UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À

L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI

COMME EXIGENCE PARTIELLE DE LA MAÎTRISE EN INGÉNIERIE

PAR

RÉMI RAJAONARIVELO

INTÉGRATION DES DONNÉES MANQUANTES EN VUE DE LA PRÉDICTION DES ÉVÉNEMENTS DE GIVRE ET DE VERGLAS

AUTOMNE 2008

Bibliothèque
Paul-Émile-Boulet

UQAC
Université du Québec
à Chicoutimi

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 acceptation 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

Les matériaux soumis à des conditions climatiques hivernales sont susceptibles de recevoir une accumulation de glace qui peut nuire à leur intégrité ou à leur fonctionnement. L'étude des phénomènes de glace atmosphérique cherche à développer une connaissance qui permettrait de prévoir et d'estimer les dangers d'un événement afin de prendre les mesures appropriées. Des recherches sont effectuées à cet effet. Cependant, aucune étude pour traiter les données manquantes n'a encore été amorcée. Or, l'intégration de ces types de données est nécessaire pour tenir compte des données disponibles et élargir le champ de vision en ce qui concerne le développement de modèle de prédiction des événements de givre et de verglas.

Le présent travail s'inscrit dans le cadre de l'analyse et de l'exploitation des données recueillies par le système d'acquisition de données SYGIVRE. En résumé, on a pu collecter des données provenant de 35 stations de mesure lors de la saison 2001-2002 et pour la saison 2003-2004, on a compté 32 stations. Ces données provenant des sites naturels ont été traitées afin de créer une base de données des tempêtes de givre et de verglas pour ces saisons d'observation. Les informations disponibles grâce à ce système sont l'ensemble des relevés horaires des valeurs des températures de l'air et des signaux des givromètres.

Afin de construire une base d'apprentissage pour le modèle de prédiction des poids d'accumulation de givre et de verglas et de retracer les tempêtes de givre et de verglas répertoriées dans la base de données SYGIVRE, un processus de modélisation des données a été amorcé. Ceci consiste en premier lieu à la recherche de méthodologie pour identifier les événements élémentaires de givre et de verglas. Ensuite, les types d'accumulation sont déterminés selon leurs caractéristiques. En tenant compte de ces types d'accumulation, une procédure basée sur l'algorithme 9-23 a été proposée pour estimer les taux d'accrétions horaires de givre et de verglas. Ceci a permis l'évaluation de leurs poids d'accumulation au fil des heures. Par la suite, on a défini la notion de tempête avant de proposer une méthode pour répertorier les tempêtes ou événements de givre et de verglas lors des saisons 2001-2002 et 2003-2004. Ce processus de modélisation de la base de données SYGIVRE se termine par le regroupement des stations de mesure. Selon les localités de ces dernières, on a distingué la région de Montréal-Québec, celle du Saguenay - Lac-Saint-Jean et celle de la Côte-Nord. D'après les résultats d'analyse des données, on a recensé 507 tempêtes de givre et de verglas. Ces tempêtes s'étendent dans l'ensemble sur une période de 3948 heures et le cumulatif du nombre de signaux de givromètre enregistrés est égal à 9167 coups. Il a été remarqué que la durée moyenne d'une tempête est égale à 8 heures et on enregistre 18 signaux de givromètre durant son occurrence.

Après avoir modélisé la base de données SYGIVRE, nous avons amorcé le développement du modèle de prédiction des poids d'accumulation de givre et de verglas lors des tempêtes identifiées par la première phase de ce travail. Nous avons choisi une architecture de réseau de neurones multicouche à rétropropagation de l'erreur. Étant donné que nous n'avons pas suffisamment de données pour effectuer l'apprentissage de ce modèle, une simulation de données de givrage atmosphérique a été effectuée pour le valider avant de le tester avec les données de la base de données SYGIVRE. La principale raison d'utilisation de ces données simulées est le fait que des tests directs sur les données réelles ne nous assurent pas la validité du modèle de prédiction quand une partie des données est manquante. Ainsi, ces tests préliminaires

nous permettront d'estimer la performance du modèle de réseau de neurones. De plus, cette simulation de données virtuelles permet non seulement de générer des données statistiquement proches des données réelles mais aussi d'avoir une représentation du phénomène de givrage atmosphérique.

En vue de poursuivre l'objectif à atteindre, une méthodologie a été élaborée afin d'intégrer les données manquantes au modèle de prédiction. En principe, toutes les observations doivent être complètes pour les deux phases de développement de réseau de neurones: apprentissage et prévision. Ces conditions ne sont pas respectées quand des valeurs sont manquantes à une ou plusieurs stations de mesure. Pour pallier ces problèmes, l'applicabilité des méthodes statistiques d'imputations a été étudiée. Trois types de données manquantes ont été considérés dans cette étude: des données manquantes au hasard, des données manquantes complètement au hasard et les autres types de données manquantes. Ainsi, des méthodes de remplacement de données ont été réalisées pour chaque groupe de stations et ont permis d'estimer des valeurs manquantes pour des courtes périodes d'observation. En ce qui concerne la résolution du problème causé par la présence de saisons manquantes, nous avons introduit la technique d'inhibition des cellules du réseau de neurones pour adapter l'algorithme d'apprentissage à ces lacunes de données.

Lors des expériences, plusieurs configurations du modèle de prédiction ont été utilisées en variant les paramètres du réseau de neurones. Ces paramètres sont le nombre de cellules dans la couche cachée, le taux d'apprentissage et le moment d'apprentissage. D'après les résultats obtenus, on démontre que la technique d'inhibition de réseaux de neurones permet d'intégrer les valeurs manquantes pour la prédiction des événements de givre et de verglas. Les tests sur les données simulées, avec et sans données manquantes, ont fait ressortir deux modèles de prédiction qui ont sensiblement les mêmes performances et qui sont fonctionnels par rapport à la prévision des tempêtes de givre et de verglas. Toutefois, quand on a varié les paramètres du modèle, on a remarqué que ce dernier ne fournit pas toujours les résultats escomptés vis-à-vis de toutes les stations composant le groupe étudié. En ce qui concerne les tests sur les données du groupe du Saguenay – Lac-Saint-Jean, les mêmes constatations ont été observées. En effet, la présence de saison manquante au niveau d'une des stations composant ce groupe n'empêche pas le modèle étudié de prédire correctement les événements de givre et de verglas aux autres stations. En résumé, les résultats des validations des modèles de prédiction des événements de givre et de verglas nous amènent à affirmer que la combinaison des méthodes d'imputation statistiques et la technique d'inhibition du réseau de neurones lors de la phase d'apprentissage permettent d'intégrer les données manquantes dans l'étude de ce phénomène. Ces modèles de prédiction sont fonctionnels et affichent de bons résultats de prédiction lorsque leurs paramètres sont bien calibrés. En effet, les perturbations engendrées par les données manquantes sont insignifiantes d'après la validation des modèles avec les données simulées et avec les données du groupe de stations de la région du Saguenay – Lac-Saint-Jean. Par ailleurs, on a constaté que ces modèles sont instables avec la variation des paramètres d'apprentissage du réseau de neurones. Cette étude a apporté des éléments de réponses aux problèmes impliqués par la présence des lacunes au niveau de la base de données SYGIVRE. De plus, elle contribue à la compréhension des problèmes reliés au suivi des événements de givre et de verglas par les interrogations qu'elle soulève et le besoin de les étudier.

ABSTRACT

Materials exposed to winter climatic conditions could get an ice accumulation which can damage their integrity or operation. Icing events often cause mechanical and electrical disturbances to the Hydro-Quebec electric transport system. Indeed, the air lines of electric power represent a high-risk structure category because they go through isolated areas, making prevention and maintenance very difficult. The atmospheric ice phenomena study seeks to develop knowledge in order to estimate or even predict dangerous events and therefore, take a suitable action. Researches have been done for that purpose. However, no methodology was carried to treat the missing data, even if integration of this kind of data is necessary to take all the available information into consideration and to stretch the way for the development of atmospheric icing model prediction.

This work is under the analysis and exploitation of provided data by the data acquisition system called SYGIVRE. The available information in this system is the whole of time statements of air temperature values and icing rate meter signals of measuring sites. In this study, those natural sites data were treated in order to create a white frost and glaze storm database for the 2001-2002 and 2003-2004 observing seasons. Thus, we have been able to collect data from 35 measuring sites during the 2001-2002 season. For the 2003-2004 season, we counted 32 stations.

The first step was to start a data transformation process to retrieve the recorded white frost and glaze storm events from the SYGIVRE database. A methodology research to identify the elementary events of cold precipitations was first set up. Then, the types of precipitation are given according to their characteristics. To take into account these kinds of precipitation, a procedure based on 9-23 algorithm is proposed to consider the hourly accretion rates of white frost and glaze. It enables the evaluation of their hourly weight accumulation in time. Finally, we define the storm concept before proposing a way to index all white frost and glaze events during the 2001-2002 and 2003-2004 seasons. The data processing method ends by regrouping the measuring sites according to their locations. We singled out the Montreal-Quebec, Saguenay - Lac-Saint-Jean and Côte-Nord areas. According to data analysis, we registered 507 white frost and glaze storms. These storms occurred during a period of 3948 hours and the cumulative number of recorded icing rate meter signals is 9167. It was noticed that a storm average duration is eight hours and we recorded 18 icing rate meter signals during its occurrence.

After the SYGIVRE database setup, we started to develop the icing accumulation weight prediction model. The feed-forward neural network architecture was chosen to build this model. Later on, a data simulation of atmospheric icing was realized to ratify the prediction model before applying it to the real data of the SYGIVRE network. The main reason to use these simulated data is that direct tests on real data do not certify the prediction model validity when data parts are missing. Therefore, these preliminary tests allow us to estimate the performance of the neurons network model. This virtual simulation permits us to generate a statistically similar data in comparison to the SYGIVRE database. Also, it allows us to figure out the atmospheric icing phenomenon.

To achieve the main goal of this study, a methodology has been elaborated in order to integrate the missing data to the prediction model. As a matter of principle, all observations have to be complete for the two phases of neurons network development: learning and forecast. Those conditions are not respected when there are some missing values in one or several of the measuring sites. To compensate for those problems, the applicability of existing statistical imputation methods has been studied. Three types of missing data were considered in this study: randomly missing data, completely random missing data and other kinds of missing data. Thus, methods to replace the missing values have been achieved on data of each station group which allowed us to estimate them for short observation periods. Regarding the resolution of the problem caused by the missing seasons, a revision of this learning algorithm model was carried out. We applied the neurons technical inhibition in order to integrate these missing seasons in the learning phase of the prediction model.

According to experience results, a combination of statistical imputation methods and the revision of the neurons network learning algorithm enable the missing data integration for the prediction model development. Tests on simulated data, with or without missing data, brought out two prediction models which have significantly the same performance and are functional for white frost and glaze storms forecast. However, we have seen that they do not always provide the anticipated results when there are variations in their learning parameters, i.e.: the number of cells in the hidden layer, the learning rate and the learning moment. Regarding testing on the data of the Saguenay - Lac Saint-Jean stations group, the same conclusions were observed. Indeed, a missing season at a station of this group does not prevent the studied model to correctly predict white frost and glaze storms in other stations. In short, the validation results of white frost and glaze events prediction models allow us to confirm that statistical imputation methods and the learning algorithm revision make it possible to integrate missing data in this phenomena study. Those models are functional and give good prediction results when their parameters are well gauged. Indeed, disturbances generated by missing data are negligible according to these models validation on simulated data and on the stations group data in Saguenay - Lac Saint-Jean area. However, we observed that those models are unstable with the variation of the neuron networks learning parameters. This study brought some partial answers to problems caused by the presence of gaps in the SYGIVRE database. Also, this work helps to understand the problems related to the real time follow-up of icing events through the questions it raises and the need to study them.

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 Hassan Ezzaidi du département des sciences appliquées.

Premièrement, je remercie vivement mon directeur de recherche, Monsieur Masoud Farzaneh, Professeur à l'Université du Québec à Chicoutimi et Titulaire de la CIGELE. Sans les appuis, encouragements et supervisions qu'il m'a témoignés, ce travail n'aurait pas pu voir le jour.

Je suis reconnaissant envers Monsieur Hassan Ezaidi pour ses précieux conseils en matière de développement de réseau de neurones ainsi que la révision de la présentation du texte de ce mémoire.

Également, j'adresse mes remerciements à Monsieur Konstantin Savadjiev pour m'avoir donné l'occasion de profiter de ses connaissances dans le domaine de traitements de données provenant des sites naturels et les diverses informations qu'il m'a communiquées.

Mes remerciements vont pareillement au Professeur Dyugu Kocaefe, Directrice du programme de maîtrise en ingénierie, pour ses orientations pédagogiques et au Professeur Louis Houde pour ses contributions au démarrage de ce projet.

Finalement, ces remerciements ne seraient pas complets sans mention des soutiens de ma famille, mes collègues de travail et amis. Les discussions et les encouragements ont été une grande source de l'énergie investie dans ce travail.

TABLE DES MATIERES

RÉSUMÉ i
ABSTRACT iii
REMERCIEMENTS v
TABLE DES MATIERES vi
LISTE DES FIGURES x
LISTE DES TABLEAUX xii
LISTE DES ANNEXES xiv
INTRODUCTION 2
1.1 Problématique 4
1.2 Objectifs 6
1.2.1 Objectif principal 6
1.2.2 Objectifs spécifiques 6
1.3 Méthodologie 7
REVUE DE LITTÉRATURE 11
2.1 Études des données de givrage 12
2.2 Historique des travaux précédents 14
2.3 Insertion de la présente étude 23
MODÉLISATION DE LA BASE DE DONNÉES SYGIVRE 26
3.1 Description de la base de données SYGIVRE 27
3.1.1 Stations de mesure 28
3.1.2 Givromètre 30
3.2 Sélection des variables de givrage étudiées 31
3.2.1 Variables mesurées 33
3.2.1.1 Température de l'air 33
3.2.1.2 Signal du givromètre 33
3.2.1.3 Autres variables 34
3.2.2 Variables calculées 35
3.2.2.1 Taux d'accrétion 35
3.2.2.2 Poids d'accumulation 35

3.2.3 Sélection des variables étudiées 36
3.3 Identification des événements élémentaires de givre et de verglas 37
3.3.1 Caractéristiques générales du givre et du verglas 39
3.3.2 Définition d'un événement élémentaire de givre et de verglas 40
3.3.3 Procédure d'identification des événements élémentaires de givre et de verglas 42
3.4 Identification des types d'accumulation de givre et de verglas 42
3.5 Estimation des taux horaires d'accrétion de givre et de verglas 46
3.6 Estimation des poids d'accumulation 47
3.7 Identification des tempêtes de givre et de verglas 48
3.7.1 Définition d'une tempête de givre et de verglas 48
3.7.2 Procédures d'indentification des tempêtes de givre et de verglas 49
3.7.2.1 Considération des temps de latence 50
3.7.2.2 Considération des types d'événements élémentaires 51
3.7.2.3 Considération des sévérités des événements élémentaires 53
3.7.3 Algorithme général 54
3.8 Regroupement des stations de mesure 55
3.9 Résumé 57
DÉVELOPPEMENT DU MODÈLE DE PRÉDICTION 60
4.1 Les réseaux de neurones multicouches 61
4.1.1 Principes de base 62
4.1.2 Techniques d'optimisation 63
4.2 Architecture générale du modèle de prédiction 64
4.3 Les données d'apprentissage du modèle de prédiction 65
4.4 Apprentissage du modèle de réseau de neurones 68
4.5 Les paramètres d'apprentissage des réseaux de neurones 69
4.6 Les fonctions de performances 69
4.6.1 Somme quadratique des erreurs 70
4.6.2 Pouvoir prédictif d'un modèle 70
4.7 Génération de données d'apprentissage virtuelles 71
4.7.1 Simulation des états des stations 71
4.7.2 Simulation des températures de l'air 73

4.7.3 Simulation des taux de précipitations 74
4.7.4 Simulation des taux d'accrétion horaires 75
4.7.5 Calculs des poids d'accumulation de givre et de verglas 75
4.8 Résumé 77
INTÉGRATION DES DONNÉES MANQUANTES 80
5.1 Traitements statistiques des données manquantes 81
5.1.1 Types de données manquantes 81
5.1.1.1 Données Manquantes au Hasard (DMH) 82
5.1.1.2 Données Manquantes Complètement au Hasard (DMCH) 82
5.1.1.3 Autre Type de Données Manquantes (ATDM) 83
5.1.2 Méthodes de traitement des données manquantes 84
5.1.2.1 Rejet des valeurs manquantes 84
5.1.2.2 Rejet des paires de valeurs aberrantes 85
5.1.2.3 Remplacement par la moyenne 85
5.1.2.4 Remplacement par une régression linéaire 86
5.1.2.5 Autres méthodes d'imputation 86
5.2 Révision de l'algorithme d'apprentissage du modèle de réseau de neurones 88
5.2.1 Algorithme standard d'apprentissage 88
5.2.2 Problèmes relatifs aux données manquantes 92
5.2.3 Inhibition des cellules correspondantes aux saisons manquantes 94
5.4 Résumé 97
RÉSULTATS ET INTERPRÉTATIONS 100
6.1 Résultats des analyses et des traitements de données 101
6.1.1 Statistique générale des données de givrage 2001-2002/2003-2004 101
6.1.2 Statistique des événements élémentaires de givre et de verglas 102
6.1.3 Statistique des températures de l'air 103
6.1.4 Statistique des signaux de givromètres 105
6.1.5 Statistique des types d'accumulation 107
6.1.6 Statistique des tempêtes de givre et de verglas 109
6.1.7 Statistique des poids d'accumulation de givre et de verglas 114
6.1.8 Statistique des données manquantes 115

6.2 Validation du modèle de prédiction avec les données simulées 118
6.2.1 Résultats des tests sans données manquantes 119
6.2.1.1 Nombre de cellules J égal à 5 119
6.2.1.1.1 Sommes des erreurs quadratiques 119
6.2.1.1.2 Prédiction des poids d'accumulation de givre et de verglas 120
6.2.1.2 Nombre de cellules J égal à 6 122
6.2.1.2.1 Sommes des erreurs quadratiques 122
6.2.1.2.2 Prédiction des poids d'accumulation de givre et de verglas 122
6.2.1.3 Nombre de cellules J égal à 7 123
6.2.1.3.1 Sommes des erreurs quadratiques 123
6.2.1.3.2 Prédictions des poids d'accumulation de givre et de verglas 124
6.2.2 Résultats des tests avec des données manquantes 126
6.2.2.1 Nombre de cellules J égal à 5 126
6.2.2.1.1 Sommes des erreurs quadratiques 126
6.2.2.1.2 Prédiction des poids d'accumulation de givre et de verglas 126
6.2.2.2 Nombre de cellules J égal à 6 127
6.2.2.2.1 Sommes des erreurs quadratiques 127
6.2.2.2.2 Prédiction des poids d'accumulation de givre et de verglas 128
6.2.2.3 Nombre de cellules J égal à 7 129
6.3 Validation du modèle de prédiction avec la base de données SYGIVRE 130
6.3.1 Résultats des tests avec les données du groupe 2 130
6.3.1.1 Nombre de cellules J égal à 5 131
6.3.1.2 Nombre de cellules J égal à 6 132
6.3.1.3 Nombre de cellules J égal à 7 133
6.3.2 Résultats des tests avec les données du groupe 1 et du groupe 3 135
CONCLUSION ET RECOMMANDATIONS 138
7.1 Conclusion générale 138
7.2 Recommandations 144
RÉFÉRENCES 147
ANNEXES 151

LISTE DES FIGURES

Figure 2-1: Modèle de prédiction des tempêtes de givre et de verglas. 20
Figure 2-2: Modèle de prédiction des poids d'accumulation de givre et de verglas. 20
Figure 3-1: SYGIVRE, un système d'acquisition des données de givrage. 28
Figure 3-2: Lieux géographiques des stations (1992-1998). 29
Figure 3-3: Le givromètre et son emplacement. 30
Figure 3-4: Processus de transformation de la base des données SYGIVRE. 37
Figure 3-5: Tempêtes de givre et de verglas au sein d'une station de mesure 49
Figure 4-1: Un réseau de neurones multicouches 61
Figure 4-2: Un neurone formel avec ses paramètres. 62
Figure 4-3: Architecture générale des modèles de réseau de neurones pour la prédiction des événements de givre et de verglas 64
Figure 4-4: Propagation temporelle des tempêtes au sein d'un groupe de stations 73
Figure 5-1: Algorithme d'apprentissage par rétropropagation de l'erreur 90
Figure 5-2: Inhibition des cellules correspondantes à une saison manquante. 95
Figure 6-1: Variation des températures moyennes des événements de givre et de verglas, Groupe 2, Saison 2001-2002 104
Figure 6-2: Variation des températures minimales des événements de givre et de verglas, Groupe 2, Saison 2001-2002 105
Figure 6-3: Variation des cumulatifs des nombres des coups de givromètre durant les événements de givre et de verglas, Groupe 2, Saison 2001-2002 106
Figure 6-4: Valeur maximale du cumulatif des signaux de givromètre, Groupe 2, Station 17, Événement 81, Saison 2003-2004 107
Figure 6-5: Prédiction des événements de givre et de verglas avec les données simulées sans données manquantes, J=5 121
Figure 6-6: Prédiction des événements de givre et de verglas avec les données simulées sans données manquantes, J=6 123
Figure 6-7: Prédictions des événements de givre et de verglas avec les données simulées sans données manquantes, J=7 125

Figure 6-8: Prédiction des événements de givre et de verglas avec les données simulées avec données manquantes, J=5 127
Figure 6-9: Prédictions des événements de givre et de verglas avec les données simulées avec données manquantes, J=6 128
Figure 6-10: Prédictions des événements de givre et de verglas, Groupe 2, J=5, saisons 2003-2004 132
Figure 6-11: Prédictions des événements de givre et de verglas, Groupe 2, J=7, saison 2003-2004 134
Figure 6-12: Prédictions des événements de givre et de verglas, Groupe 2, J=7, saison 2001-2002 135
Figure I-1: Prédictions des événements de givre et de verglas avec les données simulées avec données manquantes, J=7. 162
Figure J-2: Prédictions des événements de givre et de verglas, J=6 164

LISTE DES TABLEAUX

Tableau 3-1: Données brutes provenant d'une station de mesure. 32
Tableau 3-2: Données disponibles à une station de mesure, saison 2001-2002. 38
Tableau 3-3: Données disponibles à une station de mesure, saison 2003-2004 38
Tableau 3-4: Un événement élémentaire de givre et de verglas, station 1, saison 2001-2002. 41
Tableau 3-5: Les types d'accumulation horaires d'un événement élémentaire. 43
Tableau 3-6: Les types d'accumulation correspondant au scénario 1. 44
Tableau 3-7: Les types d'accumulation correspondant au scénario 2. 45
Tableau 3-8 : Les types d'accumulation correspondant au scénario 3. 45
Tableau 3-9: Les types d'accumulation correspondant au scénario 4. 45
Tableau 3-10: Poids d'accumulation de givre et de verglas, station 1, saison 2001-2002 48
Tableau 3-11: Tempête de givre et de verglas, station 1, saison 2001-2002. 55
Tableau 3-12: Groupes de stations du réseau SYGIVRE 2001-2004. 56
Tableau 4-1: Événements de tempête de givre et de verglas, groupe 1, saison 2001-2002..... 67
Tableau 4-2: Températures de l'air, groupe 1, saison 2001-2002 67
Tableau 4-3: Poids d'accumulation de givre et de verglas, groupe 1, saison 2001-2002..... 68
Tableau 4-4: Données simulées de précipitations givrantes. 76
Tableau 6-1: Effectifs des événements élémentaires, Groupe 2, Saisons 2001-2002/2003-2004 102
Tableau 6-2: Analyse descriptive des variations des températures moyennes, Groupe 2, Saison 2001-2002/2003-2004 103
Tableau 6-3 : Analyse descriptive des cumulatifs des signaux de givromètre durant les événements de givre et de verglas, Groupe 2, Saison 2001-2002/2003-2004 105
Tableau 6-4: Répartition des types d'accumulation, Station Mont-Bélair, 2001-2002/2003-2004 108
Tableau 6-5: Répartition des types d'accumulation, Groupe 2, Saisons 2001-2004 108
Tableau 6-6: Analyse descriptive des tempêtes de givre et de verglas, saison 2001-2002/2003-2004 110
Tableau 6-7: Analyse descriptive des tempêtes de givre et de verglas, saison 2001-2002 110
Tableau 6-8: Analyse descriptive des tempêtes de givre et de verglas, saison 2003-2004 111

Tableau 6-9: Analyse descriptive des tempêtes de givre et de verglas, groupe 2, saison 2001-2002 111
Tableau 6-10: Analyse descriptive des tempêtes de givre et de verglas, groupe 2, saison 2003-2004 111
Tableau 6-11: Analyses descriptives des poids d'accumulation de givre et de verglas, saison 2001-2002 114
Tableau 6-12: Analyses descriptives des poids d'accumulation de givre et de verglas, saison 2003-2004 114
Tableau 6-13: Données de températures manquantes, groupe 2 116
Tableau 6-14: Sommes des erreurs quadratiques, J = 5 119
Tableau 6-15: Sommes des erreurs quadratiques, J = 6 122
Tableau 6-16: Sommes des erreurs quadratiques, J = 7 124
Tableau G-1: Analyse descriptive des tempêtes de givre et de verglas, groupe 1, saison 2001-2002 158
Tableau G-2: Analyse descriptive des tempêtes de givre et de verglas, groupe 1, saison 2003-2004 158
Tableau G-3: Analyse descriptive des tempêtes de givre et de verglas, groupe 3, saison 2001-2002 159
Tableau G-4: Analyse descriptive des tempêtes de givre et de verglas, groupe 3, saison 2003-2004 159
Tableau H-5: Données de températures manquantes, groupe 1 160
Tableau H-6: Données de températures manquantes, groupe 3 160
Tableau I-7: Sommes des erreurs quadratiques, J = 5. 161
Tableau I-8: Sommes des erreurs quadratiques, J = 6. 161
Tableau I-9: Sommes des erreurs quadratiques, J = 7. 162
Tableau J-10: Sommes des erreurs quadratiques, J = 5. 163
Tableau J-11: Sommes des erreurs quadratiques, J = 6. 163
Tableau J-12: Sommes des erreurs quadratiques, J = 7. 164

LISTE DES ANNEXES

ANNEXE A: Détermination des types d'accumulation de givre et de verglas 152
ANNEXE B: Algorithme 9-13-20-23, estimation des taux d'accrétion de givre et de verglas.. 153
ANNEXE C: Détermination du type d'accumulation d'un événement élémentaire 154
ANNEXE D: Vérification de la sévérité d'une tempête de givre et de verglas..... 155
ANNEXE E: Algorithme d'identification des tempêtes de givre et de verglas..... 156
ANNEXE F: Algorithmes pour la simulation de données virtuelles..... 157
ANNEXE G: Statistique des tempêtes de givre et de verglas du groupe 1 et du groupe 3 158
ANNEXE H: Statistique des données manquantes du groupe 1 et du groupe 3..... 160
ANNEXE I: Résultats des tests avec les données manquantes..... 161
ANNEXE J : Résultats des tests avec les données du groupe 2 163

CHAPITRE 1

INTRODUCTION

CHAPITRE 1

INTRODUCTION

Plusieurs événements historiques révèlent les menaces perpétuelles causés à notre environnement par les phénomènes météorologiques ou climatologiques. On peut évoquer par exemples les cyclones, les tornades et les tempêtes de verglas. En portant notre attention sur les accumulations givrantes ; on constate que les régions froides comme le Canada, la Sibérie, la Norvège, la France et la Chine sont les plus concernées par ces phénomènes. Au Québec particulièrement, les tempêtes de verglas et de givre imposent des pressions énormes sur un vaste réseau de transport électrique. En effet, il y a plus de 140 000 km de lignes électriques aériennes et la sévérité des dommages d'équipements causés par ces phénomènes à chaque saison hivernale est loin d'être négligeable.

Face à cette situation, La Chaire industrielle sur le givrage atmosphérique des équipements des réseaux électriques (CIGELE), en partenariat avec Hydro-Québec depuis 1997, a comme objectif principal de trouver des nouvelles techniques pour rendre plus robuste la résistance des réseaux électriques face aux accumulations givrantes. Pour atteindre ce but, le programme de recherche de la CIGELE est constitué de plusieurs orientations d'études dont *le Traitement des données provenant des sites naturels et modélisations probabilistes* en est une. Il y a aussi le *Développement de méthodes de prévention des accumulations de glace et de neige*.

Le présent ouvrage touche notamment ces deux axes de recherche en se penchant sur les données de givrage fournies par Hydro-Québec ainsi qu'en développant un modèle de prédiction des tempêtes ou événements de givre et de verglas.

Plusieurs études ont précédé celle-ci et elles ont permis l'exploration de la discipline reliée au phénomène de givrage. Des vulgarisations scientifiques et des éclaircissements sur la compréhension des événements de givre et de verglas sont les premiers résultats tangibles de ces études. D'ailleurs, celles-ci ont permis l'obtention de résultats encourageant et prometteur. En dépit de tout, aucun modèle réaliste n'existe jusqu'à présent pour prédire quand et où s'abattra une tempête de verglas et de givre, d'où l'initiative de pousser plus loin la quête des connaissances dans ce domaine. En premier lieu, nous désirons décortiquer les informations cachées derrières les données de givrage, ce qui nous permettra de saisir plus clairement ces phénomènes. En second lieu, nous voulons approfondir la méthodologie concernant la modélisation et la simulation des événements de givre et de verglas en tenant compte des caractéristiques de ces données.

1.1 Problématique

Les ouvrages soumis à des conditions climatiques hivernales sont susceptibles de recevoir une accumulation de glace pouvant nuire à leur intégrité ou à leur fonctionnement. Ce phénomène provoque souvent des perturbations mécaniques et/ou électriques au réseau de transport électrique d'Hydro-Québec.

L'étude des phénomènes de glace atmosphérique cherche à développer une connaissance qui permettrait de prévoir et d'estimer les dangers d'un événement afin de prendre les mesures appropriées pour y faire face [5] [11] [20]. Grâce à la base de données SYGIVRE, il est possible d'extraire et d'étudier les historiques passés des événements de givre et de verglas à chaque station. En effet, SYGIVRE est un système informatisé de gestion d'informations d'Hydro-Québec sur les givrages atmosphériques [14] [17] [40]. Mis en place depuis 1992, des données météorologiques et climatologiques y sont recueillies par une trentaine de stations de mesures couvrant l'ensemble de la province du Québec.

Ainsi, des travaux de recherche ont été réalisés. Dans un premier temps, on assiste à la modélisation des événements de givre et de verglas. Ensuite, on note le développement de modèles de prédiction de ces derniers en se basant sur les données provenant du réseau de stations de mesure SYGIVRE. En résumé, les études antérieures permettent de conclure les faits suivants : les méthodes statistiques standards et l'intelligence artificielle sont des outils potentiels pour l'étude des accumulations de givre et de verglas. Les résultats obtenus

suite à ces recherches sont cependant encore insatisfaisants. Ces problèmes sont attribués principalement à la versatilité de la nature, à la complexité des données vis-à-vis de leur nature spatio-temporelle, au caractère aléatoire des données de givrage, et surtout aux lacunes présentes dans la base de données SYGIVRE. De ces recherches surgissent les problématiques citées ci-dessous :

- des données de saisons d'observations sont manquantes à des stations de mesure du réseau SYGIVRE ;
- aucune base de données n'existe pour valider les modèles de prédictions;
- les saisons de givrage atmosphérique présentent entre elles des différences énormes en ce qui concerne la répartition, la fréquence et la durée des événements de givre et de verglas ;
- les taux de bonnes prédictions obtenus sont encore relativement insatisfaisants: les résultats sont aléatoires et certains groupes de stations présentent de meilleurs taux de bonnes prédictions par rapport à d'autres.
- on ne tient pas compte de tous les paramètres climatologiques et météorologiques, ainsi que des types d'accumulation (givre et verglas). Or, il est important de mener des études en fonction de ces derniers afin de bien saisir les phénomènes de givrage atmosphérique.

