de l'avancée des connaissances scientifiques dans le domaine. L'approche par modélisation vise à donner la possibilité d'une action interventionniste appropriée lors de grandes tempêtes de givre ou de verglas. Un modèle efficace permettrait de prédire la quantité de glace qui va se déposer sur la structure, et par conséquent, quand des conditions dangereuses de glace atmosphérique vont apparaître et menacer l'intégrité des structures.

Toutefois, le développement de modèles de glace atmosphérique est en partie menacé par le fait qu'il y a peu de données de terrain avec lesquelles ils peuvent être validés [29]. Le développement de sites de mesure a été beaucoup plus lent que la progression fantastique de la puissance de calcul des ordinateurs et de la sophistication des modèles de glace. Une autre difficulté vient du fait que la formation de la glace atmosphérique peut se produire selon différents processus en séquence ou simultanément. Au Canada, on a vérifié principalement les modèles de pluie verglaçante et pour seulement une courte durée d'accumulation [27]. C'est avec le givre de nuage que les modèles ont obtenu les meilleurs résultats. La croissance en régime sec du givre de nuage fut simulée avec succès en soufflerie. De plus, les modèles de glace atmosphérique ne simulent généralement que la première phase, celle de l'accumulation. Il y en a très peu qui décrivent les deux autres phases, soit la persistance et le délestage.

D'une façon plus globale, la recherche a surtout porté sur l'utilisation des ressources de la science de la météorologie pour la prédiction des événements des tempêtes de givres et de verglas. Certains chercheurs se sont attardés à l'aspect de la problématique concernant l'évolution temporelle des tempêtes ainsi que leur distribution spatio-géographique [8,14,28,36]. Ces études ont permis d'acquérir une compréhension plus globale du phénomène pour un territoire donné (ex: province de Québec) et d'établir des corrélations statistiques entre les événements qui surviennent en différents points géographiques.

Toutefois, bien que la météorologie ait du succès pour faire de la prédiction à grande échelle à moyen terme, elle ne permet pas d'établir des prédictions très localisées de phénomènes se déroulant sur une courte période de temps. Présentement, la prédiction météorologique à grande échelle n'est pas suffisamment précise pour pouvoir prédire à l'avance où et quand des conditions dangereuses de glace atmosphérique vont menacer l'intégrité des structures. De plus, les paramètres mesurés par les stations météorologiques ne sont pas tous directement applicables à la problématique du givrage et du verglas [29]. À cause de ces lacunes, plusieurs chercheurs ont développé des instruments mieux adaptés à la mesure des conditions givrantes. Ainsi, Hydro-Québec a en opération depuis 1973 un réseau de glacimètres passifs qui permet de recueillir de l'information quantitative. Plus tard, Hydro-Québec s'est doté d'un système de surveillance en temps réel de la glace atmosphérique qui couvre une partie de la vallée du St-Laurent (le système SYGIVRE). Des chercheurs ont développé un instrument permettant de mesurer directement l'intensité des conditions givrantes: le givromètre [29,39].

Il est évident qu'un système de surveillance en temps réel peut jouer un rôle important dans la problématique de la glace atmosphérique. Il est maintenant de plus en plus reconnu que les conséquences des désastres climatiques peuvent être drastiquement réduites par l'utilisation d'un tel système de surveillance en temps réel. Avec le développement des technologies modernes de communication et de traitement automatisé de l'information et la disponibilité d'une instrumentation

appropriée, il demeure donc important de développer des modèles fiables d'accumulation de glace qui permettent une prédiction aussi réaliste que possible des charges mécaniques avant que des dommages matériels sérieux ne surviennent.

D'autres chercheurs ont fait avancer la recherche du côté des technologies actives de déglacement. Elles pourraient permettre éventuellement de traiter la problématique de la glace atmosphérique de manière pro-active [30].

1.2 OBJECTIFS

Le présent travail de recherche vise à modéliser l'accumulation, la persistance et le délestage de la glace atmosphérique sur des conducteurs. Ce travail est effectué dans le cadre de la chaire industrielle CRSNG/Hydro-Québec/UQAC sur le givrage atmosphérique des équipements des réseaux électriques (CIGELE).

Plus spécifiquement, cette recherche a comme objectif principal d'explorer l'applicabilité potentielle de différentes architectures de réseaux de neurones à la prédiction du givrage atmosphérique.

Des modèles informatiques utilisant la modélisation fonctionnelle (basée sur des données réelles) visant à prédire l'évolution temporelle de la charge de glace

atmosphérique sur une ligne de transport en fonction de différentes variables météorologiques et de la mesure de l'intensité de conditions givrantes (tel que fourni par un givromètre) seront créés. Ces modèles seront développés à l'aide d'événements réels de givre et de verglas survenus au site instrumenté du Mont-Bélair tels qu'enregistrés automatiquement par le système de surveillance en temps réel d'Hydro-Québec SYGIVRE.

1.3 MÉTHODOLOGIE

Nous utiliserons comme données source des informations fournies par Hydro-Québec et provenant de la base de données centrale du système SYGIVRE. Seule la ligne de 315 kV sur le site du Mont-Bélair sera considérée puisque ce sont les seules données disponibles.

Les modèles informatiques créés dans le cadre de cette recherche utiliseront une technique d'intelligence artificielle pour traiter l'information connue sous l'appellation de "réseaux de neurones formels". Elle fait partie d'une branche de l'informatique appliquée: le "connexionisme". Le principe de base consiste à interconnecter ensemble un grand nombre d'éléments opératoires simples. Cet outil a été appliqué à plusieurs domaines avec succès. Les réseaux de neurones ont

généralisé également dans la recherche récente beaucoup d'intérêt pour la prédiction des phénomènes à caractère temporel. Ils semblent donc appropriés pour modéliser un phénomène complexe tel que l'accrétion de la glace et ont l'avantage de pouvoir s'adapter à de nouvelles données au fur et à mesure que celles-ci seront disponibles.

1.4 SURVOL DU MÉMOIRE

Le présent chapitre constitue l'introduction. Le chapitre deux fera une revue de la littérature sur la modélisation de la glace atmosphérique. Au chapitre trois, les rudiments des réseaux de neurones à connaître pour comprendre ce travail de recherche seront introduits. Au chapitre quatre, la provenance des données utilisées sera expliquée ainsi que les différentes architectures de réseaux de neurones qui ont été expérimentées. Le chapitre cinq contiendra les résultats expérimentaux obtenus. Nous terminerons par la conclusion au chapitre 6.

CHAPITRE 2

REVUE DE LA LITTÉRATURE

2.1 INTRODUCTION

La prédiction de l'accumulation de la glace atmosphérique sur les conducteurs des lignes de transport d'énergie électrique est un sujet relativement complexe. Une multitude d'aspects différents doivent être pris en compte (processus physiques, aspect aléatoire des tempêtes de givre, etc.) Par conséquent, le problème peut être abordé sous plusieurs angles [47].

On peut distinguer en fait trois grandes approches possibles vis-à-vis de la problématique de la prédiction. Une première approche est constituée par la modélisation à proprement parler physique qui cherche à décrire avec le plus de réalisme possible les multiples processus physiques impliqués dans le phénomène d'accrétion de la glace. Mais à cause de la complexité des phénomènes en jeu, elle nécessite souvent des approximations simplificatrices qui permettent de traiter plus aisément le problème mathématique qui en découle.

Une seconde approche est basée sur les modèles utilisant les techniques de marches aléatoires ("*random-walk models*"). Elle cherche à simuler la structure et la densité de l'accrétion de la glace formée par les impacts aléatoires de gouttelettes en état de surfusion. Il s'agit donc d'une modélisation essentiellement de nature stochastique (et balistique).

Une troisième approche emprunte la voie de la modélisation empirique. Elle consiste à créer des modèles de type “boîte noire” qui tirent profit des liens statistiques et des corrélations existant entre les variables météorologiques et les charges de glace correspondantes. À l’intérieur de cette approche, l’utilisation des réseaux de neurones est plutôt d’usage récent. Il s’agit d’une avenue nouvelle encore très peu explorée pour la problématique de la glace atmosphérique.

2.2 DÉTAILS DES APPROCHES

On élabore ici sur les différentes approches citées ci-haut, à savoir l’approche physique, l’approche stochastique et l’approche empirique.

2.2.1 APPROCHE PHYSIQUE

L’approche physique construit des modèles basés sur les équations différentielles décrivant les lois de conservation de l’énergie (énergie cinétique des précipitations, équilibre de chaleur, flux radiant, etc.), de la quantité de mouvement et de la masse. Ces modèles peuvent aussi être appelés “continus” puisqu’ils tirent leur origine d’une description de changements continus des paramètres physiques importants. Le caractère de continuité des modèles dont il s’agit ne découle pas d’un choix arbitraire du modélisateur mais de la fidélité au phénomène physique que ceux-ci tentent de reproduire analytiquement.

Ces modèles peuvent être divisés en deux principales catégories: ceux qui sont respectivement indépendants et dépendants du temps. Dans les modèles qui sont indépendants du temps [32], les calculs sont basés sur l'hypothèse que le taux de croissance initial demeurera inchangé durant la simulation. Les modèles dépendants du temps tiennent compte à divers degrés du fait que l'accumulation en train de croître change l'écoulement des fluides autour des objets, la trajectoires des gouttelettes incidentes, l'aire de surface exposée à l'impact des gouttelettes et les conditions de transfert de chaleur [52]. Malgré des progrès importants dans leur développement, la grande majorité des modèles continus exigent des hypothèses *a priori* sur la forme de l'accumulation. Celle-ci peut prendre une morphologie circulaire, elliptique ou même non circulaire [42]. Toutefois, dans l'approche continue la relation qui existe entre les conditions de glace atmosphérique et la forme d'accrétion demeure plutôt incertaine.

De manière à simuler le processus d'accrétion théoriquement, plusieurs paramètres physiques et géométriques décrivant la situation doivent être connus [37]. Par exemple, il faut au moins connaître le rayon du câble de la ligne de transport R_0 , l'intensité de précipitation incidente I , la vitesse du vent V et les paramètres de transfert de chaleur. Toutefois, une modélisation plus détaillée nécessite l'angle θ entre l'axe du câble et le vecteur vitesse du vent, ainsi que la vitesse de chute des gouttelettes V_d , de telle sorte que la vitesse d'impact V_i puisse être déterminée. À partir de V_i , le flux d'eau incident F peut être calculé en autant que le contenu en eau

de l'air W soit connu. Celui-ci peut être calculé de la façon suivante: $W = \rho_w I / V_d$, où ρ_w est la densité de l'eau. La vitesse de chute des gouttelettes V_d peut être déterminée à partir d'expressions empiriques reliant V_d à la distribution en taille des gouttelettes et I [51].

Dans des conditions de pluies verglaçantes, les gouttelettes sont si grosses que l'efficacité de collision peut être prise pour l'unité. Cependant, l'eau collectée par le câble ne gèle pas nécessairement à la surface et peut être par conséquent perdue partiellement par écoulement au sol. D'un autre côté, cette eau d'écoulement peut contribuer à la formation de glaçons. Ces glaçons à leur tour offrent une surface supplémentaire pour la collection de gouttelettes transportées par le vent, augmentant ainsi le flux incident total. La simulation de ces processus nécessite naturellement l'utilisation de la température ambiante T , température des gouttelettes d'eau, température de surface, etc.

Évaluation de modèles spécifiques

Modèle d'Imai

Imai [23] a proposé que le taux de croissance de verglas par unité de longueur d'une ligne de transport est donné par:

$$\frac{dM}{dt} = C_1 \sqrt{VR}(-T) \quad (1)$$

où C_1 est une constante, R le rayon du cylindre glacé et M la masse par unité de longueur. L'intégration de l'équation précédente donne:

$$R^{3/2} = C_2 \sqrt{V}(-T)t \quad (2)$$

où une valeur fixe pour la densité de la glace est présumée (0.9 g/cm^3) et t est évidemment le temps.

L'équation (1) est basée sur l'idée que l'intensité de formation de la glace est contrôlée par le transfert de chaleur à partir du cylindre, c'est-à-dire que le mode de formation de la glace suit le régime humide. Dès lors, dM/dt est proportionnel à $-T$ et l'intensité de précipitation I n'a alors aucun effet. Ce modèle fort simple est en principe conceptuellement correct. Cependant, des études récentes ont montré que le transfert de chaleur (coefficient C_2) est aussi affecté par la rugosité de la surface et le refroidissement évaporatif et dès lors, la croissance humide ne subsiste pas jusqu'à $-5 \text{ }^\circ\text{C}$ comme présumé par Imai.. À cause de ce fait, le modèle tend à surestimer les charges de glace sous des conditions habituelles où le flux d'eau F contrôle la formation de glace plutôt que le transfert de chaleur; par ailleurs, il sous-estime les charges de glace en conditions extrêmes parce que la valeur de C_2 est trop petite et la formation de glaçons est négligée. Cette sous-estimation est particulière-

ment sévère si la température de l'air avoisine 0 °C.

Modèle de Lenhard

Basé sur des données empiriques Lenhard [31] a proposé que le poids de glace par mètre-unité M est donné par

$$M = C_3 + C_4 H_g \quad (3)$$

où H_g est la quantité totale de précipitation durant l'événement de givrage et C_3 et C_4 sont des constantes. Il vient alors

$$\frac{dM}{dt} = C_4 I \quad (4)$$

À vrai dire, ce modèle est plutôt simpliste. Il néglige complètement tous les effets reliés au vent et à la température. Il a été démontré empiriquement que la corrélation existant entre la quantité de précipitation au sol et la charge de glace est très basse.

Modèle de Goodwin et al.

Le modèle de Goodwin et al. [18] part de l'hypothèse que toutes les gouttelettes collectées par le câble gèlent *in-situ*. Autrement dit, le régime de

croissance est sec. Dès lors, le taux d'accrétion par unité de longueur est

$$\frac{dM}{dt} = 2RwV_i \quad (5)$$

Ici, R est le rayon du cylindre, w est le contenu en eau de l'air, V_i est la vitesse d'impact de gouttelettes et M est la masse par unité de longueur.

La masse par unité de longueur au temps t égale $\pi \rho_i(R^2 - R_0^2)$ où R est le rayon du cylindre glacé, R_0 est le rayon du câble et ρ_i est la densité de la glace. Substituant dans eq. (5), on trouve

$$\frac{dR}{dt} = \frac{wV_i}{\rho_i\pi} \quad (6)$$

L'intégration donne l'épaisseur radiale de glace $\Delta R = R - R_0$ accumulée durant une période t .

$$\Delta R = \frac{wV_i t}{\rho_i\pi} \quad (7)$$

La vitesse d'impact des gouttes est

$$V_i = \sqrt{V_d^2 + V^2} \quad (8)$$

où V_d est la vitesse de chute et V est la vitesse du vent. On suppose dans ce cas-ci que le vent est directement perpendiculaire à la ligne de transport. Le contenu en eau peut être obtenu par la hauteur de précipitation liquide H_g mesurée à partir du sol durant le temps d'accrétion t selon

$$\rho_w H_g = w V_d t \quad (9)$$

où ρ_w est la densité de l'eau. En insérant éq. (8) dans éq. (7) on a

$$\Delta R = \frac{w V_d t}{\rho_i \pi} \sqrt{1 + \left(\frac{V}{V_d}\right)^2} \quad (10)$$

avec éq. (9) on trouve

$$\Delta R = \frac{\rho_w}{\rho_i} \frac{H_g}{\pi} \sqrt{1 + \left(\frac{V}{V_d}\right)^2} \quad (11)$$

qui constitue l'équation fondamentale du modèle. Cette solution analytique découle directement des hypothèses physiques de départ. Le modèle de Goodwin et al. est conceptuellement correct.

Modèle de Chainé et Castonquay

Chainé et Castonquay [5] supposent également que toutes les gouttelettes incidentes gèlent dès l'impact mais considèrent au départ une accumulation de glace de forme elliptique. Dans un tel cas, l'aire de section du dépôt de glace S_i devient selon [5]

$$S_i = \frac{\pi R_0}{2} \sqrt{H_g^2 + H_v^2} \quad (12)$$

où H_v est l'épaisseur de la couche d'eau déposée sur la surface verticale, $H_v = w V t / \rho_w$. Ces auteurs définissent ensuite un facteur de correction K constitué par l'aire de section réelle et celle donnée par l'éq. (12). Ce facteur est déterminé par une méthode empirique comme étant une fonction de R_0 et de la température de l'air t_a . Ce qui permet de trouver l'épaisseur radiale de l'accumulation de glace

$$\Delta R = \left[\frac{R_0 K}{2} \sqrt{H_g^2 + H_v^2} + R_0^2 \right]^{1/2} - R_0 \quad (13)$$

Comparant cette solution à un dépôt elliptique et définissant:

$$H = \sqrt{H_g^2 + H_v^2} \quad (14)$$

résulte:

$$\frac{\rho_v}{\rho_i} \frac{H}{\pi} = \left[\frac{R_0 K H}{2} + R_0^2 \right]^{1/2} - R_0 \quad (15)$$

En solutionnant pour K à partir de l'éq. 15, nous trouvons:

$$K = \frac{4}{\pi} \frac{\rho_w}{\rho_i} + \frac{2}{\pi^2} \left(\frac{\rho_w}{\rho_i} \right) \frac{H}{R_0} \quad (16)$$

Comme on peut se rendre compte, le facteur de correction K dépend de tous les paramètres pertinents qui affectent l'épaisseur de la glace. Ainsi, la méthode solutionne l'épaisseur de la glace à partir d'une équation qui inclut une constante dépendant elle-même de l'épaisseur de la glace. Dès lors, la méthode de Chainé et Castonguay est plutôt conceptuellement douteuse.

Modèles numériques

Plusieurs modèles de nature purement numérique ont été développés pour *simuler la formation de glace atmosphérique sur les câbles*. Cela inclut le modèle de Lozowski et al. [32], le modèle MRI [1], le modèle MEP [2], le modèle de Makkonen [34] et celui de Finstad et al. [16]. Ces modèles décrivent les différents processus physiques impliqués à des degrés variés.

L'avantage principal des méthodes numériques est que les effets dépendants du temps peuvent être incorporés et que, dès lors, des changements dans les paramètres d'entrée peuvent être pris en compte facilement. De plus, tous les modèles susmentionnés peuvent aussi simuler une accrétion de givre (accrétion en régime sec) et peuvent même détecter le régime d'accrétion par des calculs d'équilibre de chaleur.

Toutefois, ces modèles numériques offrent peu d'amélioration vis-à-vis des solutions analytiques plus simples dans l'estimation des charges de glace dues aux précipitations verglaçantes. Premièrement, le planificateur de ligne de transport est intéressé à des charges extrêmes (dans laquelle les situations de bris importants surviennent) et ne peut donc utiliser avec avantage la capacité de ces modèles de prédire de petites charges de glace près de 0°C . Deuxièmement, il a été démontré expérimentalement par Makkonen et Stallabrass [35] et numériquement par Finstad et al. [16] que le taux d'accrétion est plutôt insensible à la forme spécifique de l'accumulation en supposant qu'il n'y ait pas de perte d'eau au sol.

Bien que ces modèles numériques soient conceptuellement corrects, le modèle plus simple de Goodwin et al. peut être avantageusement utilisé dans l'estimation des charges dues à la pluie verglaçante. D'autant plus que des modèles numériques complexes peuvent contenir des problèmes de convergence de nature numérique ou encore comporter des bogues logiciels. Cela semble le cas pour les modèles MRI et MEP. Des tentatives pour prendre en compte la formation de glaçons ont été aussi faites dans la modélisation numérique pour des câbles exposés à de la pluie verglaçante.

Modèle de Makkonen

Makkonen [34] en 1984 présenta un modèle numérique dépendant du temps.

Le modèle simule l'accrétion de gouttes d'eau en état de surfusion sur les lignes de transport. Dans celui-ci, le dépôt de glace est supposé maintenir une forme cylindrique durant le processus dû à de lentes rotations du câble. Sur un objet cylindrique, l'intensité de formation de glace peut être calculée par:

$$I = \frac{2}{\pi} Envw \quad (17)$$

où E est le coefficient de collection (rapport du flux de masse des gouttes incidentes au flux que subirait la surface si les gouttes n'étaient pas déviées dans le jet d'air), n est la fraction givrante (rapport de l'intensité de formation de glace au flux de masse des gouttes d'eau incidentes), v est la vitesse du vent et w le contenu en eau de l'air.

Le coefficient de collection est déterminé par des calculs basés sur la solution numérique de Langmuir et Blodgett (1946) pour l'équation du mouvement d'une goutte dans un flux d'air. Le coefficient est corrigé semi-empiriquement pour tenir compte de la distribution en taille des gouttes. La fraction givrante dans l'équation 17 peut être calculée par l'équilibre de chaleur sur la surface de formation de la glace. Le modèle tient compte de la dimension temporelle en traitant le diamètre D du dépôt de glace comme variable avec le temps et effectue une intégration pas à pas.

Le modèle de Makkonen de 1984 fut amélioré par la suite pour tenir compte de l'impact de l'eau sur des glaçons en croissance et aussi pour simuler la croissance

de glace de structure spongieuse [37]. Ce nouveau modèle se base sur [34] mais intègre également un modèle de simulation de formation de glaçons comme sous-routine.

Évaluation des modèles

L'évaluation des modèles précédents de glace atmosphérique peut être résumée en notant que le modèle de Goodwin (éqs 11 ou 14) est conceptuellement correct alors que la plupart des autres modèles sont soit basés sur de la physique déficiente ou sont tout simplement inapplicables. On dresse dans la table 2.1 un portrait comparatif selon Makkonen [37].

Table 2.1								
Comparaison des modèles								
Modèle	Type	Inclut le vent	Inclut la température	Inclut les glaçons	Cohérence logique	Utilisable opérationnellement	Prédiction en conditions modérées	Prédiction en conditions sévères
Imai	équation analytique	oui	oui	non	oui	oui	pauvre	pauvre
Lenhard	équation empirique	non	non	non	non	oui	pauvre	pauvre
Goodwin et al.	équation analytique	oui	non	non	non	oui	bonne	raisonnable

Chainé & Castonguay	équation semi-empirique	oui	oui	non	non	oui	raisonnable	pauvre
Lozowski et al	numérique pseudo-temporel	oui	oui	non	oui	non	X	X
Makkonen	numérique temporel	oui	oui	oui	oui	oui	bonne	raisonnable
MRI	numérique temporel	non	oui	non	non	oui	pauvre	pauvre
MEP	numérique temporel	oui	oui	non	non	oui	raisonnable	pauvre
Finstad et al.	numérique temporel	oui	oui	non	oui	oui	bonne	pauvre
Szilder	analytique/marche aléatoire	non	oui	oui	oui	non	X	X

2.2.2 APPROCHE STOCHASTIQUE

Les modèles basés sur l'approche stochastique concernent principalement l'accumulation de glace à partir de pluie verglaçante sur des objets de forme cylindrique représentant un câble de garde. Les modèles qui suivent l'approche stochastique emploient une technique de modélisation basée sur le cheminement aléatoire pour simuler la formation de la glace atmosphérique combinée avec une

approche de trajectoire balistique; ces modèles sont également référés dans la littérature sous le vocable de modèles de Monte Carlo. Ils visent à produire une prédiction de la masse totale de l'accumulation ainsi que sa structure morphologique tridimensionnelle.

Ces modèles peuvent s'appliquer à un large éventail de conditions de croissance, incluant aux limites, la croissance où les gouttes gèlent automatiquement dès l'impact (croissance en régime sec) et la croissance où les gouttes se répandent par ruissellement sur la structure de glace déjà existante (croissance en régime humide).

Dans l'approche stochastique, le mouvement aléatoire de chaque goutte ou chaque ensemble de gouttes est suivi individuellement. Le modèle balistique détermine la localisation de l'impact des éléments de fluide après avoir été "lancés" et le modèle de cheminement aléatoire prédit leur mouvement ultérieur le long de la surface de la glace déjà accumulée sur l'objet. L'élément fluide peut être imaginé comme étant composé de multiples petites gouttelettes ou comme faisant partie d'une grosse goutte unique formée par coalescence.

Les modèles balistiques ont été développés par plusieurs auteurs pour simuler des croissances en régime sec. Gates et al. [17] ont examiné l'accrétion sur des cylindres fixes et Personne et al. [46] le cas de cylindres en rotation. Cette approche

a été appliquée avec succès pour prédire l'accumulation en conditions givrantes lorsque les gouttelettes incidentes en surfusion gèlent dès l'impact. Cependant, ces modèles décrivent adéquatement uniquement les cas caractérisés par de faible quantité de mouvement (faible éparpillement des gouttes) et des température froides dans lesquelles la durée de gel est extrêmement courte. Une mobilité limitée des gouttes fut proposée par Lozowski et al. [33]. Szilder [53] introduisit la méthode par marche aléatoire pour permettre la simulation des cas en régime humide où l'eau non gelée s'écoule sur la surface de la glace accumulée avant de geler. L'addition d'un élément de hasard au forme des accumulations permet une meilleure conformité avec les observations expérimentales. L'analyse bi-dimensionnelle d'accumulation sur des cylindres fut entreprise par Szilder [54]. Une version tri-dimensionnelle qui permet une prédiction d'accumulation sur des objets 3-D complexes a été développée ultérieurement par Szilder et Lowoski [55-57].

L'avantage principal de l'approche stochastique est qu'elle permet de modéliser réalistement et de manière numériquement efficace l'écoulement de l'eau le long de la surface d'accrétion. Ce type de modèles permet non seulement de calculer l'accrétion du côté de l'objet qui se trouve exposé aux précipitations incidentes mais aussi l'accrétion qui se crée en-dessous de l'objet, incluant la formation possible de multiples glaçons formés par l'écoulement. Ainsi, il n'exige pas de connaissance *a priori* de la forme de l'accrétion. Il permet également l'étude de la variabilité stochastique de la forme de l'accumulation, ce qui n'est pas le cas avec

des modèles continus déterministes conventionnels. Les modèles stochastiques donnent la possibilité de prédire différentes formes d'accumulation en relation avec des paramètres atmosphériques ambiants tels que la température de l'air et le taux de précipitations.

La densité de la glace et la structure de l'accumulation prédite par ces modèles concordent qualitativement avec les observations expérimentales pour des cylindres fixes. La couche de glace uniforme se compare bien avec celle prédite par des modèles analytiques simples. La croissance de l'accumulation sous le cylindre est aussi en accord qualitatif avec les résultats expérimentaux. La formation des glaçons prévue par l'approche utilisée se compare avec celle prédite par l'approche classique théorique. Les résultats ressortant de ces modèles tendent à montrer que la formation de glaçons est en fait fortement liée au flux de chaleur externe par convection entrant dans les équations d'équilibre thermique.