À ce stade, nous avons constaté qu'une méthodologie adéquate pour traiter les données manquantes doit être amorcée. L'intégration de ces types de données dans la phase de développement des modèles de prédiction est nécessaire pour éviter la perte

d'informations pertinentes pouvant améliorer ces derniers. Nous sommes donc amenés à continuer la percée dans ce domaine de recherche en proposant des solutions pour résoudre le principal problème connu actuellement : *les lacunes au niveau de la base de données SYGIVRE.*

1.2 Objectifs

1.2.1 Objectif principal

L'objectif principal de la présente étude est d'intégrer les données manquantes au modèle de prédiction des événements de givre et de verglas. Ce modèle de prédiction est basé sur la technique neuronale et la base d'apprentissage est construite à partir des données provenant du système d'acquisition SYGIVRE. Ainsi, une méthodologie d'analyse et de traitement des données recueillies par ce système doit être préalablement élaborée.

1.2.2 Objectifs spécifiques

Le plus grand défi de ce travail de recherche est de faire face aux problèmes rattachés aux valeurs manquantes dans la base de données SYGIVRE. Étant donné que les données des saisons de givrage entre 1992 et 1998 sont les sujets des études antérieures, les objectifs spécifiques de cette étude sont :

- étudier et traiter les nouvelles données recueillies entre les saisons d'observation 1998 et 2004;
- proposer des algorithmes d'identification des tempêtes de givre et de verglas recensées au cours de ces saisons d'observations.

- estimer ou intégrer les valeurs manquantes pour nous permettre d'avoir une base de données plus complète et plus pratique pour la modélisation neuronale des phénomènes de givre et de verglas ;
- développer des modèles de réseaux de neurones pour la prédiction des poids d'accumulation des tempêtes de givre et de verglas à l'intérieur d'un groupe de stations de mesure;
- Simuler des données virtuelles pour valider la méthodologie proposée dans cette étude, et aussi pour valider le fonctionnement des modèles de prédiction.

1.3 Méthodologie

Les données exploitées lors de cette recherche proviennent du réseau de stations de mesure SYGIVRE d'Hydro-Québec [17]. Il contient un ensemble complexe de données de givrage. De plus, des lacunes y ont été observées. Afin d'accomplir les objectifs à atteindre dans le présent travail, nous avons suivi les quatre étapes suivantes:

Étant donné que les données recueillies lors des saisons 1992-1998 sont déjà examinées lors des études précédentes [14], la première étape des travaux effectués est le processus d'analyse et de traitement des données recueillies lors des saisons de givrage 2001-2002 et 2002-2004. Une modélisation de la base de données SYGIVRE a été donc réalisée. Les données brutes recueillies par les stations de mesure sont transformées en des données de tempêtes de givre et de verglas. Ensuite, des analyses statistiques sont effectuées afin de décrire les variables étudiées.

La deuxième étape est le développement du modèle de prédiction des événements de givre et de verglas. Ce modèle est destiné à la prédiction temporelle des poids d'accumulation de givre et de verglas. En plus de l'utilisation des données traitées lors de la première étape de ce travail, cette étude se démarque en introduisant le développement d'un modèle de simulation de données de givrage. Ce modèle permet non seulement de générer des données statistiquement proches des données réelles mais aussi d'avoir la représentation presque parfaite du phénomène de givrage atmosphérique. Ainsi, nous pourrons utiliser des données d'apprentissage virtuelles pour valider le modèle de prédiction avant de tester celui-ci avec la base de données SYGIVRE.

La troisième étape du présent travail consiste à l'intégration des données manquantes. La première partie de cette étape consiste à l'étude des applicabilités des méthodes statistiques d'imputation de données manquantes. Quant à la deuxième partie, elle consiste à l'adaptation de l'algorithme d'apprentissage du modèle de prédiction. Pour ce faire, une technique d'inhibition de certains neurones composant le modèle a été appliquée dans le but d'intégrer les saisons manquantes au niveau de la base de données SYGIVRE.

Enfin, la dernière étape est la validation des modèles de réseaux de neurones en estimant leurs performances. En premier lieu, des tests sont effectués en utilisant les données simulées. La principale raison d'utilisation de ces données simulées est le fait que des tests directs sur les données provenant des sites naturels ne nous donnent pas de

certitude absolue sur la validité du nouvel algorithme intégrant les données manquantes. De cette façon, ces tests préliminaires permettent d'estimer la performance du modèle de prédiction. Enfin, des validations du modèle de prédiction sont effectuées en utilisant la base de données SYGIVRE.

CHAPITRE 2

REVUE DE LITTÉRATURE

CHAPITRE 2

REVUE DE LITTÉRATURE

Depuis longtemps, les chercheurs au sein des entreprises ou des institutions concernées par le phénomène du givrage atmosphérique se sont penchés sur plusieurs aspects concernant l'accumulation de givre et de verglas. D'après les études bibliographiques, l'historique de l'avancement des études réalisées dans ce domaine peut être résumé comme suit : d'abord, il y a eu la conception d'appareils capables de mesurer avec précision les données de givrage sur les sites naturels. Suivirent les études de corrélation statistique entres les différentes mesures ainsi que la détermination des types de distribution appropriés pour chaque variable étudiée. Au fil des ans, il est impressionnant de constater à quel point les techniques utilisées pour aborder les phénomènes de verglas et de givre ont évolué. Ceci est dû en premier lieu aux efforts fournis par les pionniers dans cette discipline de recherche. En deuxième lieu, nous savons bien ce que la technologie apporte à l'évolution de la science : facilité d'acquisition et de stockage des données et la vitesse de calculs grâce aux ordinateurs. De plus, cette évolution est évidente quand on constate la migration des études statistiques vers des études incorporant l'intelligence artificielle. Le concept d'intelligence artificielle forte désigne le projet de créer une machine capable non seulement de simuler un comportement intelligent mais aussi d'avoir une compréhension de ses propres raisonnements. Les sections suivantes présentent en détails la revue de littérature des travaux afférents à la présente étude.

2.1 Études des données de givrage

Le point de départ remonte au début des années 1980 alors qu'on assiste à la mise au point et à l'amélioration des appareils de mesure pour les données des accumulations givrantes. Grâce à ces études [5] [16] [32], deux appareils sont disponibles actuellement pour évaluer périodiquement les taux d'accrétion de givre et de verglas. Le premier appareil, appelé glacimètre (Passive Ice Meter), est de type passif et manuel, tandis que le second, appelé givromètre (Icing Rate Meter), est automatisé. Quoique le glacimètre soit encore utilisé de nos jours, on reconnaît qu'il est moins performant que le givromètre. Le réseau de transport d'Hydro-Québec utilise toujours ces appareils pour le suivi en temps réel des événements de givre et de verglas à travers le Québec.

En ce qui concerne les analyses des aspects spatiaux et temporels des données mesurées sur les sites naturels, nous pouvons constater que plusieurs chercheurs se sont penchés sur différents problèmes relatifs au givrage atmosphérique :

En 1992, Laflamme démontre dans son ouvrage qu'il est important d'étudier les événements de givrage atmosphérique en tant que distributions temporelles et spatiales afin de connaître les régions à risques [20].

En 1996, Elfashny *et al.* font des analyses statistiques sur les données issues de la base de données PIM [24]. Ces dernières couvrent 21 ans d'observations sur 180 stations de mesure. Les variables étudiées sont la vitesse du vent et le taux d'accrétion des

accumulations givrantes. Les résultats de cette recherche indiquent que le type de distribution le mieux approprié pour estimer le taux d'accrétion est le «Log Pearson Type III». En ce qui concerne la vitesse du vent, plusieurs distributions ont été proposées : «Extreme Type I, Extreme Type II, Rayleigh, 3-p Lognormal, Log Pearson Tye III».

Dans la même optique, Chouinard *et al.* [7] ont fait une recherche en 1998, sur le type de distribution statistique applicable concernant l'épaisseur de la glace autour des câbles des lignes de transmission de l'énergie électrique. Par la suite, ils ont construit une carte des régions à risque en fonction de l'intensité des événements de givrage atmosphérique.

Plus tard, Druez *et al.* ont effectué une classification des tempêtes de glace atmosphériques enregistrées sur des lignes expérimentales au Mont Valin [10] [11], au cours des saisons 1986-1998. Tous les événements ont été classifiés en deux catégories : soit le givrage en nuage et les précipitations froides. Cette répartition des tempêtes a permis d'évaluer les charges maximales sur les différents câbles. Il a été conclu que 51.2 % des tempêtes étaient des événements de givre.

Les travaux évoqués ci haut emploient principalement des approches statistiques pour analyser et modéliser les données de givrages atmosphériques. Bien que le domaine de l'intelligence artificielle a été introduit depuis les années 1950, il a fallu attendre le milieu des années 1990 pour que Ohta *et al.*, du Japon, aient l'idée de développer la première

application des modèles de réseaux de neurones pour la prédiction des événements de givre et de verglas [33]. Ce modèle estime l'éventualité d'un désastre avec 3 à 24 heures d'avance en requérant les données des dernières 24 heures. Ses variables d'entrée sont la température, le taux de précipitation et la vitesse du vent. Les résultats de cette recherche indiquaient que ce modèle était assez performant en considérant un seuil de prédiction égal à trois heures.

Suivant la même idée, McComber *et al.* ont exploré en 1998 l'approche neuronale pour prédire le taux d'accrétion des accumulations givrantes sur le site de Mont-Bélair [29]. Les variables d'entrée de ce modèle sont la température, le taux de précipitation, le nombre de signaux de givromètre et la vitesse normale du vent. Les résultats de cette recherche ont démontré que la technique utilisant des neurones donne des résultats assez impressionnants. Il a été noté également que des améliorations sont encore à découvrir et à réaliser. La comparaison des taux de prédiction obtenus par leur modèle de réseau de neurones par rapport à la méthode de régressions multiples montrait que ces deux modèles sont assez compétitifs. Les réseaux de neurones sont cependant plus adéquats en termes de temps de développement et de modélisation.

2.2 Historique des travaux précédents

D'après le survol de la littérature sur les études des données de givrage, nous avons réalisé que bon nombre de projets ont été menés dans les années 1990. Malgré tout, ceci n'a pas atténué la nécessité grandissante des besoins d'acquérir davantage d'investigations sur

les phénomènes de givrage atmosphérique. C'est la principale raison de la naissance de la Chaire industrielle sur le givrage atmosphérique des équipements des réseaux électriques (CIGELE) en 1997. Ainsi, plusieurs chercheurs au sein de ce centre de recherche ont effectué des études afin de prendre la relève dans ce domaine. Nous pouvons citer particulièrement les travaux mentionnés dans les sections suivantes.

Dans leurs activités de recherche, Savadjiev *et al.* se penchent sur une approche statistique des données provenant de la station de Mont-Bélair [35] [36] [37]. Cette station comporte une ligne électrique de 315 kV et elle est dotée de plusieurs appareils de mesures. Particulièrement, on y a installé des capteurs de force pour évaluer avec une certaine précision les taux d'accrétion des accumulations givrantes. Ainsi, ils ont exploité les données recueillies de ce site entre février 1998 et janvier 2000. Les paramètres observés sont : la température ambiante de l'air, la vitesse et la direction du vent, le taux de précipitation au sol et le nombre de signaux de givromètre. Cette étude a permis de proposer des modèles mathématiques pour estimer indirectement le taux d'accrétion des accumulations givrantes; elle présente également les caractéristiques physiques des événements de givre et de verglas. De plus, on y démontre qu'il y a des corrélations très significatives entre les différents paramètres. Ainsi, il a été conclu, par exemple, que le taux d'accrétion est plus important lorsque la température est négative et que l'accumulation de givre ou de verglas est significativement plus abondante quand la direction du vent est parallèle à l'axe de la ligne de transport électrique.

Suivant l'idée de généraliser les études des phénomènes de givre et de verglas, Guesdon *et al.* ont étudié les répartitions des événements de verglas et de givre à travers le Québec [17]. La source de données exploitée lors de cette étude provient d'un réseau de stations d'Hydro-Québec appelé SYGIVRE, les positions géographiques des stations de mesure constituant ce réseau étant données. Cette base de données réunit les informations sur les événements de givrage de 1992 à 1998. On y observe la température extérieure et l'épaisseur de la glace accumulée. Comme l'objectif principal de cette recherche était de définir les liens spatiaux et temporels entre les stations, une matrice de distance a été calculée en utilisant deux types de distance entre chaque paire de stations soit :

- la distance basée sur les fonctions de répartition des événements des stations ;
- la distance basée sur les fréquences des événements des stations sur une période donnée.

Trois grands groupes de stations sont ainsi identifiés : le premier est situé dans la région de Montréal et comporte des stations à dominance d'événements de verglas; le second dans la région de Québec avec des stations sans dominance de type d'événements et le troisième est situé dans la région de la Côte Nord avec des stations à dominance d'événements de givre. En outre, cette recherche a permis d'observer les faits suivants :

- le poids moyen d'accumulation de givre et de verglas s'élève à 209.5 g/m pour l'ensemble des stations. L'écart type est de 544.2 g/m avec un maximum égal à 977.3 g/m et un minimum égal à 7 g/m. Une accumulation peut être constituée à la fois de givre et de verglas ;

- les régions qui reçoivent les plus fortes accumulations sont la région de Québec et celle de la Côte Nord ;
- en moyenne, on retrouve pour une saison de givrage, 31 événements par station dans le cas d'accumulation de givre et 33 événements pour le verglas ;
- les épisodes de verglas durent moins longtemps que ceux de givre ;
- le poids des accumulations de verglas peut atteindre des valeurs plus élevées que celles de givre en raison de leur forte densité. En effet, la densité de verglas varie de 0.8 à 0.9 g/cm^3 alors que celle du givre lourd est comprise entre 0.6 à 0.87 g/cm^3 et le givre léger est à moins de 0.6 g/cm^3 ;
- les prédictions des partages d'événements entre les stations à partir des données d'étude et des paramètres de régressions donnent des résultats relativement satisfaisants. En effet, ceux-ci s'échelonnent de 68% à 88% de bonnes prévisions en fonction du type d'événement et des régions.

Par souci de performance et suite à cette dernière recherche, Larouche *et al.* ont entrepris une étude sur l'applicabilité d'une nouvelle méthodologie [24]. Les réseaux de neurones ont été choisis comme outil de prédiction. En effet, les applications existantes des réseaux de neuronaux dans le domaine de prédiction connaissent un succès assez fascinant [3] [4] [5] [6]. De plus, les tentatives d'exploitation de cette méthodologie pour les phénomènes de givrage donnent des résultats prometteurs [24] [28] [29]. Cette étude vise à explorer ce domaine pour modéliser les phénomènes de givrage atmosphérique. Les données étudiées proviennent uniquement du site expérimental de Mont-Bélair. L'analyse a

été effectuée seulement sur une station de mesure et sur une période de 2 ans (1998 à 2000). Dans la catégorie des architectures statiques de réseau de neurones, les modèles suivants ont été exploités : Perceptron multicouche, le réseau à fonctions de base radiales (RBF). Dans la catégorie des architectures temporelles, on distingue le réseau à réponse impulsionnelle finie (FIR) et le réseau récurrent Elman. Cette étude a permis de conclure les faits suivants pour la prédiction en temps réel, « nowcasting »:

- les réseaux avaient quelque peu de difficulté à estimer les poids d'accumulation correspondant à toutes les observations ;
- les réseaux ont plus de facilité à prédire les données d'apprentissage que les données de validation ;
- une meilleure performance est observée sur l'ensemble des données incorporant uniquement la phase d'accrétion par rapport aux données de l'événement au complet. Ceci corrobore le fait attendu postulant que les données d'accrétion présentent une plus grande homogénéité au sens statistique puisqu'elles font appel à un seul processus physique ;
- une dégradation de la performance est constatée avec l'augmentation du nombre de neurones constituant le réseau neuronal ;
- bien que le réseau FIR soit le meilleur candidat pour la tâche de prédiction du signal du capteur de force, le MLP le suit de près ;
- le réseau FIR présente une certaine tendance à surestimer les petits événements de givrage et à sous-estimer les événements plus grands.

En ce qui concerne la prédiction dans un temps futur, « forecasting », les résultats confirment les faits suivants :

- il y a une concordance plutôt bonne du signal prédit et du signal cible ;
- 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 ;
- l'historique passé du capteur de force comporte un plus grand pouvoir prédictif que le givromètre.

Basée sur les résultats obtenus lors des travaux évoqués ci haut, l'étude de Eter *et al.* consistait à modéliser la propagation spatiale des tempêtes de givre et de verglas [14]. L'architecture d'un réseau de neurones multicouche a été choisie en raison de sa performance et surtout de sa flexibilité pour modéliser le phénomène des accumulations givrantes. Deux modèles de prédiction sont ainsi proposés. Selon la figure 2-1, le premier consistait à prévoir si des événements de givre et de verglas auront lieu ou non au sein des stations issues d'un même groupe, en connaissant les états présents et passés de ces stations.

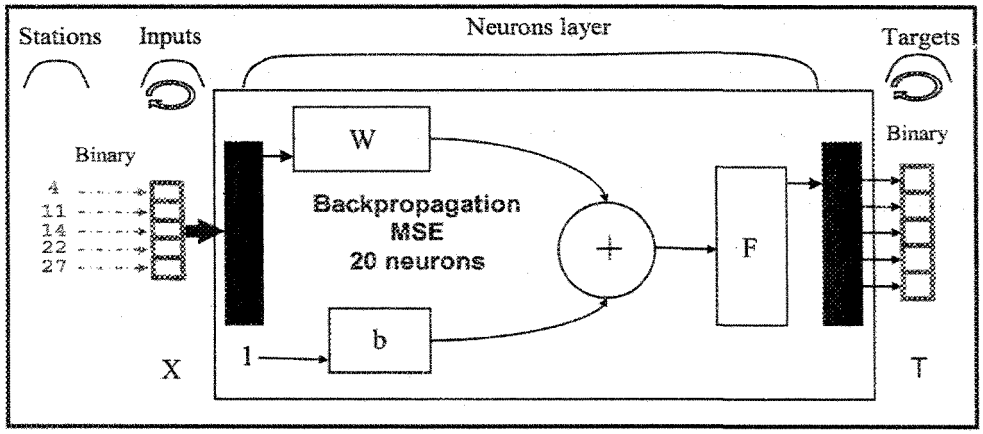

Figure 2-1: Modèle de prédiction des tempêtes de givre et de verglas.

Basé sur ce dernier, un second modèle de prédiction a été développé afin de prédire l'évolution temporelle des poids d'accumulation de givre et de verglas (voir figure 2-2).

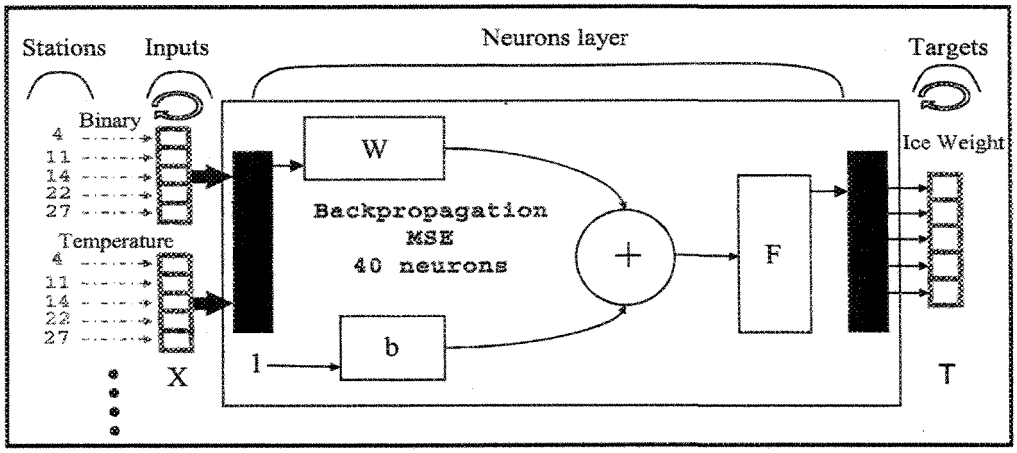

Figure 2-2: Modèle de prédiction des poids d'accumulation de givre et de verglas.

Les observations les plus importantes suite à cette étude sont les suivantes :

- les modèles tendent à donner de bonnes prédictions mais ils ont de la difficulté à détecter les grosses tempêtes ;
- les résultats de prédictions obtenus à chaque groupe de stations sont différents les uns des autres ;
- les taux moyens de mauvaises prédictions pour tous les groupes de stations sont d'environ 10 % en ce qui concerne les données d'apprentissage et 20 % pour les données de test;
- la considération de plusieurs états dans le passé ne donne pas de meilleurs résultats ;
- la propagation spatiale des tempêtes de givrage ne dépend pas seulement des états passés observés par les stations de mesure ;
- la prédiction du poids d'accumulation dépend de la relation spatiale entre les stations constituant un groupe.

Récemment, en 2007, Maralbashi S. et al. ont développé des réseaux de neurone pour la prédiction du type et du taux de glace accumulée sur les lignes aériennes de transport d'énergie électrique [28]. Dans un premier temps, plusieurs configurations de réseau neuronal multicouche ont été testées pour prédire le type de glace. L'étude est basée sur les données provenant de la station de Mont-Bélair. Les paramètres d'entrée choisis sont la température de l'air, la vitesse du vent et la dimension de gouttelette. Ainsi, il a été conclu qu'une architecture de réseau de neurones dont la couche cachée est composée de 10 neurones et les fonctions d'activation de tous les neurones sont sigmoïdales est celle qui

donne les meilleurs résultats pour la détermination du type de glace accumulée. En deuxième partie, plusieurs architectures de réseaux de neurones ont été étudiées en vue de l'estimation des taux d'accrétion de givre et de verglas. On a distingué l'architecture multicouche, le réseau d'Elman et le réseau de Jordan. Les paramètres d'entrées choisis sont la température de l'air, le taux de précipitation, le signal de givromètre, la vitesse du vent ainsi que sa direction. En rapport avec les différentes caractéristiques d'un événement de givre et de verglas, trois modèles de prédictions ont été proposés afin d'estimer les taux d'accrétions. Le premier prend en charge uniquement les données de la phase d'accumulation. Le second est généralisé pour toutes les phases d'un événement de givre et de verglas, c'est-à-dire : la phase d'accumulation, la phase de persistance et la phase de délestage. Enfin, un troisième modèle a été développé spécifiquement pour prédire les taux d'accrétions des événements de verglas. Malgré les limitations causées par l'insuffisance de données pour réaliser convenablement les tests et les validations, les résultats de cette expérimentation ont montré que le modèle de Jordan est le plus approprié pour l'estimation des glaces accumulées sur les lignes de transport aériennes d'électricité.

En résumé, les résultats de ces travaux de recherche nous confirment l'intérêt de l'application des réseaux de neurones pour la prédiction temporelle des phénomènes relatifs aux accumulations des givres et des verglas. En effet, cette technique a permis de produire plusieurs constatations précieuses au niveau des corrélations entre les données de givrage.

2.3 Insertion de la présente étude

L'ensemble des études citées dans la section précédente exploitent la base de données SYGIVRE. Elles utilisent des approches probabilistes et ont comme objectifs principaux le traitement des données de givrage atmosphérique et la modélisation de ces phénomènes.

Dans le même contexte, la présente étude est étroitement liée à ces recherches, elle est à la fois une continuation et une prolongation des travaux effectués dans ce domaine. C'est une continuation puisque cette étude se base sur les connaissances acquises grâce aux précédentes. On parle également de prolongation car notre ultime objectif est de faire un pas de plus en ce qui concerne l'avancement de ces connaissances. Cela concerne principalement la recherche d'une méthodologie pour résoudre les problèmes des données manquantes rencontrés lors des études antérieures. Il est important d'affirmer que le présent travail s'effectue parallèlement et complémentairement à toutes les études recherchant des modèles pour estimer le poids d'accumulation de givre et de verglas sur les conducteurs électriques. On peut citer par exemple le travail effectué par Savadjiev *et al.* [35] [37]. Ce dernier consiste particulièrement à estimer le taux d'accumulation de poids de givre et de verglas par une approche statistique. Aussi, on peut citer le travail de Maralbashi S. qui a proposé une méthodologie spécifique pour identifier les différents types d'accumulation ainsi que l'estimation de leurs taux d'accrétion [28].

Enfin, l'originalité du présent travail repose sur trois points. Le premier concerne le traitement des nouvelles données encore inexploitées dans les études antérieures. Le

deuxième repose sur la recherche de méthodologie pour résoudre les lacunes au niveau de la base de données. Et le dernier point est le développement des modèles de réseaux de neurones adaptés à ces données. D'où le titre de la présente étude: « *Intégration des données manquantes en vue de la modélisation de réseaux de neurones pour la prédiction des événements de givre et de verglas* ».

CHAPITRE 3

MODÉLISATION DE LA BASE DE DONNÉES SYGIVRE

CHAPITRE 3

MODÉLISATION DE LA BASE DE DONNÉES SYGIVRE

Les deux axes principaux de la présente étude sont le traitement des données provenant des sites naturels et la modélisation d'un modèle de prédiction des événements de givre et de verglas. Dans ce chapitre, nous mettons l'accent sur les différents procédés relatifs aux traitements et aux analyses des données disponibles. Cette étape est un préalable pour la poursuite des travaux relatifs au développement d'un modèle de prédiction et à l'intégration des données manquantes.

Depuis plusieurs années, Hydro-Québec est particulièrement impliqué dans l'étude des phénomènes d'accumulation de givrage atmosphérique sur l'ensemble des structures et matériaux formant son réseau de transport d'électricité dans toutes les régions du Québec. Pour cela, l'entreprise a mis en place un réseau de stations de mesure appelé SYGIVRE. Chaque station de mesure est munie d'appareils de mesure des relevés météorologiques et climatologiques. Ces relevés sont caractérisés par leurs natures spatiales et temporelles. Ils représentent le moteur et le pilier central de la présente étude. Moteur, car ces données sont l'ensemble des valeurs des mesures effectuées sur les terrains d'observation. De ce fait, elles représentent les images des événements de givre et de verglas à travers la province du Québec. Pilier central, puisque les descriptions et les analyses de ces données sont des

tâches primordiales avant de procéder au développement du modèle de prédiction fondé sur les données traitées et analysées.

Les sections suivantes décrivent en détails ce processus d'analyse et de traitement des données de givrage atmosphérique. Nous commençons par la présentation de la base de données SYGIVRE. Suivent les descriptions des données disponibles et les données étudiées. Ensuite, la prochaine section est consacrée aux différentes étapes de la transformation de la base de données pour faire ressortir les données relatives aux tempêtes de givre et de verglas aux stations de mesure SYGIVRE.

3.1 Description de la base de données SYGIVRE

SYGIVRE est un système informatisé de gestion d'informations sur les givrages atmosphériques. Appartenant à Hydro-Québec, il a été mis en place depuis 1992 afin de suivre l'évolution temporelle et spatiale des accumulations givrantes. Des données météorologiques et climatologiques y sont recueillies par une trentaine de stations de mesures qui sont réparties principalement le long du fleuve Saint-Laurent et couvrent l'ensemble de la province du Québec. Grâce à ces stations, les données de givrages sont enregistrées périodiquement et transmises au système central par des liaisons satellites, par ligne téléphonique ou par onde [40].

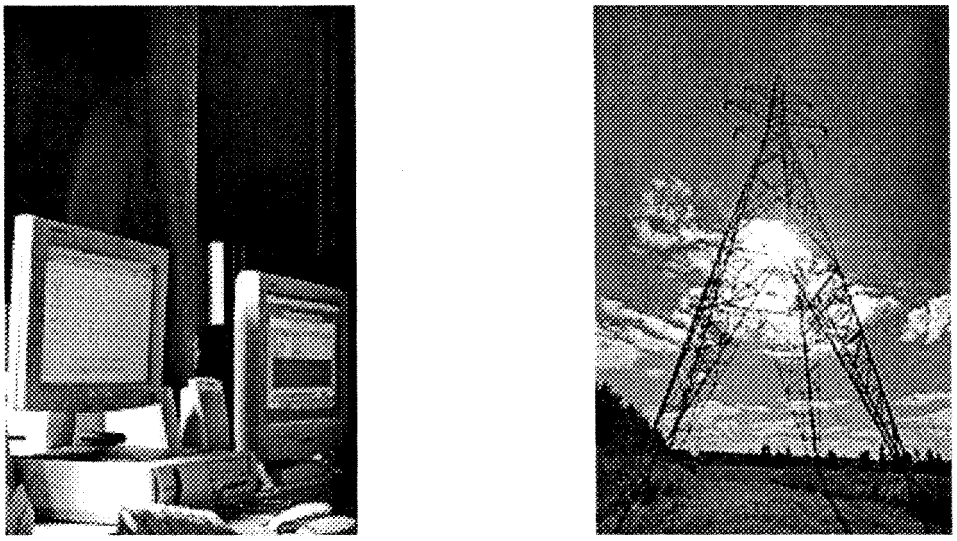

Figure 3-1: SYGIVRE, un système d'acquisition des données de givrage.

Grâce à ce système, il est possible de faire le suivi des historiques passés des événements de givre et de verglas à chaque station. Ainsi, nous pouvons désormais développer un modèle de prédiction basé sur ces données. En bref, SYGIVRE est l'ensemble des stations de mesure et des données de givrage qui y sont collectées.

3.1.1 Stations de mesure

Le réseau SYGIVRE est constitué d'une trentaine de stations de mesure [17] [20]. Celles-ci sont réparties à travers la province du Québec, à proximité des lignes de transport d'énergie électrique. Chaque station d'observation est généralement munie d'un givromètre et d'une sonde de température. De ce fait, les données recueillies rassemblent au moins des relevés météorologiques sur les températures extérieures et des relevés climatologiques sur le taux d'accrétion de givre et de verglas.

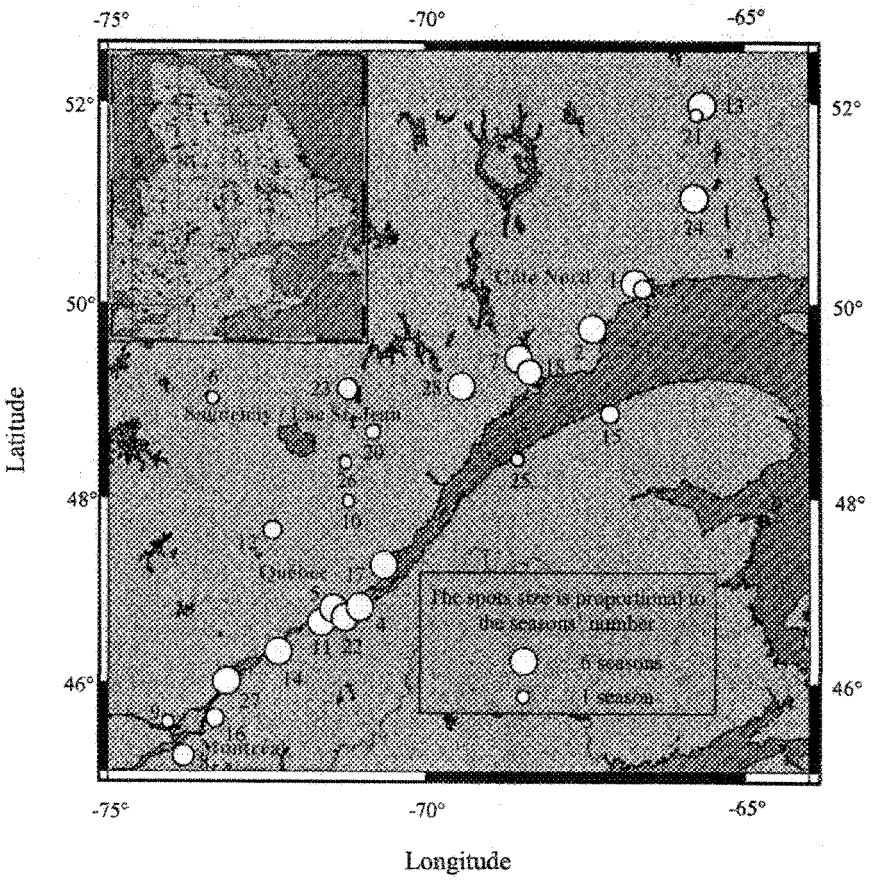

Figure 3-2: Lieux géographiques des stations (1992-1998).

Les études antérieures exploitaient les données de givrages durant les saisons hivernales entre les années 1992 et 1998. En somme, elles couvrent six saisons d'observations. Il a été remarqué qu'une station de mesure ne fournit pas nécessairement les mêmes informations en comparaison avec une autre (voir figure 3-2). En plus de cela, il arrive que certaines d'entre elles soient en pannes ou inactives pour certaines saisons d'observations. Ainsi, les nombres de saisons hivernales observées par les stations du réseau SYGIVRE sont différents.

3.1.2 Givromètre

Installé près de la ligne de transport d'énergie, le givromètre (voir figure 3-3) détecte les dépôts de givre et de verglas et il permet ainsi d'estimer l'intensité des conditions givrantes [28].

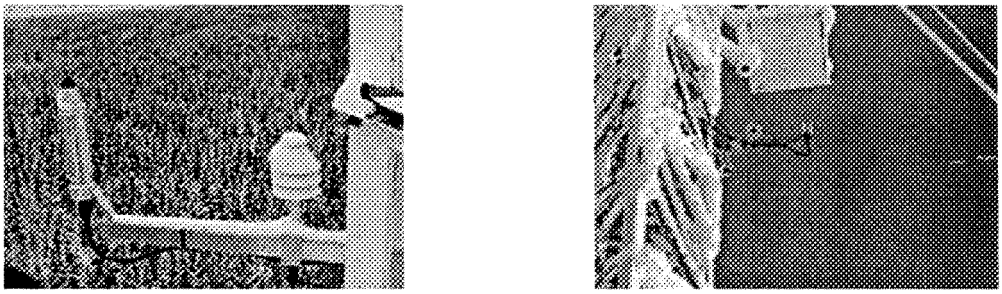

Figure 3-3: Le givromètre et son emplacement.

La sonde du givromètre est un cylindre de 6.2 mm de diamètre et de 25.4 mm de longueur. Il opère à sa fréquence de résonance, égale à 40000 Hz, lorsqu'il est libre de dépôts de glace. Lors d'une accumulation de givre et de verglas sur le givromètre, il y a un changement de la fréquence qui dépend du type d'accumulation. Dans le cas d'une accumulation de verglas ou de givre, on obtient une déviation de la fréquence. 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 la réception de ce signal, le système informatique enregistre un coup de givromètre puis repart en période d'accumulation. S'il y a accumulation d'eau, il se produit une déviation positive de la fréquence de résonance et le détecteur entre dans une phase de nettoyage par chauffage et agitation de la sonde.

3.2 Sélection des variables de givrage étudiées

À la fin d'une saison de givrage, les mesures effectuées dans une station du réseau SYGIVRE sont sauvegardées dans un fichier en format texte. Le tableau 3-1 illustre un échantillon des données de la station de Mont-Bélair durant les saisons de givrages 2001-2002. La première colonne est l'identification de la station de mesure. La deuxième est le nom de la variable mesurée. Les spécifications sur les variables mesurées sont explicitées ci-dessous, telles qu'elles sont décrites selon la source de données originale:

- DIRECT : direction du vent;
- FREVER : fréquence de la sonde de verglas ;
- HUMREL : humidité relative de l'air ;
- INVER : signal d'inversion de verglas ;
- PRECACC : précipitation accumulée;
- TEMPEXT : température extérieure de l'air;
- TENSIO : poids mesuré du conducteur ;
- VERGLAS : signal d'accumulation de verglas ;
- VITESSE : vitesse du vent ;
- VITMAX : rafale de vent ou vitesse maximale du vent.

Il est important de noter que les séries de mesures sont effectuées à intervalles réguliers. Dans le cas de la station de Mont-Bélair, cet intervalle de temps dure un quart d'heure. En effet, la troisième colonne du tableau 3-1 correspond à la date et l'heure

d'acquisition de ces données tandis que la dernière colonne correspond à la valeur de chaque variable.

Station	Variable	Date/Heure	Mesure
BELAIR	DIRECT	2001-10-10 09:30	256
BELAIR	DIRECT	2001-10-10 09:30	300
BELAIR	DIRECT	2001-10-10 09:45	297
BELAIR	FREVER	2001-11-09 23:45	39987
BELAIR	FREVER	2001-11-10 00:00	39987
BELAIR	FREVER	2001-11-10 00:15	39987
BELAIR	HUMREL	2002-03-20 14:45	100
BELAIR	HUMREL	2002-03-20 15:00	100
BELAIR	INVER	2002-06-27 03:00	0
BELAIR	INVER	2002-06-27 03:15	0
BELAIR	INVER	2002-06-27 03:30	0
BELAIR	PRECACC	2002-06-19 11:30	56
BELAIR	PRECACC	2002-06-19 11:45	56
BELAIR	PRECACC	2002-06-19 12:00	56
BELAIR	TEMPEXT	2002-04-14 03:15	5.5
BELAIR	TEMPEXT	2002-04-14 03:30	5.5
BELAIR	TENSIO	2002-06-17 15:45	0
BELAIR	TENSIO	2002-06-17 15:45	0
BELAIR	TENSIO	2002-06-17 15:45	0.43
BELAIR	VERGLAS	2002-01-02 18:30	204
BELAIR	VERGLAS	2002-01-02 18:45	204
BELAIR	VITESSE	2002-06-10 13:00	16
BELAIR	VITESSE	2002-06-10 13:15	7
BELAIR	VITESSE	2002-06-10 13:15	5
BELAIR	VITMAX	2002-05-18 15:30	8
BELAIR	VITMAX	2002-05-18 15:30	18

Tableau 3-1: Données brutes provenant d'une station de mesure.

Afin d'exploiter convenablement la base de données SYGIVRE pour la prédiction des événements de givre et de verglas, nous devons décrire les deux catégories de variables

identifiées pour les données de givrage, c'est-à-dire : les variables mesurées et les variables calculées. Ensuite, une sélection des variables d'étude doit être effectuée.

3.2.1 Variables mesurées

3.2.1.1 Température de l'air

Les stations de mesure du réseau SYGIVRE sont toutes munies d'une sonde de température. Cet instrument enregistre la valeur de la température extérieure de l'air qui est mesurée en degré Celsius (°C). Selon chaque station, des mesures sont effectuées à toutes les heures ou à toutes les demi-heures ou encore à tous les quarts d'heures. En ce qui concerne les disponibilités de cette variable, les analyses de la base de données SYGIVRE [14] [17] ont permis de constater que des valeurs de températures sont manquantes. Ces lacunes sont causées par des bris d'appareils ou d'autres raisons comme des oublis et des erreurs de transmissions. D'autre part, les analyses ont permis également d'observer que certaines valeurs de températures sont aberrantes.

3.2.1.2 Signal du givromètre

Comme les sondes de température, les givromètres sont installés aux stations de mesure du réseau SYGIVRE. Ils détectent par intervalle de temps régulier l'évolution des tempêtes de givre et de verglas. L'unité de mesure de cette variable est un nombre signifiant les coups ou signaux du givromètre. Une accumulation de précipitations froides est constatée par l'augmentation du nombre de signaux du givromètre tandis que l'absence d'accumulation est traduite par la constance de ce dernier. En réalité, la valeur initiale de cette variable est égale à zéro et elle augmente au fur et à mesure qu'il y a de l'accrétion de

givre et de verglas. Elle varie donc entre 0 et 999. Si le nombre de signaux atteint la valeur limite, elle est réinitialisée à zéro. Après les analyses de la base de données SYGIVRE, nous avons observé que certaines valeurs sont manquantes pour des intervalles de temps de la saison de givrage. Aussi, certaines valeurs ne sont pas dans les normes du principe de fonctionnement du givromètre, c'est-à-dire que : soit sa variation n'est pas croissante comme prévue soit qu'elle est discontinue à cause de l'absence de certaines valeurs.

3.2.1.3 Autres variables

Le vent est un facteur influençant l'accumulation de givre et de verglas sur les lignes de transport aériennes. C'est la raison pour laquelle sa vitesse est mesurée à certaines stations. L'appareil utilisé à cette fin est nommé « HYDROTECH » et l'unité de mesure est en kilomètre par heure (km/h). Dans le cadre de cette étude, elle n'a pas été étudiée due au fait que cette variable n'est pas disponible à toutes les stations de mesure.

Les études réalisées par les chercheurs dans le domaine du givrage ont montré que la direction du vent accélère ou pondère les taux d'accrétion de givre et de verglas sur les lignes conducteurs d'énergies électriques. Seulement, quelques stations de mesure sont donc munies d'anémomètres. La direction du vent est donc omise lors des analyses et traitements de données effectuées dans cet ouvrage malgré son importance.

Certaines stations de mesure appartenant au réseau SYGIVRE disposent des moyens et équipements pour mesurer des variables destinées à l'étude du phénomène de givrages atmosphériques. On peut énumérer comme exemples les variables suivantes : l'humidité relative de l'air, la fréquence de la sonde du verglas, le signal d'inversion du verglas et le

poids mesuré du conducteur. En effet, quelques unes d'entre elles sont des sites expérimentaux destinés particulièrement à divers travaux de recherche. Ces variables sont fort intéressantes à étudier mais vu qu'elles ne sont pas disponibles à toutes les stations, nous avons décidé de les omettre dans cette étude.

3.2.2 Variables calculées

3.2.2.1 Taux d'accrétion

D'après la littérature, le taux d'accrétion d'accumulation givrante peut être estimé à partir du nombre de coups de givromètre et du type d'accumulation [23]. Lors de cette étude, l'algorithme utilisé pour le calculer est fondé sur le principe de « l'algorithme 9-23 » [17]. La méthode utilisée pour l'estimation de cette variable est détaillée plus loin dans ce chapitre et son unité de mesure est en *g/m/h*. En ce qui concerne l'algorithme 9-23, son principe de base repose sur le fait qu'un signal du givromètre correspond à 9 grammes de dépôt de givre ou à 23 grammes de dépôt de verglas.

3.2.2.2 Poids d'accumulation

Le poids total d'accumulations givrantes à un instant donné est estimé en effectuant la sommation des taux d'accrétion depuis le début de l'accumulation jusqu'à cet instant. Il est exprimé en *g/m*. Toutefois, il est important de noter que le givromètre ne permet pas d'évaluer le poids exact des accumulations givrantes sur les lignes conductrices d'électricité même s'il est placé à une hauteur plus ou moins égale à ces dernières. Il nous donne tout simplement une estimation du poids d'accumulation. Autrement dit, il nous indique l'intensité de la précipitation courante.

3.2.3 Sélection des variables étudiées

Suite aux descriptions des données de givrage ci-dessus, il a été mentionné que certaines variables ne sont pas mesurées à certaines stations du réseau SYGIVRE. Leurs absences s'expliquent par les manques ou les bris d'appareils de mesures. Ces variables sont : la vitesse du vent, la direction du vent et l'humidité relative de l'air. Étant donnés que les taux d'accumulation et les poids d'accumulation peuvent être estimés à partir des valeurs des températures et des nombres de signaux de givromètre, les seules variables disponibles à toutes les stations sont donc à priori les suivantes: la date et l'heure d'une observation, la température de l'air, le nombre de signaux de givromètres, le taux d'accrétion de givre et de verglas et le poids d'accumulation de givre et de verglas. En somme, elles constituent les variables d'étude dans ce travail en dépit de l'importance des autres paramètres. C'est le choix le plus adéquat puisque nous avons besoin d'une banque de données exploitable et uniforme. Cette uniformité est nécessaire pour l'identification des événements élémentaires de givre et de verglas ainsi que les tempêtes de givre et de verglas lors des saisons d'observation 2001-2002 et 2003-2004.

Après l'étape de sélection des variables étudiées, les données de chaque saison de givrage sont regroupées dans un fichier Excel. La date d'observation, la température de l'air et le nombre de coups du givromètre sont les seules variables conservés. Cela constitue une première ébauche de la base de données SYGIVRE qu'on exploite pour la prédiction des événements de givre et de verglas.

3.3 Identification des événements élémentaires de givre et de verglas

Comme l'indique la figure 3-4, l'identification des événements élémentaires de givre et de verglas est une étape préliminaire au processus d'identification des tempêtes de givre et de verglas ainsi que du regroupement des stations de mesure.

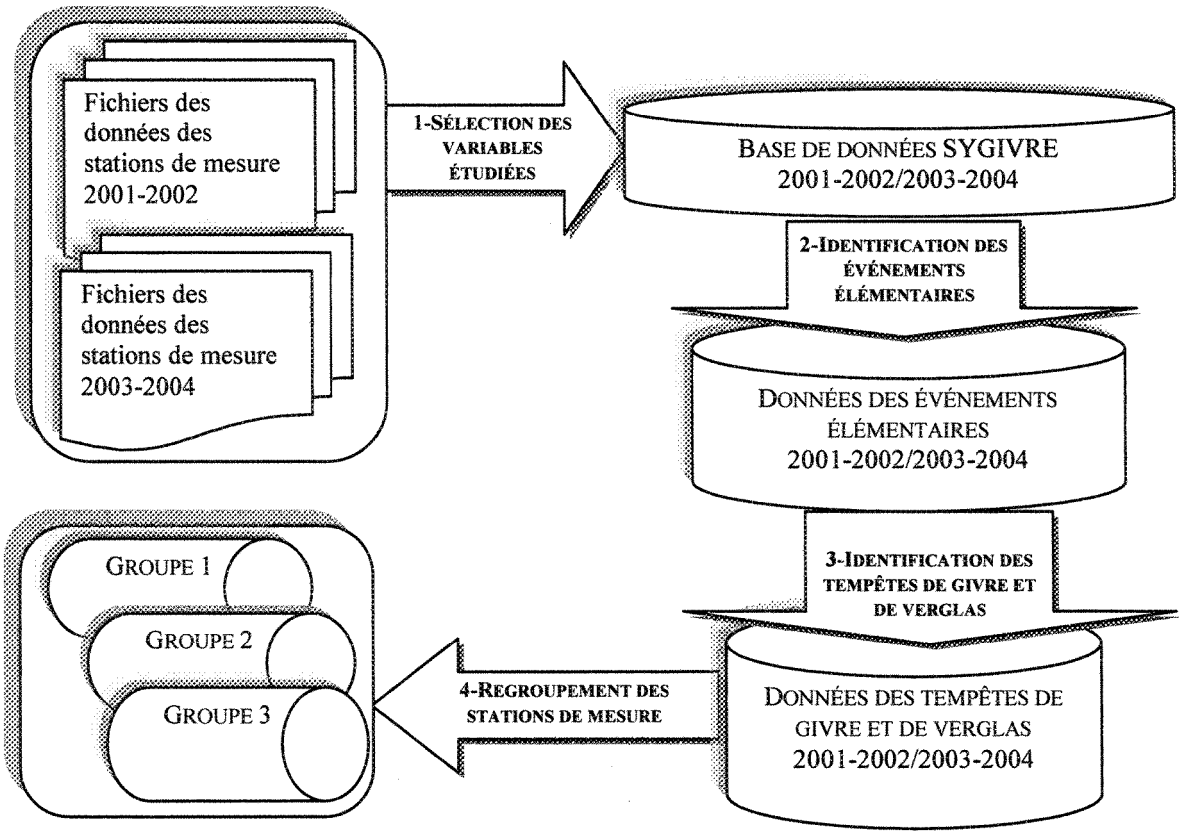

Figure 3-4: Processus de transformation de la base des données SYGIVRE.

Les tableaux 3-2 et 3-3 illustrent des échantillons des données issues de la station Jacques-Cartier lors des deux saisons d'observations 2001-2002 et 2003-2004.

Station	Date/Heure	Temp	NCG
JCARTIER	10/10/2001 9:15	7.9	58
JCARTIER	10/10/2001 9:30	7.9	58
JCARTIER	10/10/2001 9:45	8	58
JCARTIER	10/10/2001 10:00	8.4	58
JCARTIER	10/10/2001 10:15	8.9	58
JCARTIER	10/10/2001 10:30	9.7	58
JCARTIER	10/10/2001 10:45	9.7	58
JCARTIER	10/10/2001 11:00	9.6	58
JCARTIER	10/10/2001 11:15	9.6	58
JCARTIER	10/10/2001 11:30	9.5	58
JCARTIER	10/10/2001 11:45	9.7	58
JCARTIER	10/10/2001 12:00	9.7	58

Tableau 3-2: Données disponibles à une station de mesure, saison 2001-2002.

Station	Date/Heure	Temp	NCG
JCARTIER	13/08/2003 11:30	25.9	127
JCARTIER	13/08/2003 11:45	26.2	127
JCARTIER	13/08/2003 12:00	26.4	127
JCARTIER	13/08/2003 12:15	26.1	127
JCARTIER	13/08/2003 12:30	26.8	127
JCARTIER	13/08/2003 12:45	26.6	127
JCARTIER	13/08/2003 13:00	26.8	127
JCARTIER	13/08/2003 13:15	27.3	127
JCARTIER	13/08/2003 13:30	26.7	127
JCARTIER	13/08/2003 13:45	26.6	127
JCARTIER	13/08/2003 14:00	27.2	127
JCARTIER	13/08/2003 14:15	27.4	127

Tableau 3-3: Données disponibles à une station de mesure, saison 2003-2004

Jusque là, on a vu que les sources de données exploitées dans ce travail proviennent du réseau SYGIVRE. Spécifiquement, on y observe la température de l'air, le nombre accumulé du signal du givromètre. Bien que ces données constituent une banque

d'informations pertinentes pour le phénomène de givrage atmosphérique, elles sont encore muettes et inexploitables si on ne détermine pas son entité fondamentale. Cette entité est l' «événement élémentaire de givre et de verglas». D'où la nécessité de la phase d'identification de ces derniers à partir de la base de données SYGIVRE. Pour ce faire, il est primordial de chercher un moyen juste pour la caractérisation et l'interprétation des données de givrage. Les sections suivantes décrivent ce qu'on sous-entend par événements élémentaires de givre et de verglas. Ensuite, on décrit la procédure utilisée pour les identifier à partir des données des saisons d'observations 2001-2002 et 2003-2004.

3.3.1 Caractéristiques générales du givre et du verglas

En général, la formation de givre ou de verglas se produit seulement à une température inférieure ou égale à 0°C. D'après la littérature, la température de formation de verglas est comprise entre 0°C et -3°C avec une température moyenne égale à -0.99°C tandis que celle pour une accumulation de type givre est comprise entre -3°C et -15°C et la moyenne est égale à -10.71°C [35] [36]. En ce qui concerne la phase de décongélation ou délestage, les études menées sur les données de givrage du site de Mont-Bélair démontrent qu'elle est propice à une température supérieure ou égale à -0.4°C [37]. À part leurs apparences physiques, les types d'accumulation givrante diffèrent par leurs densités. Le verglas a une densité située entre 0.8 g/cm^3 et 0.9 g/cm^3. Le givre lourd a une densité située entre 0.6 g/cm^3 et 0.87 g/cm^3 tandis que celle du givre léger est de moins de 0.6g/cm^3 [17]. Particulièrement, le verglas est un dépôt de glace transparent et lisse. En général, il est

formé par la congélation de la pluie ou de la bruine sur des objets (fils, tours, sols, routes, poteaux). De plus, il est très adhésif, dense et dur. Il peut être formé également par certains brouillards qui, lorsqu'ils sont persistants, donnent des accumulations énormes. Quant à la durée de la phase d'accumulation et de persistance, elle dépend fortement des conditions météorologiques locales : la topographie (crête, vallée), la proximité des sources d'humidité, le chemin préférentiel des tempêtes et aussi des conditions atmosphériques.

3.3.2 Définition d'un événement élémentaire de givre et de verglas

On peut définir un événement élémentaire comme étant une période d'accumulation de givre et/ou de verglas. Autrement dit : il commence dès la première détection d'accumulation par le givromètre et il se termine quand celui-ci signale la fin de cette accumulation.

En général, il y a trois phases dans un événement de givre ou de verglas : la phase d'accumulation, la phase de persistance et la phase de délestage [15]. Dans cette étude et en ce qui concerne la notion d'événement élémentaire, seules les deux premières phases sont considérées vu que les signaux de givromètre ne permettent pas de détecter la phase de délestage. De plus, cette dernière peut être négligée puisqu'elle ne représente aucun danger pour le réseau de transport d'énergie électrique.

Spécifiquement, on caractérise un événement élémentaire de givre et de verglas par une série d'informations temporelles comprenant les variables étudiées. On distingue: la

station de mesure; le numéro d'identification de l'événement élémentaire, noté « Evt »; la température extérieure; le nombre de coups de givromètre, noté « NCG ». De plus, nous avons ajouté une variable calculée à partir des nombres de signaux de givromètre, cette variable étant la différence des nombres de coups du givromètre, noté « DNCG ». Ce dernier permet d'illustrer la phase d'accumulation de givre et de verglas et il nous sert également à détecter la fin d'un événement. Le tableau suivant présente le premier événement élémentaire de givre et de verglas recensé lors de la saison 2001-2002 à la station 1.

Station	Evt	Date/Heure	Temp	NCG	DNCG
1	1	30/11/2001 12:45	-0.4	58	0
1	1	30/11/2001 13:00	-0.3	59	1
1	1	30/11/2001 13:15	-0.3	59	0
1	1	30/11/2001 13:30	-0.2	60	1
1	1	30/11/2001 13:45	-0.1	60	0
1	1	30/11/2001 14:00	-0.1	61	1
1	1	30/11/2001 14:15	-0.1	61	0
1	1	30/11/2001 14:30	-0.1	62	1
1	1	30/11/2001 14:45	-0.1	62	0
1	1	30/11/2001 15:00	0	62	0
1	1	30/11/2001 15:15	0	63	1
1	1	30/11/2001 15:30	0	64	1
1	1	30/11/2001 15:45	0	65	1
1	1	30/11/2001 16:00	0.1	65	0
1	1	30/11/2001 16:15	0.2	65	0
1	1	30/11/2001 16:30	0.2	65	0
1	1	30/11/2001 16:45	0.3	65	0
1	1	30/11/2001 17:00	0.3	65	0
1	1	30/11/2001 17:15	0.2	65	0
1	1	30/11/2001 17:30	0.2	65	0
1	1	30/11/2001 17:45	0.3	65	0

Tableau 3-4: Un événement élémentaire de givre et de verglas, station 1, saison 2001-2002.

3.3.3 Procédure d'identification des événements élémentaires de givre et de verglas

Dans le cadre de ce travail, les données issues du réseau SYGIVRE sont décomposées en une série d'événements élémentaires. Sachant que la procédure d'identification des événements doit se baser sur la définition précédente, il est clair qu'on doit utiliser les signaux des givromètres pour la réaliser. C'est-à-dire que : le début d'un événement élémentaire est lié à une augmentation du nombre de signaux du givromètre. Tandis que la fin de cet événement est le moment où on ne détecte plus de changement au niveau de ces signaux. En ce qui concerne la période de persistance, nous avons supposé qu'elle peut durer au maximum deux heures de temps afin de minimiser l'erreur commise lors de cette phase de traitement de données.

Étant donné que les données d'observation proviennent de plusieurs stations de mesure pour chacune des saisons de givrage 2001-2002 et 2003-2004, l'identification des événements élémentaires est effectuée station par station. En prenant comme illustration les données de la première station (voir tableau 3-4), les événements élémentaires sont numérotés par ordre d'apparition. Ils sont caractérisés par leurs dates de début, leurs dates de fin et leurs durées. En tout, on y a recensé 15 événements élémentaires durant la saison 2001-2002. L'unité de temps pour la durée est en heures et minutes.

3.4 Identification des types d'accumulation de givre et de verglas

Un algorithme d'identification des types d'accumulation a été proposé lors de ce travail afin de prendre en compte l'impact de ces derniers sur la prédiction des poids

d'accumulation de givre et de verglas. Pour ce faire, on s'est basé tout d'abord sur le principe selon lequel une valeur de température supérieure à -3 °C est synonyme d'une accumulation de pluie de verglas tandis qu'une température inférieure à -3 °C indique une accumulation de givre. Ensuite, nous avons introduit deux autres types d'accumulation qui sont des mélanges de givre et de verglas. En effet, on peut rencontrer les deux cas suivants : l'accumulation est constituée d'un fort pourcentage de verglas ou d'un fort pourcentage de givre. Selon les caractéristiques générales du givre et de verglas, l'échelle de températures pour catégoriser ces types d'accumulation est présentée au tableau 3-5.

Type d'accumulation	**Code**	**Température (°C)**
Verglas	10	[-3,+5]
Forte probabilité de verglas	12	[-6,-3[
Forte probabilité de givre	21	[-9,-6[
Givre	20	[-45,-9[

Tableau 3-5: Les types d'accumulation horaires d'un événement élémentaire.

Ce classement est applicable pour un point ou une heure d'observation quelconque de la base de données SYGIVRE. Cependant, un événement élémentaire dure en général plusieurs heures et on assiste à une variation des valeurs des températures. Par conséquent, les types d'accumulation varient aussi en fonction du temps. On peut par exemple avoir du verglas au début d'un événement élémentaire. Ensuite, on aurait une forte probabilité de verglas quand la température diminue. Et après, on aurait une forte probabilité de givre si la

température continue à chuter avant la fin de cet événement. Selon notre point de vue, ce scénario peut se présenter contrairement à l'idée d'assister uniquement soit à une pluie verglaçante soit à un nuage de givre. Une continuité de la variation des types d'accumulation doit donc être assurée pour éviter le compromis entre la détermination des types d'accrétion horaires d'un événement. C'est la raison pour laquelle nous introduisons un deuxième facteur qui est le dernier type d'accumulation juste avant l'heure courante d'observation. Une récurrence est alors réalisée pour modéliser cette continuité de la variation des types d'accumulation et nous avons les scénarios suivants :

- scénario 1 : le dernier type d'accumulation observé est du verglas;
- scénario 2 : le dernier type d'accumulation est fort probablement un verglas;
- scénario 3 : le dernier type d'accumulation est fort probablement du givre;
- scénario 4 : le dernier type d'accumulation est du givre.

Ainsi, de nouvelles échelles de températures sont dressées pour ces différents scénarios. Les tableaux suivants présentent comment les types d'accumulation sont identifiés en tenant compte de tous ces facteurs.

Type d'accumulation	**Code**	**Température (°C)**
Verglas	10	[-6,5]
Forte probabilité de givre	21	[-9,-6[
Givre	20	[-45,-9[

Tableau 3-6: Les types d'accumulation correspondant au scénario 1.

Type d'accumulation	Code	Température (°C)
Forte probabilité de verglas	12	[-9,5[
Givre	20	[-45,-9[

Tableau 3-7: Les types d'accumulation correspondant au scénario 2.

Type d'accumulation	Code	Température (°C)
Verglas	10	[-3,5]
Forte probabilité de givre	21	[-45,-6[

Tableau 3-8 : Les types d'accumulation correspondant au scénario 3.

Type de précipitation	Code	Température (°C)
Verglas	10	[-3,5]
Forte probabilité de verglas	12	[-6,-3[
Givre	20	[-45,-6[

Tableau 3-9: Les types d'accumulation correspondant au scénario 4.

En ce qui concerne les traitements des séries d'observations, les codes sources listés à l'annexe A présentent comment cette méthode de détermination du type d'accumulation a été implémentée en VBA (Visual Basic for Application) sans tenir compte du type d'accumulation durant les dernières heures. Les implémentations des algorithmes utilisés pour les différents scénarios explicités précédemment sont semblables à ce dernier.

3.5 Estimation des taux horaires d'accrétion de givre et de verglas

Lors de cette étude, la recherche d'une méthodologie pour évaluer les taux d'accrétion de givre et de verglas a conduit à l'utilisation d'une version modifiée de la formule nommée : «Algorithme 9-23» [21]. Selon cette formule, un signal du givromètre correspond à 9 g/m de dépôt de givre ou à 23 g/m de dépôt de verglas. Bien qu'elle soit intéressante, elle ne permet pas d'estimer les taux d'accrétion si on introduit les deux autres types d'accumulation : forte probabilité de verglas et forte probabilité de givre. Nous avons alors procédé à une modification de cette formule en faisant des interpolations linéaires entre la densité de givre et celle de verglas. Ainsi, nous stipulons les deux cas additionnels suivants :

- un signal du givromètre correspond à 13 g/m de dépôt d'accumulation givrante s'il est fort probable qu'elle soit du givre. Dans ce cas, la probabilité d'occurrence du givre est supposée égale à 75% tandis que celle du verglas est égale à 25% ;
- un signal du givromètre correspond à 20 g/m de dépôt d'accumulation givrante s'il est fort probable qu'elle soit du verglas. Dans ce cas, la probabilité d'occurrence du verglas est égale à 75% tandis que celle du givre est égale à 25%.

Par conséquent, les paramètres utilisés lors de l'estimation des taux d'accrétion sont la température de l'air, le nombre des signaux du givromètre durant l'heure courante

d'observation et le type d'accumulation durant les dernières heures. La température de l'air et le type d'accumulation durant les dernières heures servent à identifier le type d'accumulation durant l'heure courante. Ensuite, le nombre de signaux du givromètre est utilisé pour calculer le taux d'accrétion en le multipliant par le coefficient correspondant au type d'accumulation. Les codes sources présentés à l'annexe C illustrent en détails cet algorithme. Nous l'avons nommé « Algorithme 9-13-20-23 ». Cette méthode d'estimation de taux d'accrétion est clairement liée aux caractéristiques physiques des verglas et des givres. Nous savons que les verglas sont plus denses que les givres et les combinaisons de ces derniers donnent différentes valeurs de densités d'accumulations givrantes. Quant aux résultats obtenus, on peut les remettre en question mais étant donné que nos approches sont probabilistes, cette formule est acceptable.

3.6 Estimation des poids d'accumulation

Pour chaque événement élémentaire, les poids d'accumulation de givre et de verglas peuvent être calculés à partir de leurs taux horaires d'accrétion. En effet, il est trivial de penser qu'il suffit de cumuler les taux horaires d'accrétion pour avoir les poids d'accumulation de givre et de verglas au cours d'un événement élémentaire. Le tableau suivant présente un événement élémentaire apparaissant à la station 1 pour la saison d'observation 2001-2002. On y distingue le numéro de l'événement, la date et l'heure d'observation, la température de l'air, le nombre de coups de givromètres, le taux d'accrétion et le poids d'accumulation de givre et de verglas.