2.2.3 APPROCHE EMPIRIQUE PAR RÉSEAUX DE NEURONES

L'accumulation de la glace atmosphérique sur les lignes de transport résulte d'une interaction complexe entre les matériaux des structures et les fluides, impliquant une dynamique plutôt difficile à modéliser avec réalisme. Les modèles physiques qui s'ensuivent nécessitent forcément l'utilisation d'hypothèses simplificatri-

ces (telles que: cible matérielle de géométrie simple, conditions ambiantes stationnaires, etc.) qui permettent de mener à bien l'analyse sans quoi la problématique deviendrait rapidement trop compliquée d'un point de vue mathématique. De plus, ce type de modèles exige, généralement pour être applicable, la connaissance exacte de nombreux paramètres physiques et météorologiques locaux qui ne sont pas toujours accessibles tels que: vitesse, orientation et niveau de turbulence du vent, teneur atmosphérique en eau liquide, dimension et distribution volumique des gouttelettes d'eau surfondues de la précipitation, température de l'air et des gouttelettes d'eau entrant en collision avec les structures matérielles. La non-disponibilité de plusieurs de ces variables implique le besoin de faire des hypothèses additionnelles pour simplifier le problème.

Dès lors, une technique empirique qui s'adapte et utilise directement les observations de "terrain" actuellement disponibles sans faire d'hypothèse *a priori* sur la nature du phénomène peut constituer une approche prometteuse et avantageuse. Une modélisation fonctionnelle basée sur des données facilement accessibles, telles que la température de l'air et l'intensité de précipitation s'avère plus pratique pour prédire les masses des accumulations de glace. L'objectif de cette approche est de créer un modèle de type "boîte noire" qui permet d'établir une corrélation statistique numérique existante entre les mesures instrumentales pertinentes et la charge correspondante de glace atmosphérique sur les lignes de transport que l'on cherche à prédire. Dans cette approche, l'intensité de givrage est mesurée ainsi que d'autres

paramètres climatiques pour établir un lien statistique (existant présumablement). L'adaptation du modèle à la réalité se fait par l'intermédiaire de "boutons d'ajustement" mathématiques dont les valeurs sont obtenues par un processus d'apprentissage à partir d'observations réelles.

À l'intérieur de cette grande famille de modèles empiriques, les réseaux de neurones constituent une nouvelle approche prometteuse. Sous plusieurs aspects, les réseaux de neurones peuvent être conçus et compris comme des extensions de techniques statistiques standards (ex: régression linéaire multiple). Les réseaux de neurones présentent un pouvoir de représentation non-linéaire très puissant qui permet de faire l'émulation de systèmes relativement complexes. Leur comportement d'entrée/sortie peut être ajusté par un processus itératif d'apprentissage. Les réseaux de neurones ont émergé comme constituant une méthodologie pratique ayant des applications réussies dans des domaines divers. Dans les années plus récentes, les techniques neuronales ont suscité beaucoup d'intérêt plus précisément pour les tâches de prédiction impliquant des processus dynamiques [61]. Leur application spécifique à la problématique du givrage atmosphérique des structures est vraiment très récente et n'a pas fait l'objet d'étude exhaustive. On peut citer à ce sujet les travaux de McComber [40,41]. Ohta et al. [45] appliquèrent les réseaux de neurones à un système intégré automatisé de prévention des désastres dus à l'accumulation de la neige sur les lignes de transport.

L'avantage principal de l'approche empirique par réseaux de neurones est qu'elle ne nécessite aucun déchiffrement analytique complexe des phénomènes physiques en jeu. Elle exige moins d'hypothèses *a priori* sur la dynamique qui régit le phénomène. Cela permet dès lors d'éviter une multitude de subtilités et de complexités mathématiques qui interviennent dans les modèles continus. Cette simplification méthodologique rend les modèles de type empirique vraisemblablement plus faciles à implanter. De plus, l'approche empirique possède comme deuxième avantage la capacité de tirer directement profit des relations existantes naturellement entre les grandeurs physiques qui sont réellement mesurables sans exiger la connaissance de "variables cachées" inaccessibles. D'une certaine manière, on pourrait dire que l'approche empirique utilise la nature elle-même comme appareil de calcul puisqu'elle produit un modèle exprimant une relation effective entre les grandeurs directement disponibles sans référence à un mécanisme exact. Cette technique permet donc d'aborder le problème sous un angle global pratique sans nécessiter d'analyse fine de détails.

Toutefois, on doit noter en toute honnêteté deux désavantages potentiels. D'abord, l'approche empirique résulte d'un processus d'apprentissage à partir de variables réelles prises sur un site d'observation donné; il est fort probable que le lien implicite statistique entre les variables exprimé par le modèle comporte une dépendance à la localité où ont été prises les observations. Cette dépendance géographique éventuelle rend difficile la transposition d'un modèle donné vers

d'autres sites instrumentés. Cet inconvénient peut cependant être vu comme un avantage si l'on considère qu'il permet une meilleure prédiction locale qu'un modèle plus général. L'importance relative de ce caractère local dépend de facteurs environnementaux difficiles à pondérer. La relation fonctionnelle incorporée dans le modèle enchevêtre inextricablement les facteurs locaux contingents et les liens phénoménologiques nécessaires découlant de la physique fondamentale des processus en question. On doit admettre cependant que seules des observations comparatives exhaustives entre plusieurs sites utilisant un modèle unique entraîné qu'une seule fois pourraient définitivement trancher la question [41]. Toutefois, bien que les paramètres explicites du modèle ne sont pas forcément transposables, la méthodologie utilisée et l'architecture fonctionnelle peuvent bien raisonnablement s'appliquer à plusieurs sites.

La deuxième faiblesse de l'approche consiste en sa dépendance absolue aux observations découlant des variables instrumentales. Cela pose deux problèmes. D'abord, les mesures instrumentales en situation de terrain comportent forcément un certain bruit de fond qui se reflétera dans le modèle. C'est une question de métrologie et d'instrumentation sur laquelle le modélisateur a peu de contrôle. Cependant, c'est un problème qui peut être considéré comme mineur et dépend de la qualité de l'instrumentation mise en place. Plus sérieusement toutefois, la puissance prédictive d'un modèle empirique dépend entièrement de la force des relations statistiques liant les variables données entre elles. Un modèle empirique

entraîné sur des observations ne pourra jamais produire une prédiction plus fiable que les variables ne le permettent par leur corrélation mutuelle. Une corrélation forte permettra l'existence d'un modèle très fiable; une corrélation faible entraînera un modèle moins solide. Le degré de déterminisme qui régit les variables dictera la fiabilité maximale que l'on peut espérer tirer du modèle.

2.3 INSERTION DU PRÉSENT TRAVAIL

La présente recherche sur l'application des réseaux de neurones pour la prédiction du givrage atmosphérique des lignes de transport s'inspire en partie des travaux initiaux effectués par McComber et al. [41,42]. Ceux-ci appliquèrent les techniques de réseaux de neurones à des données réelles de terrain provenant du système de surveillance en temps réel de la glace atmosphérique d'Hydro-Québec (SYGIVRE). Ils sélectionnèrent des données venant du site instrumenté du Mont-Bélair qui concernent des tempêtes bien délimitées dans le temps. Ils utilisèrent les variables instrumentales suivantes: le signal provenant d'un givromètre, la température ambiante de l'air, la vitesse du vent (composante transversale) et le signal du capteur de force converti en charge linéaire spécifique de glace par un algorithme. Le modèle neuronal utilisé permettait de faire une prédiction de type "*nowcasting*" (c'est-à-dire qu'il prédit la charge de glace en sortie de manière concourante dans le temps avec les entrées). Ils obtinrent des résultats raisonnables avec l'architecture

classique des réseaux de neurones. Une légère modification pour rendre l'architecture pseudo-dynamique donna de meilleurs résultats.

Le présent travail vise à continuer l'exploration encore jeune des différentes architectures de réseaux de neurones susceptibles de fournir une bonne prédiction de la charge de glace. Nous expérimenterons plus particulièrement des architectures adaptées pour traiter de l'information de nature temporelle (c'est-à-dire qui présente un caractère évolutif et/ou dynamique). Cet aspect de l'évolution temporelle est une composante importante de la problématique du processus physique de l'accrétion de la glace atmosphérique. Un réseau structuré de telle sorte qu'il permet de manière inhérente et constitutive l'inclusion de la dimension du temps donnera vraisemblablement de meilleurs résultats qu'un réseau de nature statique.

On doit noter cependant que les techniques pour traiter des sources non-stationnaires d'information ne sont pas aussi hautement développées que celles, bien établies, pour traiter les problèmes de nature statique. De manière générale, l'application des réseaux de neurones à des problèmes temporels est d'usage plus récent que l'application de ceux-ci à des problèmes comportant un caractère spatial. Dès lors, le fondement théorique mathématique des réseaux de neurones dynamiques de récentes dates ainsi que leur pouvoir de représentation en tant qu'outil prédictif sont établis avec moins de rigueur que les réseaux traditionnels.

Leurs limites éventuelles ont été moins analysées en profondeur par la communauté des chercheurs en mathématiques ou en science informatique à cause de leur complexité accrue. Il est aussi vraisemblable que les chercheurs trouveront dans l'avenir une multitude d'autres architectures pour le traitement temporel de l'information n'existant pas encore à l'heure actuelle. C'est un domaine qui demande encore à être débroussaillé.

CHAPITRE 3

LES RÉSEAUX DE NEURONES ARTIFICIELS

3.1 INTRODUCTION

Le microprocesseur Pentium d'Intel mis en service en 1993 contient 3,1 millions de transistors et consomme 16 watts de puissance. Par contraste, le cerveau humain contient de l'ordre de 10^{11} neurones, chaque neurone pouvant avoir plus de 10 000 interconnexions avec des voisins. Approximativement 20 watts au total sont nécessaires pour alimenter le cerveau. Si l'on devait identifier un transistor unique à une seule connexion synaptique cela prendrait approximativement $3 \cdot 10^8$ puces Pentium pour égaler la complexité structurelle brute du cerveau humain! Pour énergiser un tel conglomérat de silicium cela nécessiterait plus de 5 gigawatts!

Cette comparaison est cependant un peu fallacieuse car les mécanismes détaillés de traitement du cerveau demeurent en partie inconnus à ce jour. Quand bien même pourrions-nous recréer la complexité matérielle de l'organe cérébral dans un "ordinateur", comment pourrions-nous programmer celui-ci?

Que sont donc les réseaux de neurones artificiels et comment sont-ils liés à leur contrepartie biologique? Du point de vue de l'ingénierie, les réseaux de neurones sont des structures calculatoires qui tentent à reproduire des facultés cérébrales rudimentaires. Les réseaux de neurones artificiels sont composés de "neurones" vaguement modélés sur les neurones biologiques véritables. Les neurones ont des connexions aux autres neurones à travers des "synapses" modélisées formellement

par des paramètres ajustables appelés “poids synaptiques”. Ils constituent les degrés de liberté élémentaires du système. Diverses architectures de réseaux de neurones diffèrent par la manière dont les neurones sont arrangés (topologie), par la méthode par laquelle les poids sont déterminés (algorithme d'apprentissage) et par le domaine d'application visé (but ultime). Commun à toutes les structures, il y a le principe de fond qui consiste dans le fait que les poids doivent être appris plutôt que programmés directement. Un réseau est itérativement entraîné par l'intermédiaire d'une interaction avec l'environnement en présentant des exemples de paires de patrons entrée/sortie.

Haykin [21] offre la définition suivante d'un réseau de neurones:

Un réseau de neurones est un processeur distribué massivement parallèle qui a une propension naturelle pour emmagasiner la connaissance expérimentielle et à la rendre disponible pour utilisation ultérieure. Il ressemble au cerveau selon deux aspects:

- 1. La connaissance est acquise par le réseau à travers un processus d'apprentissage.*
- 2. Les forces de connexion interneurones appelées poids synaptiques sont utilisées pour l'emmagasinage de l'information.*

L'utilisation des réseaux de neurones offre les caractéristiques intéressantes suivantes:

- *Nonlinéarité.* Le neurone élémentaire est un dispositif non-linéaire. Conséquemment, l'interconnexion de plusieurs neurones ensemble résulte en un

système global lui-même non-linéaire. Cette propriété importante permet l'emploi des réseaux de neurones pour la modélisation de systèmes où existent des non-linéarités inhérentes.

- *Relation d'entrée/sortie.* Un paradigme populaire d'apprentissage est l'entraînement supervisé qui consiste à présenter au réseau des exemples d'apprentissage. Chaque exemple comporte un signal d'entrée et une réponse de sortie désirée. Les poids synaptiques sont modifiés de telle sorte à minimiser l'erreur entre la réponse voulue et la réponse effective du réseau. En alimentant le réseau avec plusieurs exemples et en répétant plusieurs fois, il finit par se stabiliser et converger. Le réseau construit donc progressivement une représentation implicite de la relation entrée/sortie.

Il est démontré mathématiquement [7] que les réseaux de neurones artificiels constituent des *approximateurs universels* puisqu'ils peuvent approximer des relations fonctionnelles entrée/sortie non-linéaires arbitraires étant donné un certain degré minimal de continuité et sous condition d'un nombre suffisant d'unités de traitement (neurones).

- *Adaptabilité.* Les réseaux de neurones artificiels ont une capacité inhérente d'adaptation en vertu de leurs connexions synaptiques ajustables. Dans un contexte d'environnement non-stationnaire, un réseau de neurones peut être

conçu pour opérer en temps réel. Cela peut constituer un atout avantageux pour le traitement adaptatif du signal par exemple.

- *Tolérance aux pannes.* Un réseau de neurones implanté dans le matériel a le potentiel d'être tolérant aux pannes dans le sens que la performance se dégrade doucement avec des conditions adverses. En effet, étant donné que l'information emmagasinée se trouve sous une forme distribuée à travers tout le réseau, un bris matériel local ne provoquera pas ainsi une panne catastrophique. Ceci peut être comparé par analogie à une image holographique dans laquelle l'information visuelle est distribuée globalement sur l'ensemble d'un substrat plutôt que localement.
- *Implantabilité VLSI.* La nature parallèle des réseaux de neurones les rendent potentiellement rapides pour certaines tâches de calcul si prise en compte. Cette même caractéristique les rend attrayants pour une implantation électronique dans le matériel utilisant la technologie du *Very-Large-Scale-Integration* (VLSI). L'avantage particulier du VLSI est de fournir un moyen d'émuler des comportements complexes dans une structure hiérarchique ce qui rend possible l'utilisation de réseaux de neurones comme outil pour les applications en temps réel. (Cependant, les implémentations informatiques usuelles par voie de logiciel sur station uniprocasseur n'utilisent que très rarement ce caractère de parallélisme).

Dans les années récentes, il y a eu une résurgence majeure de l'activité de recherche dans le domaine des réseaux de neurones. De nouvelles applications ont fait leur apparition dans divers domaines allant de la classification au contrôle robotique. Les produits commerciaux incluent: des systèmes de reconnaissance de caractère, des séquenceurs moléculaires pour la génétique, des systèmes de détection de fraude dans les transactions par cartes de crédit, des logiciels antivirus intelligents, des logiciels de prédiction de cours boursier, des systèmes de contrôle de processus, etc.

Plus près de notre problématique, les réseaux de neurones artificiels ont généré récemment de l'espoir pour la prédiction non-linéaire des séries temporelles ainsi que pour la modélisation de systèmes dynamiques complexes. Ils constituent une alternative prometteuse aux techniques traditionnelles pour le traitement temporel de l'information [61].

3.2 MODÈLE D'UN NEURONE ARTIFICIEL

En 1943, McCulloch et Pitts [43] ont proposé le premier modèle de base d'un neurone artificiel. Bien que très simple, il constitue l'élément de base des réseaux de neurones. La figure 3.1 montre un modèle de neurone.

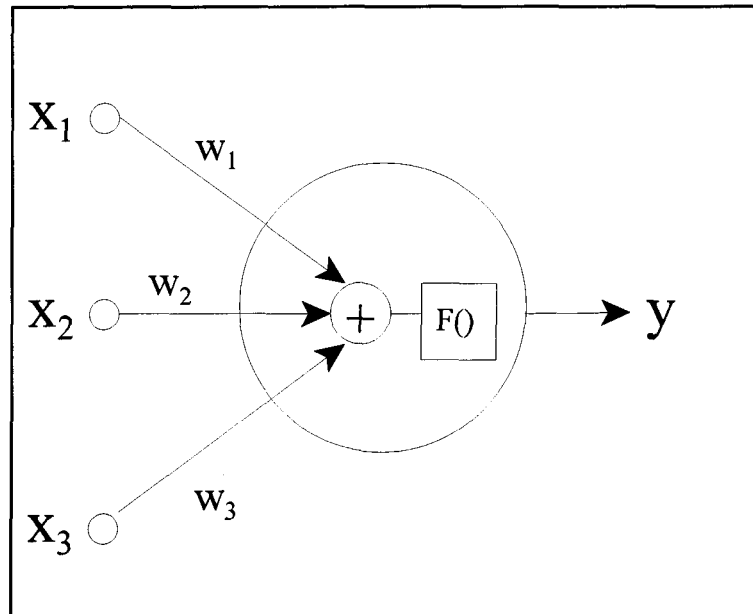


Figure 3.1: Modèle d'un neurone élémentaire

Les entrées au neurone sont notées par x_i . Il a un ensemble de synapses caractérisées par un poids synaptique qui indique la force du lien entre l'entrée et le neurone. Ceux-ci sont notés par w_i . Les signaux d'entrée sont multipliés par ceux-ci avant d'être additionnés. Un poids positif est associé à une synapse excitatrice; un poids négatif est associé à une synapse inhibitrice.

Un additionneur est utilisé pour faire la somme pondérée des entrées plus un certain décalage constant de biais θ . Le résultat est ensuite passé à travers une fonction de transfert $f()$ généralement non-linéaire qu'on appelle *fonction d'activation*. Celle-ci limite l'amplitude de sortie du neurone dans un certain intervalle borné, usuellement de $[0, 1]$ ou bien encore de $[-1, 1]$.

On peut représenter mathématiquement le traitement que fait le neurone élémentaire par les équations suivantes:

$$s = \sum_{i=1}^N w_i x_i + \theta \quad (1)$$

$$y = f(s) \quad (2)$$

Les poids w_i et θ constituent les paramètres libres du système et doivent être ajustés.

Dans les premiers travaux fondateurs de McCulloch et Pitts sur le Perceptron la fonction d'activation était une simple fonction de seuillage brutale de type $sign(x)$. Dès lors, la sortie du réseau ne pouvait prendre que des valeurs binaires (-1 ou +1). Cela était peut-être utile pour des tâches de classification. Cependant, de nos jours, il est plus usuel d'utiliser une fonction continue de forme sigmoïdale. C'est de loin la fonction la plus utilisée dans la littérature sur les réseaux de neurones artificiels (RNA). Elle a les avantages d'être continue, limitante pour la sortie ("*squashing*") et bien différentiable. Algébriquement, on la retrouve sous deux formes très voisines l'une de l'autre. Soit la fonction log-sigmoïde:

$$y(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

soit la fonction tangente hyperbolique:

$$y = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4)$$

Elles diffèrent principalement par leur intervalle de sortie, soient $[0,1]$ et $[-1,1]$ respectivement. Les formes graphiques de celles-ci sont illustrées aux figures 3.2 et 3.3.

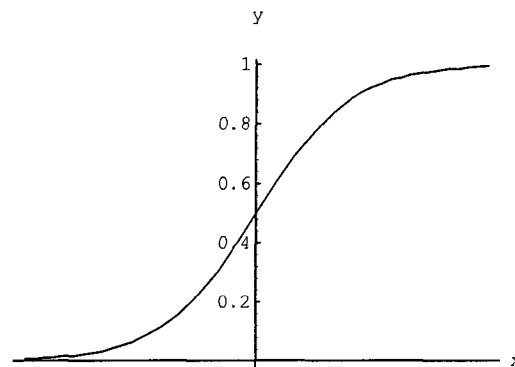


Figure 3.2: Fonction sigmoïdale

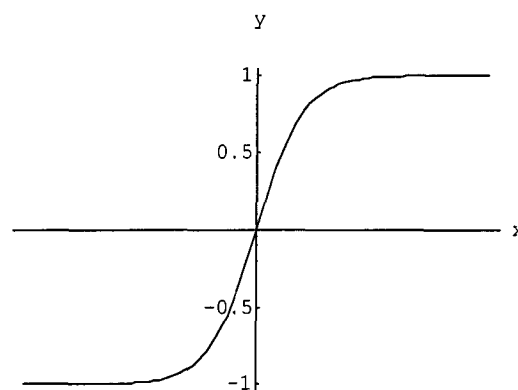


Figure 3.3: Fonction tangente hyperbolique

Dans certaines architectures de réseaux de neurones (ex: ADALINE) on retrouve plutôt une fonction de transfert purement linéaire.

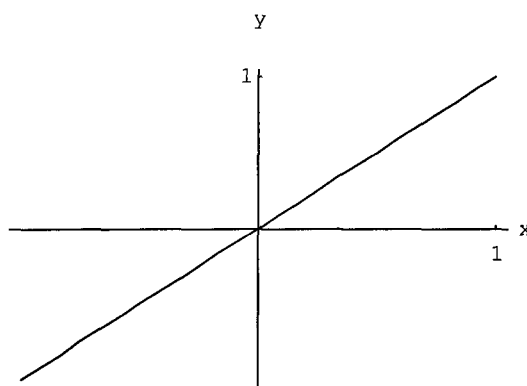


Figure 3.4: Fonction de transfert linéaire

Ce modèle formel est évidemment une approximation très grossière de la réalité biologique. Toute la dynamique complexe de la génération des influx nerveux est simplement représentée par une fonction scalaire sans mémoire $f()$. En fait, il n'y a pas de dynamique interne. Pour des poids fixes, le neurone effectue une opération statique. Ce modèle intègre plutôt les propriétés spatiales distribuées du neurone biologique. Mais bien sûr, les buts de l'ingénieur sont bien différents de ceux des neurobiologistes...

À l'instar des systèmes biologiques, les capacités calculatoires d'un neurone isolé sont plutôt limitées. De loin, les applications les plus réussies des neurones uniques sont les filtres FIR (Finite Impulse Response) adaptatifs développés en premier au début des années 1960 par Widrow et Hoff [63]. Dans ce cas, les entrées au neurone sont fournies à partir d'une ligne à délai ($s(k) = \sum w_i x(k-i)$), où k est un

indice temporel discret). Typiquement, seulement une fonction d'activation linéaire est utilisée dans ce cas. Les filtres adaptatifs constituent un outil fondamental du traitement du signal avec des applications pratiques allant de l'égalisation adaptative à l'annulation active du bruit.

3.3 TYPES DE TOPOLOGIE

Un réseau de neurones résulte de l'arrangement du modèle de base du neurone individuel dans diverses configurations. Plusieurs architectures sont donc envisageables. On doit cependant réaliser que les algorithmes utilisés pour entraîner les RNA ont une adéquation très forte avec une architecture donnée, c'est-à-dire qu'ils leurs sont intimement liés. Jusqu'à ce jour, plusieurs topologies et algorithmes d'apprentissage ont été imaginés et conçus.

On distingue trois principaux types d'architecture:

- les réseaux à propagation avant unicouches (ex: ADALINE)
- les réseaux à propagation avant multicouches (ex: Multilayer Perceptron, Radial-Basis Functions network)
- les réseaux récurrents (ex: réseau d'Elman, Dynamic Recurrent Neural Network)

Les réseaux à propagation avant sont caractérisés par le fait que les signaux ne se propagent que dans une seule direction, c'est-à-dire des noeuds d'entrée vers les noeuds de sortie. Dans le cas des réseaux unicouches, on ne retrouve qu'une seule couche de neurones soit à la sortie tel qu'illustré à figure 3.5. Il s'agit de la forme la plus simple, celle qui fut utilisée historiquement en premier jusqu'au milieu des années 1980.

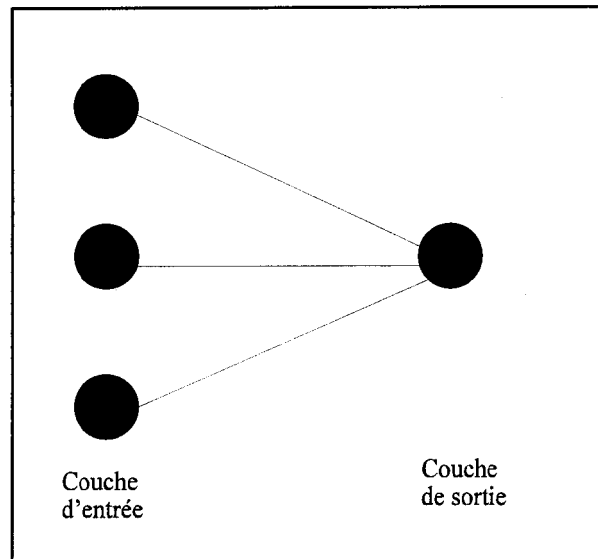


Figure 3.5: Réseau à propagation avant unicouche

La seconde classe d'architectures, soit les réseaux à propagation avant, constitue de loin la configuration la plus populaire et la plus utilisée. Elle se distingue telle qu'illustrée à la figure 3.6 par la présence d'une ou plusieurs couches cachées comportant des noeuds calculatoires internes nommés à juste titre *neurones cachés*. Leur rôle est d'intervenir entre les entrées externes et les sorties du réseau, chaque couche alimentant dans une seule direction les couches successives. En ajoutant ainsi des couches cachées, le réseau peut extraire des statistiques d'ordre supérieur et augmenter sa puissance de représentation. Cette architecture devint populaire au milieu des années 1980 lorsque Rumelhart et McClelland [48] redécouvrirent et publièrent un algorithme approprié pour l'entraîner: la rétropropagation ("*backpropagation algorithm*"). C'est toutefois à Werbos [62] qu'on doit la paternité de la

découverte originelle de la rétropropagation.

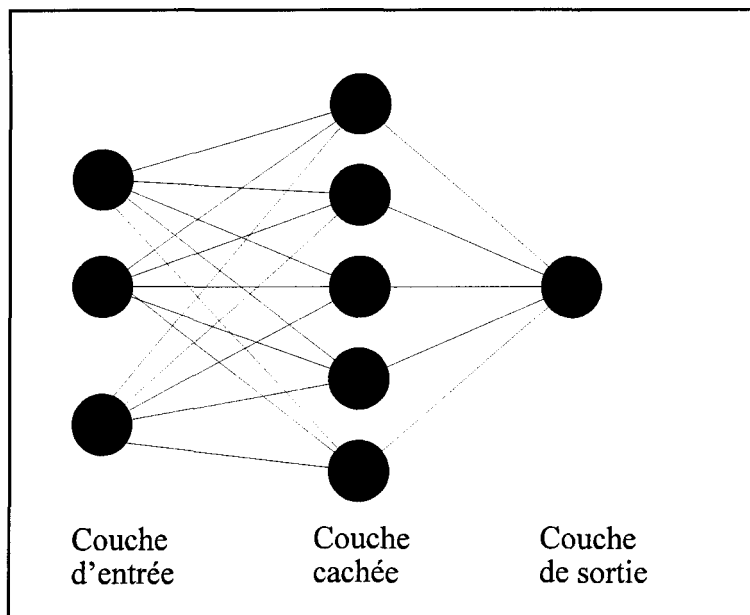


Figure 3.6: Réseau à propagation avant multicouche

La troisième classe d'architectures, soit les réseaux récurrents, se distingue des réseaux à propagation avant par le fait qu'elle contienne des boucles de rétroaction (ou liens récurrents). Une partie du signal de sortie peut revenir à l'entrée. On peut aussi imaginer de la rétroaction sur des cellules cachées comme on le voit à la figure 3.7. Ce changement structurel a un impact profond sur la capacité d'apprentissage du réseau et son fonctionnement. De plus, les boucles de rétroaction utilisent des branches composées d'élément à retard unitaire (dénnoté par z^{-1}) résultant en un comportement dynamique non-linéaire. Ainsi, le réseau peut acquérir dès lors un comportement temporel proprement dit.

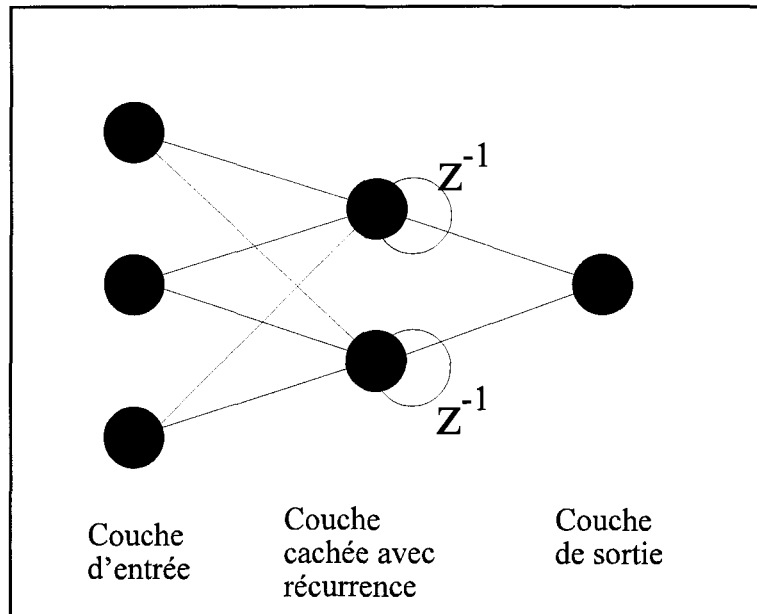


Figure 3.7: Réseau récurrent avec neurones cachés

3.4 ALGORITHME STANDARD D'APPRENTISSAGE

La relation spécifique associée à un réseau en couches est déterminée par les valeurs des poids synaptiques. Entraîner un réseau consiste donc à trouver un ensemble de poids qui produit une application entrée-->sortie désirée. Dans la plupart des cas, la relation exacte n'est pas connue d'avance. Seuls des exemples spécifiques de paires entrée/sortie peuvent être disponibles pour faire l'apprentissage.

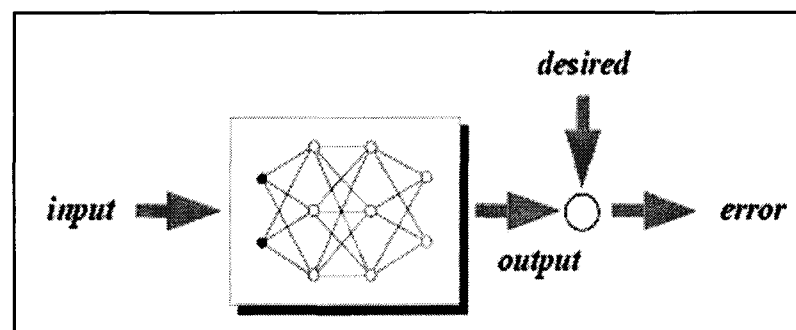


Figure 3.8: Apprentissage supervisé (tiré de ref. [58])

Dans la littérature, cette méthodologie est catégorisée sous le nom d'*apprentissage supervisé* (par opposition à l'apprentissage automatique ou non-supervisé existant dans d'autres domaines d'application telle que la reconnaissance des formes), voir la figure 3.8. Le réseau est alimenté avec un ensemble d'entraînement comportant des vecteurs d'entrée, chacun étant associé avec une sortie correspondante désirée $\{(x_1, d_1), \dots, (x_p, d_p)\}$ où les x_i sont les variables d'entrée et les d_i constituent les variables de sortie correspondantes désirées. Pour une entrée x donnée, la

différence avec la sortie désirée et la sortie effective du réseau constitue le signal d'erreur: $\mathbf{e} = \mathbf{d} - \mathbf{y}$,

où \mathbf{y} est la réponse effective du réseau en cours d'apprentissage.

L'erreur quadratique totale sur l'ensemble d'apprentissage complet est alors donnée par:

$$J = \sum_{p=1}^P \mathbf{e}_p^T \mathbf{e}_p \quad (4)$$

L'objectif de l'entraînement est de déterminer l'ensemble des poids \mathbf{w} qui minimise la fonction de coût J sous contrainte de garder la topologie du réseau fixe. Ainsi, nous pouvons voir que faire l'apprentissage du réseau de neurones revient à un problème d'optimisation numérique standard. Due à la structure non-linéaire du réseau, un algorithme stochastique de descente le long du gradient est le choix le plus populaire comme méthode d'optimisation. Après la présentation d'un échantillon provenant de l'ensemble d'apprentissage, les poids sont adaptés selon:

$$\Delta \mathbf{w} = -\mu \hat{\nabla}, \quad (5)$$

où $\hat{\nabla} = \frac{\partial e^T}{\partial \mathbf{w}}$ est la valeur instantanée du gradient de l'erreur correspondant au

patron d'entrée courant, μ est une constante qui contrôle la vitesse d'apprentissage.

La rétropropagation

Pour complètement spécifier l'algorithme d'apprentissage nous devons fournir une expression explicite de l'erreur du gradient. Les dérivées partielles correspondant à chaque poids synaptique doivent être trouvées. La rétropropagation n'est rien de plus qu'un algorithme efficace pour les trouver. Malgré tout, cela demeure la percée la plus significative dans le domaine des réseaux de neurones dans les deux dernières décennies. L'algorithme a été décrit en premier par Werbos [62] en 1974. Cependant, ce n'est qu'en 1986 lorsqu'il fut présenté par Rumelhart et McClelland [48] qu'il acquit une réelle acceptation dans la communauté scientifique.

Cas du neurone unique

Considérons d'abord un seul neurone linéaire décrit par: $y = \sum_{i=0}^N w_i x_i = w^T x$,

où $\mathbf{w}=[w_0, w_1, \dots, w_N]$ et $\mathbf{x}=[1, x_1, \dots, x_N]$. Dans ce cas, il vient:

$$\hat{\nabla} = \frac{\partial e^2}{\partial w} = 2e \frac{\partial (d - y)}{\partial w} = -2ex \quad (6)$$

et le changement de poids est alors donné par: $\Delta w = 2\mu ex$ (7)

Ceci constitue l'algorithme du Moindre Carré Moyen (LMS) très utilisé pour le

traitement adaptatif des signaux. Il sert aussi à entraîner le réseau ADALINE. Pour des données de nature statique, cet algorithme itératif est équivalent à une régression linéaire multiple et converge asymptotiquement vers la solution optimale au sens de la minimisation de l'erreur quadratique [21].

Cas d'un réseau multicouche

La rétropropagation peut être vue comme une extension de l'algorithme du LMS. Pour calculer le gradient vis-à-vis des poids, on utilise la règle de dérivation en chaîne:

$$\frac{\partial e^T}{\partial w_{ij}^l} = \frac{\partial e^T}{\partial s_j^l} \frac{\partial s_j^l}{\partial w_{ij}^l} = \delta_j^l a_i^{l-1}, \quad (8)$$

$$\text{où } \delta_j^l \equiv \frac{\partial e^T}{\partial s_j^l}.$$

Ici, l'indice "j" réfère à l'ordinal du neurone de la couche "l"; l'indice "i" quant à lui réfère à l'ordinal de la connexion afférente au neurone "j". Les termes "s" sont les jonctions sommatrices alimentées par les poids en question et les termes "a" sont les valeurs d'activation agissant comme entrées pour les poids (elles sont multipliées par ceux-ci); les w sont les valeurs de poids. Cela nous amène à l'équation de mise à jour des poids suivante:

$$\Delta w_{ij}^l = -\mu \delta_j^l a_i^{l-1} \quad (9)$$

C'est la généralisation de l'algorithme du LMS cherchée et elle demeure valide pour n'importe quel poids du réseau. Le terme d'erreur δ est associé à chaque neurone individuel, il mesure la sensibilité à une perturbation à un noeud donné. Il constitue une sorte "d'erreur locale". Les valeurs des δ sont dérivées récursivement en partant de la couche de sortie:

$$\delta_j^L \equiv \frac{\partial e^T}{\partial s_j^L} = \frac{\partial e^T}{\partial a_j^L} \frac{\partial a_j^L}{\partial s_j^L} = \frac{\partial e^T}{\partial y_j} f'(s_j^L) \quad (10)$$

où f' est la dérivée de la fonction de transfert évaluée à la valeur courante de la jonction sommatrice et L est l'indice de la couche de sortie. L'activation des neurones de sortie a été remplacée par la valeur de la sortie y .

Pour la couche de sortie, chaque neurone a une réponse désirée connue. Dès lors,

$$\frac{\partial e^T}{\partial y_j} = \frac{\partial e_j^2}{\partial y_j} = 2e_j \frac{\partial (d_j - y_j)}{\partial y_j} = -2e_j \quad (11)$$

Substituer pour l'erreur δ donne donc: $\delta_j^L = -2e_j f'(s_j^L)$. (12)

Le calcul des δ internes est compliqué par le fait qu'il n'y pas de réponse

explicite désirée dans les couches internes. Cependant, les neurones cachés influencent la réponse à la sortie par l'intermédiaire des poids par lesquels ils s'attachent aux couches supérieures. En employant encore la règle de la dérivation en chaîne:

$$\delta_i^l \equiv \frac{\partial e^T}{\partial s_i^l} = \sum_j \frac{\partial e^T}{\partial s_j^{l+1}} \frac{\partial s_j^{l+1}}{\partial s_i^l} \quad (13)$$

La sommation se fait sur toutes les branches que le neurone i de la couche l

alimente. On note que: $\frac{\partial s_j^{l+1}}{\partial s_i^l} = w_{ij}^{l+1} f'(s_i^l)$. De plus, $\frac{\partial e^T}{\partial s_j^{l+1}} = \delta_j^{l+1}$.

Faisant les substitutions nécessaires, cela conduit à la relation suivante:

$$\delta_i^l = f'(s_i^l) \sum_j \delta_j^{l+1} w_{ij}^{l+1} \quad (14)$$

Il s'agit d'une forme récursive dans laquelle les termes δ précédemment calculés sont "rétropropagés" vers l'arrière à travers les connexions synaptiques.

L'algorithme de rétropropagation pour les réseaux multicouches peut donc être résumé ainsi:

$$\Delta w_{ij}^l = -\mu \delta_j^l a_i^{l-1}$$

$$\delta_i^l = \begin{cases} -2e_i f'(s_i^L) & l = L \\ f'(s_i^l) \cdot \sum_j \delta_j^{l+1} \cdot w_{ij}^{l+1} & 1 \leq l \leq L-1 \end{cases}$$

Pour les poids de biais il faut noter que la valeur de l'activation est unitaire. On peut observer aux figures 3.9-3.10 qu'il y a une symétrie entre la phase de propagation avant de la valeur d'activation des neurones et la phase de propagation arrière avec les termes δ . Tous les calculs sont localement distribués, l'information circulant bidirectionnellement à travers les connexions synaptiques. Cet algorithme possède un coût d'exécution en $O(n)$ où n est le nombre de neurones, ce qui le rend très attrayant.

Pour résumer le processus, un patron d'entrée \mathbf{x} est fourni. L'entrée cause une propagation de l'activité des neurones résultant en une sortie \mathbf{y} . La sortie est comparée avec la réponse désirée \mathbf{d} ce qui donne un vecteur d'erreur \mathbf{e} . Le gradient de l'erreur quadratique est récursivement "rétropropagée" à travers le réseau dans la direction inverse. Les poids sont alors ajustés de manière stochastique (ou en lot) selon la descente du gradient. Ce processus est répété plusieurs fois jusqu'à ce que l'erreur atteigne un niveau acceptable.

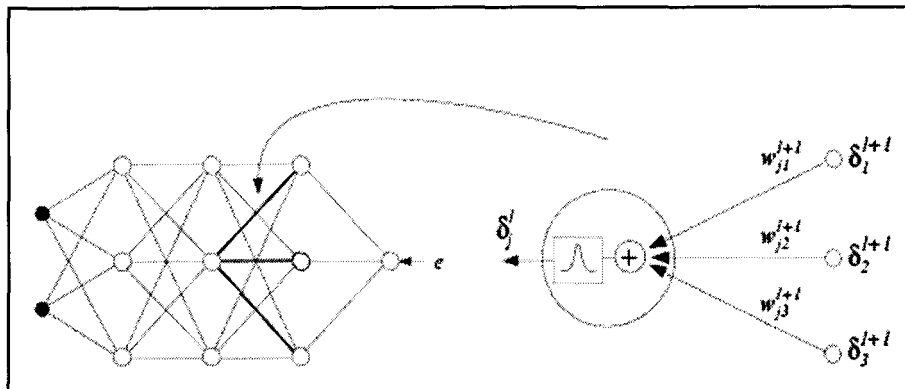


Figure 3.9: Processus de rétropropagation (tiré de ref [58])

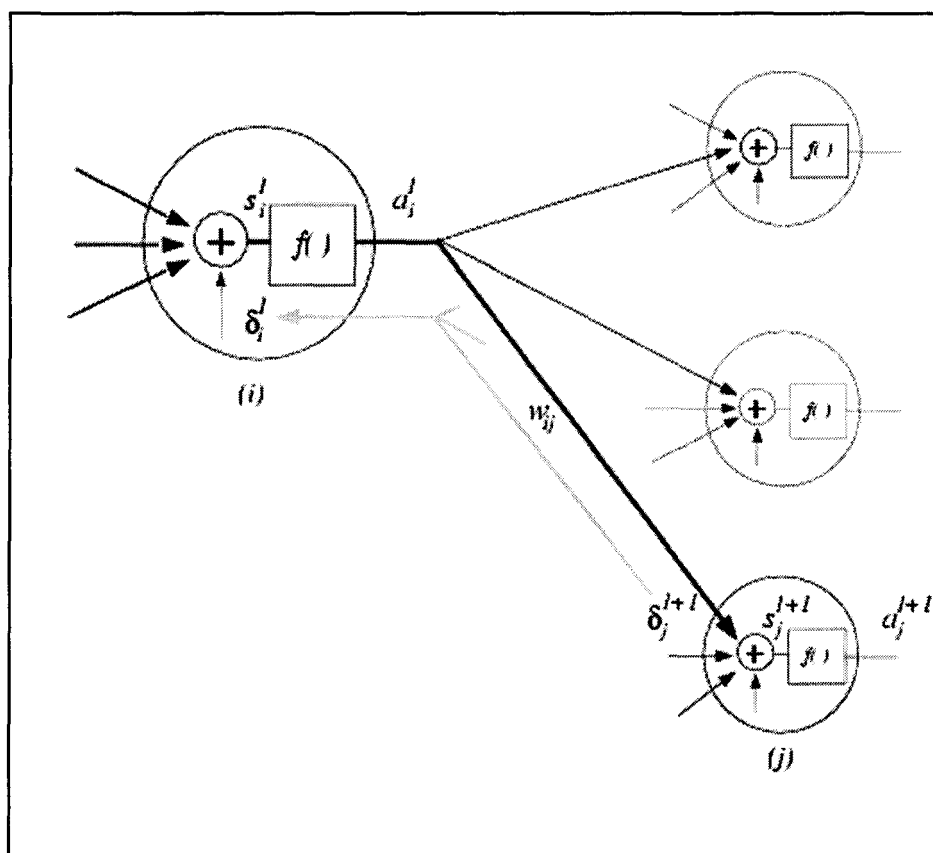


Figure 3.10: Détails du processus de rétropropagation (tiré de ref[58])

3.5 STRUCTURE PRÉDICTIVE TEMPORELLE

L'utilisation type des réseaux de neurones statiques se situe dans le traitement de formes spatiales. Par contraste, la prédiction de patrons temporels nécessite le traitement de patrons qui évoluent séquentiellement à travers le temps; la réponse à un instant particulier du temps dépend non seulement de la présente valeur de l'entrée mais aussi potentiellement des valeurs passées du processus. Ceci vient du fait que le temps est une notion ordonnée par nature.

Dans ce contexte, selon Mozer [44] un système prédictif peut être vu comme étant composé de deux parties distinctes (voir fig. 3.11). La première composante est constituée par une mémoire à court-terme qui retient des aspects de la séquence d'entrée pertinents pour la prédiction. C'est la mémoire qui a la responsabilité de tenir compte de la nature temporelle de l'information. La deuxième composante est un module prédictif qui utilise l'information contenue dans la mémoire à court-terme pour fournir une prédiction en sortie.

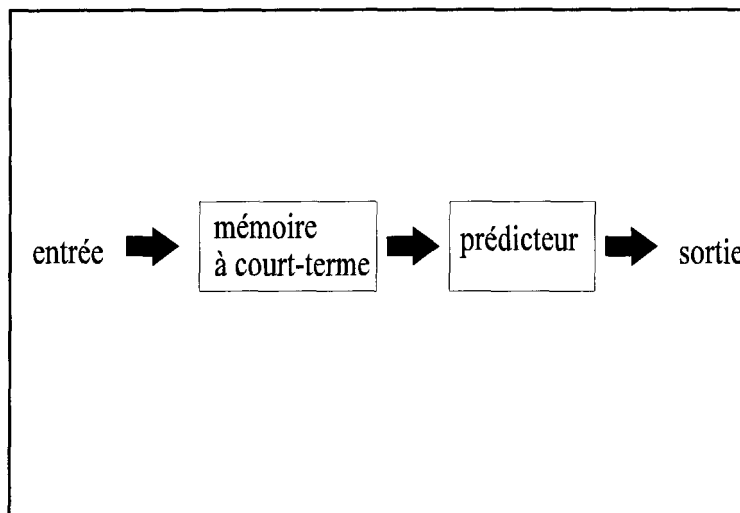


Figure 3.11: Structure temporelle prédictive

Dans le cadre des réseaux de neurones, le prédicteur sera souvent un réseau à propagation avant alors que la mémoire à court-terme pourra avoir des connexions internes récurrentes. Plusieurs architectures de mémoire ont déjà été proposées: ligne à délai, trace exponentiel, mémoire gamma, etc. La forme la plus simple de mémoire et la plus courante est sans contredit la ligne à délai. Elle consiste en une banque d'opérateurs à délai unitaire qui permet d'aller chercher un échantillonnage régulier des p valeurs les plus récentes d'une variable: $[X(t-1) X(t-2) X(t-3) \dots X(t-p)]$. Il faut noter que cette forme de mémoire possède une *résolution* et une *profondeur* constante et uniforme, c'est-à-dire qu'elle traite chaque valeur sur le même pied d'égalité (autant le passé "lointain" que le passé "proche"). Ceci termine notre survol des rudiments des réseaux neuronaux. Nous aborderons au prochain chapitre les données statistiques et les architectures spécifiquement étudiées.

CHAPITRE 4

DONNÉES D'ENTRAÎNEMENT

ET ARCHITECTURES

4.1 SOURCE DES DONNÉES

Dans le but d'améliorer la surveillance de leur propre réseau électrique, Hydro-Québec a instauré un système de surveillance en temps réel (SYGIVRE) basé sur un détecteur de glace modifié appelé le givromètre [29]. Présentement, ce système peut détecter la glace atmosphérique localement et peut donner une estimation des charges d'accumulation de glace sur les lignes de transport. Basé sur ces estimés, des avertissements sont émis selon trois niveaux de sévérité: léger, médium et sévère.

Dans le cadre du système SYGIVRE, Hydro-Québec a équipé le Mont-Bélair avec de l'instrumentation [19,39,50]. Pour faire l'entraînement de modèles à base des réseaux de neurones nous avons utilisé des données provenant de ce site de mesure du Mont-Bélair, se trouvant à 25 km au nord-ouest de Québec. Situé à une altitude de 490 m, ce site se trouve localisé dans la vallée du St-Laurent et entouré par la chaîne de montagnes des Laurentides. Les vents principaux, généralement orientés selon l'axe du fleuve St-Laurent, sont souvent chargés d'humidité. Par conséquent, le site est fréquemment le théâtre d'événements de glace atmosphérique importants. Il reçoit tous les types de glace atmosphérique: verglas, givre de nuage, neige fondante. Sa situation topogéographique et la présence de deux lignes de haute tension (315 kV et 735 kV) qui le croisent rendent ce site très approprié pour l'observation de l'influence de la glace atmosphérique (voir figure 4.1). Pour notre

travail de recherche, nous n'avons considéré que la ligne de 315 kV.

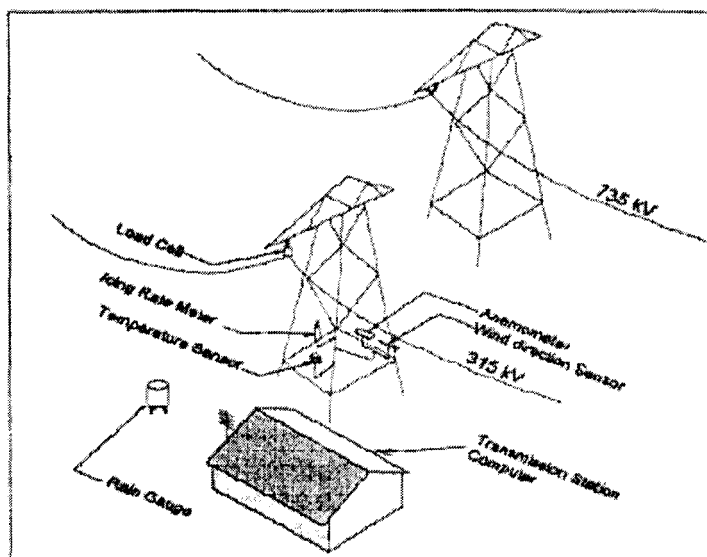


Figure 4.1: Schéma du site de mesure du Mont-Bélair

La plupart des variables météorologiques standards y sont mesurées: la température ambiante de l'air, l'humidité relative, la vitesse et la direction du vent, le taux de précipitation. Hydro-Québec a installé des capteurs de force pour mesurer les forces mécaniques dues aux poids de glace. Un système de mesure de charge développé par Ontario-Hydro, le TLNS-10, a été installé sur la ligne de 315 kV. Le capteur de force installé entre la chaîne d'isolateurs et le pylône mesure la force appliquée dans l'axe des isolateurs et deux angles de rotation suivant les axes longitudinal et latéral de la ligne. Le capteur de force donne une mesure reliée à la tension verticale mécanique. Un algorithme de calcul peut convertir la lecture directe en une masse de glace par unité de longueur (en kg/m) ou en un taux horaire

d'accrétion (kg/m/h). À noter que le signal brut du capteur n'est pas étalonné directement en newtons mais fournit une lecture qui lui est proportionnelle (transducteur linéaire).

Le site est également doté d'un givromètre développé par Hydro-Québec qui permet d'estimer l'intensité des conditions givrantes. Il enregistre l'arrivée des événements de givrage ou de verglas ainsi que leur durée. Sa sonde (Rosemount 871) est un cylindre de 6,2 mm en diamètre et de 25,4 mm de longueur. Un oscillateur force la sonde à vibrer dans le sens longitudinal, parallèle à son axe. Il est opéré à sa fréquence de résonance lorsque libre de dépôts de glace; cependant, l'accumulation de glace va causer un décalage de la résonance correspondant à une augmentation de la charge de la glace sur la sonde. Après qu'une certaine épaisseur de glace fixée d'avance se soit accumulée, la sonde est déglacée par chauffage et un nouveau cycle recommence. À l'aide d'un circuit électronique de contrôle, la sonde est calibrée pour émettre un signal pour un décalage de fréquence de 123 Hz, correspondant à 60 ± 2 mg de glace accumulée sur le cylindre.

Toutes ces informations sont enregistrées en continu et transmises à Montréal par divers moyens de communication: par lignes téléphoniques, par téléphones cellulaires et par satellites. Les données sont alors lues et traitées de manière automatique par le système de surveillance SYGIVRE qui émet des avis d'événements si nécessaire.

Puisque ces variables possèdent un caractère continu (exception faite du signal de givromètre), elle sont moyennées temporellement et fournies sur une base horaire (une moyenne est prise pour une période d'une heure). Ces données sont emmagasinées dans les serveurs centraux du système SYGIVRE.

Dans le contexte du présent travail de recherche, nous avons dû retenir certaines grandeurs physiques et en omettre d'autres. Les modèles utilisaient comme variables d'entrée la température ambiante, la vitesse normale du vent et les signaux de givromètre. Pour ce qui est du vent, nous avons seulement considéré la composante perpendiculaire (normale) à la ligne de haute-tension; la composante tangentielle apporte une contribution beaucoup plus faible à l'accumulation de la glace atmosphérique et a donc été éliminée. Quant au givromètre, le signal a été pris sous la forme de nombre de cycles (coups) par heure.

Comme variable de sortie, le signal provenant du capteur de force sur la ligne de 315 kV a été utilisé. Dans une certaine version des modèles, nous avons fait un usage direct du signal instrumental en volts qui est proportionnel à la charge mécanique totale. Dans une autre version des modèles, nous avons pris le taux métrique horaire d'accrétion tel que donné par un algorithme de conversion [49]. À cause de questions logistiques informatiques, nous n'avons pas pu faire un usage uniforme des deux méthodes sur l'ensemble de la période couverte par notre recherche. (Les données des deux versions se recoupent partiellement dans le temps

mais ne coïncident pas complètement). Le conditionnement de sortie et les questions techniques qui s'y rattachent se trouvent en dehors de la problématique fixée dans le cadre de cette étude. Pour résumer donc, les modèles suivaient le schéma de la figure 4.2.