Evt	Date/Heure	Temp	NCG	T(g/h)	P(g)
5	1/9/02 17:00	-1.9	1	23	23
5	1/9/02 18:00	-1.8	1	23	46
5	1/9/02 19:00	-1.8	0	0	46
5	1/9/02 20:00	-1.7	1	23	69
5	1/9/02 21:00	-1.7	1	23	92
5	1/9/02 22:00	-1.6	0	0	92
5	1/9/02 23:00	-1.6	0	0	92

Tableau 3-10: Poids d'accumulation de givre et de verglas, station 1, saison 2001-2002

3.7 Identification des tempêtes de givre et de verglas

Jusqu'ici, les traitements des données sont effectués en termes d'événements élémentaires à toutes les stations de mesure. Or, notre objectif final est de prédire les évolutions temporelles des tempêtes de givre et de verglas. Il est donc nécessaire de définir les liens entre les événements élémentaires d'accumulations givrantes et les tempêtes de givre et de verglas. Les sections suivantes introduisent la notion de tempêtes de givre et de verglas ainsi que la méthodologie utilisée pour les identifier à partir de la base de données d'événements élémentaires.

3.7.1 Définition d'une tempête de givre et de verglas

Comme il a été remarqué, il est indispensable de considérer la notion de « tempête » de givre et de verglas pour l'étude du phénomène de givrage atmosphérique. Dans le présent travail, on définit une tempête comme un ensemble d'événements élémentaires successifs de givre et/ou de verglas et qui sont liés intrinsèquement entre eux. Une tempête

peut donc être vue comme un ensemble constitué par une ou plusieurs accumulation(s) successive(s) de givre et de verglas.

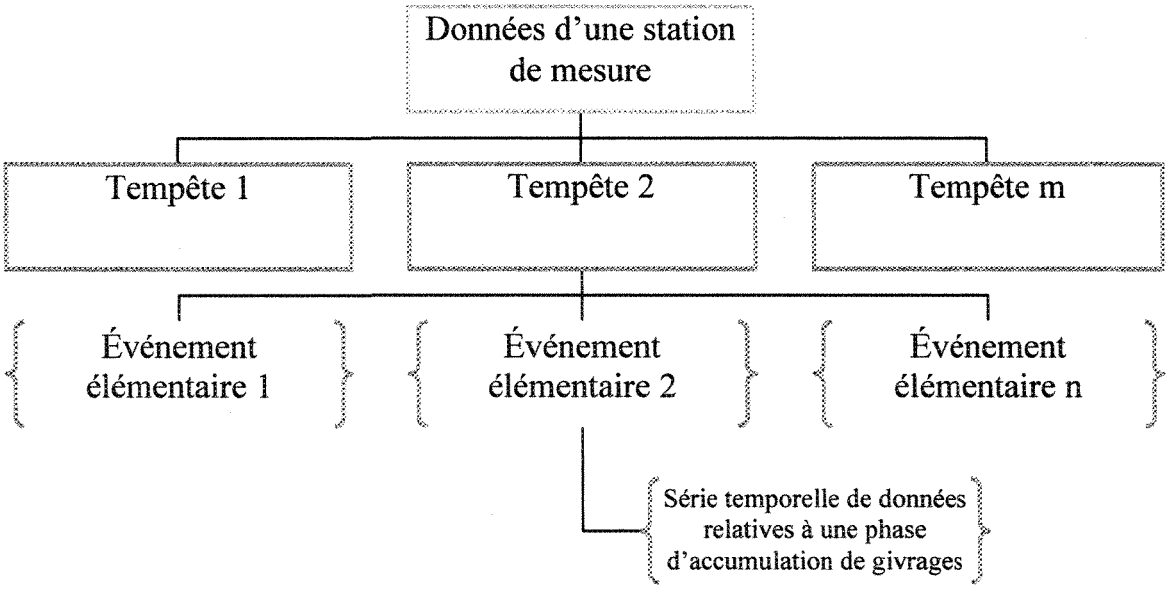

Figure 3-5: Tempêtes de givre et de verglas au sein d'une station de mesure.

Cette définition permet en premier lieu d'avoir une trace de l'évolution temporelle des événements de givre et de verglas. En second lieu, elle nous permet de modéliser ensemble les différentes phases naturelles des accumulations givrantes qui sont : les phases d'accumulation et les phases de persistance.

3.7.2 Procédures d'indentification des tempêtes de givre et de verglas

Afin de retracer les tempêtes de givre et de verglas de la base de données SYGIVRE, le principal facteur pris en compte est l'intervalle de temps entre deux

événements élémentaires successifs. On l'a appelé « temps de latence ». Le deuxième facteur est la corrélation entre les types d'accumulation enregistrés durant ces événements. En dernier, nous avons pris en considération la sévérité de la tempête en question. En effet, on peut assister soit à une forte accumulation givrante, soit à une accumulation moyenne.

3.7.2.1 Considération des temps de latence

Le temps de latence entre deux événements élémentaires successifs est obtenu en faisant une soustraction entre la date du début du deuxième événement et la date de la fin du premier. Pour identifier une tempête, on a établi la première règle suivante : « *deux événements font partie intégrante d'une même tempête si leur temps de latence est inférieur à une valeur limite qu'on doit fixer* ». Il est difficile de déterminer exactement cette valeur car, comme nous le savons, l'occurrence d'une tempête est influencée par plusieurs autres conditions telles que : la température, l'humidité relative de l'air, le vent, la pression atmosphérique, la localité de la tempête elle-même et les conditions climatiques au voisinage de la position géographique de la station étudiée. Pour remédier à cette problématique, nous avons procédé par une approche ad hoc en prenant une valeur limite minimale. La procédure utilisée vérifie en premier lieu si le temps de latence est inférieur ou égal à deux heures pour décider si un événement peut être associé à son précédent et que ces derniers font parties d'une même tempête de verglas et de givre.

3.7.2.2 Considération des types d'événements élémentaires

En ce qui concerne la corrélation entre les types d'accumulation des deux événements élémentaires, le principe de base est fort simple : « *deux événements de types différents ne peuvent pas être considérés comme partie intégrante d'une tempête* ». Ainsi, un événement de givre et un événement de verglas successifs sont considérés comme deux tempêtes distinctes de givre et de verglas. Pour pouvoir appliquer ce procédé, il est d'abord primordial de bien définir les catégories d'événements élémentaires.

Après une analyse fine des données, on a pu établir le fait qu'un événement élémentaire est défini par sa durée, sa température minimale et sa température maximale. Ces informations sont les facteurs pris en compte pour définir le type d'accumulation associé à un événement élémentaire. Le type d'événement élémentaire est différent du type d'accumulation qu'on a utilisé pour les calculs des taux d'accrétion de givre et de verglas. En fait, on peut avoir l'un des cas suivants pour un événement élémentaire quelconque : un événement de verglas, un événement à dominance de verglas, un événement à dominance de givre et un événement de givre. On sous-entend par événement à dominance de verglas un événement dont les deux types d'accumulation sont présents mais le verglas est prédominant. L'explication inverse est valide pour un événement à dominance de givre.

Le principe de base pour l'identification des types d'événement élémentaire est le même que celui utilisé pour l'indentification des types d'accumulation horaires. En effet, il suffit d'étudier les variations des valeurs de températures au sein de ces événements. Quand

la moyenne des températures mesurées au cours d'un événement est supérieure à -3 °C, on fait face à un événement de verglas. Sinon, on est en présence d'un événement de givre. Ensuite, les températures minimale et maximale sont utilisées si un événement présente un caractère dominant de givre ou de verglas. Les codes sources présentés à l'annexe C illustrent la procédure utilisée pour arriver à cette fin.

En tenant compte de ces différents types d'accumulation relatifs à un événement, nous avons ajouté les règles suivantes pour identifier les tempêtes de givre et de verglas :

- un événement de verglas et un événement à dominance de verglas sont des candidats pouvant être associés à une tempête s'ils sont successifs;
- un événement de givre et un événement à dominance de givre sont des candidats pouvant être associés à une tempête s'ils sont successifs;

Compte tenu du fait que les événements de verglas et les événements de givre ont des caractéristiques différentes, nous avons introduit dans l'algorithme d'identification des tempêtes des paramètres distincts pour ces deux cas. En effet, on reconnait particulièrement que les événements de givre durent généralement plus longtemps que les événements de verglas. Par conséquent, nous avons les scénarios suivants pour deux événements élémentaires candidats:

- deux événements candidats de verglas font partie intégrante d'une même tempête si leur temps de latence est inférieur à 4 heures;

- deux événements candidats de givre font partie intégrante d'une même tempête si leur temps de latence est inférieur à 6 heures.

3.7.2.3 Considération des sévérités des événements élémentaires

Quant à la sévérité de la tempête, une forte accumulation de givre ou de verglas est caractérisée par un grand nombre de signaux de givromètres tandis qu'une moyenne accumulation est caractérisée par un nombre moyen de signaux. Pour un événement élémentaire quelconque, la sévérité de l'accumulation peut être connue en vérifiant les nombres moyens et les nombres cumulés des signaux de givromètres. Les codes sources présentés à l'annexe D illustrent cette procédure de vérification de sévérité d'accumulation, les paramètres de la fonction étant le cumulatif des nombres de coups de givromètres et le taux moyen des coups de givromètres. Ainsi, nous avons tenu compte du fait qu'un événement présentant une forte accumulation givrante peut avoir une phase de persistance plus longue. Ce qui veut dire que le temps de latence acceptable doit être défini différemment par rapport aux autres scénarios possibles. C'est la raison pour laquelle nous avons stipulé les conditions suivantes :

- quand on est en présence d'une accumulation forte de verglas ou de givre, le temps de latence maximal acceptable est de 22 heures;
- quand on est en présence d'une accumulation moyenne de verglas ou de givre, le temps de latence maximal est de 10 heures.

Il est important de noter que les valeurs des différents paramètres de cet algorithme d'identification des tempêtes sont définies suite aux analyses des poids d'accumulation des différents essais de calculs. En effet, de mauvais calibrages de cet algorithme nous donnent des valeurs aberrantes de poids d'accumulation de givre et de verglas. Alors, en s'assurant que ces valeurs soient dans les normes, nous avons la quasi-certitude que ces paramètres sont adéquats dans le cadre de cette étude. Le chapitre des résultats et interprétations discutent en détails de ces informations.

3.7.3 Algorithme général

En résumé, l'algorithme d'identification des tempêtes de verglas et de givre au sein d'une station de mesure est décrit par les codes présentés à l'annexe A. Il tient compte des trois facteurs mentionnés précédemment : le temps de latence entre les deux événements élémentaires à intégrer, les types d'accumulation de ces événements et la sévérité de cette tempête. L'application de cet algorithme sur les données des stations pour les saisons 2001-2002 et 2003-2004 nous donne la base de données SYGIVRE destinée à la prédiction des événements de givre et de verglas. En guise d'illustration, le tableau 3-11 présente trois tempêtes apparaissant à la station 1 lors de la saison 2001-2002. On constate dans ce tableau 3-11 que les événements numéros 7, 8 et 9 sont regroupés ensemble et forment la tempête numéro 7. On remarque également que les poids d'accumulation de givre et de verglas sont cumulés d'un événement à un autre quand on est en présence d'une tempête.

SE	Evt	Date/Heure	Temp	NCG	T(g/h)	P(g)
1	1	11/30/01 13:00	-0.4	1	23	23
1	1	11/30/01 14:00	-0.2	2	46	69
1	1	11/30/01 15:00	-0.1	1	23	92
1	1	11/30/01 16:00	0.0	3	69	161
1	1	11/30/01 17:00	0.3	0	0	161
1	1	11/30/01 18:00	0.2	0	0	161
5	5	1/9/02 17:00	-1.9	1	23	23
5	5	1/9/02 18:00	-1.8	1	23	46
5	5	1/9/02 19:00	-1.8	0	0	46
5	5	1/9/02 20:00	-1.7	1	23	69
5	5	1/9/02 21:00	-1.7	1	23	92
5	5	1/9/02 22:00	-1.6	0	0	92
5	5	1/9/02 23:00	-1.6	0	0	92
7	7	2/22/02 1:00	0.0	0	0	0
7	7	2/22/02 2:00	0.0	2	46	46
7	7	2/22/02 3:00	-0.2	0	0	46
7	7	2/22/02 4:00	-0.5	0	0	46
7	8	2/22/02 6:00	-1.5	0	0	46
7	8	2/22/02 7:00	-1.6	1	23	69
7	8	2/22/02 8:00	-1.9	1	23	92
7	8	2/22/02 9:00	-2.0	0	0	92
7	8	2/22/02 10:00	-1.9	0	0	92
7	9	2/22/02 11:00	-1.9	0	0	92
7	9	2/22/02 12:00	-1.7	1	23	115
7	9	2/22/02 13:00	-1.5	0	0	115
7	9	2/22/02 14:00	-1.3	0	0	115

Tableau 3-11: Tempête de givre et de verglas, station 1, saison 2001-2002.

3.8 Regroupement des stations de mesure

Suite à l'étude de la répartition des événements de givre et de verglas à travers la province de Québec [14] [17], les compositions de chaque groupe de stations pour les saisons d'observations 2001-2002 et 2003-2004 sont présentées par le tableau 3-12.

Groupes de stations	Saison 2001-2002	Saison 2003-2004	
	No ID1	**No ID2**	**Nom de station**
Groupe 1 Montréal-Québec	24	9	Beaumont
	28	15	Poste Chateauguay
	29	16	Chenier
	31	*	Donnacona
	34	12	Gentily
	2	36	Ligne exp IREQ
	12	24	Pont Pierre-Laporte
	17	*	Sorel
Groupe 2 Québec-Saguenay Lac Saint-Jean	25	10	Mont-Bélair
	30	17	Tour de Daran
	*	25	Lac Lavoie
Groupe 3 Côte-Nord	22	5	Ste-Marguerite
	*	6	Lac St-Pierre
	23	8	Poste Arnaud
	27	14	Lac Caouette
	33	20	Eric
	9	21	Poste Montagnais
	13	32	Premio
	19	*	Virot

Tableau 3-12: Groupes de stations du réseau SYGIVRE 2001-2004.

Chaque groupe est composé de stations de mesure identifiées par leurs numéros et leurs noms. Étant donné que les traitements des données ont été effectués séparément pour ces deux saisons de givrage, une station n'a pas le même numéro d'identification pour chaque saison. Les données de certaines stations sont absentes pour une saison de givrage (*). Selon ce tableau, nous sommes en présence de trois groupes de stations. Celui de Montréal-Québec et celui de la Côte-Nord sont composés chacun de huit stations de mesure. Tandis que celui de Québec-Saguenay-Lac-Saint-Jean est composé seulement de trois stations de mesure.

3.9 Résumé

En raison des natures spatiales et temporelles des phénomènes de givrages atmosphériques, les données contenues dans la base de données SYGIVRE sont abstraites et complexes. La composante spatiale est liée à la répartition géographique des stations de mesure. La composante temporelle est rattachée aux observations effectuées périodiquement à chaque station de mesures. Afin de décortiquer ces données brutes, un ensemble de procédures a été effectué pour analyser et traiter ces dernières. Autrement dit, une vue montante des données recueillies sur les terrains a été esquissée dans le but de représenter adéquatement les tempêtes de givre et de verglas de chaque groupe de stations SYGIVRE.

Deux saisons d'observations sont étudiées dans le présent travail : la saison 2001-2002 et la saison 2003-2004. On a recensé 35 stations de mesure pour la saison de givrage 2001-2002 tandis que la saison 2003-2004 comporte 39 stations. Les données des saisons 1992-1998 ont été utilisées seulement comme base de données d'informations et à titre de références. Pour la suite de l'étude, on a mis en évidence les variables étudiées en éliminant de l'étude les données non communes à toutes les stations. Les variables éliminées sont : la vitesse du vent, la direction du vent, l'humidité relative de l'air, la fréquence de la sonde de verglas, le signal d'inversion du verglas et le poids mesuré des conducteurs. Par conséquent, les variables étudiées sont : la température de l'air, le nombre de signaux de givromètre, le taux d'accrétion et le poids d'accumulation de givre et de verglas. Après la sélection des variables étudiées, nous avons procédé à la phase d'identification des

événements élémentaires dans cette base de données. Un événement élémentaire est l'entité fondamentale de cette étude car il représente l'élément de base d'un événement ou tempête de givre et de verglas. Les signaux de givromètres sont les paramètres servant à ce recensement. Après cette étape, nous avons procédé aux calculs des poids d'accumulation des givres et des verglas de tous les événements élémentaires de la base de données. Pour y arriver, nous avons tout d'abord proposé une méthode basée sur les valeurs de températures pour définir le type d'accumulation. Quatre types d'accumulation ressortent de cette étude. On distingue : le type givre, le type à dominance de givre, le type à dominance de verglas et le type verglas. Ensuite, nous avons estimé les taux horaires d'accrétion en utilisant l'algorithme nommé « Algorithme 9-13-20-23 ». Cette procédure est basée sur un algorithme existant nommé « Algorithme 9-23 ». En résumé, elle consiste à pondérer les nombres de signaux de givromètres par 9, 13, 20 et 23 grammes selon le type d'accumulation à un instant donné. L'étape suivante est l'identification des tempêtes de givre et de verglas au sein de chaque station. Trois facteurs sont considérés pour caractériser une tempête: le premier étant le temps de latence entre ses événements élémentaires, le second est la corrélation entre les types d'accumulation de ces événements et le dernier facteur est la sévérité des accumulations observées. Ainsi, des événements élémentaires sont vus comme une seule tempête de givre et de verglas après avoir réalisé cette étape d'identification des tempêtes.

CHAPITRE 4

DÉVELOPPEMENT DU MODÈLE DE PRÉDICTION

CHAPITRE 4

DÉVELOPPEMENT DU MODÈLE DE PRÉDICTION

Dans le but de prédire les poids d'accumulation de givre et de verglas au sein des groupes de stations SYGIVRE, un modèle de réseaux de neurones est à mettre en application. Ce modèle est semblable à ceux proposés dans l'étude précédant ce présent travail [14]. Le premier consistait à vérifier si des tempêtes de précipitations givrantes auront lieu ou non au sein des stations issues d'un même groupe. L'historique passé des événements de givrage a été utilisé comme entrée de ce modèle. Basée sur ce dernier, la fonction du second modèle est de prédire l'évolution temporelle des poids d'accumulation de givre et de verglas. Dans ce chapitre, nous décrivons le modèle de réseaux de neurones utilisé pour la prédiction des événements de givre et de verglas au sein d'un groupe de stations de mesure. L'architecture choisie est un réseau de neurones multicouches à rétropropagation de l'erreur. Le développement de ce modèle est fondé sur la base de données SYGIVRE. En premier lieu, nous ne tenons pas compte des données manquantes. L'intégration de ces dernières dans la présente étude est décrite dans le prochain chapitre. Ainsi, nous débutons par une présentation générale de l'architecture de réseau de neurones choisie pour le développement du modèle de prédiction des accumulations de givre et de verglas. Ensuite, nous précisons quelles sont les données d'apprentissage utilisées pour construire ce modèle et les paramètres pris en compte. Enfin, nous présentons la méthode

de validation utilisée pour mesurer son fonctionnement. Particulièrement, on fait référence à la méthode de génération des données d'apprentissage virtuelles.

4.1 Les réseaux de neurones multicouches

Actuellement, il existe plusieurs variantes d'architecture de réseaux de neurones. La plus couramment utilisée s'appelle «réseau multicouche». Nous avons opté pour ce modèle en tant qu'outil de prédiction à cause de sa performance [14] [24] [28]. Un réseau multicouche est organisé en niveaux ou couches qui sont connectés aux unités du niveau suivant et seulement à celles-là. Le réseau est donc sans rétroaction, comporte une couche d'entrée, une ou plusieurs couches de cellules cachées et une couche de sortie. Les cellules cachées sont celles qui ne sont ni à l'entrée, ni à la sortie du réseau.

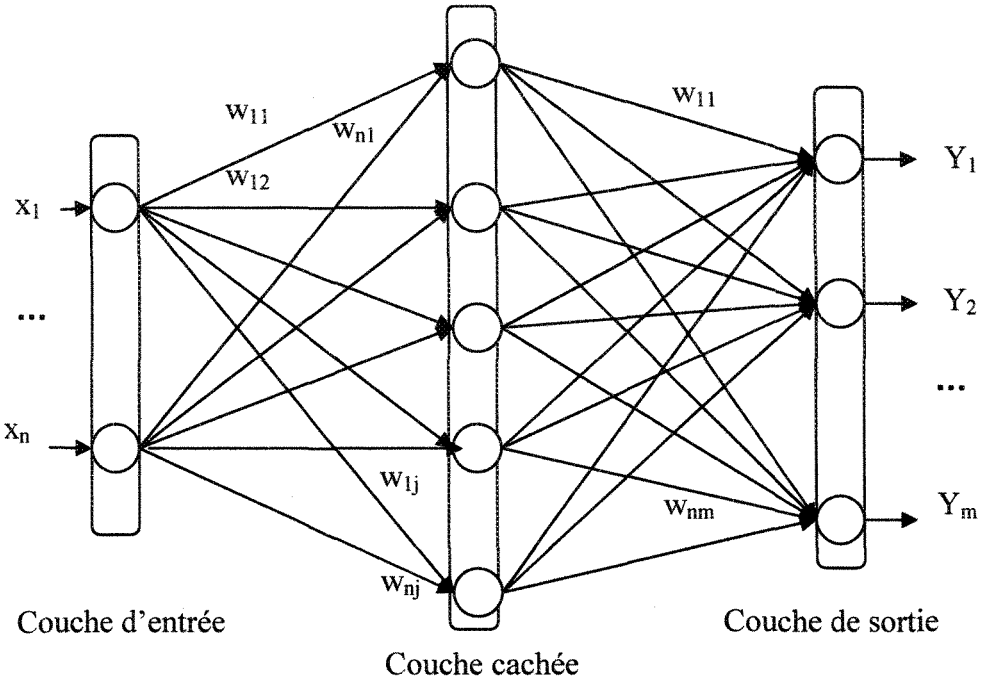

Figure 4-1: Un réseau de neurones multicouches

L'intérêt porté dans cette recherche et dans les études précédentes aux réseaux de neurones tire sa justification des propriétés fascinantes qu'ils possèdent et qui devraient permettre de modéliser et d'étudier un phénomène aussi complexe que le givrage atmosphérique. On peut citer parmi ces propriétés leur parallélisme, leur capacité d'adaptation, leur capacité de généralisation et leur mémoire distribuée [1] [3] [6]. Ces propriétés sont inspirées en principe du fonctionnement du système nerveux de l'homme.

4.1.1 Principes de base

Dans le cas d'une architecture de réseau de neurones multicouches, chaque unité ou neurone i calcule à sa sortie une grandeur $a_i(t)$ en fonction de l'état d'activation $e_i(t)$.

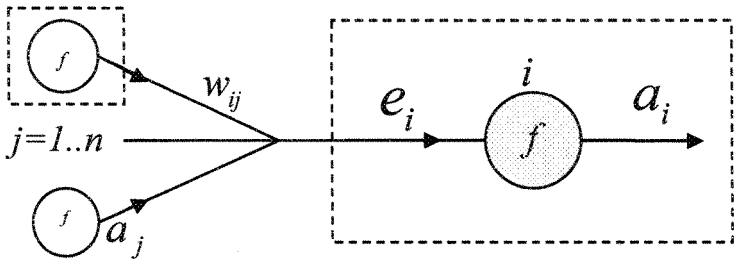

Figure 4-2: Un neurone formel avec ses paramètres.

Selon la figure 4-2 et la formule 4.1, un neurone i est connecté à un nombre fini de neurones j. Pour chaque neurone i, la loi d'activation f doit être une fonction différentiable et croissante appelée fonction d'activation.

$$e_i(t) = \sum_j w_{ij} a_j(t) + \theta_i \quad \text{, avec : } \quad a_i(t) = f\left[e_i(t)\right] \qquad (4.1)$$

Ainsi, on obtient à la sortie du réseau multicouches un ensemble de sorties $\{a_i(n)\}$ qui sont des fonctions non linéaires des entrées de ce modèle.

Dans le cas d'un problème d'approximation de fonction, la fonction de coût à minimiser est généralement de la forme :

$$Q(n) = \frac{1}{2}\sum_i [a_i(n) - s_i(n)]^2 \qquad (4.2)$$

$\{a_i(n)\}$ sont les sorties obtenues avec le réseau de neurones et $\{s_i(n)\}$ sont les sorties espérées.

4.1.2 Techniques d'optimisation

D'après la littérature [6], les réseaux de neurones affichent une meilleure performance au niveau de leur capacité de généralisation si les conditions suivantes sont respectées :

- les entrées et les sorties ont un bon degré de corrélation;
- les données d'apprentissage sont assez nombreuses pour représenter la fonction à estimer.

En général, il n'existe pas de méthodologie bien précise pour déterminer le nombre optimal de neurones dans la couche cachée. Néanmoins, on suppose qu'elle dépend intrinsèquement du nombre de neurones à l'entrée du réseau, de la taille des données d'apprentissage, de la complexité de la fonction à calculer approximativement, du degré de

bruit présent dans les données, de l'architecture choisie et enfin de l'algorithme d'apprentissage.

4.2 Architecture générale du modèle de prédiction

Selon les données disponibles à chaque station de mesure et en tenant compte de la notion de tempête, l'architecture générale des modèles de réseaux de neurones est représentée par la figure 4-3.

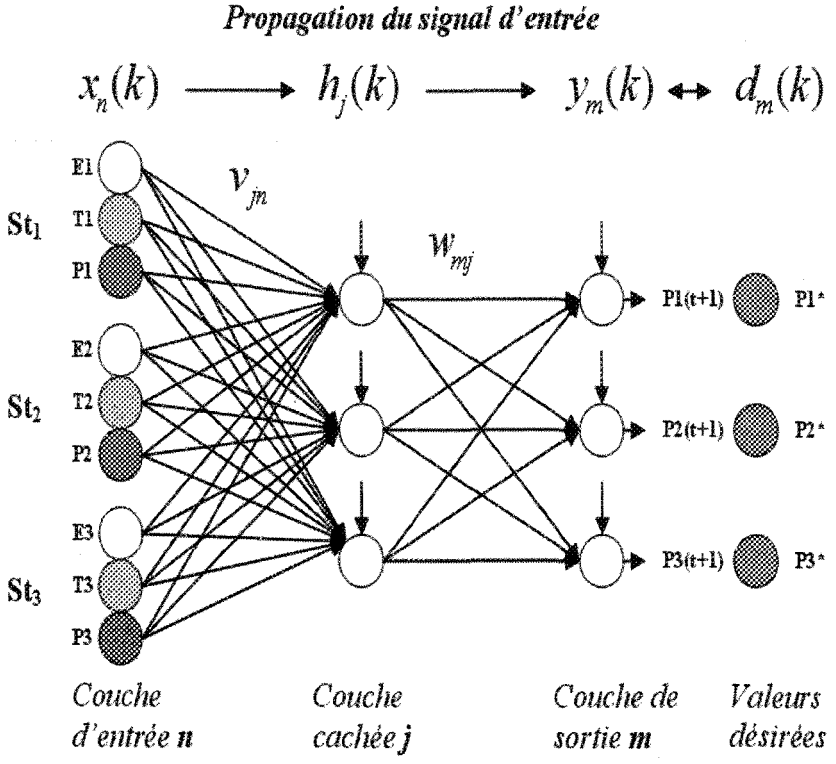

Figure 4-3: Architecture générale des modèles de réseau de neurones pour la prédiction des événements de givre et de verglas

Les variables d'entrée sont: les états des stations d'un groupe « *E* », les températures extérieures moyennes «*T* » de l'air observées dans ces stations et les poids d'accumulations de givre et de verglas « *P* » calculés à partir des taux d'accrétions. Les variables de sortie sont les poids d'accumulation de givre et de verglas « *P(t+1)* » à un temps t+1 si les variables d'entrée sont prises à un temps t. Les variables désirées sont les poids d'accumulation « *P** » selon les observations.

L'état « E » est une variable indiquant si une tempête a lieu ou non à une station. Une valeur égale à 0 signifie qu'il n'y a pas de tempête, une valeur égale à 1 signifie qu'il y a une tempête, et une valeur négative égale à -1 signifie qu'on ne peut pas affirmer l'occurrence d'une tempête à cette station. Les valeurs des températures ambiantes de l'air et les poids d'accumulation sont pris tels qu'ils sont à partir des données des tempêtes de givre et de verglas répertoriées à chaque groupe de stations.

En prenant comme exemple le groupe du Saguenay – Lac-Saint-Jean, les trois stations de mesure composant ce groupe font ressortir une architecture de réseau de neurones à neuf cellules dans sa couche d'entrée et trois cellules dans sa couche de sortie.

4.3 Les données d'apprentissage du modèle de prédiction

Afin de prédire les événements de givre et de verglas, nous avons besoin d'un ensemble de données d'apprentissage pour entrainer le modèle de réseau de neurones. Les variables d'entrée du modèle de prédiction à développer dans cette étude sont : l'événement

de givre et de verglas, la température de l'air et le poids d'accumulation de givre et de verglas.

Pour les variables d'états des stations, le tableau d'événements (tableau 4-1) présente pour chaque station les occurrences de tempêtes de givre et de verglas. Ces données sont extraites du groupe de stations de la région de Saguenay Lac Saint-Jean pour la saison 2001-2002. Les valeurs zéros signifient qu'il y a absence d'accumulation de givre et de verglas et les valeurs non nulles signifient leur présence. Un numéro est assigné à chaque tempête selon son ordre d'apparition dans une station. Selon ces spécifications, on peut observer dans le tableau 4-2 qu'aucune tempête de givre et de verglas n'est enregistrée aux stations 24, 28, 31. Pour les autres stations, une ou plusieurs tempêtes sont repérées et leurs durées peuvent être calculées à partir de leurs dates de début et leurs dates de fin. On a par exemple une tempête commençant le 11/28/2001 0:00 A.M. d'une durée de trois heures à la station 29. Ensuite, une deuxième tempête d'une durée de cinq heures est repérée à partir de midi le 11/29/2001.

En ce qui concerne les variables de températures et de poids d'accumulation de givre et de verglas, les tableaux 4-2 et 4-3 nous illustrent respectivement les variations des valeurs de ces dernières à chaque station d'un groupe et à chaque heure d'observation. Les zones ombrées correspondent aux occurrences des tempêtes de givre et de verglas. Il est important de noter que les données illustrées par ces tableaux ne présentent aucune donnée

manquante. En effet, il n'y a aucune saison manquante au groupe de stations 1 durant la saison 2001-2002.