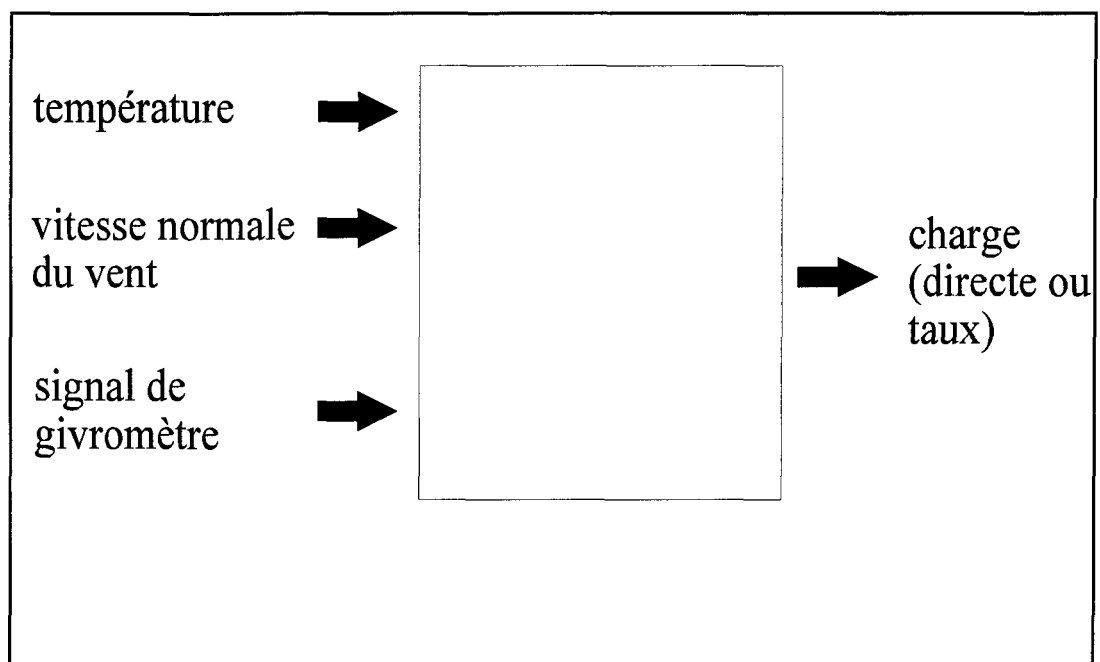


Figure 4.2: Schéma global du modèle

4.2 PRÉ-TRAITEMENT

Avant de fournir les données aux modèles à base de réseaux de neurones, celles-ci ont nécessité un certain pré-traitement qui consistait en une sélection volontaire de données pertinentes et des transformations préparatoires pour les adapter aux techniques “connexionistes”.

Pour les modèles utilisant le signal instrumental du capteur de force, une base de données expérimentales a été créée en utilisant 33 événements de givrage qui survinrent durant la période couvrant de février 1998 à avril 1999 inclusivement. Ce qui totalise environ 1250 points. Dans ce cas, il s'agissait des seules données disponibles. Pour les modèles basés sur le taux d'accrétion, nous avons pu inclure de manière additionnelle des données concernant l'année 2000 ce qui permit de créer une deuxième base de données distincte de la première.

La définition d'un “événement” de glace atmosphérique et sa délimitation temporelle étaient entièrement basées sur les avis automatiques émis par Hydro-Québec par l'intermédiaire du système SYGIVRE. Un événement de givre ou de verglas et son processus physique peuvent être divisés en trois phases de nature distincte: la phase d'accrétion (accumulation croissante de la charge de glace), phase de persistance (état stable du poids de glace) et phase de délestage (perte graduelle et/ou chute de la glace). En considération de ce fait, une seconde série de données

a été construite dans laquelle on ne retrouve par conséquent que la phase d'accrétion. Celle-ci contenait environ 520 points (pour la période 1998-1999). Chaque base de données fut ensuite sous-divisée en deux ensembles de données dans le but d'entraîner et de valider les modèles neuronaux. Dans ce cas, la règle 75/25 fut utilisée, c'est-à-dire que 75% des données servirent à l'apprentissage et 25% servirent à effectuer la validation.

Après sélection, les données furent transformées. D'abord, sur le serveur SYGIVRE originel, la vitesse du vent se trouvait sous la forme *vitesse/azimut*. Un simple calcul trigonométrique permit d'extraire la composante normale de la vitesse du vent (connaissant l'orientation absolue de la ligne de 315 kV); la composante tangentielle a été éliminée parce qu'elle apporte une contribution négligeable au phénomène d'après [40].

Le signal brut du givromètre subit également une transformation. Sous la forme originelle dans SYGIVRE, un compteur existe qui maintient un décompte cumulatif des coups de givromètre reçus; celui-ci se réinitialise automatiquement au bout de 1000 impulsions. Pour obtenir le taux d'impulsion horaire, on a dû soustraire la différence entre deux états successifs du compteur pour un intervalle d'une heure (ex: $\text{taux} = 570 (\text{compteur à } 9\text{h}00) - 560 (\text{compteur à } 8\text{h}00) = 10 \text{ impulsions / heure}$).

Une transformation préalable fut également appliquée au signal brut du capteur de force. Pour chaque événement individuel, celui-ci fut décalé d'une valeur constante de telle sorte à enlever exactement un quelconque décalage initial au début de l'événement. Cela fut fait pour homogénéiser les conditions de départ au début de chaque événement. De plus, il semble que sur toute la période (février 1998 - avril 1999) d'où les données furent tirées, le calibrage du capteur de force ait changé une fois. Cette transformation fut appliquée pour combattre aussi un éventuel changement systématique du niveau de base du signal instrumental du capteur de force. Pour ce qui est du taux métrique d'accrétion, nous n'avons pas fait d'ajustement spécifique puisqu'il représente la dérivée temporelle de la charge.

Enfin, chaque variable a été normalisée à l'intervalle de variation unitaire [0..1] par une simple transformation affine. Il s'agit d'une transformation très classique et universellement utilisée pour les réseaux de neurones qui permet d'obtenir une convergence plus rapide et une meilleure adaptation. Cette normalisation assure que les variables d'entrée et les variables de sortie soient de l'ordre de l'unité, auquel cas on s'attend à ce que les poids synaptiques du réseau soient également de l'ordre de l'unité. Les poids peuvent alors recevoir une initialisation aléatoire appropriée avant de commencer l'apprentissage. Sans ce changement d'échelle linéaire, on se trouverait dans l'obligation de trouver une solution pour les poids dans laquelle certains poids auraient des valeurs dramatiquement différentes des autres [3, 21].

4.3 ANALYSE STATISTIQUE

On présente dans cette section une analyse statistique sommaire des données utilisées pour faire l'entraînement des modèles. Cette analyse n'a pas été poussée très loin mais peut donner une idée de la variabilité et de la nature des données et du degré de relation existant entre les variables. La définition de notre problématique n'incluait pas une recherche statistique profonde mais visait plutôt à faire une étude comparative de différents modèles à base de réseaux de neurones.

D'abord, suivent les graphiques montrant l'évolution des différents paramètres (température, vitesse normale du vent, nombre d'impulsions de givromètre, signal du capteur de force, taux d'accrétion). Il faut noter que la délimitation des événements n'est pas identifiée, c'est-à-dire que tous les événements sont juxtaposés. De ce fait, il y a évidemment des discontinuités dans les signaux présentés. On montre l'événement complet et non pas seulement la phase d'accrétion.

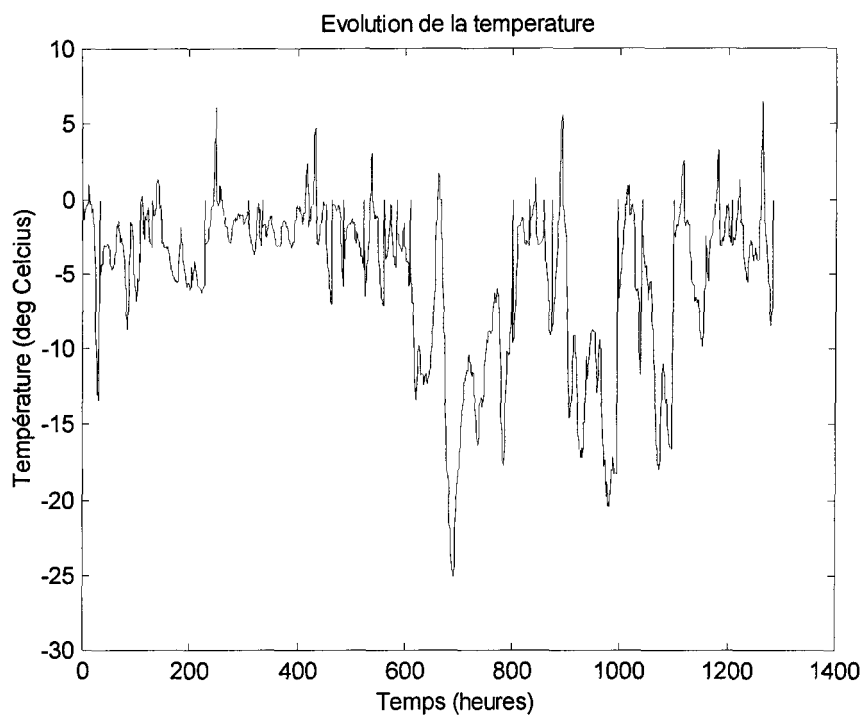


Figure 4.3: Graphique de la température en fonction du temps

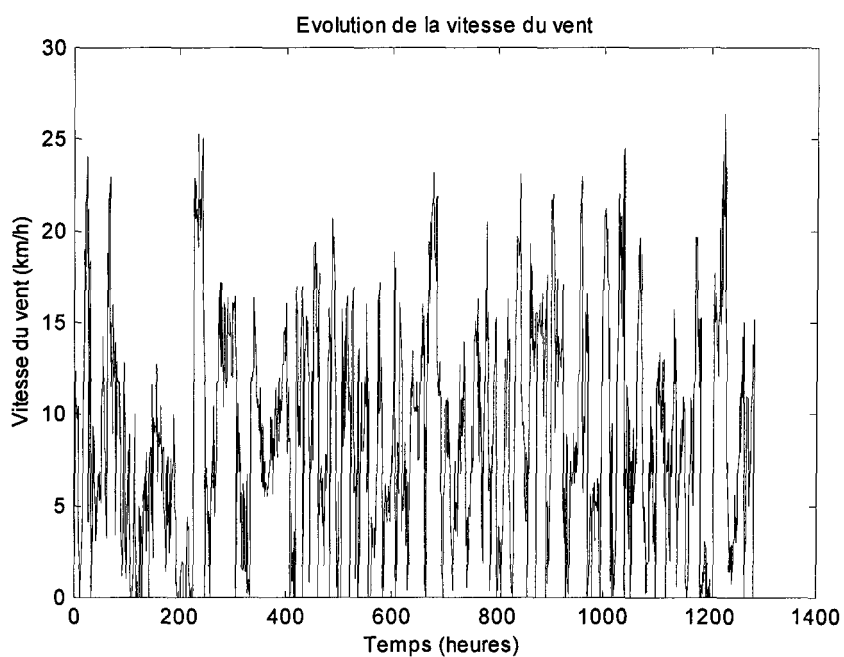


Figure 4.4: Graphique de la vitesse normale du vent en fonction du temps

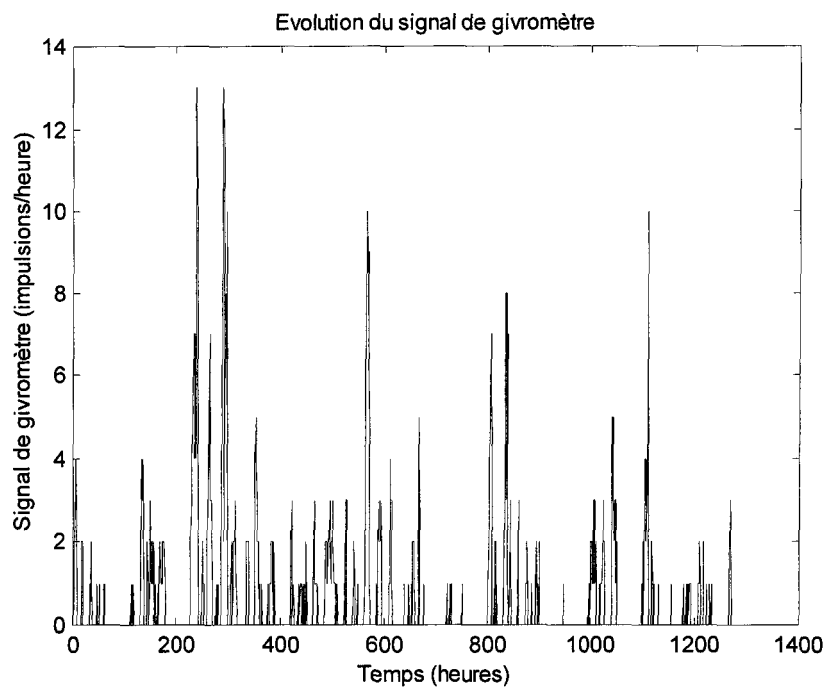


Figure 4.5: Graphique du signal de givromètre en fonction du temps

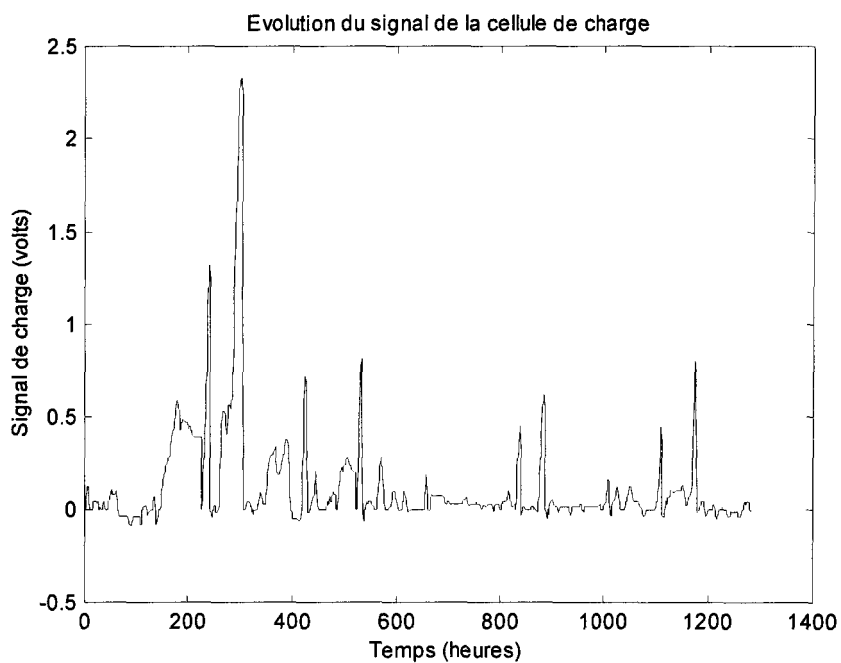


Figure 4.6: Graphique du signal du capteur de force en fonction du temps

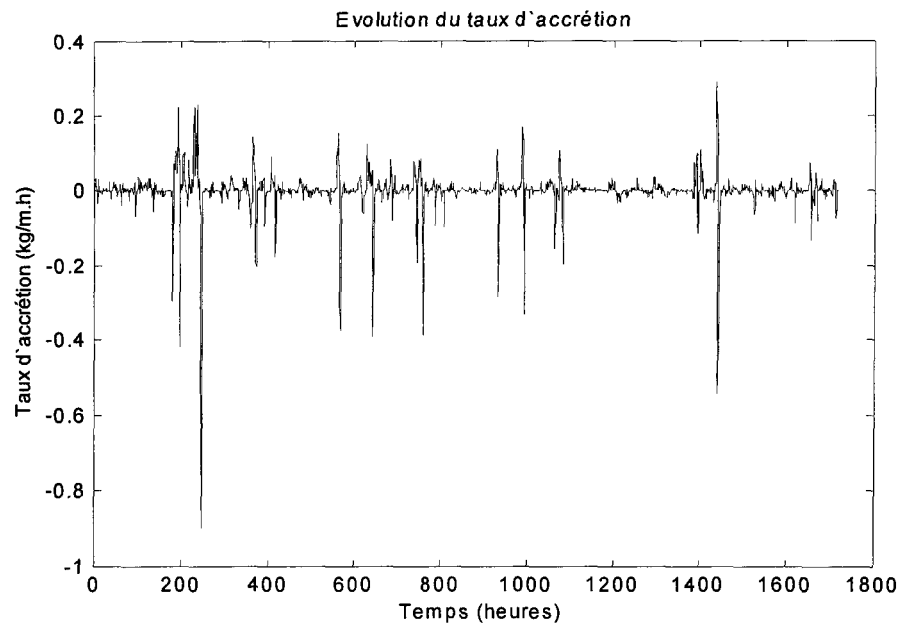


Figure 4.7: Graphique du taux d'accrétion en fonction du temps

Pour avoir une idée globale plus juste du lien statistique unissant les variables, il est commode de représenter les données par un nuage de points. Puisque notre problème possède une dimensionnalité supérieure à 3, nous ne pouvons pas tout présenter dans un seul graphique. Par contre, on peut visualiser le signal de charge ainsi que le taux d'accrétion avec chacune des variables d'entrée formant ainsi six couples, voir les figures 4.8-4.12.

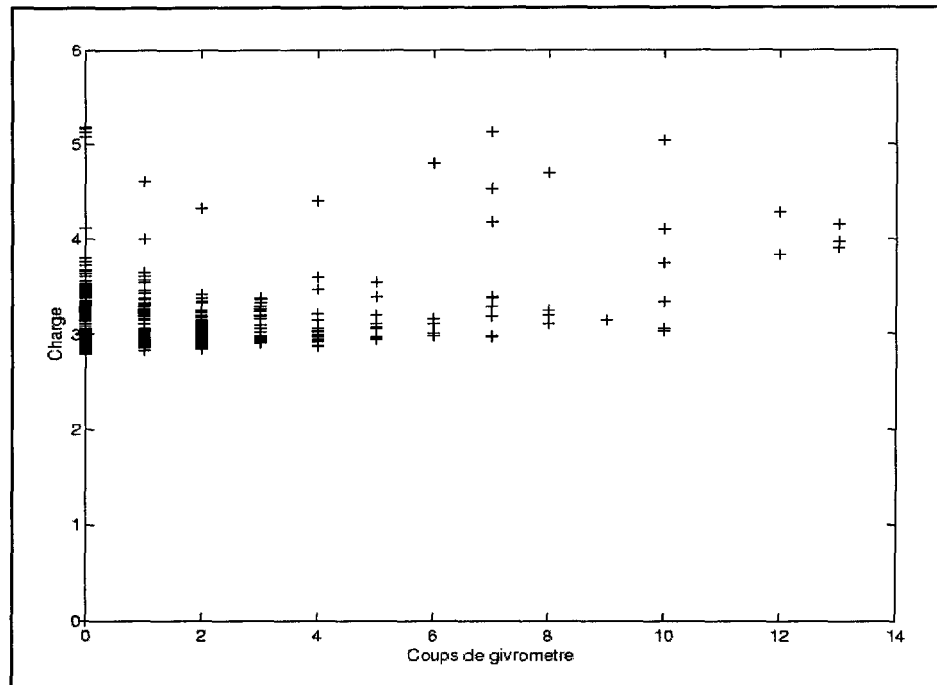


Figure 4.8: Nuage de points du signal de charge avec le signal de givromètre (échelle verticale en volts)

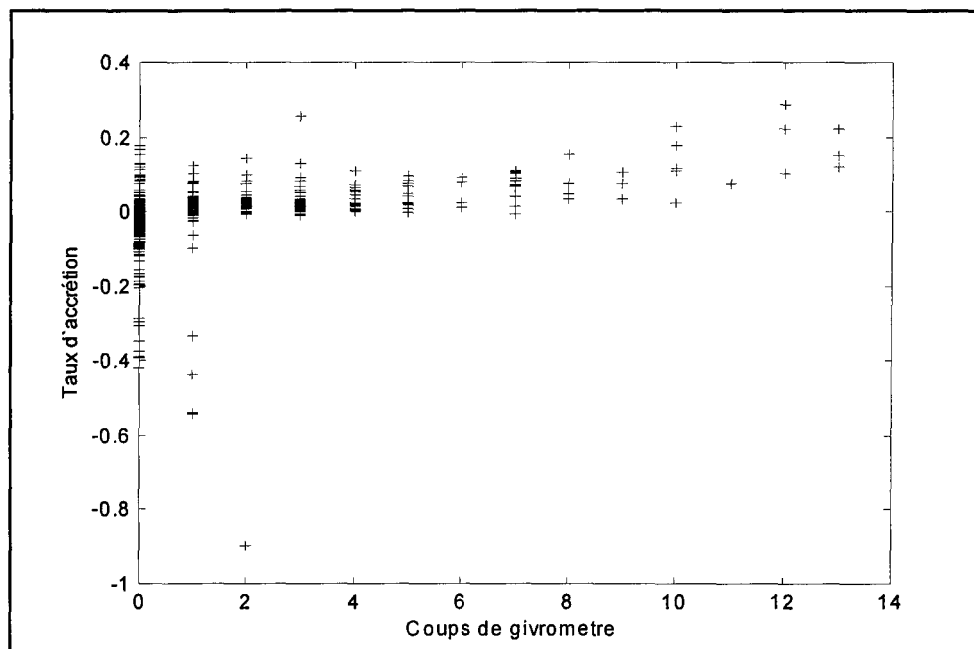


Figure 4.9: Nuage de points du taux d'accrétion avec le signal de givromètre (échelle verticale en kg/m/h)

On peut observer dans les deux figures précédentes que les événements à faible coups de givromètre sont plus fréquents que ceux avec un fort taux. De plus, il n'y a pas de différence substantielle entre le signal du capteur de force et le taux d'accrétion vis-à-vis du givromètre.

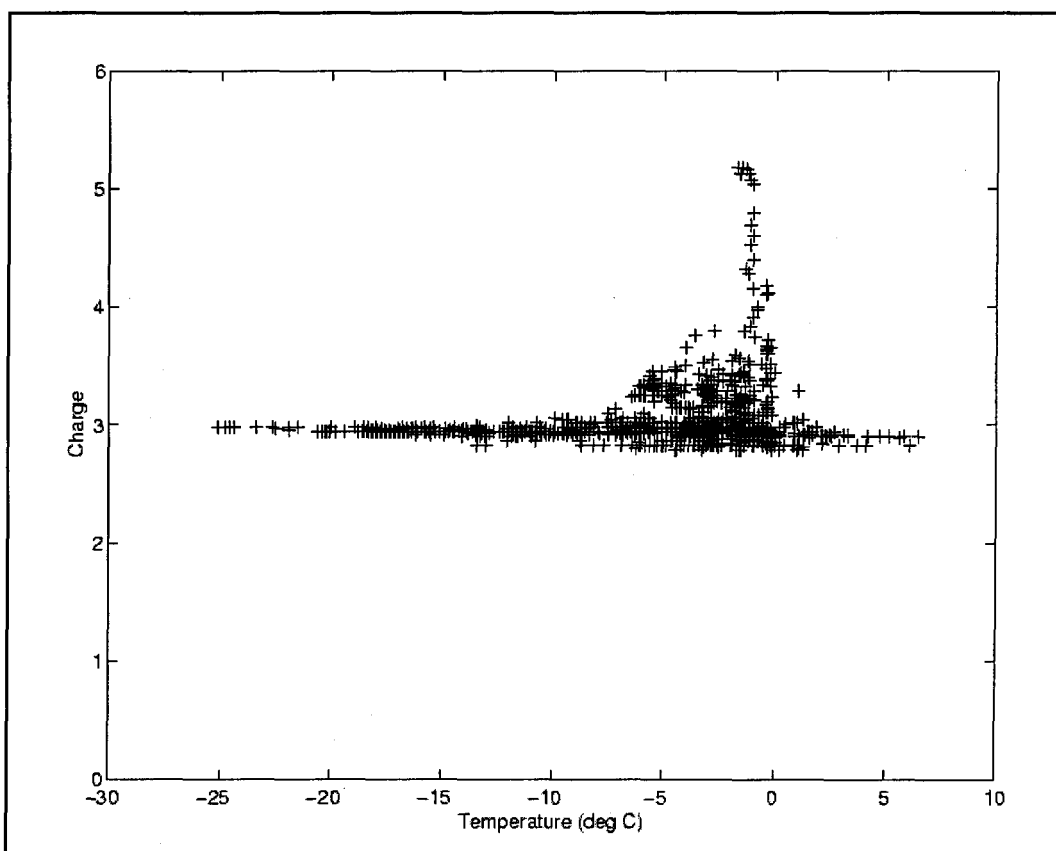


Figure 4.10: Nuage de points du signal de charge avec la température (échelle verticale en volts)

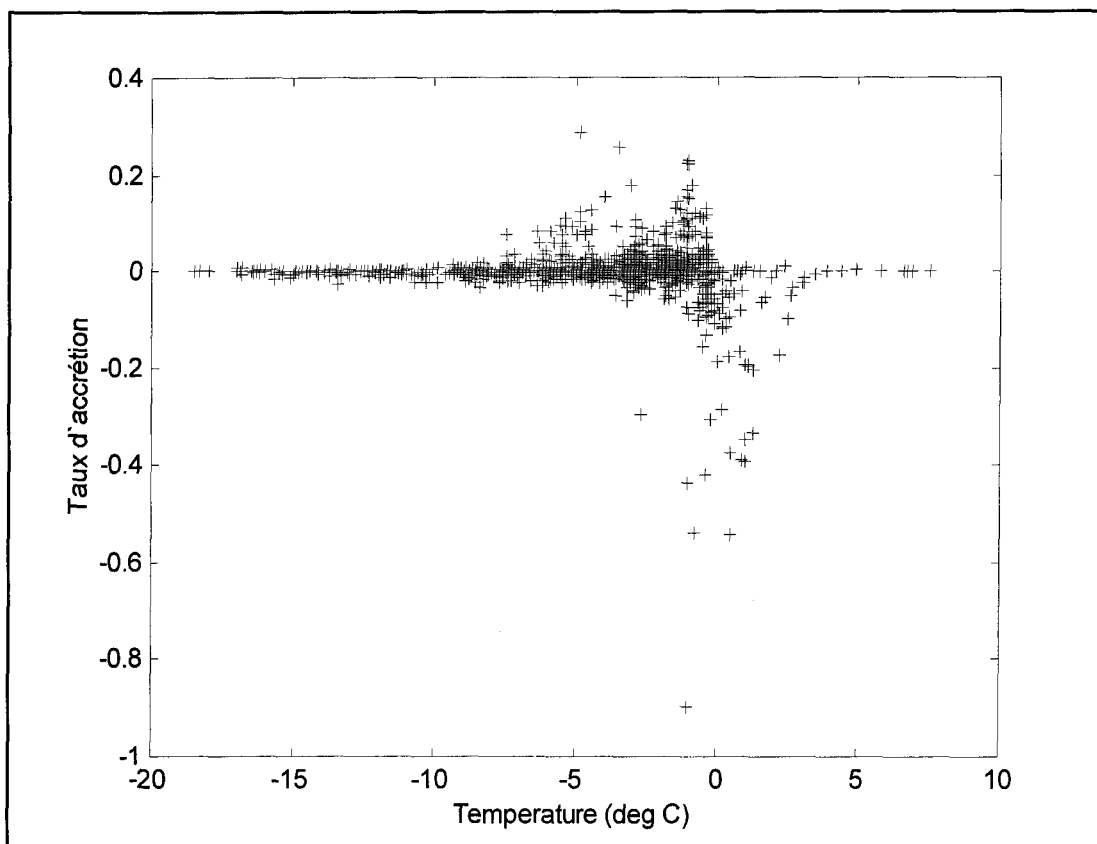


Figure 4.11: Nuage de points du taux d'accrétion avec la température (échelle verticale en kg/m/h)

Aux figures 4.10 et 4.11, on peut noter une très forte concentration de points dans la zone de température entre 0°C et -7°C . Pour des froids plus intenses, il ne semble pas qu'il y ait d'accumulation appréciable puisque le signal de charge demeure à son niveau de base. Ceci se confirme par le taux d'accrétion qui affiche un groupement autour de la valeur nulle pour les mêmes températures. Dans la figure 4.11, on observe un nombre appréciable de points dans la plage négative d'accrétion autour de l'abscisse $T=0$, ceci correspond à du délestage.

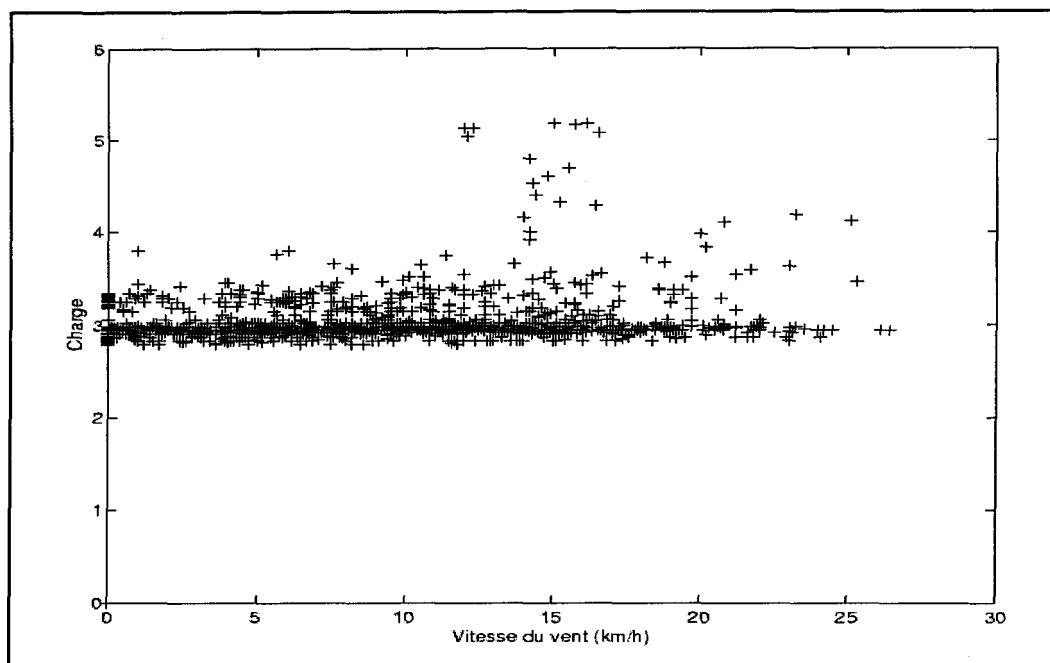


Figure 4.12: Nuage de points du signal de charge avec la vitesse normale du vent (échelle verticale en volts)

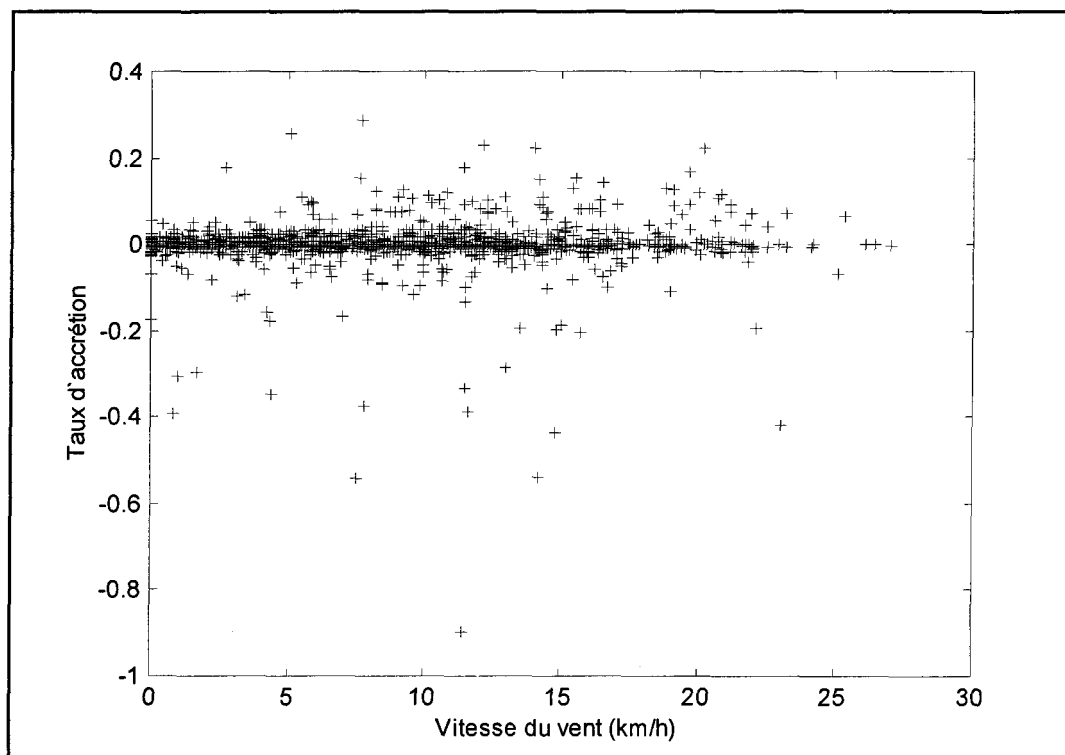


Figure 4.13: Nuage de points du taux d'accrétion avec la vitesse normale du vent (échelle verticale kg/m/h)

Il ne semble pas exister de lien flagrant entre le signal de charge et la vitesse normale du vent (figures 4.12 et 4.13). On peut noter toutefois une valeur "critique" de vitesse normale de vent autour de laquelle il y a une augmentation de la charge, soit 15 km/h. Le taux d'accrétion ne présente pas non plus de lien statistique évident à l'oeil. En fait, l'éparpillement des points semble plutôt indiquer que la vitesse normale du vent soit encore moins corrélée que dans le premier cas.

On peut aussi caractériser la variabilité des paramètres précédents par des statistiques standards. Une information utile à cet égard pour cette étude est la corrélation existant entre les variables qui nous permet d'estimer le degré de déterminisme entre celles-ci; celle-ci est donnée dans les tableaux qui suivent pour les différentes bases de données.

Tableau 4.1			
Matrice de corrélation normalisée des paramètres pour l'ensemble avec la phase d'accrétion et le signal du capteur de force			
	vitesse normale du vent	givromètre	capteur de force
température	-0,047	0,18	0,16
vitesse normale du vent	X	0,15	0,27
givromètre	X	X	0,33
capteur de force	X	X	X

Tableau 4.2			
Matrice de corrélation normalisée des paramètres pour l'ensemble avec l'événement complet et le signal du capteur de force			
	vitesse normale du vent	givromètre	capteur de force
température	0,01	0,23	0,19
vitesse normale du vent	X	0,1	0,14
givromètre	X	X	0,35
capteur de force	X	X	X

On peut se rendre compte que le signal de charge est corrélé avec un certain degré aux trois variables d'entrée en question. Pour ce qui est de la phase d'accrétion, le givromètre et la vitesse normale du vent ont une corrélation presque aussi importante avec la charge (le givromètre possédant une certaine supériorité à cet égard); la température suit en troisième. Pour l'événement complet, c'est le givromètre qui donne une meilleure indication de la charge, la vitesse normale du vent et la température ayant une importance alors comparable.

Malgré ces coefficients non nuls qui montrent une certaine tendance, il faut cependant observer que le degré de corrélation est somme toute encore faible: 35% et moins. Avec ces statistiques, il nous est difficile d'imaginer que l'on puisse établir un modèle entièrement déterministe qui fournirait une prédiction vraiment exacte. On note aussi que le signal de givromètre est corrélé de manière non négligeable à la température et à la vitesse normale du vent.

Tableau 4.3 Matrice de corrélation normalisée des paramètres pour l'ensemble avec la phase d'accrétion et le taux d'accrétion			
	vitesse normale du vent	givromètre	taux d'accrétion
température	-0,05	0,14	0,08
vitesse normale du vent	X	0,32	0,40
givromètre	X	X	0,60
taux d'accrétion	X	X	X

Tableau 4.4 Matrice de corrélation normalisée des paramètres pour l'ensemble avec l'événement complet et le taux d'accrétion			
	vitesse normale du vent	givromètre	taux d'accrétion
température	-0,05	0,17	-0,07
vitesse normale du vent	X	0,16	0,02
givromètre	X	X	0,34
taux d'accrétion	X	X	X

Ici, le taux d'accrétion est fortement corrélé à la réponse du givromètre en phase d'accrétion. Le vent joue un rôle important aussi alors que la température affiche un effet moindre. Dans le cas des événements complets, les corrélations sont moins fortes. De manière surprenante, le vent a une corrélation presque nulle dans ce cas avec le taux d'accrétion. Nous pouvons postuler que cela est attribuable à la phase de délestage qui "brouille" le lien statistique par rapport au premier cas.

4.4 ARCHITECTURES EXPÉRIMENTÉES

On décrit ici les architectures expérimentées lors de ce travail de recherche. Les modèles à base de réseaux de neurones que nous avons implémentés peuvent être placés très globalement dans deux catégories différentes: les architectures statiques qui ne tiennent compte que de l'organisation spatiale mais qui négligent la "composante" temps, les architectures temporelles qui incorporent la dimension temporelle dans le traitement de l'information. Les premières sont plus connues et mieux solidement implantées dans la littérature parce qu'elles appurent à l'origine. Les secondes sont moins bien comprises et ont reçu l'attention des chercheurs depuis une époque plus récente que les architectures statiques. Étant donné leur plus grande complexité, leurs possibilités et leurs limitations sont moins bien circonscrites.

Dans la catégorie des architectures statiques, nous avons implémenté le Perceptron multicouches (MLP: Multilayer Perceptron) et le réseau à fonctions de base radiales (RBF: Radial-Basis Functions network). Dans la catégorie des architectures temporelles, nous avons expérimenté le réseau à réponse impulsionnelle finie (FIR: Finite-Impulse Response) et le réseau récurrent Elman. À titre de comparaison, nous avons aussi inclus dans notre étude le réseau ADALINE qui est en fait équivalent à une régression linéaire multiple.

Commune à toutes ces architectures, le nombre de neurones d'entrée était fixé à trois (nombre de variables d'entrée) et il n'y avait qu'un neurone de sortie (charge ou taux d'accrétion). Le nombre de neurones dans la couche intermédiaire fut varié.

4.4.1 Architectures statiques

4.4.1.1 Perceptron Multicouches (MLP)

C'est le réseau de neurones le plus couramment rencontré. Il comprend des couches multiples (couche d'entrée, une ou plusieurs couches cachées, couche de sortie). Les signaux se propagent de l'entrée vers la sortie (direction de propagation avant). La fonction de transfert est de type sigmoïdale (log-sigmoïde). Pour effectuer l'apprentissage nous avons utilisé une modification de l'algorithme standard de rétropropagation appelé la "rétropropagation avec terme de moment". Le moment permet au réseau de répondre non seulement au gradient local mais aussi aux tendances récentes de la surface d'erreur (hypersurface de la fonction de coût vis-à-vis des paramètres libres du réseau, c'est-à-dire les poids synaptiques). Agissant comme un filtre passe-bas, le moment permet au réseau d'ignorer des petits détails de la surface réduisant dès lors la probabilité que l'algorithme d'apprentissage ne reste coincé dans un minimum local non globalement optimal. Le moment peut être ajouté à l'apprentissage par rétropropagation en gardant une mémoire du changement

de poids le plus récent. L'amplitude de cet effet peut-être sélectionnée par la "constante de moment". Typiquement, elle prend la valeur $mc=0,9$. Nous avons conservé cette valeur standard. Selon le nouvel algorithme, la formule d'apprentissage devient alors:

$$\Delta w_{ij}^l = mc \cdot \Delta w_{ij}^l + (1 - mc) \mu \delta_j^l a_i^{l-1} \quad (1)$$

4.4.1.2 Réseau à Fonctions de Base Radiales (RBF)

Le RBF possède aussi une topologie de type propagation avant à l'instar du Perceptron Multicouches. Le principe du réseau de fonctions à bases radiales dérive de la théorie de l'approximation fonctionnelle. Il s'agit en fait d'une décomposition radiale de l'espace d'entrée. Étant données des paires (x_i, y_i) on cherche une fonction f de la forme suivante:

$$y_i = f(\vec{x}) = \sum_{i=1}^K c_i h(\|\vec{x} - \vec{t}_i\|) \quad (2)$$

où h est une fonction d'interpolation radiale, les \vec{t}_i sont les centres nécessitant une sélection; les coefficients c_i constituent les paramètres libres qui doivent être optimisés; les x_i sont les vecteurs d'entrée. h est appliqué à la distance euclidienne entre les centres d'interpolation et les vecteurs d'entrée. Habituellement, une fonction qui possède un maximum de sensibilité à une distance nulle mais qui décroît

asymptotiquement vers les distances infinies est choisie. Dans notre cas, nous avons fait usage de la fonction gaussienne qui constitue le choix le plus usuel dans ce contexte.

On peut voir que cette forme mathématique peut s'interpréter comme un réseau de neurones dont la couche cachée est constituée de neurones radiaux alors que la couche de sortie est purement linéaire. Dans notre travail, le nombre de neurones dans la couche cachée a été progressivement augmenté un à la fois par un algorithme constructif qui choisit automatiquement les centres appropriés des fonctions à bases radiales. Le nombre maximal de neurones cachés permis à l'algorithme a été choisi comme paramètre de conception. Il faut aussi noter qu'un autre paramètre ajustable était disponible dans l'algorithme qu'on appelle la "constante d'échelle". Elle permet de fixer la taille effective de l'espace couvert par chaque fonction de base; conceptuellement, elle s'assimile à un rayon au sens géométrique. Nous avons choisi 0,1 comme valeur qui nous semblait sensée étant donné que les variables avaient été préalablement normalisées à l'intervalle unitaire [0..1].

4.4.2 Architectures dépendantes du temps

4.4.2.1 Réseau à Réponse Impulsionnelle Finie (FIR)

Ce réseau fut introduit par Wan [58-60] en 1993. Il ressemble au Perceptron multicouches à propagation avant standard excepté que chaque synapse est remplacée par un filtre adaptatif FIR linéaire tel qu'illustré à la figure 4.14. Le filtre FIR forme une somme pondérée des valeurs passées des entrées.

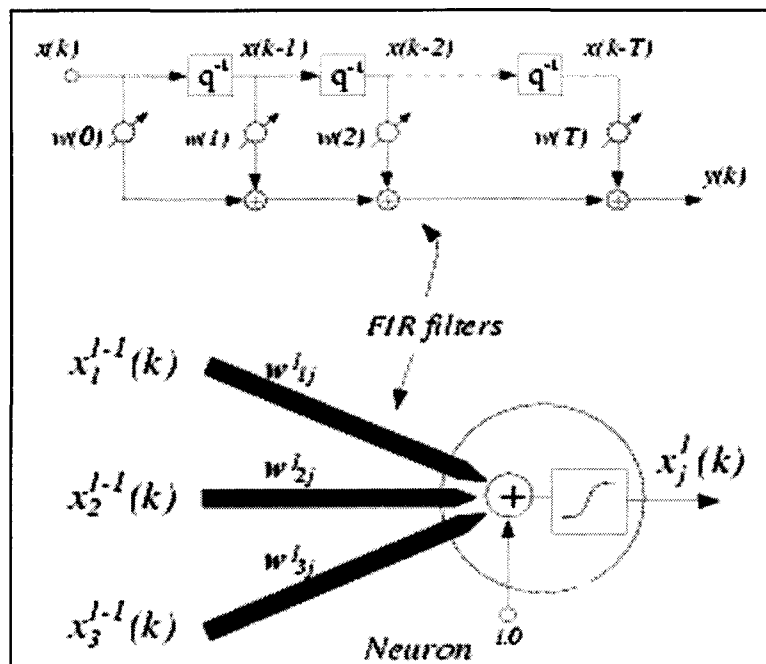


Figure 4.14: Architecture du réseau FIR (tiré de ref. [59])

Le corps du neurone reçoit les signaux filtrés et passe la somme à travers une fonction non-linéaire limitante. Les neurones sont arrangés en couches successives avec des connexions de filtres synaptiques. Les coefficients des filtres FIR constituent les paramètres libres du système que l'on doit optimiser pour minimiser l'erreur de prédiction. L'entraînement du réseau s'effectue par l'intermédiaire d'une modification de l'algorithme de rétropropagation standard appelée *la rétropropagation temporelle* dans laquelle les termes d'erreur sont symétriquement rétro-filtrés dans la direction inverse à travers le réseau. Il s'agit en fait d'une extension naturelle au domaine temporel de l'algorithme standard.

On peut concevoir ainsi que ce réseau remarquable effectue un traitement temporel complexe d'une manière complètement distribuée par le truchement des filtres FIR se trouvant à chaque synapse. La prise en compte de la dimension temporelle se trouve contenue dans la structure intrinsèquement globale du réseau et non seulement par un quelconque artifice d'entrée.

Dans notre expérience, le réseau ne possédait qu'une seule couche cachée. Pour les deux premières couches (couche d'entrée, couche cachée), la fonction de transfert utilisée était la tangente hyperbolique alors que les neurones de sortie utilisaient une fonction linéaire. Ceci constitue le choix le plus approprié pour une tâche de prédiction régressive par opposition à une tâche de classification et/ou de catégorisation. L'ordre de délai des filtres FIR fut choisi à 3:2 (ordre 3 à l'entrée,

ordre 2 pour la couche cachée). Ces paramètres ont été sélectionnés quelque peu empiriquement. On doit comprendre toutefois qu'il est inutile de fixer la profondeur du délai des filtres à un degré beaucoup plus élevé que les corrélations croisées existant entre les variables. Les valeurs passées lointaines ont vraisemblablement moins d'impact prédictif que les valeurs plus récentes. Étant donné la nature volatile de notre problématique (tempête de givre), nous avons considéré qu'un délai effectif de $5=3+2$ pas de temps (5 heures) était entièrement suffisant. De plus, puisque les événements de givre atmosphérique ont une certaine délimitation dans le temps, nous ne pouvions pas nous permettre de prendre des ordres de délai sensiblement plus élevés.

4.4.2.2 Réseau récurrent d'Elman

Elman [11] a proposé en 1990 une architecture de réseau de neurones récurrente pour la tâche de prédiction de phonèmes. Le réseau d'Elman comprend des connexions récurrentes des neurones cachés vers une couche d'unités contextuelles constituées de délais unitaires. Ces unités contextuelles emmagasinent les sorties des neurones cachés pour un pas de temps et alimentent les neurones d'entrée avec celles-ci. Les neurones cachés alimentent aussi la couche de sortie.

Le rôle des unités contextuelles est de fournir au réseau une mémoire dynamique qui est évidemment utile pour une tâche de prédiction. Selon Haykin [21] et Hertz [22], grâce à la présence de rétroaction, les neurones cachés continuent à recycler l'information dans le réseau à travers de multiples pas de temps et peuvent dès lors découvrir des représentations abstraites de l'information temporelle. Ce réseau récurrent ne fait donc pas une copie conforme simpliste des données du passé mais possède la faculté d'effectuer un traitement non-trivial.

Dans notre expérience, la couche cachée avait une fonction de transfert sigmoïdale alors que la couche de sortie était linéaire. L'algorithme d'apprentissage suit celui décrit par Elman.

CHAPITRE 5

RÉSULTATS EXPÉRIMENTAUX

5.1 INTRODUCTION

On présente maintenant les résultats obtenus avec les diverses architectures. Toutes les architectures ont été uniformément entraînées et validées sur les mêmes ensembles de données tels que décrits dans le chapitre 4. On rappellera qu'il y avait une base de données pour un événement de givre complet et une autre pour la phase d'accrétion uniquement. De plus, il y avait une base de données contenant le signal instrumental du capteur de force direct et une autre qui contenait le taux d'accrétion. En accord avec la méthodologie des réseaux de neurones, chaque base a été divisée en un ensemble d'apprentissage et un ensemble de validation. On a utilisé 75% des points pour faire l'entraînement et les points restants pour la validation.

Pour chaque architecture, nous avons varié les paramètres de conception pour observer l'effet sur la performance de prédiction. Dans le cadre de ce travail, nous avons étudié l'impact du nombre de neurones dans la couche cachée. C'est le paramètre par excellence qui permet de sélectionner à volonté la complexité intrinsèque du modèle implémenté par une architecture de réseaux de neurones donnée.

5.2 CRITÈRE D'ERREUR

Pour comparer les architectures entre elles, il est nécessaire de définir un critère d'erreur ou de performance de prédiction. L'entraînement proprement dit des réseaux de neurones nécessite également de faire un tel choix. Usuellement, la fonction de coût la plus utilisée pour faire l'apprentissage est la somme quadratique des termes d'erreur. D'une manière analogue, il est raisonnable de choisir dans ce cas comme fonction de performance de prédiction la racine carrée de l'erreur quadratique moyenne ("*Root Mean Square Error*"). C'est le choix le plus souvent utilisé.

Cependant, à notre sens, ce critère d'erreur possède un certain défaut. Puisqu'il s'exprime dans les mêmes unités que la sortie, son ordre de grandeur peut varier. Nous avons opté pour un critère plus "stable" qu'on appelle le "*Normalized Mean Square Error*" (NMSE). Ce critère fut retenu officiellement dans un workshop international sur l'analyse des séries temporelles [61]. Il s'agit en fait du rapport de la variance de prédiction sur la variance naturelle de la variable-cible à prédire. Une valeur de NMSE=1 correspond à la prédiction de la moyenne statistique. En termes mathématiques, on le définit ainsi:

$$NMSE = \frac{1}{\sigma^2} \frac{1}{N-1} \sum (y_i - \hat{y}_i)^2 \quad (1)$$

5.3 VARIATION DES PARAMÈTRES

Suivent les résultats obtenus pour la variation du nombre de neurones cachés pour chacune des architectures expérimentées telles que décrites précédemment dans le chapitre 4. Nous présentons les résultats pour chaque architecture individuellement. De plus, nous montrons autant les résultats pour l'ensemble de validation que pour l'ensemble d'apprentissage.