	Groupe de stations 1							
Date/Heure	**24**	**28**	**29**	**31**	**34**	**2**	**12**	**17**
11/10/2001 19:00	0	0	0	0	0	0	1	0
11/10/2001 20:00	0	0	0	0	0	0	1	0
11/28/2001	0	0	2	0	3	0	0	0
11/28/2001 1:00	0	0	2	0	3	2	0	0
11/28/2001 2:00	0	0	2	0	0	2	0	0
11/29/2001 12:00	0	0	3	0	0	3	0	0
11/29/2001 13:00	0	0	3	0	0	3	0	0
11/29/2001 14:00	0	0	3	0	0	3	0	0
11/29/2001 15:00	0	0	3	0	0	3	0	2

Tableau 4-1: Événements de tempête de givre et de verglas, groupe 1, saison 2001-2002

	Groupe de stations 1							
Date/Heure-Station	**24**	**28**	**29**	**31**	**34**	**2**	**12**	**17**
11/10/2001 19:00	0.40	2.80	0.50	-1.00	-1.40	0.10	-0.15	0.10
11/10/2001 20:00	0.60	2.60	0.20	-0.80	-1.20	0.20	0.20	0.20
11/28/2001	-1.30	0.00	-1.03	-2.40	-1.18	-0.60	-1.50	-1.30
11/28/2001 1:00	-1.60	-0.70	-1.18	-2.70	-0.95	-0.68	-1.70	-1.40
11/28/2001 2:00	-1.40	-0.60	-1.18	-2.60	-1.50	-0.63	-1.60	-1.50
11/29/2001 12:00	-4.70	0.00	-4.00	-5.90	-4.90	-2.63	-4.80	-3.50
11/29/2001 13:00	-4.40	1.20	-2.68	-5.60	-4.90	-1.55	-4.50	-2.90
11/29/2001 14:00	-4.00	2.00	-1.80	-5.30	-4.80	-0.38	-4.30	-2.60
11/29/2001 15:00	-4.00	1.40	-1.98	-5.30	-4.60	-0.23	-4.10	-2.65

Tableau 4-2: Températures de l'air, groupe 1, saison 2001-2002

	Groupe de station 1							
Date/Heure-Station	**24**	**28**	**29**	**31**	**34**	**2**	**12**	**17**
11/10/2001 19:00	0	0	0	0	0	0	23	0
11/10/2001 20:00	0	0	0	0	0	0	46	0
11/28/2001	0	0	23	0	23	0	0	0
11/28/2001 1:00	0	0	46	0	46	23	0	0
11/28/2001 2:00	0	0	69	0	0	46	0	0
11/29/2001 12:00	0	0	40	0	0	23	0	0
11/29/2001 13:00	0	0	120	0	0	115	0	0
11/29/2001 14:00	0	0	220	0	0	184	0	0
11/29/2001 15:00	0	0	440	0	0	253	0	23

Tableau 4-3: Poids d'accumulation de givre et de verglas, groupe 1, saison 2001-2002.

En ce qui concerne les valeurs des différentes variables constituant les données d'apprentissage, elles ont été échelonnées entre -1 et 1 afin d'éviter l'oscillation des poids synaptiques des modèles à construire. Cette technique est couramment utilisée dans la phase d'apprentissage des réseaux de neurones.

4.4 Apprentissage du modèle de réseau de neurones

L'apprentissage du modèle de réseau de neurones peut être considéré comme le processus de la mise à jour de ses poids synaptiques afin que le réseau puisse relier les entrées aux sorties. Il existe plusieurs types et règles d'apprentissage (supervisé, non supervisé). Dans ce travail, on a utilisé l'apprentissage supervisé et l'algorithme de rétropropagation par la mise à jour des poids du modèle [3] [6].

4.5 Les paramètres d'apprentissage des réseaux de neurones

En principe, un modèle de réseau de neurones est construit par essais. Des paramètres lui sont attribués arbitrairement. Ensuite, une phase d'apprentissage est effectuée pour lui permettre de lire les liens entre l'ensemble des données d'entrée ou données d'apprentissage. Essentiellement, trois variables sont à prendre en compte lors de cette phase d'apprentissage. On distingue le nombre de cellules dans la couche cachée *J*, le taux d'apprentissage *mc* et le moment d'apprentissage *mu*. En premier lieu, le nombre de cellules dans la couche cachée doit varier afin de sélectionner le modèle optimal permettant de généraliser la propagation spatiale et temporelle des événements de givre et de verglas. Il s'agit d'une technique d'optimisation par construction ou par essais. En ce qui concerne le taux d'apprentissage et le moment d'apprentissage, ces deux paramètres permettent de résoudre les problèmes d'oscillations des poids synaptiques du réseau multicouches quand on applique l'algorithme de rétropropagation de l'erreur. Plusieurs modèles ayant des paramètres différents peuvent donc être étudiés dans le but de trouver le meilleur modèle dédié à un groupe de stations de mesure.

4.6 Les fonctions de performances

Après avoir réalisé une phase d'apprentissage, deux fonctions de performances sont prises en compte pour évaluer les fonctionnalités du modèle de prédiction. La première est basée sur les erreurs commises par le modèle de prédiction, elle est nommée « *somme quadratique des erreurs*». La seconde fonction est le « *pouvoir prédictif du modèle* ». Ce

dernier est basé sur la fonctionnalité du modèle de prédiction vis-à-vis des résultats de prédiction.

4.6.1 Somme quadratique des erreurs

Les erreurs à la sortie du réseau de neurones se définissent comme les différences entre ses sorties et les sorties désirées vis-à-vis d'un signal d'entrée quelconque. Pour un signal d'entrée, la somme des erreurs quadratiques se définit comme la somme des carrés des erreurs des cellules de la couche de sortie. Ainsi, la somme quadratique des erreurs pour toutes les données d'apprentissage, notée *SSE*, est obtenue par la sommation totale des sommes des carrés des erreurs correspondant à toutes ces données. Son unité est le g^2/m^2.

4.6.2 Pouvoir prédictif d'un modèle

On dit qu'un modèle de réseau de neurones a un pouvoir prédictif acceptable par rapport à un groupe de stations s'il prédit correctement les évolutions temporelles des tempêtes de givre et de verglas. À l'inverse, on dit que le pouvoir prédictif d'un modèle de réseau de neurones est insignifiant s'il ne prédit pas comme attendu les événements à l'une de ces stations. En fait, il est inacceptable d'avoir une station non prise en compte par le modèle de prédiction, d'autant plus que le but de cette étude est la prédiction des accumulations de givre et de verglas à l'intérieur de chaque groupe de stations de mesure SYGIVRE.

4.7 Génération de données d'apprentissage virtuelles

Dans le but de créer une base de données externe et indépendante des données de givrage provenant du réseau SYGIVRE, une simulation de données virtuelles a été réalisée. Sachant que les données d'apprentissage des modèles de réseaux de neurones sont constituées par les variables caractérisant les tempêtes de givre et de verglas, le modèle de simulation doit reproduire ces variables. La première étape est alors la génération des variables d'états de chaque station pour une période de temps choisie arbitrairement. La propagation spatiale des événements de givre et de verglas à l'intérieur d'un groupe doit être assurée pour refléter la corrélation existante entre les stations. Ensuite, les valeurs des températures et les taux de précipitation de l'air sont générés. Enfin, les poids d'accumulation de givre et de verglas sont calculés en fonction des ces deux dernières variables. Les sections suivantes décrivent les différentes étapes effectuées pour obtenir cette base de données virtuelle.

4.7.1 Simulation des états des stations

La première étape de la création des données de givrage est la simulation des états des stations « ***Ei*** » composant un groupe donné. Ces variables indiquent les occurrences des tempêtes au sein de chaque station. En ce sens, le modèle de simulation doit générer les durées des tempêtes « ***Di*** » et les décalages de temps « ***Dti*** » des tempêtes propagées entre les stations. Après avoir étudié les répartitions de ces deux variables selon la base de données SYGIVRE, nous avons opté de les simuler de telle manière que leur distribution

statistique suive une loi normale. Comme l'indique le tableau 4-4, nous connaissons les espérances mathématiques et les écarts-types de ces paramètres.

Saison	2003-2004	2001-2002
Moyenne	8	7
Médiane	5	5
Écart-Type	12	6
Variance	146	41
Maximum	176	55
Effectif	455	522

Tableau 4-4: Statistique descriptive des durées des tempêtes, saisons 2001-2004

Ces valeurs d'écart-type et de moyenne ont été prises comme paramètres initiaux lors de la génération des données. Ensuite, elles ont été calibrées en fonction des résultats finaux. Dans l'exemple illustré à la figure 4-4, nous sommes en présence d'un groupe composé de trois stations de mesure. Les deux premières tempêtes durent 8 heures de temps tandis que la troisième dure 11 heures. Le décalage de temps entre les deux premiers événements est égal à une heure et il est égal à 4 heures pour les deux derniers. L'algorithme pour générer les états des stations est présenté à l'annexe F.

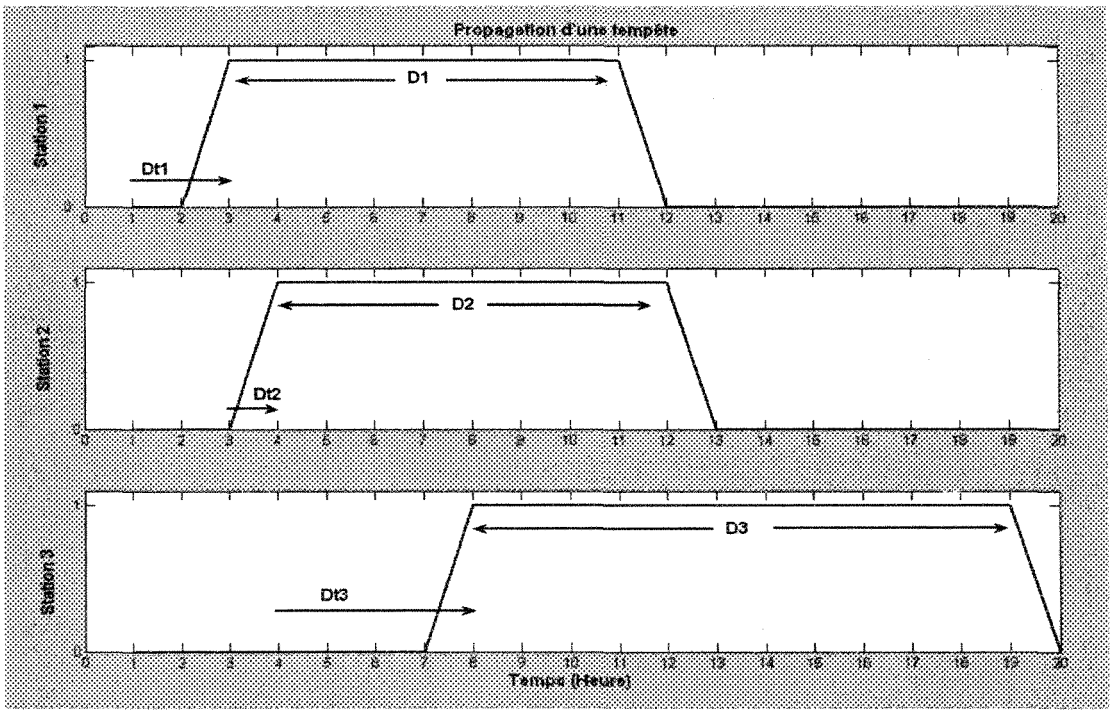

Figure 4-4: Propagation temporelle des tempêtes au sein d'un groupe de stations

4.7.2 Simulation des températures de l'air

Contrairement aux états des stations, la variation des valeurs des températures est complexe à modéliser [12]. De plus, l'étude des données de température provenant de la base de données SYGIVRE nous amène à conclure qu'elle ne peut pas être présentée par une distribution normale. Nous avons alors opté pour une solution ad-hoc dont le principe consiste simplement à l'utilisation des données de températures enregistrées lors des saisons d'observation 2001-2002 et 2003-3004. Les valeurs de températures sont donc tirées de la base de données 2001-2004 au lieu de les générer aléatoirement. Notons que cette variable est statistiquement indépendante des états des stations. Par contre, un tirage

d'une série de valeurs de températures à partir de la base de données SYGIVRE correspond à une tempête simulée de givre et de verglas.

4.7.3 Simulation des taux de précipitations

En ce qui concerne les valeurs des taux de précipitations, nous avons décidé de faire référence aux données provenant du site du Mont-Bélair pour l'étude de leurs variations [18] [36]. Après l'étude de variation de cette variable, nous avons considéré qu'elle est distribuée normalement avec une espérance mathématique et un écart-type. Comme la méthode de génération des états des stations, les valeurs de ces paramètres sont préalablement fixées pour chaque station.

Une fois les paramètres de chaque distribution de taux de précipitations fixés, nous avons simulé une corrélation entre les paires de stations afin de refléter l'interdépendance des stations de mesure. Ainsi, la valeur moyenne des taux de précipitation de la première station est générée aléatoirement. Le taux moyen de précipitation de la deuxième station est une fonction linéaire de celui de la première station, tandis que celui de la troisième station dépend des valeurs de précipitations moyennes des deux premières stations. Les codes sources présentés à l'annexe F décrivent l'algorithme utilisé pour ces calculs de précipitation moyennes pour trois stations de mesure.

4.7.4 Simulation des taux d'accrétion horaires

Après avoir généré toutes les variables requises (***Ei***, ***Ti***, ***Pi***), les taux d'accrétion horaire correspondant aux séries de données sont calculés en fonction des valeurs de températures et des taux de précipitation. En référence aux études réalisées sur l'accrétion de givre et de verglas sur les lignes conductrices d'Hydro-Québec [31] [35] [36], plusieurs formules sont proposées afin d'estimer les taux d'accrétion et les poids d'accumulation. Toutefois, les utilisations de ces formules sont contraignantes à cause de leurs paramètres et de leurs conditions d'application. À titre d'exemple, une équation établie en étudiant les données d'une station de mesure particulière n'est pas directement applicable pour une autre station de mesure.

Dans le cadre de cette simulation de données, une fonction linéaire a été choisie pour générer les valeurs des taux d'accrétion (voir annexe F). Le choix des coefficients dans cette formule d'approximation est partiellement arbitraire. En effet, nous devons les calibrer pour obtenir des données non aberrantes et statistiquement proches des données du réseau SYGIVRE. Cette méthode est tout à fait valide puisque notre but principal est de générer des données d'apprentissage virtuelles mais non pas de modéliser les distributions statistiques des paramètres des données de givrage atmosphérique.

4.7.5 Calculs des poids d'accumulation de givre et de verglas

Pour terminer la simulation des données des tempêtes de givre et de verglas, les poids d'accumulation des précipitations (***PAi***) sont calculés en cumulant les valeurs des taux d'accrétion horaire. Le tableau 4-4 illustre un échantillon des données simulées lors de

cette étude. On observe clairement, à travers les trois premières colonnes, la propagation spatiale et temporelle d'une tempête au sein des trois stations.

E1	E2	E3	T1	T2	T3	P1	P2	P3	PA1	PA2	PA3
0	0	0	-0.99	-0.96	-1.74	0	0	0	0	0	0
0	0	0	-1.96	-1.56	-1.33	0	0	0	0	0	0
1	0	0	-1.37	-1.61	-1.32	3	0	0	41	0	0
1	1	0	-1.55	-1.33	-1.67	2	7	0	79	61	0
1	1	0	-1.88	-1.51	-1.73	3	4	0	125	109	0
1	1	0	-2.15	-2.04	-1.60	4	6	0	179	171	0
1	1	0	-2.06	-1.97	-1.55	3	8	0	226	243	0
1	1	1	-1.99	-2.07	-1.46	3	5	5	273	300	52
1	1	1	-2.19	-2.12	-1.43	5	5	6	332	358	109
1	1	1	-2.32	-2.07	-1.90	7	6	6	401	421	170
1	1	1	-2.13	-2.25	-1.93	4	6	5	455	485	226
0	1	1	-1.89	-2.28	-2.12	0	4	6	0	540	290
0	0	1	-0.62	-1.98	-2.06	0	0	5	0	0	347
0	0	1	-0.22	-1.38	-2.09	0	0	5	0	0	405
0	0	1	0.21	-1.13	-2.30	0	0	4	0	0	460
0	0	1	0.72	-0.47	-2.45	0	0	5	0	0	521
0	0	1	1.55	-0.31	-2.29	0	0	3	0	0	570
0	0	1	1.74	-0.20	-2.30	0	0	5	0	0	630
0	0	1	2.51	-0.08	-1.75	0	0	7	0	0	695

Tableau 4-4: Données simulées de précipitations givrantes.

4.8 Résumé

En résumé, un réseau de neurones agit comme une boîte noire, il réalise une fonction complexe et non linéaire de ses entrées pour calculer ses sorties; ses performances peuvent être mesurées en faisant des approches statistiques entre les sorties calculées et les sorties exactes.

Pour la présente étude, on peut dire que le réseau de neurones multicouches est un bon outil pour traiter les données manquantes et pour réaliser la prédiction des événements de givre et de verglas. L'entrée d'un modèle de prédiction pour un groupe de stations est un ensemble de *n-uplets* correspondant aux variables d'entrée des *n* stations constituant ce groupe. Chaque *n-uplet* est formé de trois. On distingue la variable d'état, la température extérieure et le poids d'accumulation à cette station. Quant à la sortie du réseau de neurones, on retrouve un ensemble de *n* valeurs de prédiction de poids d'accumulation. De ce fait, ce modèle prédit les poids d'accumulation de givre et de verglas à partir des conditions connues aux stations de mesure d'un groupe.

Pour mesurer les fonctionnalités de chaque modèle de réseau de neurones développé dans le présent travail, deux fonctions de performance ont été utilisées. La première, appelée « somme des erreurs quadratiques », est basée sur les erreurs de prédiction de ce modèle. Un modèle de prédiction d'événements de givre et de verglas est optimal si sa somme d'erreurs quadratiques est minimale. La seconde fonction de performance, appelée « pouvoir prédictif du modèle », est basée sur les résultats de prévision du modèle en

question. Ce dernier critère de performance s'appuie sur le fait que tous les événements à chaque station d'un groupe doivent être prédits avec succès pour pouvoir affirmer que le modèle est fonctionnel.

Enfin, la phase de validation des modèles de réseaux de neurones pour prédire les tempêtes de givre et de verglas est subdivisée en deux étapes distinctes. La première consiste à valider le modèle de prédiction avec des données simulées. L'utilisation des données simulées permet de mesurer adéquatement leur performance et de s'assurer de la propreté de la base de données étudiée. Quant à la deuxième étape, elle consiste à utiliser les données provenant de la base de données SYGIVRE afin de tester la fonctionnalité du modèle de prédiction.

CHAPITRE 5

INTÉGRATION DES DONNÉES MANQUANTES

CHAPITRE 5

INTÉGRATION DES DONNÉES MANQUANTES

En principe, la modélisation d'un modèle de prédiction pour les tempêtes de givre et de verglas nécessite l'existence d'une base de données exempte de lacunes. Ces dernières nous empêchent l'utilisation directe des modèles de réseaux de neurones. D'un autre côté, les données brutes ne peuvent pas être exploitées pour la prédiction. Effectivement, la construction de modèles neuronaux procède de l'analyse de données et non de la magie : les variables doivent être explicatives et les exemples doivent être suffisants. C'est la raison pou laquelle il a été nécessaire de modéliser les données collectées par les stations de mesure avant d'attaquer le développement du modèle de prédiction. Lors des réalisations de ces processus, des problèmes sont constatés au niveau de la base de données SYGIVRE. Principalement, on y observe des données d'observations manquantes en ce qui concerne les données saisonnières. Aussi, certaines stations n'ont aucune observation pour certaines saisons. Ce chapitre décrit les natures des problèmes engendrés par ces lacunes ainsi que les techniques utilisées pour les résoudre. En premier lieu, nous avons fait usage de méthodes statistiques d'imputation. En second lieu, une amélioration ou adaptation de l'algorithme d'apprentissage standard du modèle de prédiction a été effectuée afin de prendre en compte la présence de saisons manquantes.

5.1 Traitements statistiques des données manquantes

La présence de données manquantes dans une base de données est un sérieux problème dans plusieurs domaines de recherche. Dans le but d'appréhender la méthodologie adéquate pour l'intégration de ces types des données, une étude de faisabilité des méthodes d'imputations des valeurs manquantes a été réalisée pour ce projet. Les procédures d'imputation pour les données manquantes sont utilisées depuis plus de 50 ans par les statisticiens. « Imputer » veut dire combler les valeurs manquantes par des valeurs prédites ou estimées. L'objectif de ces méthodes n'est pas de rechercher la plus grande précision ou de décrire les données de la meilleure façon possible. Les objectifs sont plutôt de refléter correctement l'incertitude des valeurs manquantes, de préserver les aspects importants des distributions des variables de la base de données et de préserver les relations importantes entre ces variables.

Ainsi, on présente des méthodes généralement utilisées pour traiter les problèmes de données manquantes. En premier lieu, nous commençons par introduire les types de données manquantes et leurs définitions générales. Ensuite, les différentes méthodes pour les traiter sont exposées au fur et à mesure qu'on avance dans les prochaines sections.

5.1.1 Types de données manquantes

Lorsqu'on se trouve confronté à des données manquantes, la première question à se poser est : leur apparition est-elle due au hasard ou, au contraire, à des raisons déterministes? La réponse à cette question oriente de manière fondamentale la prise en compte des données manquantes. Nous allons donc préciser les termes de cette question en

introduisant les définitions de Little et Rubin [25] [26]. D'après ces derniers, on distingue trois types de données manquantes : les données manquantes au hasard, les données manquantes complètement au hasard et les autres types de données manquantes.

5.1.1.1 Données Manquantes au Hasard (DMH)

On dira qu'une donnée manque au hasard si l'occurrence de son absence est indépendante de la valeur qu'elle prend [2] [7]. Autrement dit, les données manquantes pour une même variable peuvent être considérées comme un sous-échantillon aléatoire de l'échantillon initial. Si les données manquantes sont de type DMH, leur prise en compte ignore les raisons de leur absence et doit s'appuyer sur les relations entre les variables.

Dans le cadre de cette étude, des valeurs manquantes pour une courte période de temps sont classées comme des données manquantes au hasard. Les variables concernées sont la température de l'air et le nombre de signaux du givromètre. En effet, leurs absences sont dues à des erreurs accidentelles, c'est-à-dire : des pannes ou des disfonctionnements des appareils de mesure pendant ces périodes de temps. D'autre part, il y a aussi les erreurs fortuites, c'est-à-dire : des erreurs humaines. Dans tous ces cas, elles sont considérées comme des données manquantes au hasard car leurs absences sont indépendantes des valeurs qu'elles prennent.

5.1.1.2 Données Manquantes Complètement au Hasard (DMCH)

On dira qu'une donnée manque complètement au hasard si l'occurrence de son absence est indépendante de toutes les valeurs que prend l'individu qui présente cette donnée manquante [2] [7]. C'est-à-dire: l'ensemble des données manquantes complètement

au hasard constitue un sous-échantillon aléatoire des valeurs prises par l'échantillon initial. Par contre, l'ensemble des données manquantes au hasard ne constitue pas nécessairement un sous-échantillon aléatoire des valeurs initiales. Contrairement aux données manquantes de type DMH, les données manquantes de type DMCH doivent être prises en compte en ignorant les raisons de leur absence. Par conséquent, on peut ignorer les raisons entre les variables.

En pratique, il est habituellement difficile de rencontrer le cas de données manquantes complètement au hasard. Ce sont les données manquantes au hasard qui sont plus souvent rencontrées. Dans le cadre de cette étude, aucune donnée ne manque complètement au hasard.

5.1.1.3 Autre Type de Données Manquantes (ATDM)

On englobe dans ce type les données manquantes dont la distribution n'est pas aléatoire et dont les estimations ne sont pas faisables à partir des autres variables présentes dans la base de données [2] [7]. Si les données manquantes ne sont pas de type DMH et DMCH, leur prise en compte doit s'appuyer sur les raisons de leur absence.

Dans ce travail, les valeurs manquantes pour une longue période de temps sont classées dans la catégorie des autres types de données manquantes. Peu importe les raisons de leurs absences, les estimations de ces données sont impossibles à partir des autres variables de la base de données. Ainsi, si toutes les données d'une saison entière ne sont pas présentes à une station de mesure quelconque, aucune méthode d'imputation ne peut être effectuée.

5.1.2 Méthodes de traitement des données manquantes

Cette section présente les méthodes les plus connues pour manipuler des données manquantes. Certaines d'entre elles consistent à rejeter les éléments constituant le sous-ensemble de données manquantes et d'autres consistent à remplacer ces données manquantes par des valeurs approximatives. Cette liste n'est pas exhaustive, mais elle couvre certaines des approches les plus largement utilisées.

5.1.2.1 Rejet des valeurs manquantes

Si dans une analyse particulière, une occurrence présente des données manquantes pour n'importe quelle variable utilisée, la méthode de rejet consiste à omettre cette occurrence de l'analyse. Cette approche est mise en application par défaut dans certains logiciels éditeurs de solutions décisionnelles et analytiques tels que SAS et SPSS.

Après les analyses des données provenant du réseau SYGIVRE, on a constaté que des valeurs de températures et des valeurs de signaux de givromètres sont manquantes pour certaines plages de données horaires. Si les valeurs de signaux de givromètres sont absentes, les lignes correspondantes à ces dernières sont supprimées de la base de données. La première raison est l'impossibilité des calculs des taux horaires d'accrétion sans les signaux de givromètres. La deuxième est tout simplement d'éviter des erreurs éventuelles. Si les valeurs de températures sont absentes pour une plage de données quelconques, des analyses sont d'abord effectuées par rapport à la durée de ces lacunes avant de prendre la décision de rejeter l'ensemble des informations rattachées à ces lacunes. En effet, la méthode de remplacement par la moyenne et celle par la régression linéaire peuvent être

appliquées dans certaines circonstances. Les sections suivantes introduisent ce en quoi consistent ces dernières.

5.1.2.2 Rejet des paires de valeurs aberrantes

Pour des variables conjointes, cette méthode effectue l'analyse sur les paires de données disponibles. On garde uniquement les données explicatives pour l'étude en question et les paires non corrélées sont rejetées des données d'analyse. Cette approche est disponible dans un certain nombre de logiciels tels que SAS et SPSS.

Des suppressions de paires de valeurs aberrantes ont été effectuées lors des processus d'analyse et de traitement de données quand les valeurs de températures et les valeurs des nombres de signaux de givromètres ne présentent pas de corrélations significatives. Un ou plusieurs signaux de givromètre observé à une température supérieure à 10 degré Celsius est un exemple concret de valeurs aberrantes. Or, cette situation est présente plusieurs fois dans la base de données SYGIVRE. Un deuxième exemple est la présence d'un grand cumulatif de signaux de givromètres en l'espace d'une heure. L'ensemble des valeurs recueillies présentant ces anomalies a été aussi rejeté.

5.1.2.3 Remplacement par la moyenne

La méthode de remplacement par la moyenne substitue les données manquantes par la valeur moyenne calculée à partir des données disponibles. Cette méthode a été appliquée lors des traitements de données pour des valeurs de températures manquantes pendant des courtes périodes de temps. Sachant que le risque d'avoir une forte variation des valeurs de températures est minimal pour un petit laps de temps, la température moyenne est calculée

à partir des valeurs de températures observées juste avant et juste après la présence de lacunes de données. Ce principe fort simple permet d'avoir une juste estimation des températures manquantes.

5.1.2.4 Remplacement par une régression linéaire

La méthode de remplacement par une régression linéaire consiste à développer une équation de régression basée sur des données disponibles. Cette méthode a été appliquée dans les cas des valeurs de températures absentes ponctuellement, c'est-à-dire : pour une donnée manquante pour une heure d'observation seulement. Dans ce cas, l'équation de régression se traduit par l'interpolation linéaire entre les valeurs de températures juste avant et juste après cette heure d'observation. Évidemment, cela est réalisable à condition que ces valeurs existent. Vu que cette approche a tendance à augmenter la corrélation entre les variables, une amélioration peut être apportée en ajoutant un caractère aléatoire sur le modèle de régression obtenu.

5.1.2.5 Autres méthodes d'imputation

Plusieurs autres méthodes sont connues dans le domaine de traitement de données manquantes. On peut citer, relativement au présent travail, les méthodes d'imputation « Hot deck » et « Cold deck ».

La méthode d'imputation « Hot deck » identifie l'occurrence qui est plus proche du cas avec une valeur absente et ensuite substitue la valeur manquante par la valeur trouvée dans cette occurrence. La première étape est l'identification de l'occurrence qui ressemble le plus à l'occurrence présentant une donnée manquante. Cela peut amener dans certaines

analyses à introduire des notions de distances entre occurrences. Une fois cette étape réalisée, la valeur manquante est imputée par le champ correspondant. Bien que cette méthode semble être intéressante pour les traitements de valeurs manquantes au niveau de la base de données SYGIVRE, son application s'avère inadéquate dû à la complexité des relations entre les variables d'études. En effet, une même variation de température peut correspondre à plusieurs scénarios de variations des signaux des givromètres. De plus, nous savons pertinemment que les phénomènes de givrages présentent des caractères complexes du point de vue spatial et du point de vue temporel. Il est alors impossible de trouver un algorithme ou une méthodologie précise pour déterminer qu'une occurrence d'informations de givrage (température et signal du givromètre) est plus proche d'une autre présentant une donnée manquante.

Quant à la méthode d'imputation « Cold deck », elle ressemble à la méthode d'imputation « Hot deck ». La différence est le fait que l'analyste impute la valeur manquante par une valeur dérivée d'une source extérieure ou d'une recherche antérieure. Comme la méthode d'imputation « Hot deck », l'imputation « Cold deck » s'avère aussi un choix intéressant. Cependant, elle n'a pas été appliquée pour traiter les données manquantes dans ce travail pour les mêmes raisons présentées ci-haut. De plus, nous n'avons pas une source de données extérieure et les traitements des données lors des recherches antérieures sont différents de ceux proposés dans la présente étude.

5.2 Révision de l'algorithme d'apprentissage du modèle de réseau de neurones

D'après notre recherche, aucun algorithme d'apprentissage de réseaux de neurones ne permet d'intégrer les valeurs manquantes. Par définition, intégrer les valeurs manquantes lors de la phase d'apprentissage signifie l'utilisation d'un ensemble vide de données pour une station de mesure quelconque. Théoriquement, un tel processus biaise le modèle de réseau de neurones à construire et alterne ses paramètres, c'est-à-dire : ses poids synaptiques. Il est cependant nécessaire dans la réalisation de ce projet. Ceci a pour objectif d'éviter le rejet d'une importante quantité d'informations disponibles. Une recherche de méthodologie pour réviser l'algorithme standard d'apprentissage de réseaux de neurones a été alors effectuée. Le but recherché lors de cette révision est de trouver une technique intégrant les données manquantes tout en minimisant les probabilités de fausses prédictions du modèle de réseau de neurones. Afin d'illustrer la méthodologie suivie pour arriver à cette fin, nous exposons dans les sections suivantes l'algorithme standard d'apprentissage. Par la suite, les solutions pour la révision de ce dernier sont présentées.

5.2.1 Algorithme standard d'apprentissage

Pour apprendre les corrélations entre les données de givrage, le réseau de neurones utilise la technique de rétropropagation de l'erreur afin d'ajuster ses paramètres et pour généraliser les relations entre ses variables d'entrée et ses variables de sortie. Dans cette section, nous présentons l'algorithme standard d'apprentissage du réseau de neurones.

Étant donné un groupe de stations de mesure, nous avons les termes suivants :

St_N : Station de mesure;

$x_n(k)$: Signal d'entrée du réseau de neurones *"input"*;

$h_j(k)$: Signal de sortie de la couche cachée;

$y_m(k)$: Signal de sortie de la couche de sortie *"output"*;

$d_m(k)$: Signal désiré selon les observations *"target"*;

v_{jn} : Poids synaptiques entre la couche d'entrée et la couche cachée;

w_{mj} : Poids synaptiques entre la couche cachée et la couche de sortie;

$\delta_m(k)$: Signal d'erreur de la couche de sortie;

$\delta_j(k)$: Signal d'erreur de la couche cachée;

N , J , M : Nombres des cellules ou neurones respectivement dans les trois couches du réseau de neurones.

Le signal d'entrée $x_n(k)$ du réseau de neurones est composé des différentes variables de chaque station de mesure St_N. Le signal de sortie $h_j(k)$ est l'ensemble des poids d'accumulation de givre et de verglas prédits par ce modèle à chaque station. Le signal désiré $d_m(k)$ est l'ensemble des poids d'accumulation à prédire selon les données d'apprentissage. Les poids synaptiques v_{jn} et w_{mj} sont les paramètres du modèle de réseau de neurones durant la phase d'apprentissage. Les signaux d'erreurs $\delta_m(k)$ et $\delta_j(k)$ sont des valeurs calculées en fonction du signal de sortie et du signal d'entrée. Enfin, nous

considérons ici que les nombres de cellules dans les trois couches du réseau de neurones sont respectivement égaux à N, J, M. La figure 5-1 illustre le principe de base de cet algorithme.

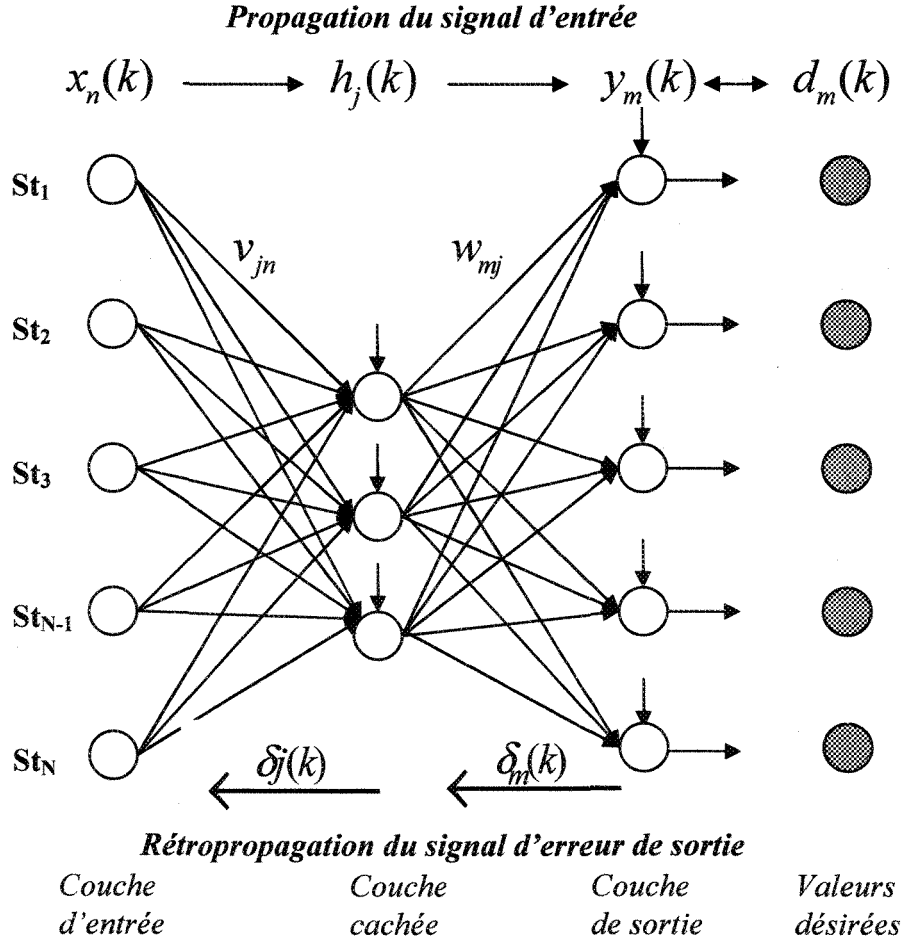

Figure 5-1: Algorithme d'apprentissage par rétropropagation de l'erreur.

En ce qui concerne l'algorithme d'apprentissage proprement dit, la procédure est subdivisée en trois étapes distinctes. La première étape consiste à l'initialisation des poids synaptiques, la deuxième consiste au parcours de la base d'apprentissage et la dernière est

l'évaluation des critères d'apprentissage afin de décider la poursuite de l'itération. Ces étapes sont formulées par l'algorithme présenté ci-dessous :

Étape 1
Initialisation des poids W et V. Leurs valeurs sont dans l'intervalle $[-\varepsilon, \varepsilon]$, avec $\varepsilon \leq 1$.

Étape 2
Parcours de la base d'apprentissage, pour $k = 1, 2, ..., K$

Propagation du signal d'entrée:

a) La forme X_k est présentée à l'entrée du réseau.

b) Une propagation directe est réalisée :

$$h_j(k) = f(\sum_{n=0}^{N} v_{jn} x_n(k))$$

$$y_m(k) = f(\sum_{j=0}^{J} w_{mj} h_j(k)) \qquad (5.1)$$

c) Calcul de la somme des erreurs quadratiques instantanées $E(k)$:

$$E = E + E(k) = E + \frac{1}{2}\sum_{m=1}^{M} (d_m(k) - y_m(k))^2 \qquad (5.2)$$

d) Calcul du signal d'erreur de sortie:

$$\delta_m(k) = (d_m(k) - y_m(k)) f'(\sum_{j=0}^{J} w_{mj} h_j(k))$$

$$\delta_m(k) = (d_m(k) - y_m(k)) y_m(k)(1 - y_m(k)) \qquad (5.3)$$

Rétropropagation du signal d'erreur de sortie :

e) Rétropropagation de $\delta_m(k)$ pour obtenir le signal d'erreur $\delta_j(k)$ associé à chaque neurone j de la couche cachée :

$$\delta_j(k) = h_j^{'}(k)\sum_{m=1}^{M}\delta_m(k)w_{mj}$$

$$\delta_j(k) = h_j(k)(1-h_j(k))\sum_{m=1}^{M}\delta_m(k)w_{mj} \qquad (5.4)$$

f) Modifications des poids de la couche cachée et de la couche de sortie :

$$\Delta w_{mj} = \eta\delta_m(k)h_j(k) \qquad (5.5)$$

$$\Delta v_{jn} = \eta\delta_j(k)x_n(k) \qquad (5.6)$$

Étape 3
Évaluation des critères d'arrêt de l'apprentissage :

$$t = t+1, E = \frac{E}{K}$$

Si ($E \le E_{\min}$ ou $t \ge I_{\max}$) *Arrêt de l'apprentissage.*

Sinon $E = 0$ *et Aller à l'Étape 2.*

5.2.2 Problèmes relatifs aux données manquantes

Si une composante du signal d'entrée est manquante, plusieurs problèmes se présentent au niveau de l'algorithme d'apprentissage. On constate que la cellule correspondant à la valeur manquante se trouve en état d'inactivation vis-à-vis du réseau de neurones. Aussi, les modifications des poids synaptiques v_{jn} dépendent des valeurs des signaux d'entrée (équation 5.6). En outre, nous n'avons pas une valeur pour comparer ou ajuster la sortie du réseau de neurones si une valeur désirée « target » est manquante. Toujours due à cette lacune, nous ne pouvons pas calculer convenablement le signal

d'erreur de sortie δ_m (équation 5.3). Or, les modifications des poids synaptiques v_{jn} et w_{mj} dépendent de ce signal d'erreur de sortie (équations 5.5 et 5.6). Il est alors nécessaire de rechercher une procédure pour modifier adéquatement les poids de connections v_{jn} et w_{mj} dans l'algorithme standard d'apprentissage du réseau de neurones. Pour résoudre ces problèmes, des études de cas des situations possibles de données manquantes doivent être effectuées. Le premier scénario est le cas d'une valeur manquante dans le signal d'entrée. Le deuxième est le cas d'une valeur manquante dans le signal d'entrée et dans le signal de sortie. Enfin, le troisième scénario présente le cas d'une lacune de données durant une longue période d'observation ou pour une saison de givrage.

Nous somme en présence de valeur manquante en entrée si seulement une valeur d'entrée est absente parmi les données d'apprentissage ou dans le cas de plusieurs valeurs manquantes pour une courte période de temps. Pour ce scénario, la solution pour intégrer les valeurs manquantes est d'appliquer une des méthodes d'imputation statistiques : soit la méthode de remplacement par la moyenne, soit la méthode d'interpolation linéaire. Remarquons que le remplacement de l'entrée par la valeur nulle 0 n'est pas convenable puisque la valeur zéro n'est pas une valeur neutre vis-à-vis de l'ensemble des données d'apprentissage.

Si on est simultanément en présence d'une donnée manquante en entrée et d'une valeur désirée d_m manquante, le problème par rapport à l'algorithme d'apprentissage se

situe au niveau de la cellule dédiée à la station présentant une valeur prédite (d_m) manquante. Dans ce cas, la solution est l'annulation du signal de sortie δ_m. Autrement dit, on suppose que nous avons la bonne valeur de sortie. Ainsi, les poids synaptiques w_{mj} ne seront pas modifiés puisque leurs variations sont proportionnelles au signal d'erreur de sortie δ_m selon l'équation 5.3.

Enfin, si l'absence de valeurs manquantes est constatée sur une longue période, il est plus adéquat d'ignorer ces lacunes lors de la phase d'apprentissage et de se baser seulement sur les données d'entrées-sorties disponibles. Cela nous conduit à l'application de la technique d'inhibition des cellules du réseau de neurones.

5.2.3 Inhibition des cellules correspondantes aux saisons manquantes

D'après les études de cas présentées précédemment, aucune méthode d'imputation ne permet de traiter la présence de saisons manquantes dans la base de données SYGIVRE. Nous sommes donc amenés à proposer une méthodologie permettant de les traiter adéquatement. C'est ainsi que nous proposons la technique de l'inhibition rétrograde des cellules du réseau de neurones. Cette technique se révèle être appropriée pour traiter la présence de saisons manquantes. Elle est fréquemment utilisée dans le domaine de l'intelligence artificielle pour simuler le comportement du cerveau humain face à une absence d'informations dont il recherche. En principe, l'atténuation temporaire du modèle de réseau de neurones par l'inhibition de certaines cellules stimule une forme

d'apprentissage nommée « potentialisation à long terme ». Son effet est local et permet aux neurones de se déconnecter brièvement de leurs voisins pour encoder une information. Ainsi, nous subdivisons le modèle en deux sous-réseaux de neurones distincts. Le premier sous-réseau est constitué par les stations avec des données ou exemples d'apprentissage et le second rassemble les stations présentant des saisons manquantes. Par conséquent, les poids synaptiques entre les cellules d'entrée manquantes et la couche cachée sont non modifiés ou inhibés lors de l'apprentissage des saisons manquantes. De même pour les poids synaptiques entre la couche cachée et les cellules de sortie manquantes. La figure 5-4 illustre cette méthodologie.

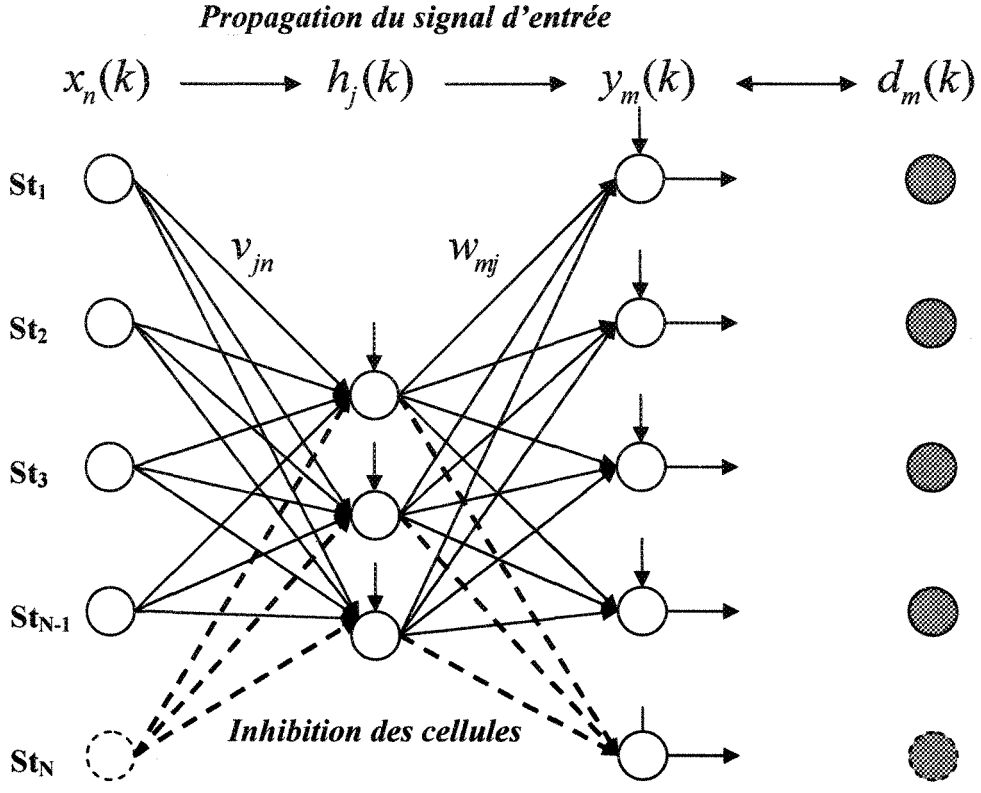

Figure 5-2: Inhibition des cellules correspondantes à une saison manquante.

En observant la figure précédente, on constate bien que l'inhibition des cellules est particulièrement adaptée pour les stations présentant des saisons manquantes. Grâce à cette technique, nous pouvons désormais exploiter le maximum d'informations et généraliser l'étude des tempêtes de givre et de verglas. Aussi, cela nous permet théoriquement d'avoir un modèle plus flexible et fonctionnel face aux lacunes dans la base de données SYGIVRE. En prenant comme exemple le cas du groupe 1, on observe que deux stations de mesure n'ont aucune observation lors de la saison 2003-2004. Par conséquent, le modèle de réseau de neurones pour ce groupe de stations se comporte comme un modèle à 8 entrées-sorties quand on lui apprend les données d'apprentissage de la saison 2001-2002. Tandis qu'il se comporte comme un réseau à 6 entrées-sorties lors de l'apprentissage des données de givrage de la saison 2003-2004.

	Groupe de stations 1							
Date/Heure-Station	**9**	**15**	**16**	*****	**12**	**36**	**24**	*****
11/4/2003 15:00	0.50	-2.10	-2.40	NAN	NAN	-2.05	0.50	NAN
11/4/2003 16:00	0.50	-2.00	-2.30	NAN	NAN	-2.03	0.50	NAN
11/4/2003 17:00	0.50	-1.80	-2.05	NAN	NAN	-1.70	0.50	NAN
11/4/2003 18:00	0.70	-1.28	-1.83	NAN	NAN	-1.53	-0.20	NAN
11/4/2003 19:00	-0.60	-0.80	-2.20	NAN	NAN	-1.00	-2.20	NAN
11/4/2003 20:00	-1.60	-0.60	-1.90	NAN	NAN	-0.90	-2.70	NAN
11/4/2003 21:00	-2.50	-0.08	-1.60	NAN	NAN	-0.50	-2.50	NAN
11/4/2003 22:00	-1.60	0.40	-1.05	NAN	NAN	-0.05	-2.80	NAN
11/4/2003 23:00	-1.20	1.20	-0.43	NAN	NAN	0.20	-2.30	NAN
11/5/2003 2:00	-0.50	1.40	0.00	NAN	NAN	0.40	-1.55	NAN
11/5/2003 3:00	-0.10	1.70	0.20	NAN	NAN	0.60	-1.15	NAN
11/5/2003 4:00	0.10	2.70	0.20	NAN	NAN	0.90	-0.58	NAN
11/5/2003 5:00	0.00	3.00	0.50	NAN	NAN	0.80	-0.48	NAN
11/5/2003 6:00	0.20	5.50	0.60	NAN	NAN	0.80	-0.20	NAN
11/5/2003 7:00	0.40	6.70	0.70	NAN	NAN	3.80	0.08	NAN

Tableau 5-1: Saison manquante, groupe 1, saison 2003-2004

5.4 Résumé

Étant donné que les données manquantes au niveau de la base de données sont les principales problématiques à résoudre, le terme « data mining » est le mot qui résume les principaux travaux réalisés lors de la phase d'intégration des données manquantes. En effet, nous avons exposé dans ce chapitre les résultats de recherche de méthodologies pour résoudre ou traiter les lacunes au niveau de la base des données SYGIVRE.

En premier lieu, on a introduit les types de données manquantes qu'on peut rencontrer. Il y a trois types distincts : les données manquantes au hasard (DMH), les données manquantes complètement au hasard (DMCH) et les autres types de données manquantes (ATDM). Statistiquement, seules les données manquantes de type DMH doivent être prises en compte en étudiant les relations entre les variables existantes. De ce fait, on a conclu qu'une saison manquante ne peut pas être imputée en utilisant les données disponibles. On a conclu également que les données manquantes pour des courtes périodes de temps de la base de données SYGIVRE sont de types DMH et qu'elles peuvent être estimées en étudiant les relations entre les variables disponibles.

En second lieu, plusieurs méthodes d'imputation sont étudiées afin d'intégrer les valeurs manquantes de températures et de signaux de givromètres de la base des données des tempêtes de givre et de verglas. Le rejet de ces valeurs est la première solution unanime quand les signaux de givromètres sont absents pour des périodes d'observation tandis que

les valeurs de températures manquantes sont sujettes à des analyses avant d'être rejetées. En second lieu, les occurrences d'informations présentant des valeurs aberrantes sont rejetées également. Troisièmement, on a introduit la méthode de remplacement par la moyenne des valeurs de températures adjacentes pour palier aux lacunes échelonnées sur des courtes durées d'observations. Enfin, on a appliqué la méthode de remplacement par régression linéaire pour imputer les valeurs manquantes ponctuelles. Des études de faisabilité d'autres méthodes d'imputation sont effectuées aussi dans cet ouvrage. On distingue : la méthode « hot deck » et la méthode « cold deck ». Malheureusement, les applications de ces approches dans le cadre de cette étude ne sont pas réalisables vis à vis des données de givrages. Celles-ci sont dues à la nature complexe de ces données et à l'absence d'une source de données externe.

Enfin, nous avons révisé l'algorithme d'apprentissage du modèle de réseau de neurones afin de l'adapter à la présence des données manquantes. Le principe de base de cette révision s'inspire de la technique d'inhibition des cellules composant le réseau de neurones. L'information des formes à apprendre intervient lors de la mémorisation ou phase d'apprentissage et produit des perturbations ponctuelles sans toutefois perturber l'organisation générale du réseau de neurones. Cet algorithme est particulièrement utile pour les groupes de stations présentant des saisons manquantes puisque ce type de lacunes n'est pas traitable avec les méthodes d'imputation statistiques.

CHAPITRE 6

RÉSULTATS ET INTERPRÉTATIONS

CHAPITRE 6

RÉSULTATS ET INTERPRÉTATIONS

Suite à la réalisation des différentes phases de la méthodologie adoptée au cours de cette étude, ce chapitre présente en détails les interprétations des résultats obtenus. La première phase englobe la modélisation de la base de données SYGIVRE. Ensuite, nous avons amorcé le développement du modèle de prédiction des poids d'accumulation de givre et de verglas. Cette étape incluait la simulation de données d'apprentissage virtuelles afin de valider le fonctionnement de ce modèle. Enfin, la dernière étape concernait l'intégration de données manquantes en utilisant les méthodes d'imputation statistique et l'approche neuronale avec une légère modification de l'algorithme d'apprentissage.

Relativement à ces étapes, nous débutons ce chapitre par la présentation des analyses descriptives de la base de données SYGIVRE, c'est-à-dire les statistiques des différentes variables étudiées et les statistiques des tempêtes de givre et de verglas. Nous présentons également les statistiques des données manquantes recensées au sein des groupes de stations de mesure : la région du Saguenay – Lac-Saint-Jean, celle de Montréal-Québec et celle de la Côte-Nord. Par la suite, la validation des modèles de réseaux de neurones avec les données simulées est présentée. Les résultats des tests sans les données manquantes sont d'abord exposés avant d'examiner les tests avec des données manquantes.

Enfin, la validation du modèle de prédiction des tempêtes de givre et de verglas est présentée avec les données de la base de données SYGIVRE.

6.1 Résultats des analyses et des traitements de données

La phase des traitements et des analyses des données consista à transformer les données brutes en des données de tempêtes de givre et de verglas. Pour y arriver, nous avons vu qu'il a fallu procéder étape par étape, en commençant par l'analyse générale des données. Ensuite, les variables d'étude ont été identifiées. Quant aux événements élémentaires et aux tempêtes de givre et de verglas, ils ont été recensés grâce aux procédures d'identification que nous avons proposé dans ce travail de recherche. Les résultats de ces analyses et traitements de données sont exposés dans les sections qui suivent. Les données recueillies au sein du groupe du Saguenay – Lac-Saint-Jean ont été particulièrement choisies comme base d'analyse en raison du nombre de stations qui la composent. Aussi, il y a le fait que l'objectif général de cette étude est l'intégration des données manquantes en vue de la prédiction des événements de givre et de verglas.

6.1.1 Statistique générale des données de givrage 2001-2002/2003-2004

La base de données SYGIVRE est constituée par l'ensemble des informations provenant des stations de mesure. Dans le cadre du processus d'analyse et de traitement de données, les données des deux saisons d'observation de précipitations givrantes 2001-2002 et 2003-2004 ont été exploitées. Des analyses et des traitements de données appartenant à chaque station de mesure ont été effectués afin de répertorier les tempêtes de givre et de verglas. Ensuite, les données de chaque groupe ont été étudiées pour le développement du

modèle de prédiction. Le premier groupe est composé de 8 stations de mesure localisées dans la région de Montréal-Québec. On y distingue les stations de Beaumont, Poste Chateauguay, de Chénier, de Gentilly, celle située sur la ligne expérimentale de l'IREQ, celle du Pont Pierre-Laporte et celle de Sorel. Le deuxième groupe est composé seulement de 3 stations de mesure qui sont localisées dans la région de Québec-Saguenay – Lac-Saint-Jean. Les stations de Mont-Bélair, de la Tour de Daran et celle du Lac Lavoie composent ce deuxième groupe. Enfin, le troisième est composé de 8 stations qui sont réparties dans la région de la Côte-Nord. Au total, 19 stations de mesure sont prises en compte après la phase de regroupement de ces dernières.

6.1.2 Statistique des événements élémentaires de givre et de verglas

Les événements élémentaires de givre et de verglas sont les entités de base des tempêtes de givre et de verglas (voir chapitre 3). Le tableau suivant présente les effectifs des événements recensés au sein des stations de mesure composant le groupe du Saguenay – Lac-Saint-Jean durant les saisons d'observation 2001-2002 et 2003-2004.

Station	2001-2002	2003-2004
Mont-Bélair	93	45
Tour de Daran	137	92
Lac Lavoie	*	49
Sous-total	230	186
Total	416	

Tableau 6-1: Effectifs des événements élémentaires, Groupe 2, Saisons 2001-2002/2003-2004

Selon ces informations, 416 événements élémentaires ont été enregistrés au sein du groupe 2 pour l'ensemble des deux saisons 2001-2002 et 2003-2004. On constate qu'il y a eu plus d'événements lors de la saison 2001-2002 que lors de la saison 2003-2004. Aucun événement élémentaire n'a été enregistré à la station du Lac Lavoie pour la saison 2001-2002 car aucune donnée n'y a été recueillie lors de cette période. Ceci est défini comme une saison manquante.

6.1.3 Statistique des températures de l'air

On sait pertinemment que les occurrences des accumulations de givre et de verglas dépendent fortement des valeurs de la température extérieure de l'air. Il est donc intéressant d'interpréter la variation de ces dernières vis à vis des événements élémentaires recensés lors de cette étude. Pour le groupe de Québec-Saguenay – Lac-Saint-Jean (groupe 2), le tableau suivant a été obtenu après l'analyse descriptive des valeurs des températures moyennes des événements enregistrés dans ce groupe de stations.

	2001-2002	**2003-2004**
Moyenne	-6.78	-5.82
Médiane	-5.53	-3.26
Écart-type	5.72	6.31
Variance	32.70	39.79
Minimum	-28.55	-28.52
Maximum	0.79	0.65
Effectif	230	186

Tableau 6-2: Analyse descriptive des variations des températures moyennes, Groupe 2, Saison 2001-2002/2003-2004

D'après le tableau 6-2, on peut affirmer que la moyenne des températures moyennes des 416 événements de givre et de verglas au sein du groupe 2 est égale à -6.78°C. L'écart-type est près de 6°C. La valeur minimale est égale à -28°C et la valeur maximale est égale à 0.79°C. Le graphe suivant illustre en détails la variation des valeurs de températures moyennes des 230 événements élémentaires de givre et de verglas, enregistrés lors de la saison 2001-2002.

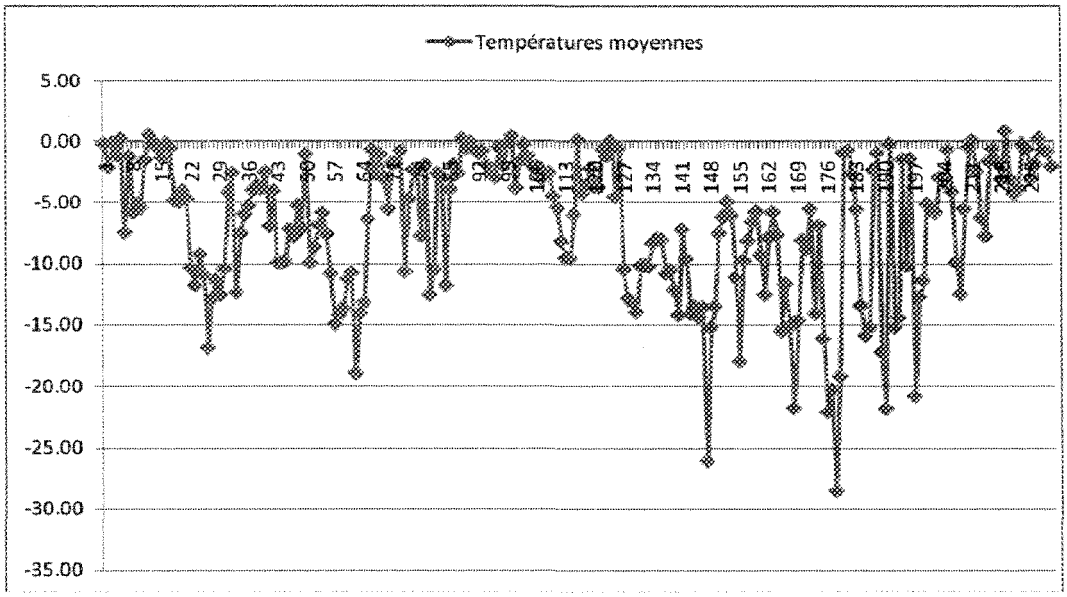

Figure 6-1: Variation des températures moyennes des événements de givre et de verglas, Groupe 2, Saison 2001-2002

À partir de ce graphique de la variation des températures moyennes des événements de givre et de verglas (figure 6-1), on constate qu'un événement de givre et de verglas se produit généralement à une température située entre 0°C et -20°C. Quant à la variation des valeurs des températures minimales des événements de givre et de verglas (figure 6-2), elle mène également à cette constatation. En effet, ces deux graphes ont la même allure puisque les événements élémentaires ne durent pas longtemps.

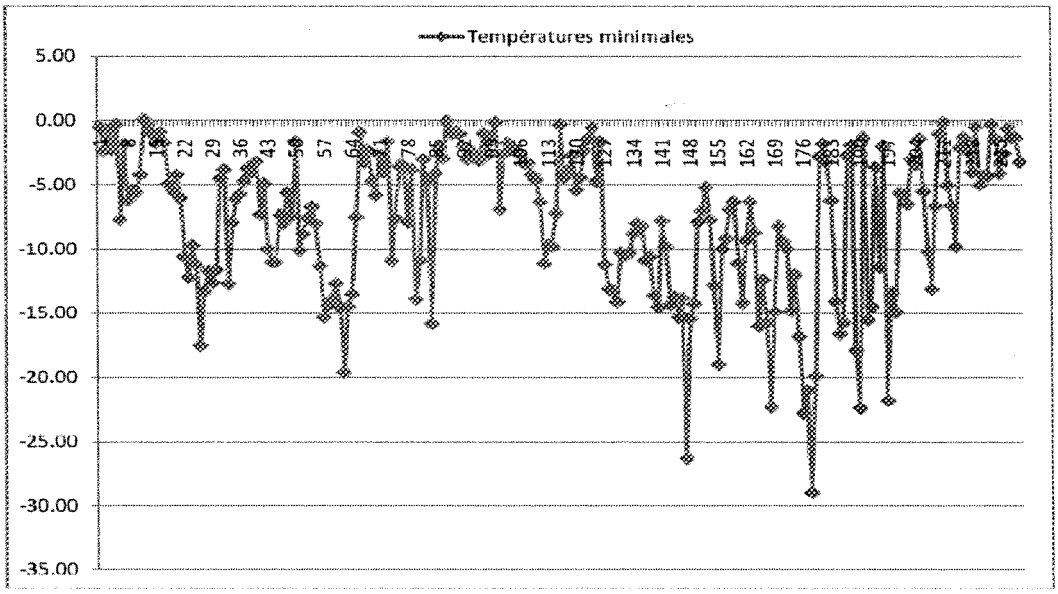

Figure 6-2: Variation des températures minimales des événements de givre et de verglas, Groupe 2, Saison 2001-2002

6.1.4 Statistique des signaux de givromètres

En ce qui concerne la sévérité des événements élémentaires répertoriés lors des processus de traitements des données, le tableau 6-3 et la figure 6-3 illustrent la variation des cumulatifs des signaux de givromètre durant les événements au sein du groupe de Québec-Saguenay – Lac-Saint-Jean (Groupe 2).

	2001-2001	**2003-2004**
Moyenne	7	10
Médiane	1	2
Écart-type	14	21
Minimum	1	1
Maximum	115	198
Somme	1550	1834
Effectif	230	186

Tableau 6-3 : Analyse descriptive des cumulatifs des signaux de givromètre durant les événements de givre et de verglas, Groupe 2, Saison 2001-2002/2003-2004

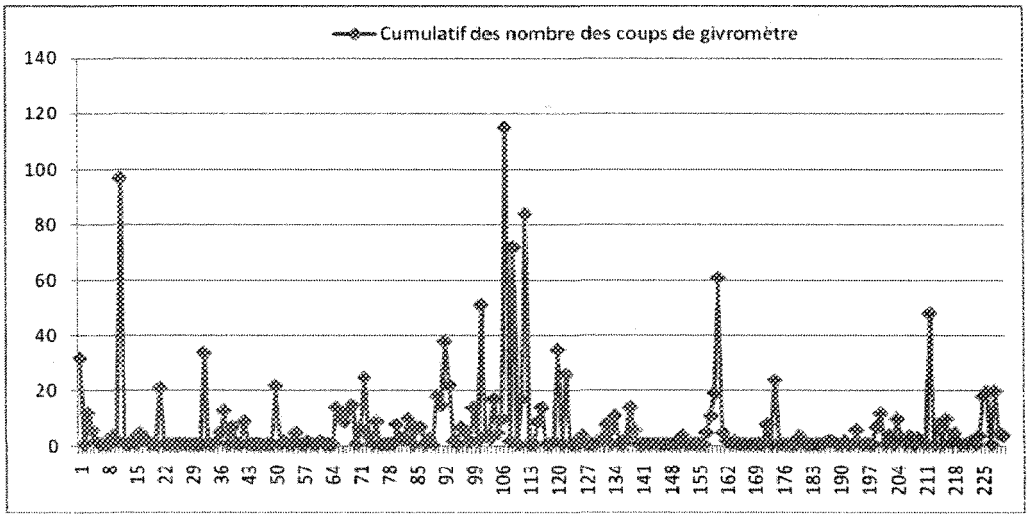

Figure 6-3: Variation des cumulatifs des nombres des coups de givromètre durant les événements de givre et de verglas, Groupe 2, Saison 2001-2002

En moyenne, on compte environ plus de 7 coups de givromètre. En considérant qu'il y a seulement deux types de précipitation et si on applique l'algorithme 9-20, cette valeur moyenne correspond au moins à 63 g/m d'accumulation de givre ou à 140 g/m de dépôts de verglas sur les lignes de transport d'énergie électrique. La valeur maximale du cumulatif de coups de givromètre enregistrée au sein de ce groupe de stations est égale à 198. Cette valeur correspond à un événement apparaissant le 02/01/2004 à 06 :00 P.M (voir figure 6-4). Ce dernier a duré environ 26 heures selon notre analyse. La température moyenne était de -13°C. Comme l'indique le graphe, il y avait une forte accumulation de givre et de verglas si on se fie aux signaux détectés par le givromètre à la station 17 du groupe 2. Les 198 coups de givromètre sont équivalents à environ 1800 g/m de dépôts de givre ou 4000 g/m de verglas.