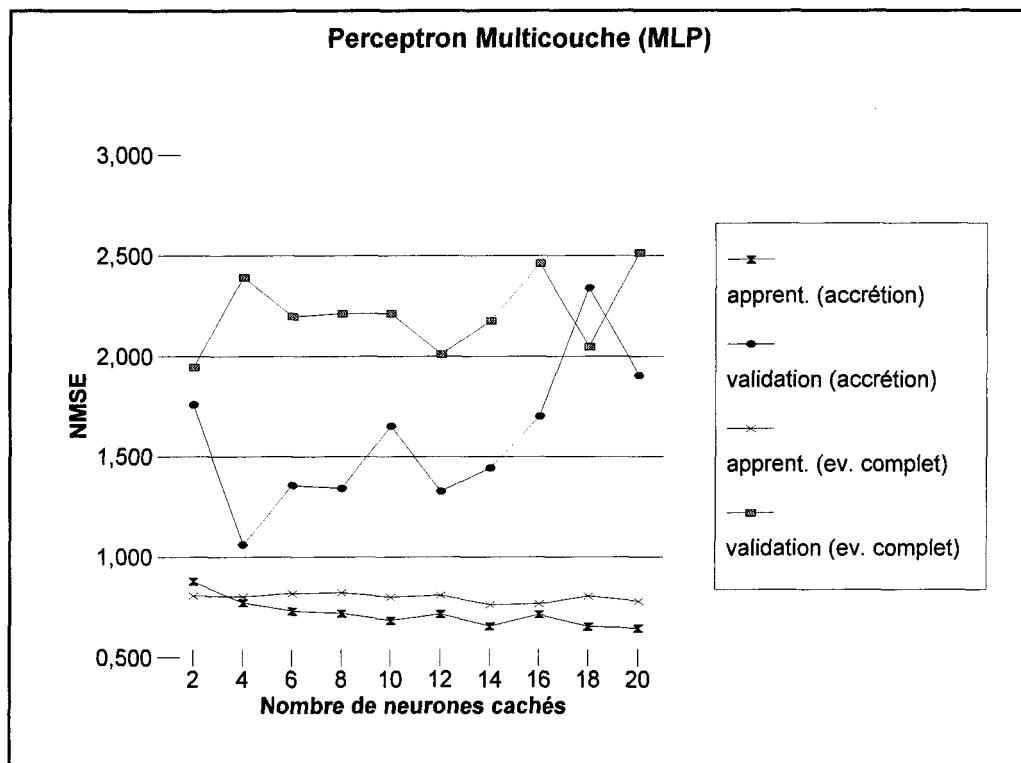


Figure 5.1: Effet du nombre de neurones cachés pour le Perceptron Multicouche avec le signal du capteur de force

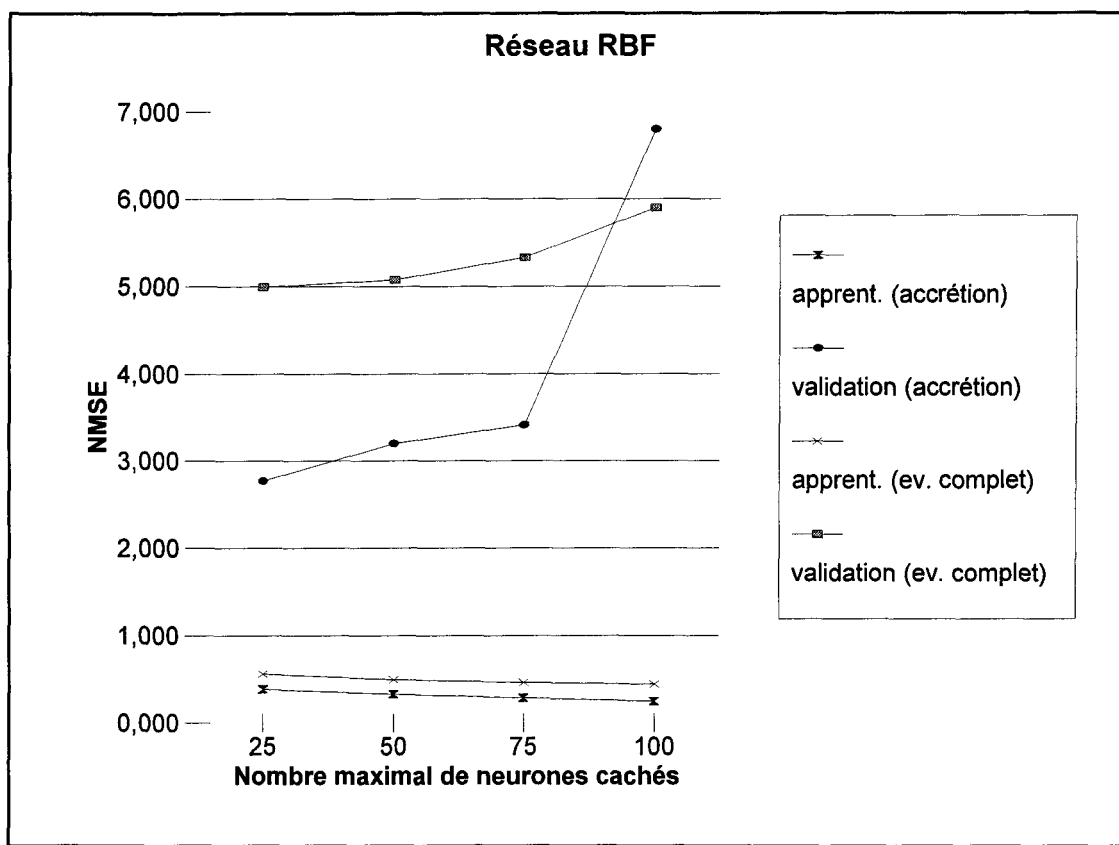


Figure 5.2: Effet du nombre de neurones cachés pour le réseau RBF avec le signal du capteur de force

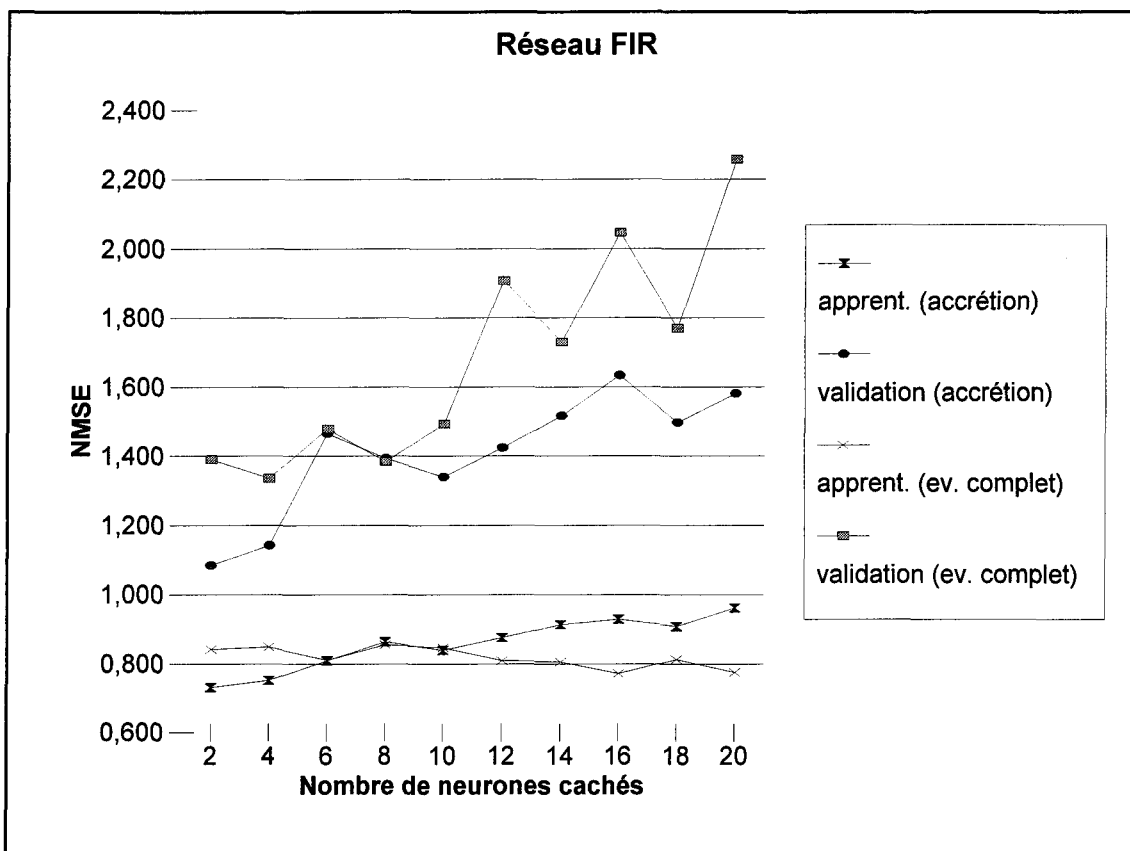


Figure 5.3: Effet du nombre de neurones cachés pour le réseau FIR avec le signal du capteur de force

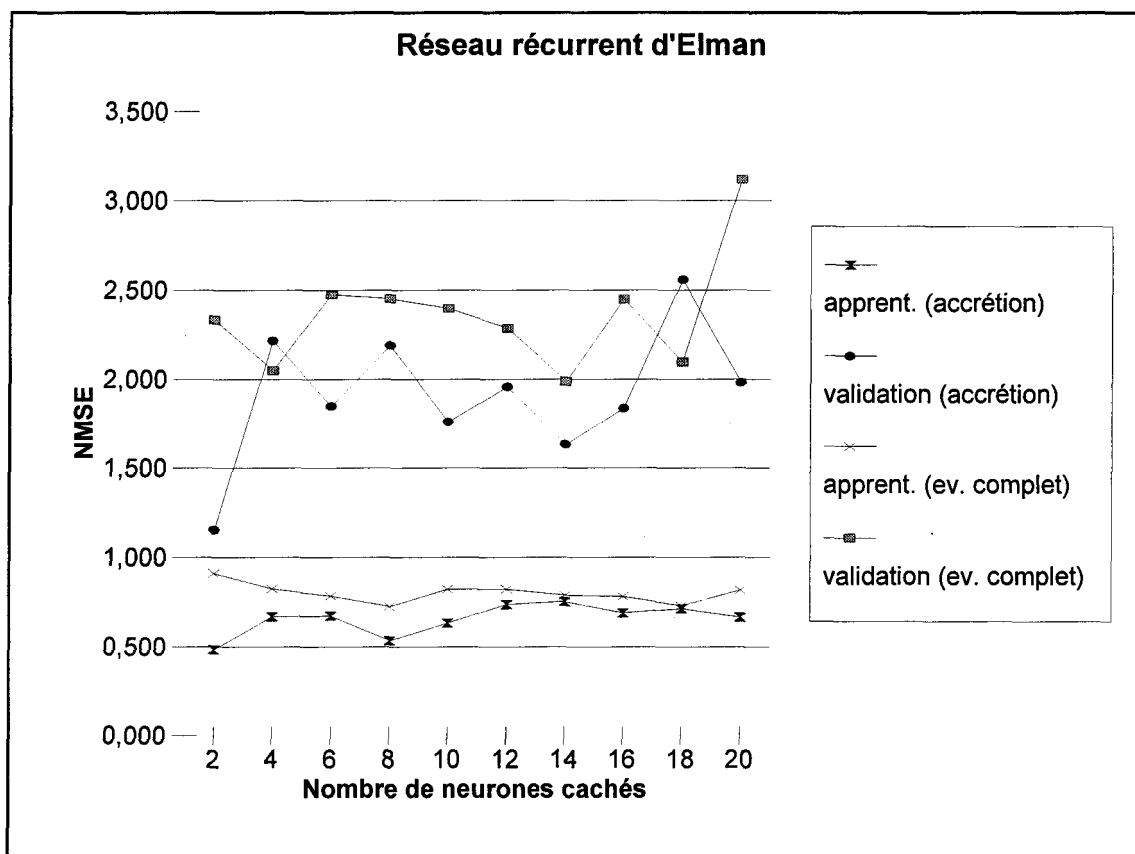


Figure 5.4: Effet du nombre de neurones cachés pour le réseau d'Elman avec le signal du capteur de force

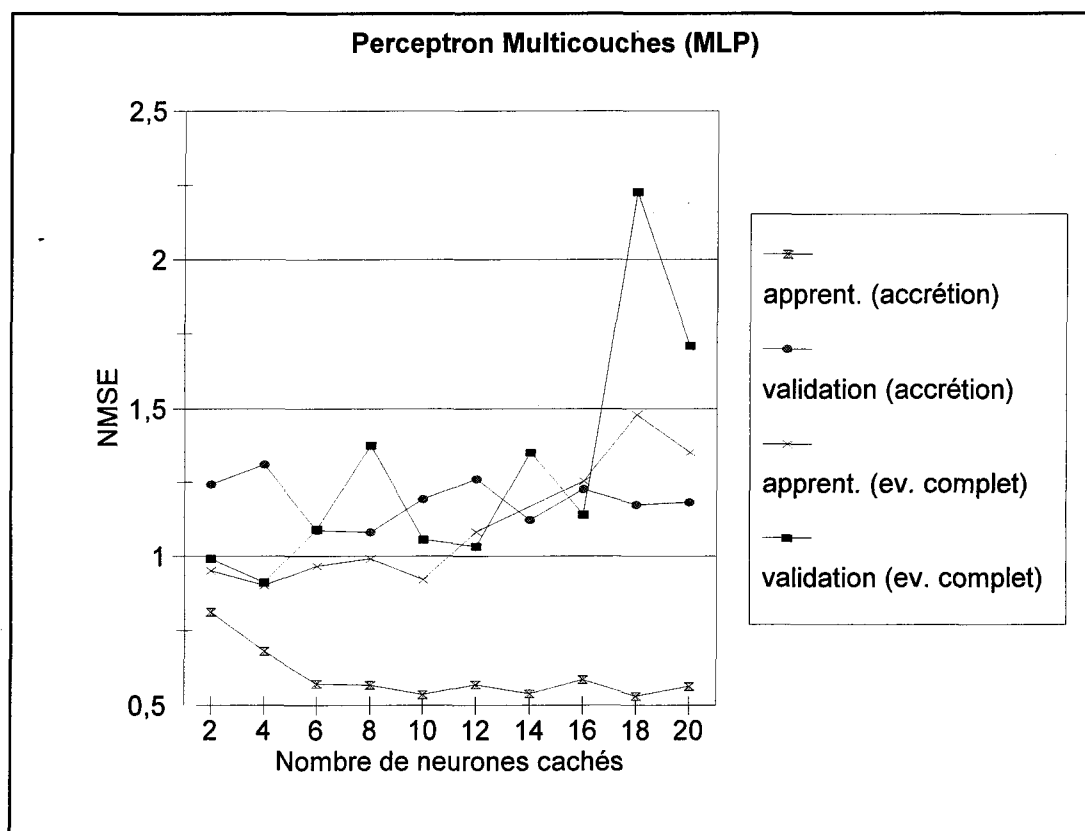


Figure 5.5: Effet du nombre de neurones cachés pour le Perceptron Multicouches avec le taux d'accrétion

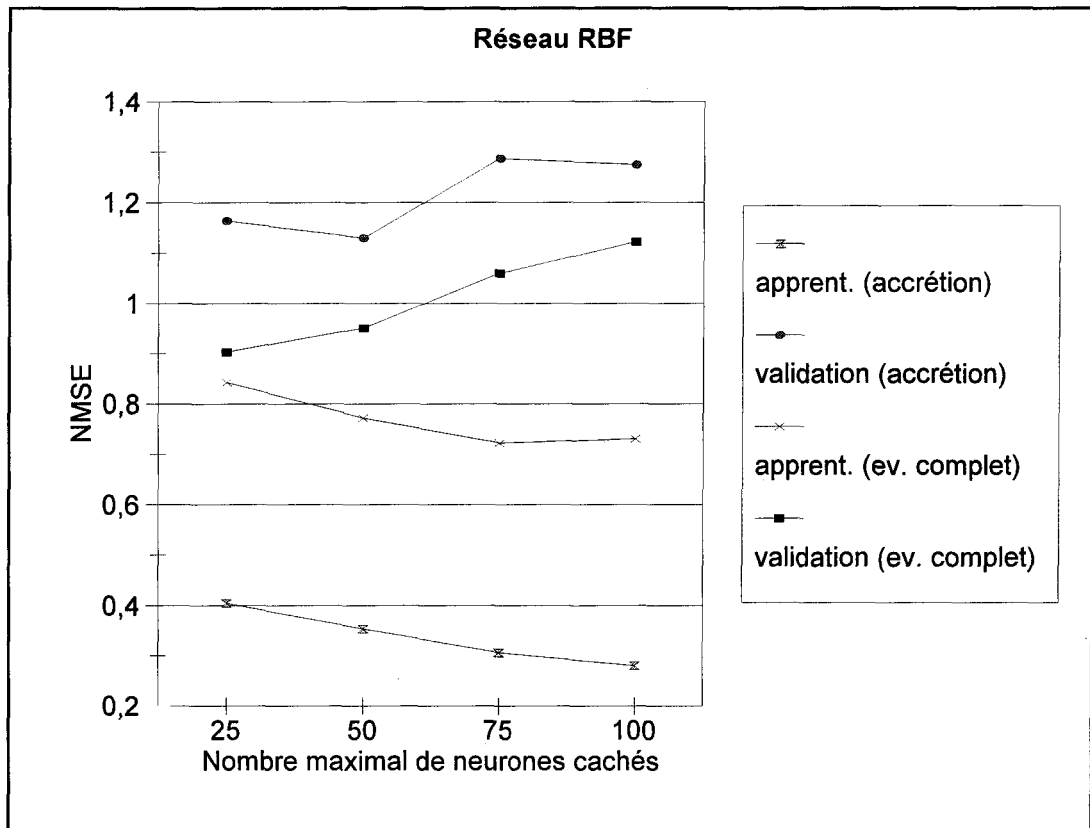


Figure 5.6: Effet du nombre de neurones cachés pour le réseau RBF avec le taux d'accrétion

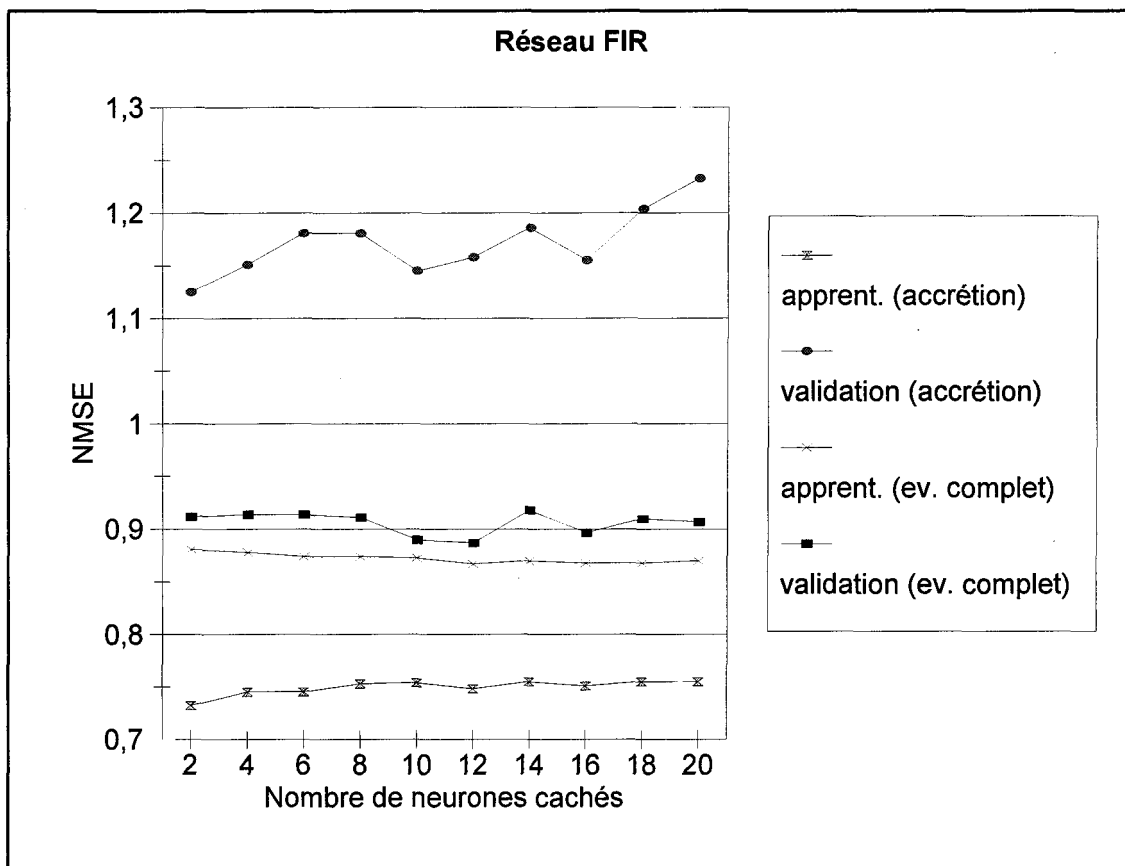


Figure 5.7: Effet du nombre de neurones cachés pour le réseau FIR avec le taux d'accrétion

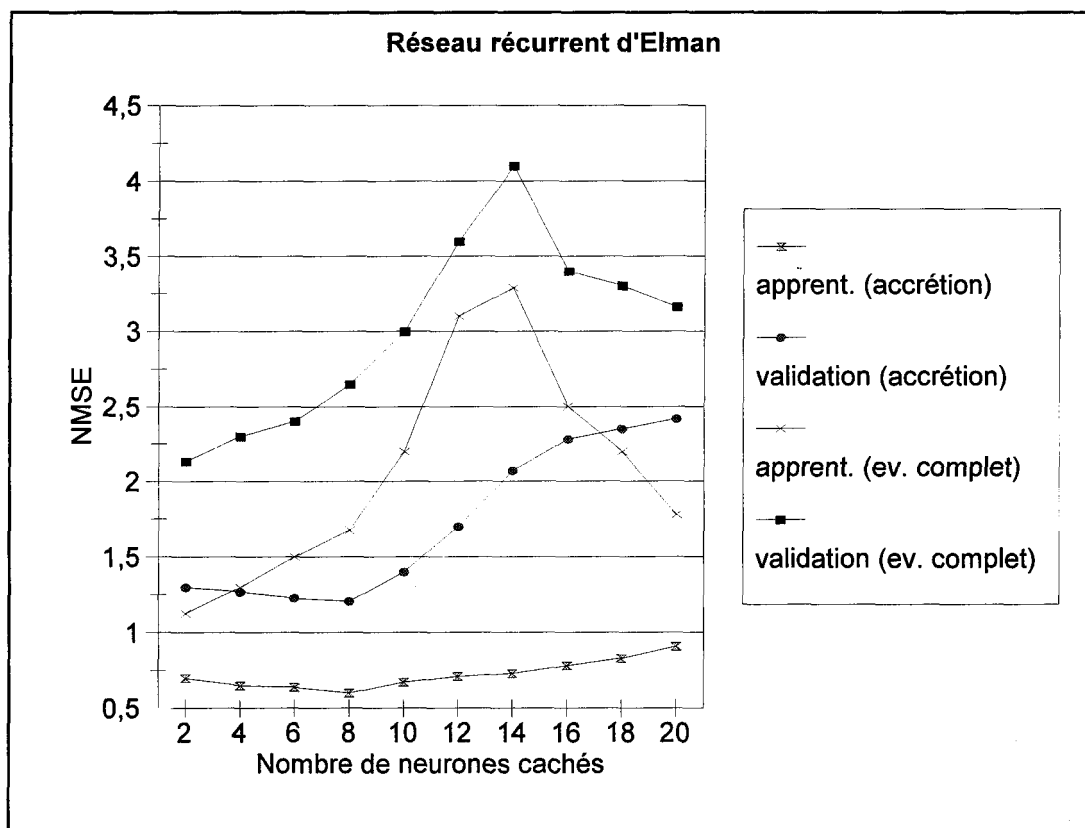


Figure 5.8: Effet du nombre de neurones cachés pour le réseau d'Elman avec le taux d'accrétion

On peut observer deux faits notables qui ne peuvent trop surprendre. D'abord, les réseaux ont plus de facilité à prédire les données d'apprentissage que les données de validation. C'est un résultat somme toute évident. Toutefois, la démarcation relativement visible sur les graphiques précédents entre ces deux ensembles indique que les réseaux avaient quelque peu de difficulté à généraliser. Une deuxième observation qui ressort systématiquement est la meilleure performance des réseaux sur l'ensemble incorporant uniquement la phase d'accrétion que l'événement de givrage complet. Ceci corrobore le fait attendu que les données d'accrétion présentent une plus grande homogénéité au sens statistique puisqu'elles font appel à un seul processus physique. Par contraste, un événement de givrage complet implique des processus variés à travers les différentes phases qui le composent (accrétion, persistance, délestage). D'une manière générale, les réseaux de neurones ont de la difficulté à traiter des données hétérogènes provenant de phénomènes aux mécanismes distincts. Si on leur présente des données hétéroclites comportant une grande variabilité statistique, ils ont tendance à afficher un comportement qui se situe dans la moyenne.

On peut aussi constater dans certains cas une dégradation de la performance avec l'augmentation du nombre de neurones cachés, particulièrement pour les ensembles de validation. Bien que difficile à expliquer, on peut postuler que l'utilisation de modèles d'ordre élevé causait un problème de généralisation (surentraînement).

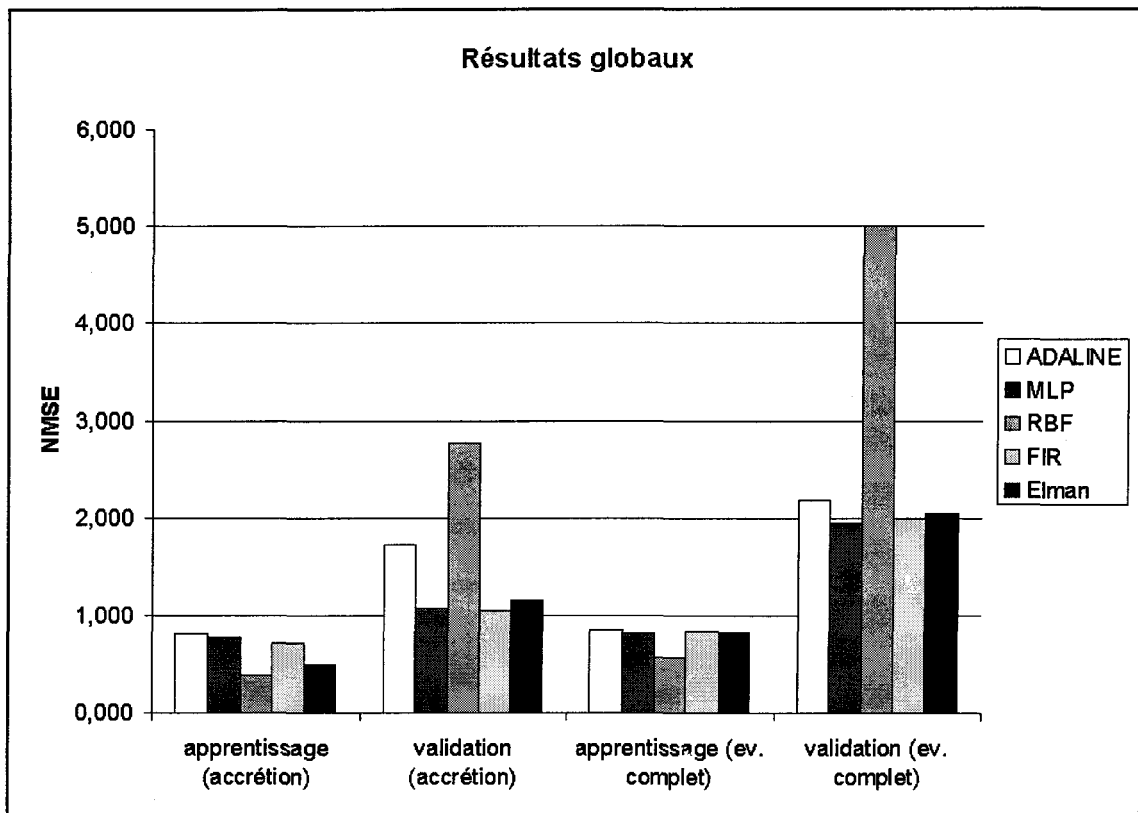


Figure 5.9: Résumé des résultats de prédiction avec le signal du capteur de force

Pour chaque type de réseau, l'architecture donnant la meilleure performance telle que mesurée sur l'ensemble de validation a été retenue. Une comparaison globale des réseaux de neurones se trouve résumée aux figures 5.9-5.10. Dans le but d'établir une ligne de base de comparaison, le réseau de neurone linéaire ADALINE fut aussi inclus. Bien qu'employant un algorithme itératif adaptatif (LMS), il est entièrement équivalent à une régression linéaire multiple [21].

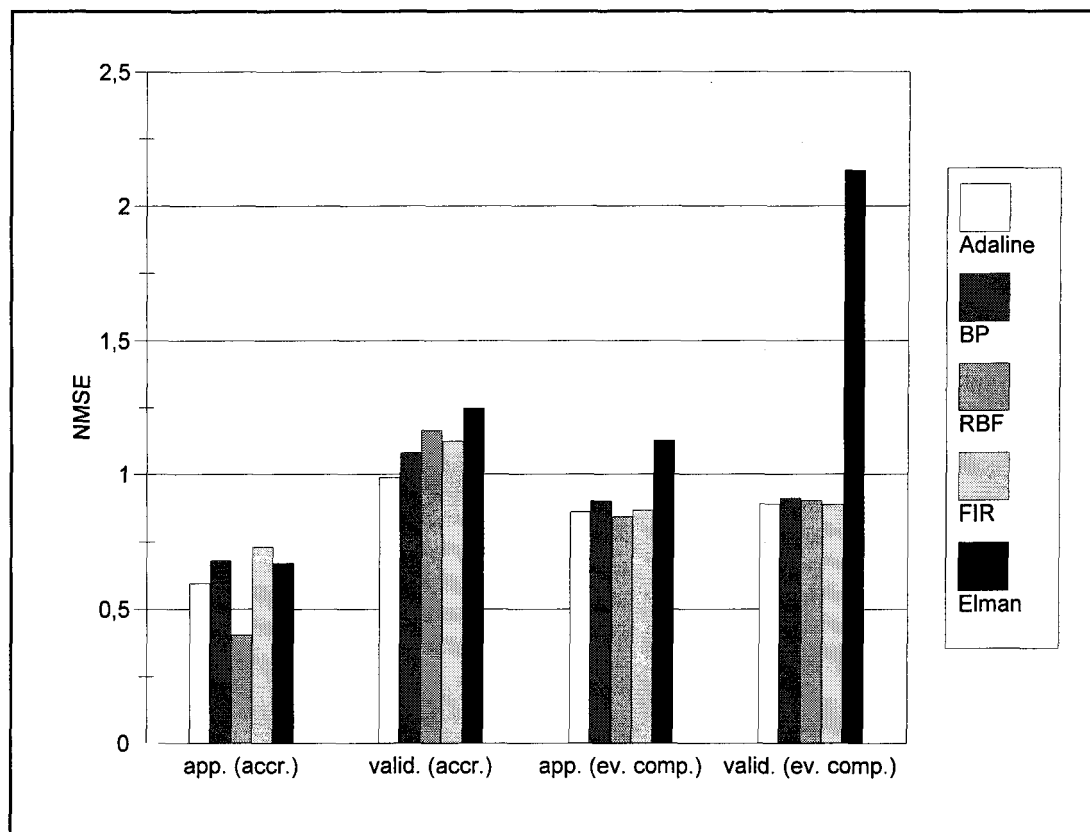


Figure 5.10: Résumé des résultats de prédiction avec le taux d'accrétion

Dans le cas de la figure 5.9 qui concerne les tests avec le signal du capteur de force, le réseau FIR démontre une performance globale comparable au MLP telle que mesurée sur les ensembles de validation. Il est suivi par le réseau d'Elman dans le cas de la phase d'accrétion. On doit garder cependant à l'esprit qu'un critère synthétique comme le NMSE ne peut être considéré comme parfait selon tous les points de vue. On peut aussi noter qu'il y a plus de différence entre les modèles neuronaux dans les données concernant l'événement complet de givrage que pour les données de phase d'accrétion (pour l'ensemble de validation). De plus, les modèles neuronaux font généralement moins d'erreur que la régression linéaire; peut-être marginalement pour les ensembles d'apprentissage mais la différence est plus prononcée dans les ensembles de validation. Ceci donne une certaine indication de l'utilité des modèles neuronaux par rapport aux techniques linéaires simplistes. Les ensembles d'apprentissage présentent comparativement un comportement moins démarqué. Il semble que la plupart des architectures apprennent avec une performance comparable pour la base d'apprentissage.

La figure 5.10 concerne les tests avec le taux d'accrétion comme signal de sortie. On doit concevoir que cette variable n'a pas la même nature que le signal du capteur de force direct mais qu'elle lui est reliée par une opération de différentiation par rapport au temps. Ce changement de variable a bien sûr modifié les propriétés des données et influencé le comportement des réseaux.

Dans le cas des ensembles avec événement complet, on peut se rendre compte qu'il y a très peu de démarcation entre les différentes architectures. Cette observation peut s'appliquer autant à l'ensemble d'apprentissage que de validation. Il est dès lors plutôt hasardeux de déclarer une architecture comme étant notoirement supérieure à une autre. Toutefois, on peut aussi noter que les réseaux performant mieux de manière générale qu'avec le signal du capteur de force (excepté l'architecture d'Elman qui présente une difficulté singulière). Les ensembles ne comportant que la phase d'accrétion montrent un comportement moins uniforme. Paradoxalement, les réseaux de neurones avaient alors dans cette situation plus de difficulté à effectuer une prédiction. Telle que mesurée sur l'ensemble de validation en phase d'accrétion, la performance globale affiche un avantage marginal par rapport aux modèles utilisant le signal du capteur de force comme sortie. On peut toutefois remarquer que les modèles performaient mieux avec l'ensemble d'apprentissage. La différence est beaucoup moins notable dans le cas des ensembles avec événement complet. La question peut surgir de la robustesse des comparaisons précédentes. En faisant des cycles répétés d'entraînement/validation pour chacune des architectures, nous évaluons à environ 0,1 en absolu l'incertitude sur la performance des architectures selon le critère NMSE. Cette incertitude a été évaluée à partir du degré de variabilité des indices de performance. À l'intérieur de cette fourchette, il est plutôt malaisé de faire une comparaison vraiment objective lorsque les comportements des architectures sont très semblables.

5.4 COURBES DE PRÉDICTION

Comme noté précédemment, le réseau FIR présente une performance comparable pour le signal du capteur de force avec le MLP. Pour donner une idée plus précise de son comportement effectif, les courbes de prédiction sur les données de validation de l'ensemble de la phase d'accrétion sont présentées à la figure 5.13. Les réseaux ADALINE et MLP sont aussi inclus à titre de comparaison (figs. 5.11-12).

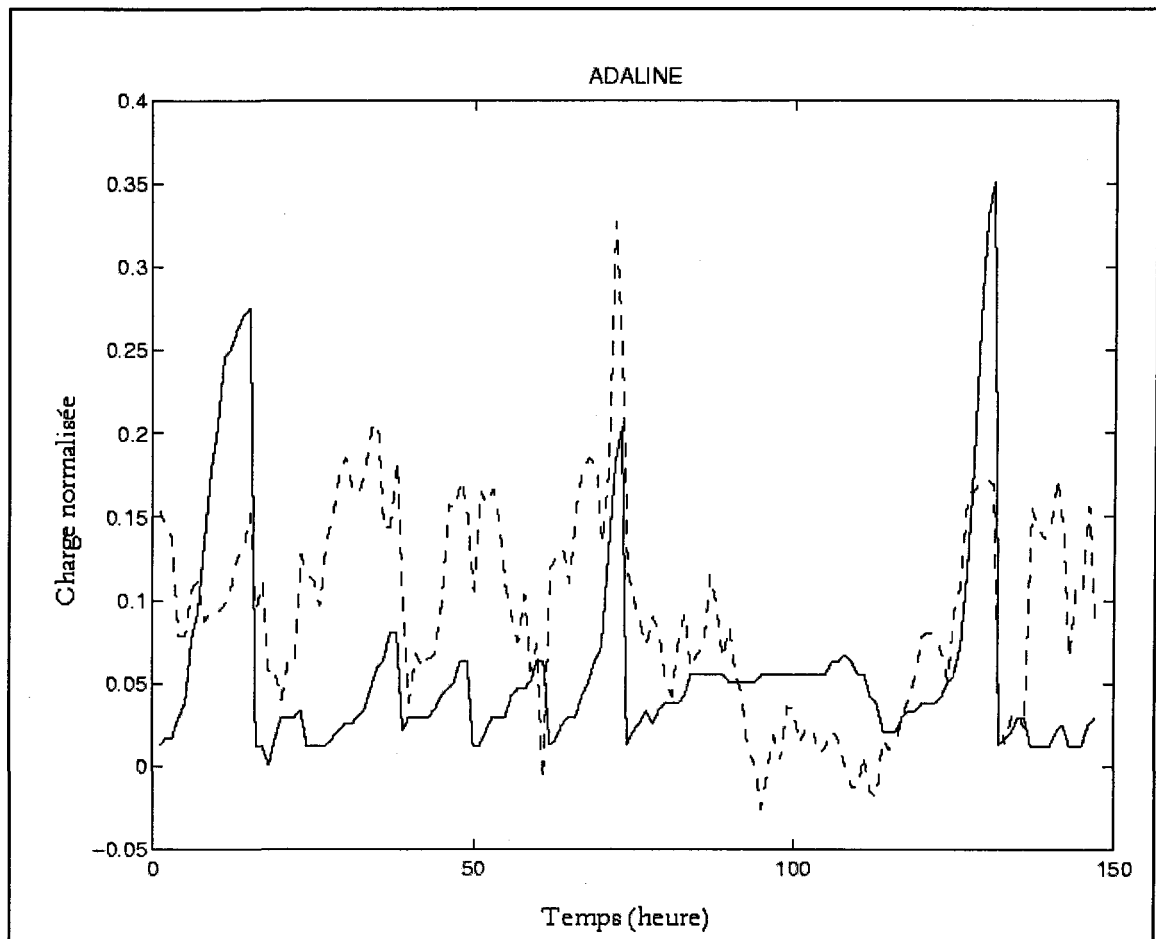


Figure 5.11: Courbes de prédiction pour l'ADALINE avec le signal du capteur de force en phase d'accrétion (courbe solide: donnée réelle; courbe pointillée: prédiction)

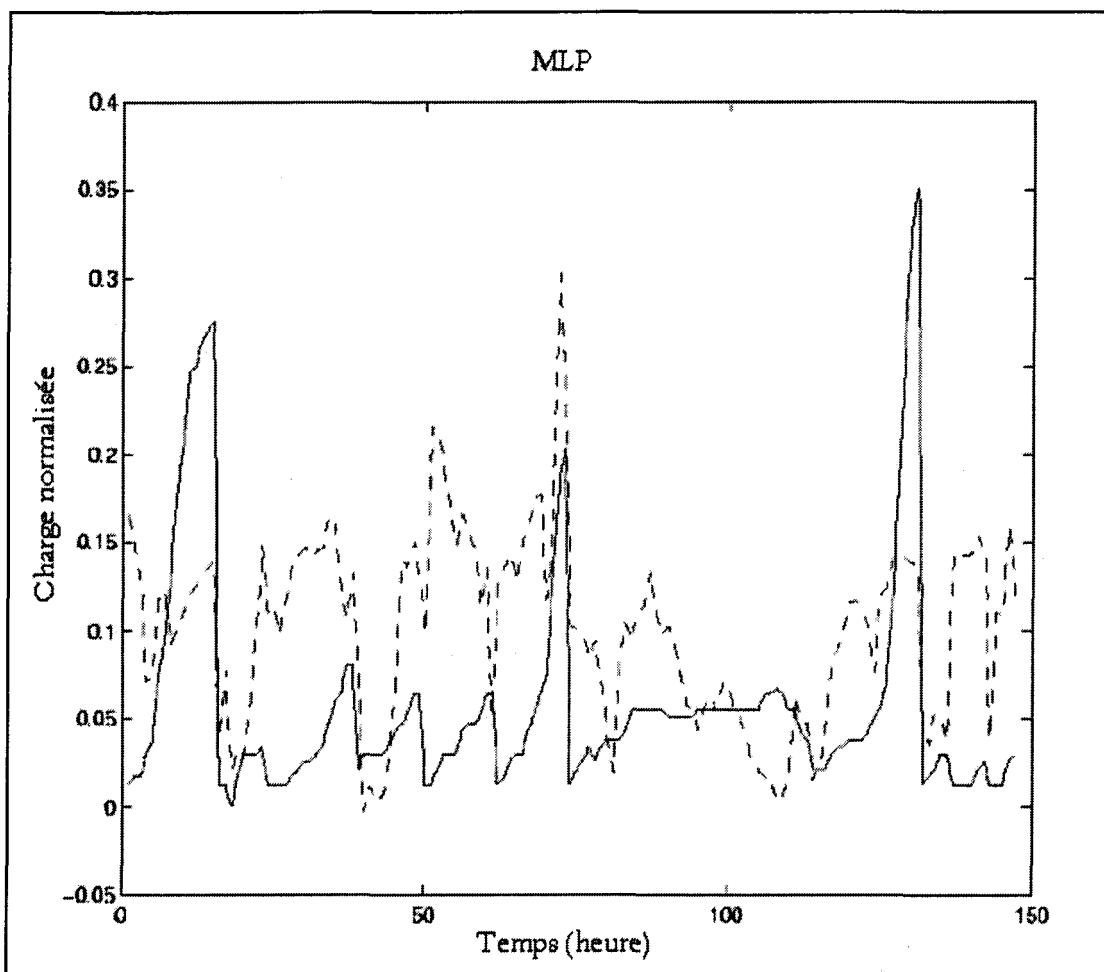


Figure 5.12: Courbes de prédiction pour le Perceptron Multicouches avec le signal du capteur de force en phase d'accrétion (courbe solide: donnée réelle; courbe pointillée: prédiction)

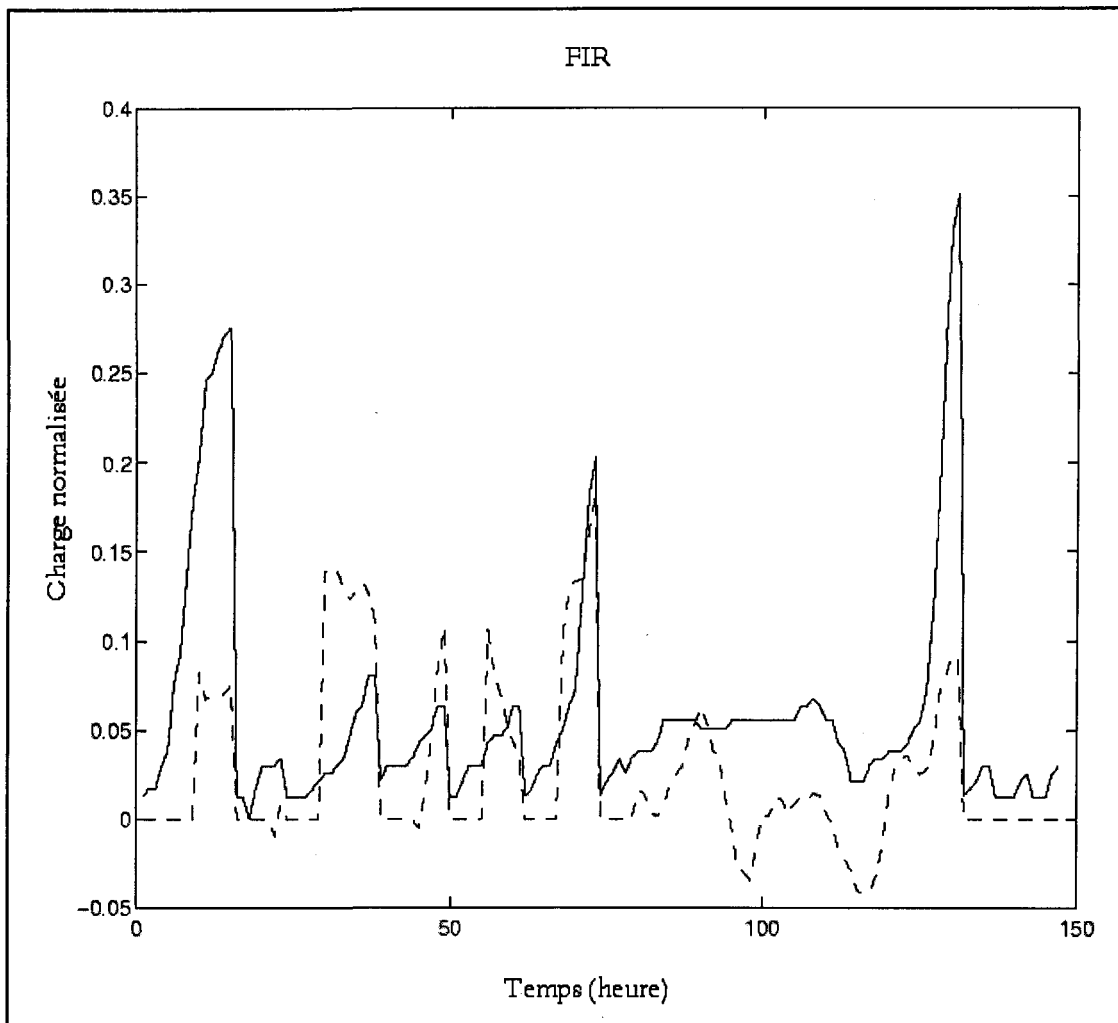


Figure 5.13: Courbes de prédiction pour le réseau FIR avec le signal du capteur de force en phase d'accrétion (courbe solide: donnée réelle; courbe pointillée: prédiction)

Bien que le réseau FIR soit le meilleur candidat pour la tâche de prédiction pour le signal du capteur de force, basé sur le critère du NMSE, le MLP le suit quand même de près. Une comparaison peut être faite à l'aide des graphiques précédents. Ils montrent que le MLP a un comportement plutôt similaire à l'ADALINE. Le réseau FIR présente un comportement visiblement différent. Le nombre de fausses alarmes est

très bas par comparaison aux autres réseaux. En fait, l'ADALINE et le MLP ont une sortie quelque peu bruyante qui semble peu corrélée avec les données qu'ils sont supposés prédire. Toutefois, le réseau FIR marque une certaine tendance à surestimer les petits événements de givrage et à sous-estimer les événements plus grands. Ceci peut suggérer que la taille des données d'apprentissage était trop faible. L'introduction de discontinuités à la délimitation de deux événements successifs a dû être prise en considération également durant l'apprentissage puisque les réseaux de neurones dépendants du temps sont généralement adaptés pour traiter de longues séquences temporelles continues. Il s'agit d'une complication qui n'a pas eu de solutions vraiment satisfaisantes dans le cadre de cette recherche. De plus, la sortie nulle observable au début de chaque événement est un artifice causé par la méthodologie de validation. Elle n'apparaîtrait pas normalement pour une application continue en temps réel du réseau.

Il est pensable qu'on pourrait donner aux modèles neuronaux une meilleure sensibilité aux événements de givrage majeurs en choisissant avec précaution les données utilisées pour l'apprentissage. En filtrant de manière sélective les données, une meilleure réponse pourrait être vraisemblablement obtenue mais avec l'inconvénient de réduire encore davantage la quantité de données utiles disponibles.

5.5 COURBES DE PRÉDICTION (MODE DE PRÉDICTION FUTURE)

Dans les tests précédents, les modèles neuronaux opéraient toujours dans un mode de prédiction appelé le “*nowcasting*”. La sortie du système était prédite en simultanéité avec les entrées, c’est-à-dire que les modèles implémentaient une fonction non-linéaire $F(*)$ telle que:

$$\hat{y}(n) = F(x1_n, x2_n, x3_n) \quad (2)$$

Il est à noter le fait important que nous n'utilisons pas l'historique passée de la variable de sortie pour prédire son évolution future. Dans la littérature, ces modèles sont référés sous le vocable de “modèles entrée-sortie”.

Au cours de cette étude surgit la question de l'impact éventuel de l'historique de la sortie sur la performance de prédiction. Pour répondre à cette question, nous avons tenté une autre forme de modèle avec le réseau FIR dans lequel la sortie actuelle est utilisée comme variable d'entrée supplémentaire pour prédire sa valeur un pas de temps dans le futur. Dans ce cas, le réseau implémentait un modèle mathématique sous la forme suivante:

$$\hat{y}(n + 1) = F(x1_n, x2_n, x3_n, y_n) \quad (3)$$

Pour faire l'entraînement, les valeurs véritables de la sortie furent utilisées. Il en est de même pour la prédiction, c'est-à-dire qu'à chaque pas de temps, c'est la sortie réelle qui servait à établir la prédiction future et non pas l'estimé sortant du modèle pour le pas de temps précédent. Dans la littérature on réfère à ces modèles sous le nom de "modèles d'identification série-parallèles". L'utilisation rétroactive de l'estimé du modèle forme ce que l'on appelle des "modèles d'identification parallèles". Des exemples avec la première méthodologie sont présentés aux figures 5.14-15.

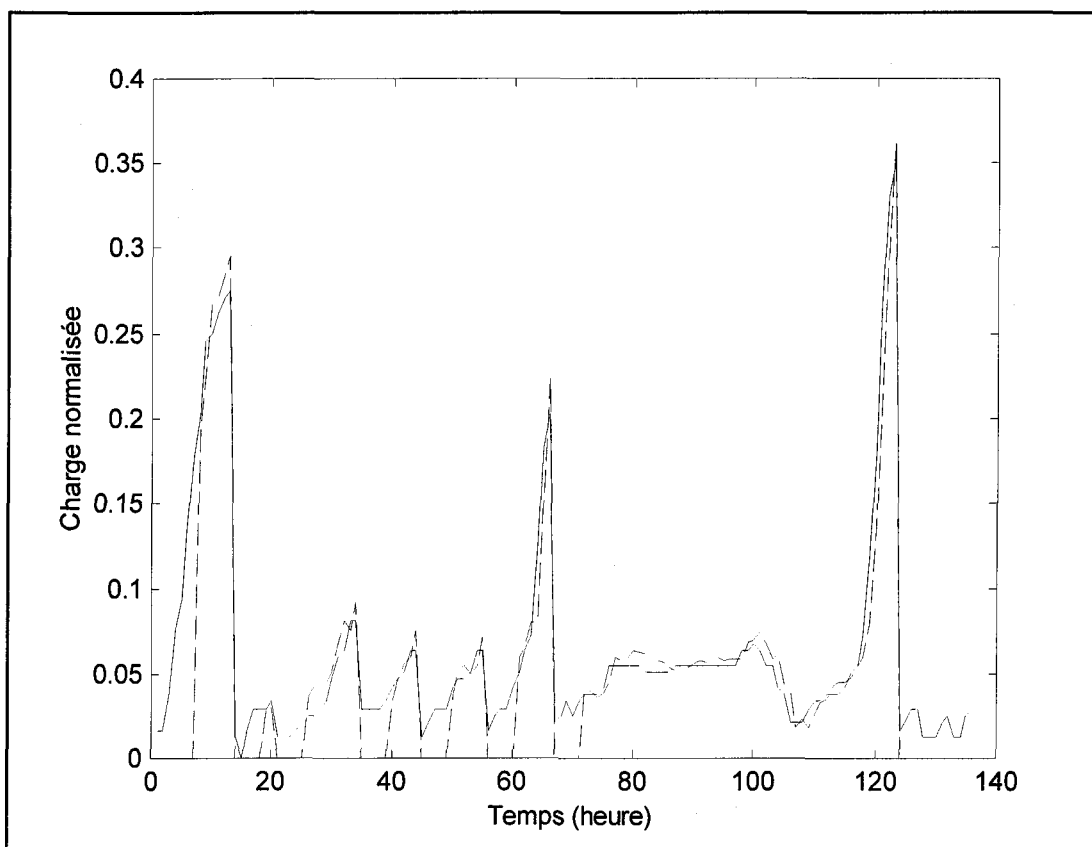


Figure 5.14: Prédiction d'un pas de temps d'avance pour le réseau FIR avec le capteur de force (courbe solide: donnée réelle; courbe pointillée: prédiction)

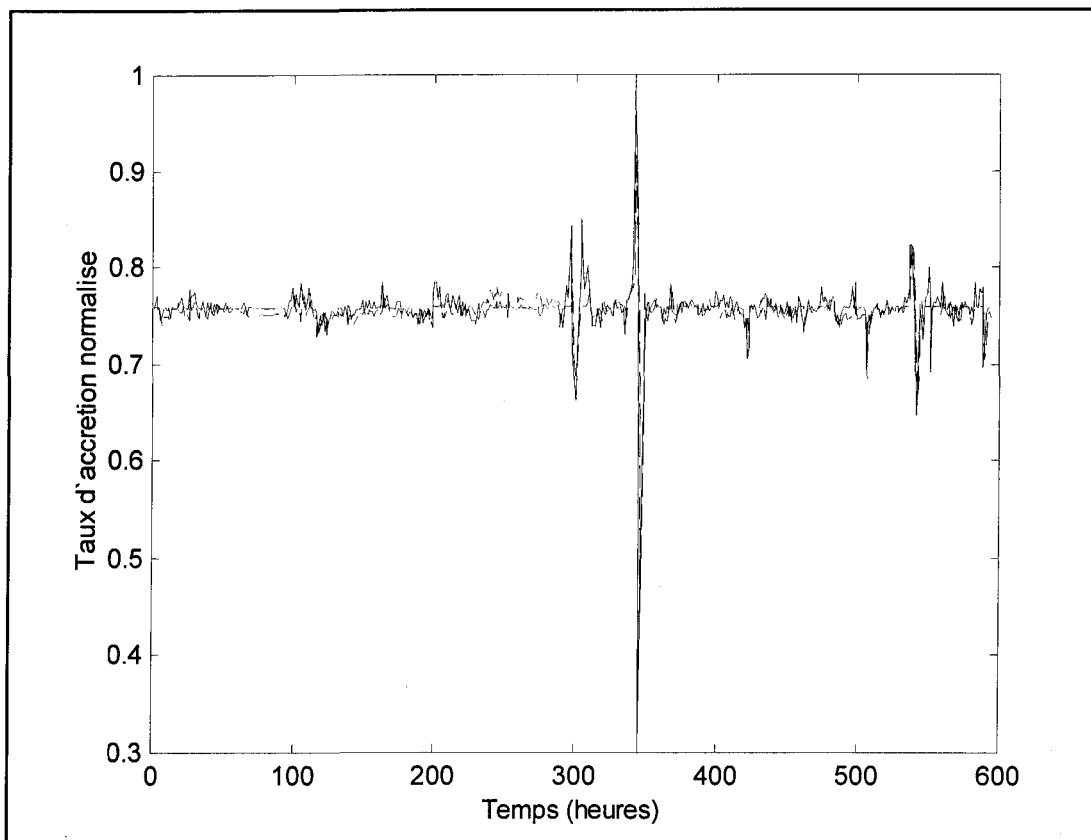


Figure 5.15: Prédiction d'un pas de temps d'avance pour le réseau FIR avec le taux d'accrétion (courbe solide: donnée réelle; courbe pointillée: prédiction)

On peut voir une concordance plutôt bonne de la prédiction et du signal-cible. Cela porte à croire que l'historique de la sortie comporte un grand pouvoir prédictif, c'est-à-dire que l'état actuel est fortement corrélé avec les états antérieurs.

Des expérimentations faites avec la configuration "identification parallèle" n'ont pas donné de résultats probants toutefois. Le comportement du réseau montrait une tendance marquée à l'instabilité dans ce cas. C'est une voie qui n'a donc pas été poursuivie.

Pour vérifier la capacité prédictive du nouveau modèle et la pertinence du capteur de force vis-à-vis du signal de givromètre pour le pouvoir prédictif nous avons fait d'autres tests. Les résultats ont montré que l'historique passé du capteur de force comporte un plus grand pouvoir prédictif que le givromètre (figures 5.16-17). Ce fait ne doit pas trop étonner; en effet, une identification de système en "aveugle" est un problème intrinsèquement beaucoup plus difficile que la continuation simple d'une série temporelle.