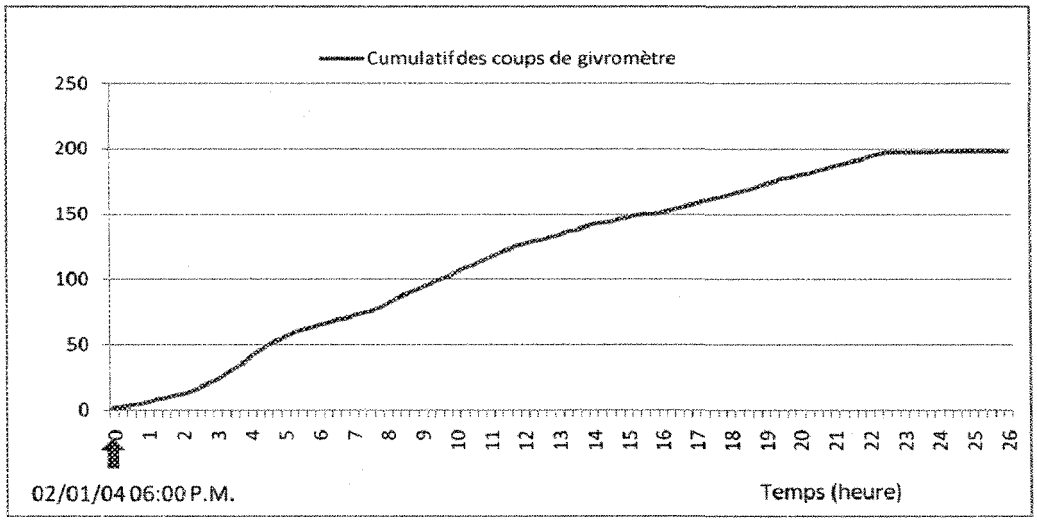

Figure 6-4: Valeur maximale du cumulatif des signaux de givromètre, Groupe 2, Station 17, Événement 81, Saison 2003-2004

6.1.5 Statistique des types d'accumulation

On a stipulé lors de cette étude qu'on peut faire face à quatre types d'accumulation. En effet, on distingue les types suivants : verglas (10), dominance de verglas (12), givre (20) et dominance de givre (21). En étudiant les données de la station du Mont-Bélair, on constate qu'il y a plus d'accumulation de givre que de verglas durant la saison 2001-2002. Quoi qu'on assiste à une proportionnalité égale à 53% de verglas contre 46% de givre durant la saison 2003-2004, le tableau suivant confirme le fait que le givre de nuage est prédominant à la station de Mont-Bélair.

Saison	2001-2002		2003-2004	
Type d'accumulation	**Fréquence**	**Cumulatif %**	**Fréquence**	**Cumulatif %**
Verglas: 10	26	27.96%	20	44.44%
Dom. de verglas: 12	8	36.56%	4	53.33%
Givre: 20	56	96.77%	18	93.33%
Dom. de givre: 21	3	100.00%	3	100.00%

Tableau 6-4: Répartition des types d'accumulation, Station Mont-Bélair, 2001-2002/2003-2004

Pour la suite de l'analyse des types d'accumulation, nous avons étudié les événements recensés au sein du groupe de Québec-Saguenay – Lac-Saint-Jean. Le tableau suivant présente la statistique obtenue lors de cette analyse.

	10		12		20		21	
Station/Saison	2001-02	2003-04	2001-02	2003-04	2001-02	2003-04	2001-02	2003-04
Mont-Bélair	26	20	8	4	56	18	3	3
Tour de Daran	37	30	7	8	83	52	10	2
Lac Lavoie	*	22	*	5	*	17	*	5
Sous-total	63	72	15	17	139	87	13	10
Total	135		32		226		23	
Cumulatif %	32.45		7.69		54.33		5.53	

Tableau 6-5: Répartition des types d'accumulation, Groupe 2, Saisons 2001-2004

En observant ce tableau, on remarque que les événements de givre sont plus fréquents que les événements de verglas dans la région de Québec-Saguenay – Lac-Saint-Jean (Groupe 2). En effet, les résultats des classifications montrent qu'il y a environ 60% de givre et 40% de verglas. En détails, nous avons les répartitions suivantes : 32% des événements du groupe 2 sont de type verglas, 8% sont de type à dominance de verglas, 54% sont de type givre et 6% sont de type à dominance de givre. Les symboles astérisques

* à la station du Lac Lavoie signifient qu'aucune donnée n'y a été collectée lors la saison 2001-2002.

6.1.6 Statistique des tempêtes de givre et de verglas

La prédiction des tempêtes de givre et de verglas au sein d'un groupe de stations nécessite la connaissance de la distribution des tempêtes à chaque station de mesure. Cette section présente les analyses statistiques effectuées pour décrire leurs occurrences durant les saisons d'observation 2001-2002 et 2003-2004. Les indices ou mesures estimées pour caractériser les tempêtes observées à toutes les stations de mesure sont les suivantes :

- NbSE : nombre de tempêtes;
- DT : durée totale des tempêtes;
- CNCGT : cumulatif total des nombres de coups de givromètre enregistrés durant les tempêtes;
- DM : durée moyenne des tempêtes;
- DMAX : durée maximale des tempêtes;
- CNCGM : cumulatif moyen des nombres de coups de givromètre enregistrés durant les tempêtes;
- CNCGMAX : cumulatif maximal des nombres de coups de givromètre enregistrés durant les tempêtes.

D'un point de vue général, les traitements et les analyses de données ont permis de faire une vue d'ensemble des fréquences d'apparition et de la sévérité des tempêtes de givre et de verglas pour chaque saison d'observation ainsi que pour chaque groupe de stations. Le

tableau 6-6 présente l'effectif global de ces tempêtes lors des saisons de givrage 2001-2002 et 2003-2004. D'après l'algorithme d'identification de tempête proposé lors de cette étude, on a au total 507 tempêtes de givre et de verglas. Leur durée totale est égale à 3948 heures et le cumulatif des nombres de signaux de givromètre enregistrés est égal à 9167 coups. En moyenne, chaque tempête dure environ huit heures et on enregistre 18 signaux durant son occurrence.

Saison	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
2001-2002	260	1752	4449	7	49	17	153
2003-2004	247	2196	4718	9	176	19	205
Total/Moyenne/Max	507	3948	9167	8	176	18	205

Tableau 6-6: Analyse descriptive des tempêtes de givre et de verglas, saison 2001-2002/2003-2004

En subdivisant cette dernière statistique en termes de groupes de stations, les deux tableaux suivants permettent de voir les fréquences d'occurrence de tempêtes au sein de chaque groupe de station. En effet, on constate que chaque groupe se comporte différemment non seulement entre eux mais aussi d'une saison à l'autre.

	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Groupe 2	81	650	1476	8	49	18	117
Groupe 1	57	203	387	4	12	7	37
Groupe 3	122	900	2586	7	36	18	153
Total/Moyenne/Max	260	1752	4449	7	49	17	153

Tableau 6-7: Analyse descriptive des tempêtes de givre et de verglas, saison 2001-2002

	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Groupe 2	76	761	1780	10	103	23	205
Groupe 1	33	188	267	5	17	8	29
Groupe 3	138	1246	2671	9	176	19	195
Total/Moyenne/Max	247	2196	4718	9	176	19	205

Tableau 6-8: Analyse descriptive des tempêtes de givre et de verglas, saison 2003-2004

En allant en profondeur, l'étude du groupe 2 (Québec-Saguenay – Lac-Saint-Jean) a permis de mettre en évidence les résultats d'analyses suivants (Tableaux 6-9 et 6-10). Le premier concerne la saison d'observation 2001-2002 et le second est rattaché à la saison 2003-2004.

Station	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Mont-Bélair	31	244	534	8	27	17	97
Tour de Daran	50	406	942	8	49	19	117
Lac Lavoie	*	*	*	*	*	*	*
Total/Moyenne/Max	81	650	1476	8	49	18	117

Tableau 6-9: Analyse descriptive des tempêtes de givre et de verglas, groupe 2, saison 2001-2002

Station	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Mont-Bélair	16	166	354	10	25	22	89
Tour de Daran	34	396	889	12	103	26	205
Lac Lavoie	26	199	537	8	36	21	161
Total/Moyenne/Max	76	761	1780	10	103	23	205

Tableau 6-10: Analyse descriptive des tempêtes de givre et de verglas, groupe 2, saison 2003-2004

D'après ces statistiques, la région de Québec-Saguenay – Lac-Saint-Jean a été le théâtre d'une quantité semblable de tempêtes de givre et de verglas durant ces deux saisons. En effet, il y a eu 81 tempêtes dont la durée totale est égale à 650 heures durant la saison

2001-2002. Durant la saison 2003-2004, il y en a eu 76 et la durée totale est égale à 761 heures.

En ce qui concerne les durées des tempêtes apparaissant lors de ces deux saisons, on observe que la durée moyenne est d'environ 8 heures pour la saison 2001-2002 et de 10 heures pour la saison 2003-2004. Le record maximal atteint a été d'une durée de deux jours (49 heures) pour la saison 2001-2002 et d'une durée de quatre jours et sept heures (103 heures) pour la saison 2003-2004.

À propos de la sévérité de ces tempêtes, on peut déduire d'après ces résultats d'analyses que le nombre de coups de givromètres moyen au cours d'une tempête de givre et de verglas est égal à peu près à 20 coups pour le groupe de stations de Québec-Saguenay – Lac-Saint-Jean. On observe 18 coups par tempête lors de la saison 2001-2002 et 23 coups par tempête lors de la saison 2003-2004. Le nombre maximal de signaux de givromètre est observé lors de la saison 2003-2004. Ce dernier est égal à 205 coups et est l'équivalent d'une accumulation de 1845 grammes de dépôt de givre ou 4100 grammes de verglas. Ceci est considéré comme une tempête d'une grande sévérité pour les lignes de transport d'énergie électrique.

Au sein du groupe de stations de Québec-Montréal, on a recensé moins de tempêtes de givre et de verglas durant les deux saisons en dépit de son nombre de stations (voir Annexe G). En effet, il y a eu respectivement 57 et 33 tempêtes de givre et de verglas

durant la saison 2001-2002 et 2003-2004. En outre, la durée moyenne de ces tempêtes a été inférieure à celle observée au sein du groupe 2 (Québec-Saguenay – Lac-Saint-Jean). On constate qu'elles durent 4 heures en moyenne pour la saison 2001-2002 et 5 heures pour la saison 2003-2004. De plus, le cumulatif moyen du nombre de coups de givromètre par tempête pour ce groupe est aussi inférieur à celui du groupe 2. Le nombre moyen de signaux de givromètre par tempête est respectivement égal à 7 coups et 8 coups seulement pour la saison 2001-2002 et la saison 2003-2004. Enfin, le fait le plus remarquable est au niveau du nombre de coups de givromètre maximal. Pour les deux saisons de givrage, le cumulatif maximal est égal à 37. Ceci montre qu'aucune tempête d'une grande sévérité n'a été répertoriée au sein du groupe de stations de Québec-Montréal.

Pour le groupe de stations de la Côte-Nord, les résultats d'analyse permettent d'affirmer que les données recueillies lors des saisons de givrage suivent à peu près la même distribution. On comptait respectivement 122 et 138 tempêtes de givre et de verglas pendant les saisons 2001-2002 et 2003-2004. La durée moyenne de ces tempêtes est égale à 7 heures pour la première saison et 9 heures pour la deuxième. La durée moyenne des tempêtes de ce groupe est donc sensiblement inférieure à celle du groupe de Québec-Saguenay – Lac-Saint-Jean. La même constatation peut être faite en ce qui concerne le cumulatif moyen de nombre de signaux de givromètre par tempête. En effet, il y a eu en moyenne 20 signaux de givromètres à chaque tempête de givre et de verglas répertoriée dans ce groupe de stations. Enfin, on remarque que le cumulatif de signaux de givromètre

varie beaucoup d'une station à une autre. La valeur minimale de ce dernier est égale à 15 coups tandis que sa valeur maximale atteint 195 coups.

6.1.7 Statistique des poids d'accumulation de givre et de verglas

D'après les études antérieures, chaque groupe de station a ses propres caractéristiques en termes de fréquence d'occurrence de tempêtes ainsi que de leur sévérité. Afin de valider cette affirmation, on a analysé la distribution des poids finaux d'accumulation de givre et de verglas des tempêtes enregistrés aux trois groupes de stations de mesure.

	Groupe 2	Groupe 1	Groupe 3
Moyenne	361	142	420
Médiane	182	100	207
Écart-type	456	132	536
Variance	208264	17296	287252
Minimum	45	46	45
Maximum	2691	743	2970
Somme	29262	8113	60083
Effectif	81	57	122

Tableau 6-11: Analyses descriptives des poids d'accumulation de givre et de verglas, saison 2001-2002

	Groupe 2	Groupe 1	Groupe 3
Moyenne	460	164	418
Médiane	229	115	146
Écart-type	618	136	621
Variance	382487	18534	385550
Minimum	45	46	46
Maximum	3646	667	4485
Somme	34966	5414	56790
Effectif	76	33	138

Tableau 6-12: Analyses descriptives des poids d'accumulation de givre et de verglas, saison 2003-2004

D'après ces résultats d'analyse, le poids cumulatif moyen de givre et de verglas pour chaque tempête est d'environ 360 g au sein du groupe Québec-Saguenay – Lac-Saint-Jean (Groupe 2) lors de la saison 2001-2002. Ce dernier a augmenté jusqu'à 460 g au cours de la saison 2003-2004. Sachant que les effectifs des tempêtes enregistrés lors de ces deux saisons sont à peu près égaux, on peut dire que les tempêtes de givre et de verglas ont été plus sévères en 2003-2004 en comparaison avec celles de 2001-2002. En effet, le poids maximum est respectivement égal à 2691 g et 3646 g au cours de la saison 2001-2002 et la saison 2003-2004. En outre, on remarque que le premier groupe de stations a un faible poids moyen par rapport aux autres groupes au cours des deux saisons d'observation en plus d'avoir un petit nombre de tempêtes de givre et de verglas. Enfin, le troisième groupe de stations a une moyenne de 420 g d'accumulation à chaque tempête pour les deux saisons de givrage. Le nombre des tempêtes étant quasiment le même : 122 en 2001-2002 et 138 en 2003-2004, on peut dire que ce groupe de stations a présenté les mêmes caractères pour ces deux saisons de givrage.

6.1.8 Statistique des données manquantes

Nous présentons dans cette section les rapports entre les données manquantes et les données disponibles au sein des groupes de stations.

Pour le groupe de Québec-Saguenay – Lac-Saint-Jean, les informations correspondantes aux tempêtes de givre et de verglas sont équivalentes à 678 heures d'observation lors de la saison 2001-2002 et 649 heures durant la saison 2003-2004 (voir

tableau 6-13). Au total, 19% des valeurs de températures sont manquantes dans ce groupe de stations. En effet, 767 valeurs de températures sont manquantes sur les 3981 (1327 multiplié par 3) valeurs dont nous devons disposer pour l'ensemble de ces trois stations de mesure.

Station	2001-2002 (/678 heures)	2003-2004 (/649 heures)	2001-2004 (/1327 heures)	Pourcentage (%)
Mont-Bélair	60	8	68	5
Tour de Daran	8	12	20	2
Lac Lavoie*	678	1	679	51
Total (/Groupe)	746	21	767	19

Tableau 6-13: Données de températures manquantes, groupe 2

D'après ce tableau, on remarque que la station du Lac Lavoie n'a aucune donnée disponible pour la saison 2001-2002. Ce cas est l'exemple typique de ce qu'on appelle une « saison manquante ». Nous n'avons pas pour cette station ni de valeur de température, ni de valeur de poids d'accumulation durant cette période d'observation. On peut également observer l'absence de données au niveau de la station de Mont-Bélair durant les deux saisons d'observation. En effet, il y a respectivement 60 valeurs manquantes et 8 valeurs manquantes de températures sur 678 heures et 649 heures d'observation. Ce qui donne un total de 68 valeurs manquantes de températures sur 1327 heures d'observation. On a alors 5% de valeurs manquantes à la station de Mont-Bélair. En ce qui concerne la station de la Tour de Daran et celle du Lac Lavoie, elles manquent respectivement 2% et 51% de valeurs de températures.

En ce qui concerne le groupe de la région de Québec et de Montréal, le taux de pourcentage des données manquantes est de 14% (voir annexe H). Étant donné que le nombre de tempêtes de givre et de verglas répertoriées dans ce groupe est faible par rapport aux autres groupes, le nombre d'heures total d'observation est seulement de 299 heures quand on combine les données de toutes les stations sous forme de table croisée. On remarque également l'absence d'une saison durant la période de 2001-2002.

Enfin, le groupe de la région de la Côte-Nord présente des lacunes de l'ordre de 14% puisqu'il y a deux saisons manquantes : celle due à l'absence de données à la station du Lac Saint-Pierre lors de la saison 2001-2002 et la seconde concerne l'absence des données de la station Virot en 2003-2004. On remarque que les taux de pourcentage des valeurs de températures manquantes de toutes les autres stations sont faibles, dû au nombre élevé de tempêtes de givre et de verglas enregistrées au sein de ce groupe.

Il est important de noter que ces valeurs manquantes ont été traitées en appliquant les méthodes d'imputation décrites dans la présente étude. Ainsi, on a pu obtenir un ensemble de données plus complet grâce à ces techniques d'imputation. Quant aux lacunes correspondantes aux saisons manquantes, le modèle de réseau de neurones les prend en charge grâce à la technique d'inhibition.

6.2 Validation du modèle de prédiction avec les données simulées

Afin de valider l'algorithme d'apprentissage utilisé par les modèles de réseaux de neurones pour la prédiction des accumulations de givre et de verglas, des tests sont effectués avec les données simulées.

Le premier test consistait à utiliser 2008 heures de données simulées de givrage sans données manquantes lors de la phase d'apprentissage du modèle de réseau de neurones. Pour simplifier le développement, nous avons mis en scène un groupe composé de trois stations de mesure virtuelles. Ceci fait référence au groupe de stations de Québec-Saguenay – Lac-Saint-Jean qui est composé également de trois stations. Ensuite, les mêmes données, c'est-à-dire les données d'apprentissage, sont réutilisées pour vérifier si le modèle est capable de prédire les événements de givre et de verglas.

Le second test consistait à simuler la présence d'une saison manquante au sein d'une station de mesure. Pour ce faire, nous avons doublé le nombre d'heures de données de givrage de la première et de la troisième station. Quant à la deuxième station, nous avons supposé qu'il y a une saison manquante. En somme, 1999 heures de données sont manquantes à la deuxième station et 4006 heures de données sont disponibles aux deux autres. Par la suite, le test proprement dit a été effectué et nous avons vérifié si le modèle est toujours capable de prédire les événements de givre et de verglas aux deux autres stations et si des perturbations sont induites par la présence de ces valeurs manquantes.

6.2.1 Résultats des tests sans données manquantes

Pour observer le comportement des modèles de prédiction développés lors de cette étude, nous avons varié le nombre de cellules *J* dans la couche cachée de 5 à 7. Les critères considérés pour vérifier leurs performances sont la somme des erreurs quadratiques et le pouvoir prédictif du modèle en question. Ainsi, les sections suivantes présentent les résultats de ces expériences.

6.2.1.1 Nombre de cellules J égal à 5

6.2.1.1.1 Sommes des erreurs quadratiques

En variant le taux d'apprentissage *mc* et le moment d'apprentissage *mu*, des modèles de prédiction des événements de givre et de verglas sont obtenus. Le tableau 6-14 présente les sommes des erreurs quadratiques finales de ces modèles après les avoir appris des données simulées sans données manquantes.

mc|mu	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	60	60	924	60	60	60	60	60	60
0.1	882	60	234	60	60	60	60	60	60
0.2	60	1758	591	60	60	60	60	60	60
0.3	60	60	60	60	60	60	60	60	60
0.4	60	60	306	532	**42**	60	60	60	60
0.5	710	60	60	60	60	60	60	60	60
0.6	287	60	60	60	60	60	78	60	60
0.7	60	60	60	60	60	60	49	60	187
0.8	60	60	60	60	60	60	60	1781	1890
0.9	60	60	60	60	60	60	60	60	60

Tableau 6-14: Sommes des erreurs quadratiques, J = 5

D'après ce tableau, la valeur minimale de la somme des erreurs quadratiques est obtenue quand le taux d'apprentissage *mc* est égal 0.4 et quand le moment d'apprentissage *mu* est égal 0.5. Visiblement, on remarque que la valeur moyenne est égale à 60 g^2/m^2 si on ne tient pas compte des valeurs extrêmes. Ceci signifie que les modèles neuronaux ont tendance à avoir la même performance vis-à-vis de la fonction de coût à minimiser, peu importe les valeurs du taux d'apprentissage et du moment d'apprentissage. Les valeurs extrêmes correspondent à des modèles instables et non adéquats pour la prédiction des événements de givre et de verglas aux trois stations de mesure.

6.2.1.1.2 Prédiction des poids d'accumulation de givre et de verglas

Après avoir fixé le nombre de cellules dans la couche cachée à 5 et effectué les tests correspondant aux paramètres du modèle de neurones, nous avons vu que le meilleur résultat est obtenu quand le taux d'apprentissage et le moment d'apprentissage sont respectivement égaux à 0.4 et 0.5. Pour ce modèle optimal, le graphique suivant présente l'allure des courbes de prédiction (en pointillé) et les courbes des événements de givre et de verglas à prédire.

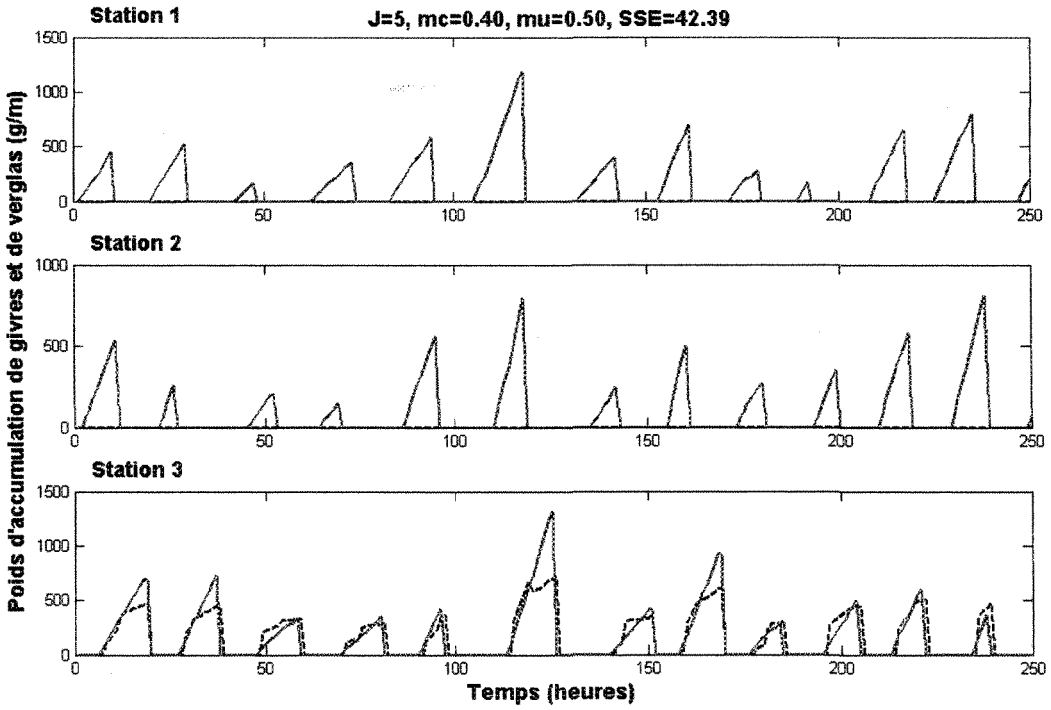

Figure 6-5: Prédiction des événements de givre et de verglas avec les données simulées sans données manquantes, J=5

La somme des carrés des erreurs pour les 2008 heures est égale à 42.39 g^2/m^2. Ces résultats semblent être très satisfaisants en termes de taux de bonnes prédictions. Or, nous constatons d'après cette figure que le modèle ne prédit pas les poids d'accumulation de givre et de verglas aux deux premières stations. D'après nos observations, ce modèle à 5 cellules prédit une seule station de mesure à la fois en raison du degré de complexité des relations entre les trois stations. Nous devons donc augmenter le nombre de cellules dans la couche cachée J pour qu'il puisse généraliser les valeurs des poids d'accumulation à prédire au sein de ces trois stations.

6.2.1.2 Nombre de cellules J égal à 6

6.2.1.2.1 Sommes des erreurs quadratiques

Avec un réseau multicouche à 6 cellules dans la couche cachée, les résultats des tests effectués en variant le taux d'apprentissage et le moment d'apprentissage sont résumés dans le tableau 6-15. On y observe que la valeur minimale des sommes des erreurs quadratiques est égale à 22 g^2/m^2. En comparaison avec les résultats obtenus en ayant 5 cellules dans la couche cachée, on peut dire qu'il y a une nette amélioration de la performance du modèle étudié.

mc|mu	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	**22**	60	60	60	60	47	60	59	60
0.1	60	60	60	68	60	56	42	60	60
0.2	59	60	60	60	47	64	44	60	53
0.3	60	60	60	60	60	60	60	60	60
0.4	60	1719	43	55	60	60	1849	60	60
0.5	60	1110	60	60	60	60	60	60	60
0.6	60	60	60	60	60	60	43	60	60
0.7	60	60	60	60	60	60	55	60	60
0.8	55	60	60	49	60	60	60	60	60
0.9	47	60	60	60	60	60	60	60	3679

Tableau 6-15: Sommes des erreurs quadratiques, J = 6

6.2.1.2.2 Prédiction des poids d'accumulation de givre et de verglas

En ce qui concerne les résultats de prédictions, le modèle à 6 neurones dans la couche cachée nous donne des résultats satisfaisants pour les trois stations de mesure. En effet, les courbes de prédictions de ce modèle suivent correctement les courbes désirées d'après les observations (figure 6-6). La première conclusion de cette phase de validation

est donc le fait suivant: «*Quand aucune donnée n'est manquante et lorsque le nombre de cellules dans la couche cachées est fixé à 6 cellules, la révision de l'algorithme d'apprentissage du réseau de neurones nous donne un modèle de prédiction fonctionnel et valide vis à vis des données simulées*».

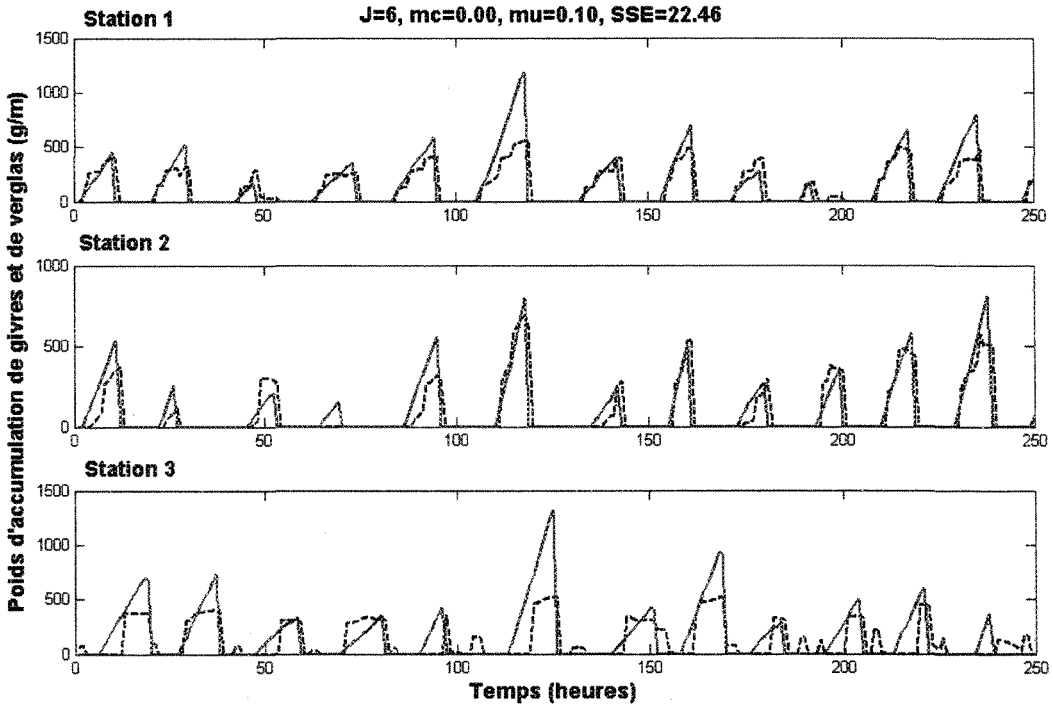

Figure 6-6: Prédiction des événements de givre et de verglas avec les données simulées sans données manquantes, J=6

6.2.1.3 Nombre de cellules J égal à 7

6.2.1.3.1 Sommes des erreurs quadratiques

En fixant le nombre de neurones dans la couche cachée à 7, les résultats des tests effectués en variant les deux paramètres d'apprentissage *mc* et *mu* sont présentés dans le

tableau 6-20. En comparaison avec le modèle à 6 cellules, il y a une dégradation de performance. En effet, on y remarque que la valeur minimale des sommes de ses erreurs quadratiques est passée de 22 g^2/m^2 à 28 g^2/m^2. Cependant, quant à la valeur moyenne, elle est restée constante pour les trois cas considérés (J=5, J=6 et J=7) et est égale à 60 g^2/m^2, peu importe le nombre de cellules dans la couche cachée.

mc|mu	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	1888	44	60	60	53	60	60	60	60
0.1	58	54	60	60	60	60	60	60	47
0.2	60	60	60	60	60	47	60	60	60
0.3	780	60	60	60	60	42	60	46	60
0.4	60	1262	60	48	60	35	60	60	60
0.5	60	43	60	**28**	60	60	60	60	60
0.6	873	60	38	44	60	35	60	60	60
0.7	60	73	60	60	60	40	60	60	1876
0.8	60	60	60	60	60	60	60	60	60
0.9	47	60	60	60	60	60	1849	60	60

Tableau 6-16: Sommes des erreurs quadratiques, J = 7

6.2.1.3.2 Prédictions des poids d'accumulation de givre et de verglas

Après avoir effectué le test du modèle optimal à 7 cellules avec les données d'apprentissage sans données manquantes, nous avons constaté que ce modèle de réseau de neurones ignore les entrées de la deuxième station mais prédit correctement les poids d'accumulation de givre et de verglas pour les deux autres stations. Nous rencontrons ici deux situations contradictoires puisqu'il y a un sur-apprentissage pour la station 1 et la

station 3 tandis que la deuxième station est sous-estimée. Selon nos analyses, il y a un sur-apprentissage car le nombre de cellules dans la couche cachée dépasse le degré de complexité des relations entre les trois stations. En ce qui concerne la sous-estimation de la station 2, il n'y a aucune explication tangible et concrète à part le fait que l'algorithme d'apprentissage a conduit à deux minimums locaux parmi les trois recherchés.