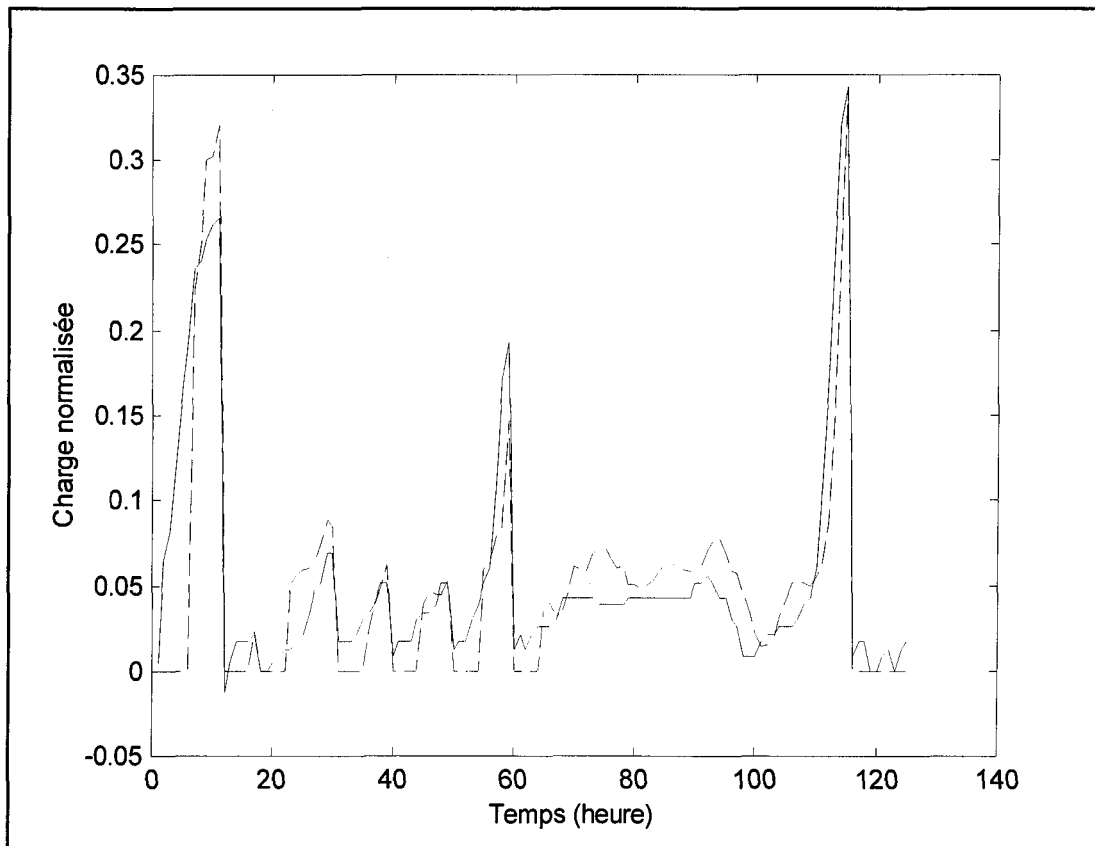


Figure 5.16: Prédiction de deux pas de temps d'avance pour le réseau FIR avec le signal du capteur de force mais sans le givromètre (courbe solide: donnée réelle; courbe pointillée: prédiction)

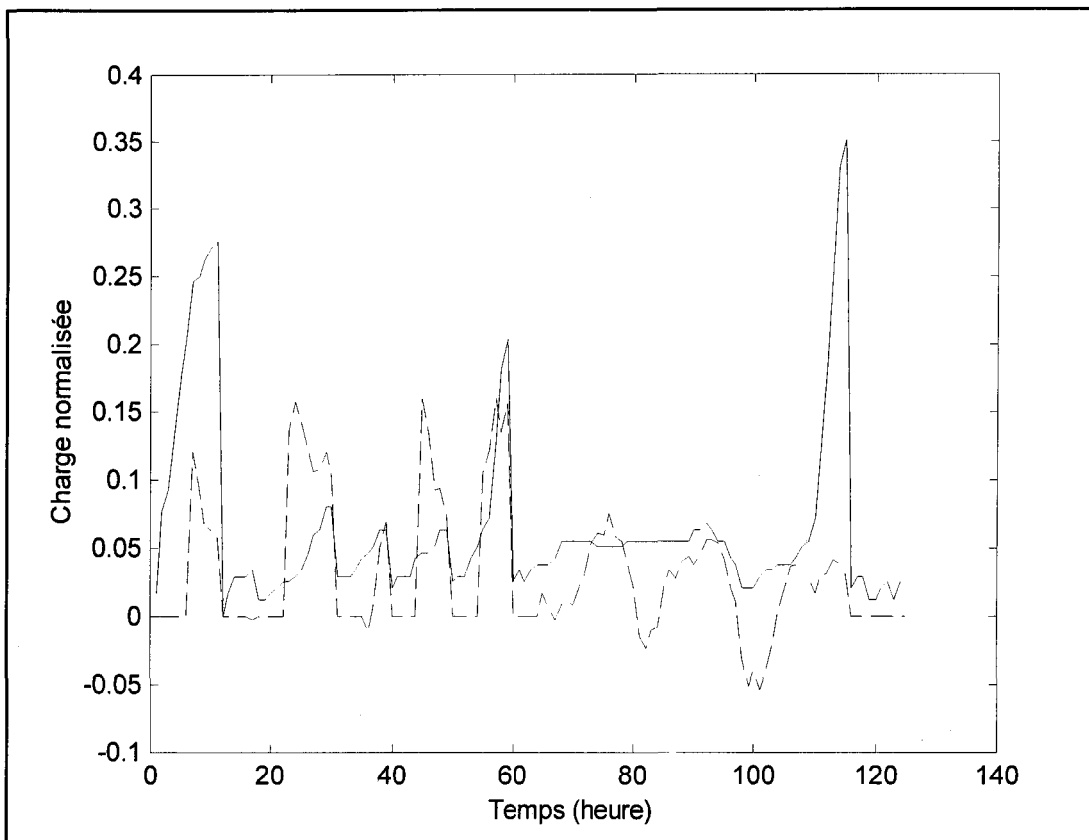


Figure 5.17: Prédiction de deux pas de temps d'avance pour le réseau FIR avec le givromètre mais sans le signal du capteur de force (courbe solide: donnée réelle; courbe pointillée: prédiction)

D'autres tests furent menés pour vérifier la profondeur de l'horizon prédictif. Il fut observé qu'à partir de trois pas de temps commencent à apparaître des divergences notables (figure 5.18). Un ou deux pas de temps semblent donc le meilleur compromis dans le présent contexte. Dans une application orientée éventuellement vers le traitement en temps réel, l'outil prédictif idéal serait un modèle d'identification parallèle puisque cela n'exigerait pas dans ce cas-là l'utilisation de capteur de force partout où le modèle se trouverait en opération.

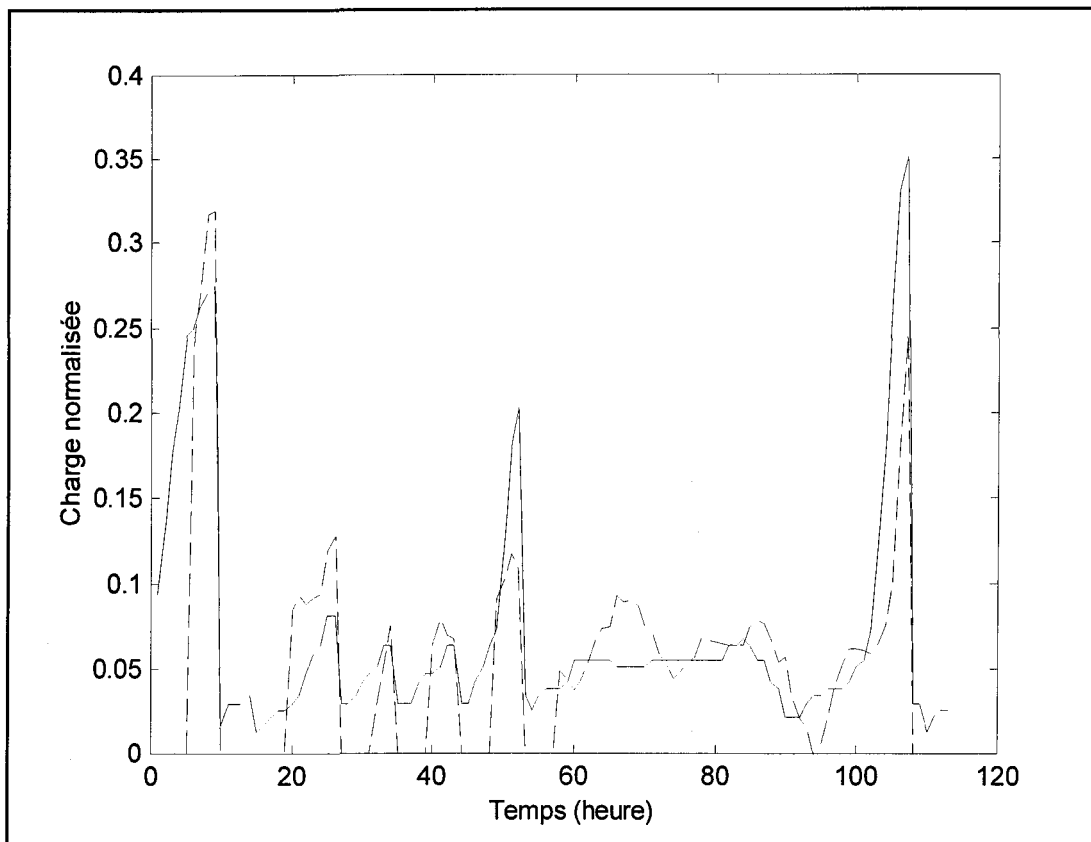


Figure 5.18: Prédiction de trois pas de temps d'avance pour le réseau FIR avec le capteur de force (courbe solide: donnée réelle; courbe pointillée: prédiction)

Toutefois, dans le cadre de cette étude, nous n'avons pas réussi à produire sous cette configuration parallèle un modèle qui donnait des prédictions vraiment satisfaisantes. Les modèles série-parallèles semblent donc plus prometteurs. Cependant, ils possèdent évidemment l'inconvénient de nécessiter plus de ressources instrumentales.

CHAPITRE 6

CONCLUSION

CONCLUSION

Dans le cadre d'une approche de prédiction en mode "nowcasting" de la charge de glace atmosphérique sur les conducteurs basée sur l'utilisation exclusive des signaux de givromètre, de la vitesse normale du vent et de la température comme paramètres d'entrée, les réseaux de neurones semblent être une approche prometteuse pour la tâche définie par la problématique. Les résultats obtenus montrent un avantage notable de l'utilisation des réseaux de neurones comme outil prédictif par rapport à des techniques statistiques plus élémentaires telle que la régression linéaire multiple par exemple. Bien que de complexité accrue, ils affichent une plus grande versatilité.

Cette recherche visait à identifier les architectures les mieux adaptées pour la prédiction de la glace atmosphérique. La problématique de prédiction comporte un aspect temporel inhérent. L'expérimentation avec des architectures de réseaux de neurones orientées vers le traitement temporel de l'information (ex: FIR) tend à donner de meilleurs résultats. Ceci indique donc que la dimension temporelle ne peut être complètement ignorée. Nous pouvons affirmer que les architectures temporelles sont plus adéquates que les architectures purement statiques.

Toutefois, la problématique d'identification de système que constituait cette tâche de prédiction demeure tout de même difficile. Dans ce contexte, une prédiction

précise point par point semble plutôt hors de portée. L'absence d'un déterminisme absolu entre les variables d'entrée et de sortie dans les données réelles utilisées doit être postulée. En fait, les analyses de corrélation statistiques présentées. En fait, n'importe quel modèle empirique ne peut produire de prédictions meilleures que les corrélations existantes déjà entre les variables elles-mêmes le permettent. En ce sens, les réseaux de neurones artificiels ne constituent certainement pas une exception.

Par ailleurs, une seconde approche basée sur l'utilisation additionnelle de l'historique passé du signal du capteur de force pour établir une prédiction future ("forecasting") semble une avenue fort avantageuse. Cette tâche de prédiction se ramène en fait à un problème de continuation d'une série temporelle qui est intrinsèquement plus facile qu'une identification de système. Dans ce contexte, les expérimentations numériques faites au cours de cette recherche ont donné des résultats définitivement encourageants. Les modèles arrivaient à faire une prédiction d'avance très raisonnable qui pourrait être utilisée. Un résultat important qui ressort également de ces modèles est le fait que l'historique du capteur de force possède un pouvoir prédictif plus fort que celui du signal de givromètre.

Les réseaux de neurones ont besoin d'un nombre important de données pour être entraînés adéquatement et pour atteindre une convergence statistique satisfaisante. Toutefois, dans la présente étude, le nombre de points de données disponibles était malheureusement limité. Ceci nous limite quelque peu par

conséquent dans nos conclusions. Malgré tout, les résultats obtenus indiquent que les réseaux de neurones peuvent être d'une assistance utile dans la prédiction de l'accrétion de la glace atmosphérique.

Notre problématique était également compliquée par le fait que nous avons utilisé des données de terrain bruitées. D'un point de vue de méthodologie scientifique, il aurait été préférable d'appliquer d'abord les modèles neuronaux à des données expérimentales de laboratoire où l'environnement et les causes d'erreurs sont bien contrôlés avant de passer à des données de terrain plus erratiques.

Nous recommandons donc que d'autres expériences avec des bases de données beaucoup plus importantes soient menées. On aurait avantage également à faire une étude du comportement des modèles sur des données de laboratoire en plus des données de terrain. Il est également envisageable que d'autres architectures de réseaux de neurones puissent être expérimentées. En particulier, il serait souhaitable de chercher des architectures plus spécialement adaptées pour traiter des séquences temporelles discontinues. La présence de discontinuités n'a pas trouvé de solution entièrement satisfaisante au courant de cette étude. Enfin, une autre direction dans laquelle la recherche peut se poursuivre concerne l'applicabilité des modèles neuronaux à plusieurs sites de givre atmosphérique. On aurait intérêt à savoir si les modèles entraînés sur un site particulier peuvent être transposés à d'autres sites.

BIBLIOGRAPHIE

- [1] Anon, "Ontario Hydro Wind and Ice Loading Model". Meteorology Research Inc. Report MRI 77 FR-1496, 1977.
- [2] Anon, "Climatological Ice Accretion Modelling". Meteorological and Environmental Planning Ltd. and Ontario Hydro, Canadian Climate, Centre Report No. 84-10, Atmospheric Environment Service, Downsview, 1984, 195 pp.
- [3] Bishop C.M., "Neural Network for Pattern Recognition", Oxford Press, New York, 1995.
- [4] Boyer A. et Meale J., "Insulator Flashovers under Icing Conditions on the Ontario Hydro 500 kV Transmission Lines System". CEA, Spring Meeting, Montreal, Canada, Mars 1988.
- [5] Chainé P.M. et Castonguay G., "New Approach to Radial Ice Thickness Concept Applied to Bundle-like Conductors". Industrial Meteorology-Study IV, Environment Canada, Toronto, 1974, 11 pp.
- [6] Chislom W. et Kuffel J., "Performance of Insulation Coating under Contamination and Icing Conditions". CEA, Spring Meeting, Vancouver, B.C., Canada, March 1995.
- [7] Cybenko G. "Approximation by Superpositions of a Sigmoidal Function", Mathematics of Control, Signals, and Systems, Vol. 2, 1989, pp. 303-314.
- [8] Druetz J., McComber P., Farzaneh M., "Atmospheric Icing on a Test Power Line". Transactions of the Canadian Society for Mechanical Engineering, C.S.M.E.,

- Vol. 24, No. 2, 2000, pp. 359-374.
- [9] Druez J., McComber P., Farzaneh M., "Correlation Between Measurement of an Ice Detector and the Mass of Ice Accreted on Two Different Size Conductors". *Revue Canadienne de Génie Civil*, Vol. 26, No. 6, 1999, pp. 869-875.
- [10] Elfashny K.N.G., Chouinard L.E., Nguyen V.T.V., Laflamme J.N., "Characteristics of Icing Events in Quebec". Eighth International Workshop on Atmospheric Icing of Structures, Reykjavik, Islande, 1998, pp. 537-542.
- [11] Elman J.L., "Finding Structure in Time". *Cognitive science*, Vol. 14, 1990, pp. 179-211.
- [12] Farzaneh M., "Ice Accretion on H.V. Conductors and Insulators and Related Phenomena". Invited article, *Philosophical Transactions, The Royal Society*, Londres, No. 358, 2000, pp. 1-35.
- [13] Farzaneh M., Kiernicki J., "Flashover Problems Caused by Ice Build-up on Insulators". *IEEE Electrical Insulation Magazine*, Invited Paper, March 1995, Vol. 11, No. 2, pp. 5-17.
- [14] Farzaneh M., Savadjiev K., Druez J., "Icing Event Occurrence in Quebec: Statistical Analysis of Field Data". *International Journal of Offshore and Polar Engineering*, Vol. 11, No. 1, Mars 2001, pp. 61-67.
- [15] Farzaneh M., Zhang J., "Modelling of DC Arc Discharge on Ice Surfaces". *IEEE Proceedings - Generation, Transmission and Distribution*, Vol. 147, No. 2, March 2000, pp. 81-86.
- [16] Finstad K., Fikke S. et Ervik M., "A Comprehensive Deterministic Model for

- Transmission Line Icing Applied to Laboratory and Field Observations". Fourth International Workshop on Atmospheric Icing of Structures, 1988, pp. 227-231.
- [17] Gates E.M., Liu A., Lozowski E.P., "A Stochastic Model of Atmospheric Rime Icing". *J. Glaciol.*, Vol. 34 No. 12, 1988, pp. 26-30.
- [18] Goodwin E.J., Mozer J.D., Di Gioia A.M. Jr, Power B.A., "Predicting Ice and Snow Loads for Transmission Lines". First International Workshop on Atmospheric Icing of Structures, 1983, pp. 267-273.
- [19] Hardy C., Brunelle J., Chevalier J., Manoukian B., Vilandr  R., "Telemonitoring of Climatic Loads on Hydro-Quebec 735 kV Lines". Seventh International Workshop on Atmospheric Icing of Structures, Chicoutimi, Canada, 1996, pp. 169-174.
- [20] Havard D.G. et Pon C.J., "Analysis of Galloping Conductor Field Data". Eighth International Workshop on Atmospheric Icing of Structures, Reykjavik, Islande, 1998.
- [21] Haykin S., "Neural Networks - A Comprehensive Foundation", Prentice-Hall, New York, 1999.
- [22] Hertz J., Krogh A. et Palmer R.G., "Introduction of the Theory of Neural Computation", Addison-Wesley, Redwood City, 1991.
- [23] Imai I. "Studies on Ice Accretion". *Researches on Snow and Ice*, No. 1, 1953, pp. 35-44
- [24] Jones K.F., "The Effect of Horizontal and Torsional Coupling on Vertical Galloping". Sixth International Workshop on Atmospheric Icing of Structures,

- Budapest, Hongrie, 1993, pp. 143-148.
- [25] Kawai R. , "AC Flashover Test at Project UHV on Ice-covered Insulators". IEEE Trans. on Power App. and Syst., Vol. PAS-89 No, 8, 1970, pp. 1800-1804.
- [26] Khalifa M. et Morris R., "Performance of Line Insulators under Rime Ice". IEEE Transac. on Power App and Syst., Vol. PAS-86 No 6, 1967, pp. 696-698.
- [27] Krishnasamy S.G., Tabatabai M. , Kastelein M., "A Pilot Field Project to Evaluate Icing Models in Ontario Canada". Sixth International Workshop on Atmospheric Icing of Structures, Budapest, Hongrie, 1993, pp. 85-97.
- [28] Laflamme J., "Spatial Variation of Extreme Values in the Case of Freezing Rain Icing". Sixth International Workshop on Atmospheric Icing of Structures, Budapest, Hongrie, 1993, pp. 19-23.
- [29] Laflamme J., "Icing Rate Measurements: a Key Way of Estimating Ice Loads on Structures". Seventh International Workshop on Atmospheric Icing of Structures, Chicoutimi, Canada, 1996, pp. 175-180.
- [30] Laforte J.L., Allaire M.A., Laflamme J., Beaudoin P., "An Up-date on Overhead Cable De-Icing Technology". Eighth International Workshop on Atmospheric Icing of Structures, Reykjavik, Islande, 1998, pp. 215-220.
- [31] Lenhard R.W., "An Indirect Method for Estimating the Weight of Glaze on Wires". Bull. Amer. Meteor. Soc., Vol. 36, 1955, pp. 1-5.
- [32] Lozowski E.P., Stallabrass J.R. et Hearty PF, "The Icing of an Unheated, Non-Rotating Cylinder. Part I: A Simulation Model.". J. Clim. Appl Meteorol., Vol. 22 No.12, 1983, pp. 2053-2062.

- [33] Lozowski E.P., Brett M., Tait N., Smy T., "Simulating Giant Hailstone Structure with a Ballistic Aggregation Model". *Q J R Meteorol Soc*, Vol. 117 No. 498, 1991, pp. 427-431
- [34] Makkonen L. "Modelling of Ice Accretion on Wires". *J. Climate and Appl. Meteor.*, Vol. 23, 1984, pp. 929-939.
- [35] Makkonen L. et Stallabrass J.R., "Ice Accretion on Cylinder and Wires", National Research Council of Canada, DME Report TR-LT-005, 1984. 50 pp.
- [36] Makkonen L. et Ahti K., "Climatic Mapping of Ice Loads Based on Airport Weather Observations". *Sixth International Workshop on Atmospheric Icing of Structures*, Budapest, Hongrie, 1993, pp. 13-17.
- [37] Makkonen L., "Modeling Power Line Icing in Freezing Precipitation". *Seventh International Workshop on Atmospheric Icing of Structures*, Chicoutimi, Canada, 1996, pp. 195-200.
- [38] McComber P., Druetz J. et Laflamme J., "A Comparison of Selected Models for Estimating Cable Icing". *Sixth International Workshop on Atmospheric Icing of Structures*, Budapest, Hongrie, 1993, pp. 41-46.
- [39] McComber P., "The Icing Rate Meter, an Instrument to Evaluate Transmission Line Icing". *Seventh International Workshop on Atmospheric Icing of Structures*, Chicoutimi, Canada, 1996, pp. 159-168.
- [40] McComber P., de Lafontaine J., Laflamme J., "A Neural System to Estimate Transmission Line Icing". *Eighth International Workshop on Atmospheric Icing of Structures*, Reykjavik, Islande, 1998, pp. 101-106.

- [41] McComber P. et al., "Estimation of Transmission Line Icing at Different Sites Using a Neural Network", Ninth International Offshore and Polar Engineering Conference, Vol. II, Brest, France, 1999, pp. 599-606.
- [42] McComber P., "A Non-Circular Accretion Shape Freezing Rain Model for Transmission Line Icing.", Ninth International Workshop on Atmospheric Icing of Structures, Chester, Royaume-Uni, 2000, pp. I-VIII.
- [43] McCulloch W.S. et Pitts W., "A Logical Calculus of the Ideas Immanent in Nervous Activity". Bulletin of Mathematical Biophysics Vol. 5, pp. 115-133.
- [44] Mozer M. C., "Neural Net Architectures for Temporal Sequence Processing", in Proceedings of the NATO Advanced Research Workshop on Time Series Prediction and Analysis, Addison-Wesley, Santa Fe, New Mexico, 1994.
- [45] Ohta H. et al., "Application of Disaster Warning System Due to Snow Accretion on Power Lines Using Neural Networks". Seventh International Workshop on Atmospheric Icing of Structures, Chicoutimi, Canada, 1996, pp. 149-154
- [46] Personne P., Couroue C., Isaka H., "Theoretical Study of Air Inclusion on Rotating Cylinders". Fifth International Workshop on Atmospheric Icing of Structures, Tokyo, Japon, 1990, pp. A2-6.
- [47] Poots G., "Ice and Snow Accretion on Structures". J. Wiley, New York, États-Unis, 1996.
- [48] Rumelhart D.E. et McClelland J.L., Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, Cambridge, MIT Press, États-Unis, 1986.

- [49] Savadjiev K., Latour A. et Paradis A., "Estimation of Ice Accretion Weight from Field Data on Overhead Transmission Line Cables". Seventh International Workshop on Atmospheric Icing of Structures, Chicoutimi, Canada, 1996, pp. 159-168.
- [50] Savadjiev K., Farzaneh M., Druetz J., "Analysis and Interpretation of Icing Rate Meter and Load Cell Measurements on the Mt Bélair Icing Site", Ninth International Offshore and Polar Engineering Conference, Vol. II, Brest, France, 1999, pp. 607-611.
- [51] Stallabrass J.R., "Aspects of Freezing Rain Simulation and Testing". First International Workshop on Atmospheric Icing of Structures, 1983, pp. 67-74.
- [52] Szilder K., Lowoski E.P., Gates E.M., "Modelling Ice Accretion on Non-Rotating Cylinders - the Incorporation of Time Dependence and Internal Heat Conduction". Cold Reg Sci Technol, Vol 13 No 2, 1987, pp 177-191.
- [53] Szilder K., "The Density and Structure of Ice Accretion Predicted by a Random-Walk Model". Q J R Meteorol Soc, Vol. 119 No. 513, 1993, p. 907-924.
- [54] Szilder K., "Simulation of Ice Accretion on a Cylinder Due to Freezing Rain". J Glaciol, Vol. 40 No. 136, 1994, pp. 586-594.
- [55] Szilder K. et Lozowski E.P., "Simulation of Icicle Growth Using a Three-Dimensional Random Walk Model". Atmosph Res, Vol. 36, 1995, pp. 243-249.
- [56] Szilder K., Lozowski E.P., "Three-Dimensional Modelling of Ice Accretion Micro Structure". Seventh International Workshop on Atmospheric Icing of Structures, Chicoutimi, Canada, 1996, 1996, pp. 60-63.

- [57] Szilder K., Lozowski E.P., Farzaneh M., "Morphogenetic Modelling of Wet Ice Accretions on Transmission Lines as a Result of Freezing Rain.". *International Journal of Offshore and Polar Engineering*, Vol. 11., No. 1, Mars 2001, pp.16-22.
- [58] Wan E., "Finite Impulse Response Neural Networks with Applications in Time Series Prediction", Thèse de Ph.D., Stanford University, États-Unis, 1993.
- [59] Wan E., "Modeling Nonlinear Dynamics with Neural Networks: Examples in Time Series Prediction". Fifth Workshop on Neural Networks: Academic /Industrial/ NASA/Defense, WNN93/FNN93, San Francisco, États-Unis, Novembre, 1993, pp. 327-332.
- [60] Wan E., "Time Series Prediction by Using a Connectionist Network with Internal Delay Lines", in *Proceedings of the NATO Advanced Research Workshop on Time Series Prediction and Analysis*, Addison-Wesley, Santa Fe, New Mexico, 1994.
- [61] Weigend A.S. et Gershenfeld N., *Proceedings of the NATO Advanced Research Workshop on Time Series Prediction and Analysis*, ed. Addison-Wesley, Santa Fe, New Mexico, 1994.
- [62] Werbos, P.J., "Beyond regression: New tools for Prediction and Analysis in the Behavioral Sciences". Thèse de Ph.D., Harvard University, Cambridge, États-Unis, 1974.
- [63] Widrow B. et Hoff M.E. Jr, "Adaptive Switching Circuits", IRE WESCON convention record, 1960, pp. 96-104.