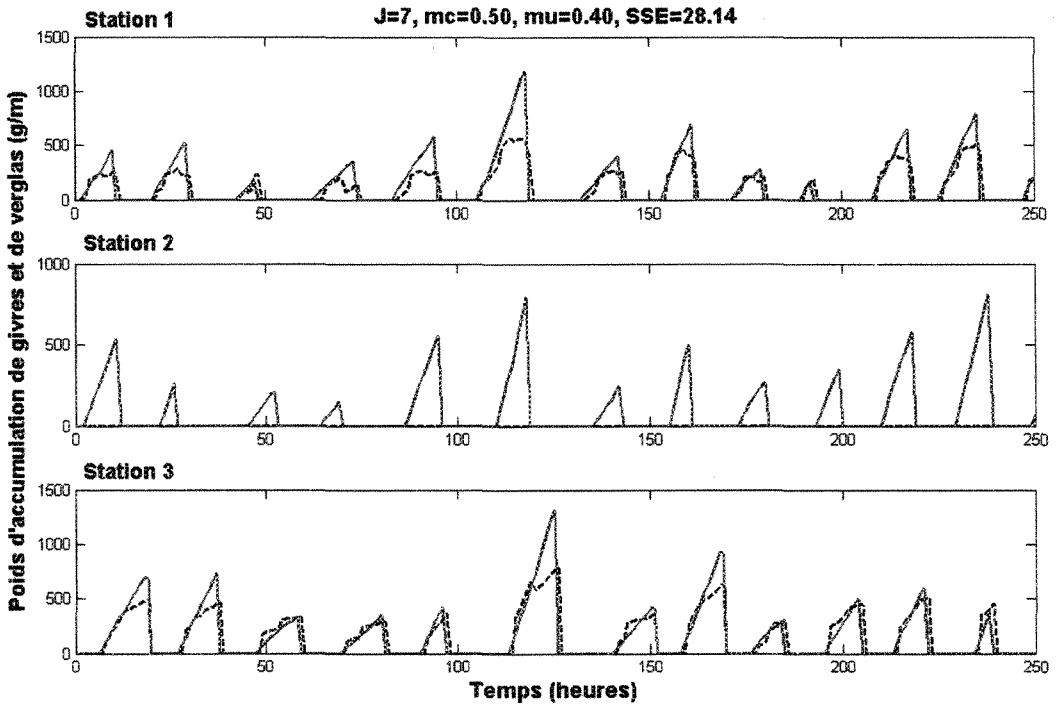

Figure 6-7: Prédictions des événements de givre et de verglas avec les données simulées sans données manquantes, J=7

6.2.2 Résultats des tests avec des données manquantes

En répétant les mêmes expériences réalisées précédemment, les tests avec des données manquantes ont donné des résultats assez semblables. Le nombre de cellules dans la couche cachée a été varié de 5 à 7; ensuite, les différentes combinaisons possibles des paramètres d'apprentissages ont été testées pour trouver le réseau de neurone optimal correspondant à chaque architecture.

6.2.2.1 Nombre de cellules J égal à 5

6.2.2.1.1 Sommes des erreurs quadratiques

Quand le nombre de cellules dans la couche cachée a été fixé à 5, le réseau de neurones optimal donnant la valeur minimale des sommes des erreurs quadratiques correspond à un taux d'apprentissage égal à 0.10 et un moment d'apprentissage égal à 0.30 (Tableau I-7 de l'annexe I). Cette valeur minimale est égale à 64.35 g^2/m^2. En outre, on remarque que la valeur moyenne est passée de 60 g^2/m^2 à 100 g^2/m^2 puisqu'on a ajouté des données d'apprentissage incluant les données manquantes.

6.2.2.1.2 Prédiction des poids d'accumulation de givre et de verglas

Avec 5 cellules dans la couche cachée, le modèle de réseau de neurones obtenu suite aux expériences réalisées dans cette phase de validation ne donne pas les résultats escomptés puisqu'il ne prédit aucune des deux premières stations de mesure. Les poids d'accumulation de givre et de verglas à la troisième station sont les seuls à être bien estimés. Cela signifie que le nombre de cellules dans la couche cachée est insuffisant pour généraliser les propagations spatiales et temporelles des événements de givre et de verglas.

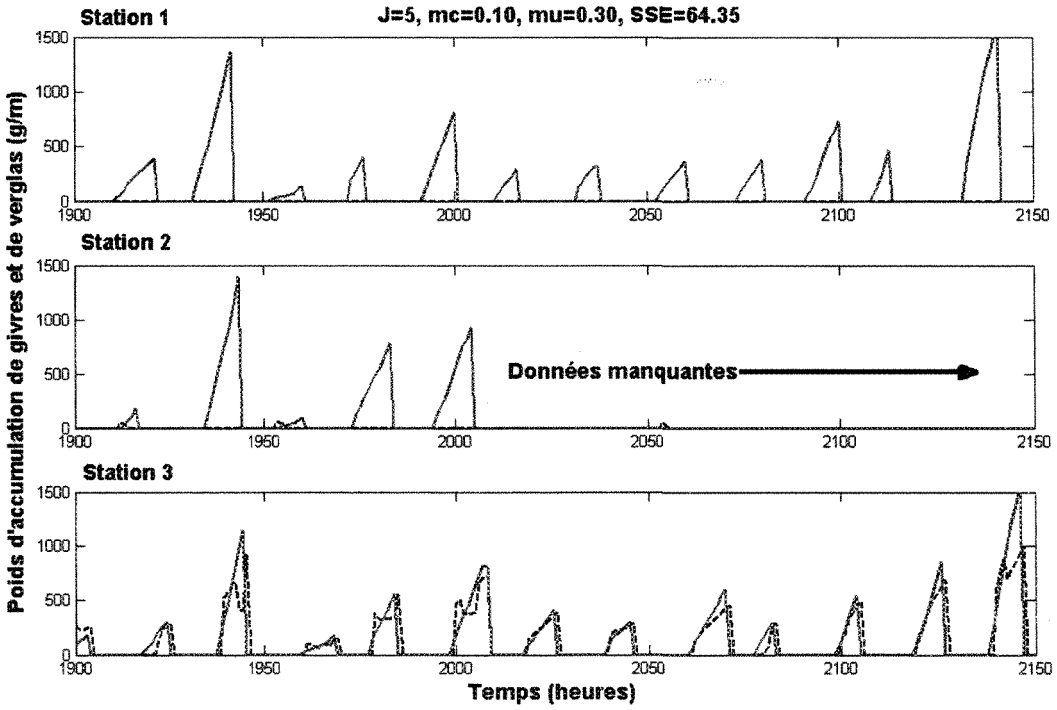

Figure 6-8: Prédiction des événements de givre et de verglas avec les données simulées avec données manquantes, J=5

6.2.2.2 Nombre de cellules J égal à 6

6.2.2.2.1 Sommes des erreurs quadratiques

En fixant le nombre de cellules dans la couche cachée à 6, la somme des erreurs quadratiques à la sortie du modèle de réseau de neurones est diminuée pour atteindre la valeur minimale qui est égale à 23.91 g^2/m^2. Cette valeur surgit quand le taux d'apprentissage et le moment d'apprentissage sont respectivement égaux à 0.00 et 0.50. On peut alors s'attendre à avoir des résultats satisfaisants relativement aux prédictions des poids d'accumulation de givre et de verglas aux trois stations mises en scène lors de cette

phase de test. Le tableau I-4 illustre les différentes valeurs des sommes des erreurs quadratiques correspondant aux différents essais réalisés pour ce modèle.

6.2.2.2.2 Prédiction des poids d'accumulation de givre et de verglas

En simulant les données d'apprentissage avec ce modèle de réseau de neurones, nous remarquons que les événements de givre et de verglas aux trois stations de mesure sont bien appris et qu'ils sont prédits avec succès.

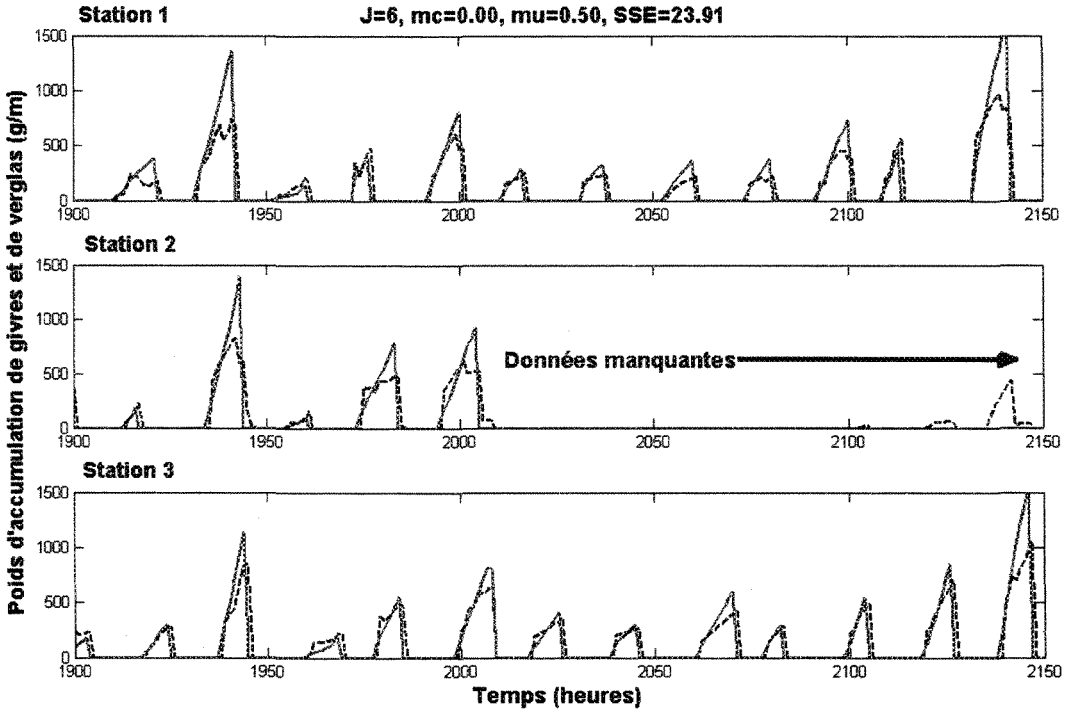

Figure 6-9: Prédictions des événements de givre et de verglas avec les données simulées avec données manquantes, J=6

En ce qui concerne les données manquantes à la deuxième station, les réseaux de neurones les ignorent complètement en continuant à prédire uniquement les événements aux deux autres stations. Ceci nous amène à la deuxième conclusion de cette phase de

validation de modèle de réseau de neurones : « *l'algorithme d'apprentissage intégrant les données manquantes est valide pour les données simulées quand le nombre de cellules dans la couche cachée est égal à 6* ».

6.2.2.3 Nombre de cellules J égal à 7

En augmentant le nombre de cellules dans la couche cachée du modèle de réseau de neurones, les résultats obtenus sont semblables à ceux obtenus quand le nombre de cellules de la couche cachée est égal à 6. Les tempêtes de givre et de verglas aux trois stations de mesure sont bien prédites même s'il y a des données manquantes à la deuxième station. Les résultats des tests effectués sont donnés à l'annexe I : voir le tableau I-9 et la figure I-1. Selon ces résultats, on peut conclure encore une fois que l'algorithme d'apprentissage intégrant les données manquantes est valide et fonctionnel pour prédire les poids d'accumulation de givre et de verglas à un groupe de stations de mesure. Ce modèle à 7 cellules est aussi fonctionnel que le modèle à 6 cellules. Cependant, il a une légère baisse de performance quand on compare la somme des erreurs quadratiques de ses signaux de sortie.

6.3 Validation du modèle de prédiction avec la base de données SYGIVRE

6.3.1 Résultats des tests avec les données du groupe 2

Après avoir validé l'algorithme d'apprentissage intégrant les données manquantes avec les données simulées, nous avons réalisé des tests en utilisant la base de données SYGIVRE. Pour ce faire, nous avons choisi les données relatives au groupe de stations de Saguenay – Lac-Saint-Jean puisqu'elles correspondent abstraitement aux données des expériences effectuées précédemment. En fait, nous voulions comparer les résultats obtenus en prenant des modèles ayant le même nombre de cellules dans leurs couches d'entrée et dans leurs couches de sortie. Sachant que ce groupe comporte trois stations de mesure comme les modèles étudiés préalablement, le choix d'utiliser ses données lors de cette phase de validation des modèles neuronaux est alors le plus logique et le plus approprié.

Ainsi, les données de givrage provenant du groupe du Saguenay – Lac-Saint-Jean sont prises pour valider le modèle de prédiction des événements de givre et de verglas. Nous avons vu qu'il y a deux saisons d'observation. Les données acquises lors de la saison 2001-2002 présentent des lacunes à la station du Lac Lavoie tandis que celles de 2003-2004 sont complètes pour les trois stations de mesure : Mont-Bélair, Tour de Daran et Lac Lavoie. En tout, les données d'apprentissage de ce groupe de stations s'étendent sur 1327 heures de données de tempêtes de givre et de verglas.

6.3.1.1 Nombre de cellules J égal à 5

Étant donné qu'il y a trois stations de mesure, l'architecture du premier modèle testé est un réseau de neurones comportant 9 cellules dans la couche d'entrée et 3 cellules dans la couche de sortie. Quant à la deuxième couche, nous avons commencé par fixer arbitrairement le nombre de cellules à 5. En variant les paramètres d'apprentissage, la somme minimale des erreurs quadratiques est égale à 22.14 g^2/m^2. Elle est obtenue quand le taux d'apprentissage et le moment d'apprentissage sont respectivement égaux à 0.9 et 0.07. Les valeurs des sommes des erreurs quadratiques pour les différents paramètres d'apprentissage sont présentées en annexe par le tableau J-10. D'après les tests effectués, les tempêtes de givre et de verglas à la station de Mont-Bélair et à la station du Lac Lavoie sont ignorées par ce modèle. Quant aux événements à la station de la Tour de Daran, ils sont prédits avec succès. Les courbes de prédictions à ces stations sont présentées par la figure 6-10.

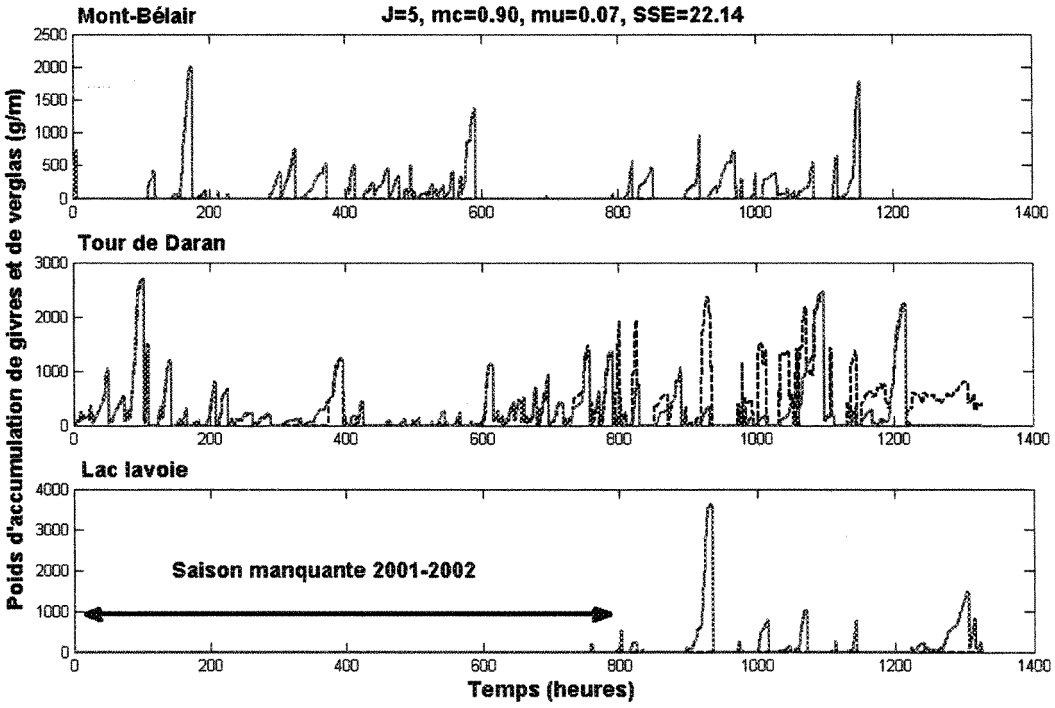

Figure 6-10: Prédictions des événements de givre et de verglas, Groupe 2, J=5, saisons 2003-2004.

6.3.1.2 Nombre de cellules J égal à 6

En fixant le nombre de cellules dans la couche cachée à 6, on constate que les erreurs de prédictions du modèle de réseau de neurones ont diminué considérablement. En fait, la valeur minimale des sommes des erreurs quadratiques est passée de 22.14 g^2/m^2 à 15.77 g^2/m^2 par rapport à celle obtenue en fixant le nombre de cellules de la couche cachée à 5. Comme l'indique le tableau J-11, cette valeur est obtenue quand le taux d'apprentissage et le moment d'apprentissage sont respectivement égaux à 0.90 et 0.03.

Après avoir testé le réseau de neurones avec les données d'apprentissage, on remarque que ce modèle prédit seulement les tempêtes de givre et de verglas à la station de

Mont-Bélair et à la station du Lac Lavoie (voir figure J-2). Quant à la station de la Tour de Daran, le modèle se comporte de telle manière que les entrées à cette station sont complètement ignorées. Les résultats sont alors insatisfaisants puisque notre objectif est de prédire les propagations spatiales et temporelles des tempêtes dans ce groupe de trois stations.

6.3.1.3 Nombre de cellules J égal à 7

Avec une configuration à 7 cellules dans la couche cachée, les erreurs à la sortie du réseau de neurones sont minimales. La somme des erreurs quadratiques des données de tests est égale à 5.44 g^2/m^2 quand le taux d'apprentissage et le moment d'apprentissage sont respectivement égaux à 0.70 et 0.03, voir tableau J-12.

D'après les résultats de simulation avec le réseau de neurones optimal dédié au groupe du Saguenay – Lac-Saint-Jean, on remarque que les prédictions des tempêtes de givre et de verglas aux trois stations sont bien précises (voir figure 6-11).

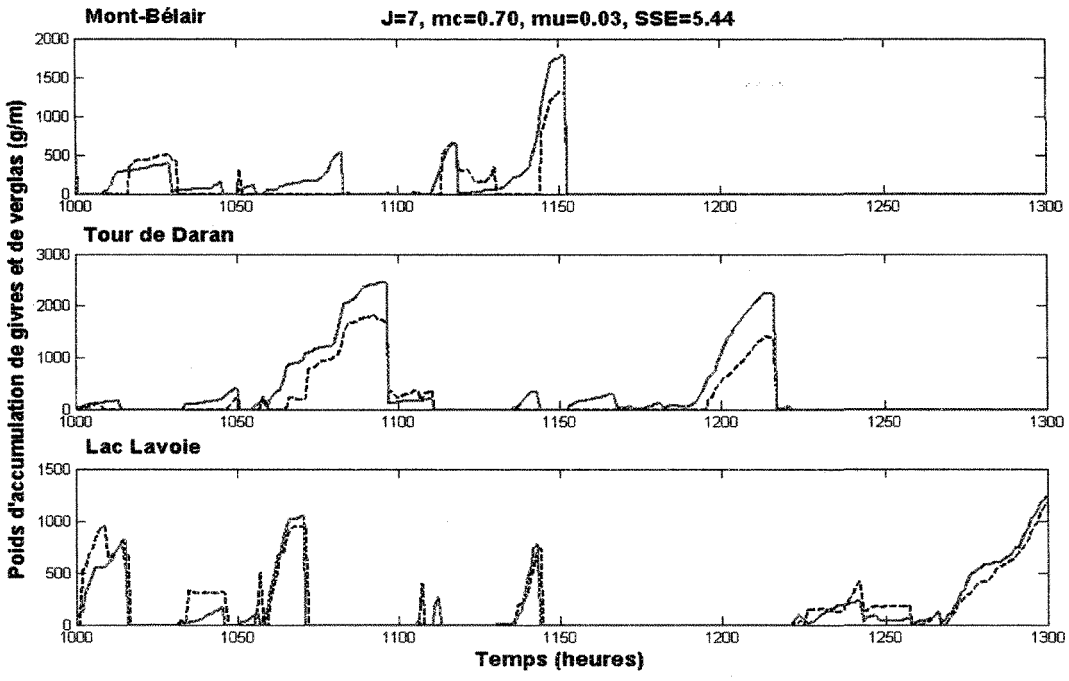

Figure 6-11: Prédictions des événements de givre et de verglas, Groupe 2, J=7, saison 2003-2004

Même si des données sont manquantes à la station du Lac Lavoie pour la saison d'observation 2001-2002, le modèle testé continue de prédire fidèlement les événements se produisant aux deux autres stations (voir figure 6-12). Ces résultats nous permettent donc d'affirmer que « *l'algorithme d'apprentissage intégrant les données manquantes est valide et fonctionnel pour le groupe du Saguenay – Lac-Saint-Jean* ».

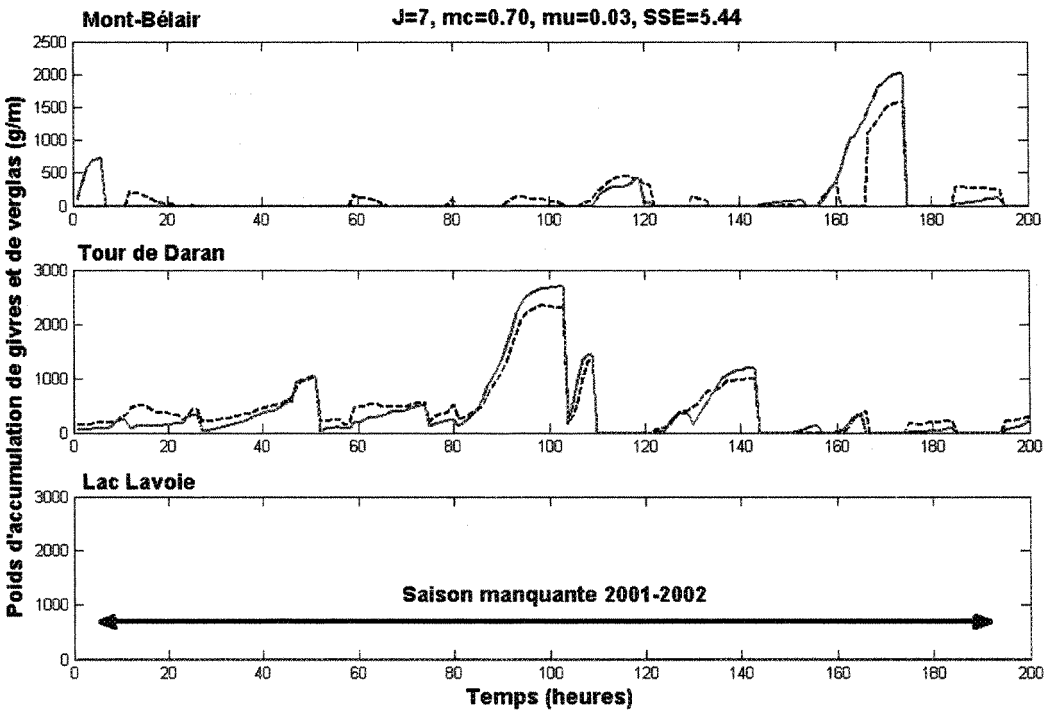

Figure 6-12: Prédictions des événements de givre et de verglas, Groupe 2, J=7, saison 2001-2002

6.3.2 Résultats des tests avec les données du groupe 1 et du groupe 3

En ce qui concerne les validations du modèle de prédiction des événements de givre et de verglas pour le groupe de Montréal-Québec et le groupe de la Côte-Nord, aucune expérience n'a été réalisée puisque ces groupes sont composés d'un nombre plus élevé de stations de mesure. Étant donné que chacun d'eux est constitué de 8 stations, cela implique que leur configuration minimale doit être un réseau de neurones à trois couches ayant 24 cellules dans la couche d'entrée, 8 cellules dans la couche de sortie et un nombre à déterminer de cellules dans la couche cachée. De plus, on doit rechercher les paramètres

d'apprentissage permettant d'obtenir les modèles capables de prédire avec succès les propagations spatiales et temporelles des événements de givre et de verglas à l'intérieur de ces groupes de stations. En effet, nous sommes en présence de huit courbes de prédictions à surveiller au lieu d'en avoir seulement trois comme les expériences qui précèdent. Alors, la réalisation des tests pour ces deux groupes sera plus complexe que ceux réalisés dans le présent ouvrage. De plus, il est possible de prédire que cela nécessitera au moins trois fois plus d'essais pour obtenir les paramètres adéquats relatifs à ces modèles. La validation de l'algorithme d'apprentissage intégrant les données manquantes avec les données de ces deux groupes n'est donc pas nécessaire puisque nous pouvons affirmer pertinemment qu'il est fonctionnel et valide pour un groupe constitué de trois stations de mesure.

CHAPITRE 7

CONCLUSION ET RECOMMANDATIONS

CHAPITRE 7

CONCLUSION ET RECOMMANDATIONS

7.1 Conclusion générale

Nombreuses sont jusqu'à présent les études réalisées sur les phénomènes de givrage atmosphérique. En ce qui concerne le suivi en temps réel des événements de givre et de verglas à travers la province de Québec, les études effectuées au sein de la CIGELE ont porté fruit en proposant des modèles de prédiction à cet effet. Tout a commencé par l'étude des répartitions des événements de givre et de verglas, ce qui a permis de subdiviser la province en plusieurs régions où sont localisés les groupes de stations de mesure d'Hydro-Québec. Ensuite, des études ont été réalisées afin de développer les premiers modèles de réseaux de neurones destinés à la prédiction de ces événements. Le plus grand problème rencontré lors de ces travaux est la présence des lacunes au niveau de la base de données de givrage SYGIVRE. C'est la raison pour laquelle le présent travail a été amorcé afin d'intégrer ces valeurs manquantes en vue du développement d'un modèle de prédiction. Tel que cité dans le titre du présent projet, celui-ci porte sur le traitement des données manquantes en vue de la prédiction de l'évolution temporelle des tempêtes de givre et de verglas. Ainsi, l'objectif principal de cette étude est d'intégrer les données manquantes des groupes de stations SYGIVRE au modèle de prédiction. Ce modèle sert à prédire les poids d'accumulation de givre et de verglas lors des occurrences des tempêtes. Selon les définitions établies dans ce travail, un événement est présenté comme une tempête ou une

suite d'événements élémentaires. Quant à l'événement élémentaire, il est vu comme une période d'accumulation de givre et de verglas.

Concernant le réseau SYGIVRE, les données exploitées dans ce travail couvrent deux saisons d'observation. Lors de la saison 2001-2002, on a pu recueillir des données provenant de 35 stations de mesure. Pour la saison 2003-2004, on a compté exactement 32 stations de mesure. Les informations contenues dans cette base de données sont l'ensemble des relevés horaires des valeurs des températures de l'air et les signaux des givromètres. Plusieurs problèmes sont connus et ont été surmontés pour mener à terme la phase de transformation des données brutes. Ces problèmes sont dus à la complexité des phénomènes de givrage atmosphérique et au manque d'information permettant de les expliciter et de les résoudre adéquatement. En premier lieu, on a constaté que les variables disponibles à toutes les stations sont seulement les températures extérieures et les signaux de givromètre. En second lieu, la méthode utilisée pour identifier les événements élémentaires de givre et de verglas est de nature déterministe. En effet, nous nous sommes basés sur les valeurs des signaux de givromètre pour repérer les débuts et les fins des périodes d'accumulation selon les observations enregistrées à toutes les stations de mesure. Or, nous savons pertinemment que l'évolution d'une précipitation de givre et de verglas n'est pas forcément une accumulation. On peut assister à plusieurs périodes de délestage entre la détection d'un premier signal de givromètre et la détection du dernier signal indiquant la fin de l'événement élémentaire. De plus, plusieurs facteurs influencent l'accumulation de givre et de verglas sur les lignes de transport aériennes d'énergie

électriques. Alors, cette méthode d'identification des événements élémentaires peut être remise en question. Cependant, l'objectif principal de cette étude est de trouver une méthodologie pour intégrer les données manquantes au modèle de prédiction des tempêtes de givre et de verglas. L'obtention d'un ensemble d'événements élémentaires qui n'est pas une reproduction exacte des événements réels n'est donc pas un obstacle majeur pour la réalisation de ce travail. D'autre part, les résultats de la méthode d'estimation des taux horaires d'accrétion de givre et de verglas sont relativement représentatifs des événements élémentaires recensés au cours de cette étude. On peut constater que les valeurs obtenues ne sont pas théoriquement exemptes d'erreurs. L'algorithme utilisé est nommé « algorithme 9-13-20-23 ». Ce dernier repose sur un paramètre correspondant au type de précipitation durant l'heure écoulée: soit un dépôt de givre, soit un dépôt de verglas ou encore un dépôt à dominance de givre ou de verglas; le deuxième paramètre est le nombre de signaux de givromètre. Ensuite, des calculs se basant sur un algorithme conçu par Hydro-Québec, appelé « algorithme 9-23 », sont effectués pour évaluer les taux d'accrétion durant l'acquisition en cours.

Après avoir procédé à une classification de chaque événement élémentaire selon son type d'accumulation, l'algorithme d'identification des tempêtes a été conçu en se basant sur les facteurs suivants: les temps de latence entre les événements élémentaires, les corrélations entre les types d'événement et la sévérité des tempêtes étudiées. Nous avons recensé 507 tempêtes de givre et de verglas lors des saisons d'observation 2001-2002 et 2003-2004. Leur durée totale est égale à 3948 heures et le cumulatif du nombre de signaux

de givromètre enregistrés est égal à 9167 coups. En moyenne, chaque tempête dure environ huit heures et on enregistre 18 signaux de givromètre durant son occurrence.

Quant aux problèmes de données manquantes, le présent travail a permis de conclure qu'ils peuvent être résolus en ayant recours aux méthodes d'imputation statistique. Bien connues dans plusieurs domaines de l'ingénierie, ces méthodes sont également efficaces pour le problème au niveau des données de givrage atmosphérique. Les résultats de cette étude le prouvent puisque des valeurs manquantes de températures ont été estimées grâce à des méthodes de remplacement comme la méthode d'imputation utilisant l'interpolation linéaire et le remplacement par la moyenne. D'après les résultats finaux, les imputations de données manquantes n'influencent pas significativement les comportements du modèle de prédiction des poids d'accumulation de givre et de verglas, et ce grâce à la résistance aux bruits et à la capacité de généralisation du réseau de neurones. De plus, les problèmes dus à la présence des saisons manquantes ont été résolus par l'utilisation de la technique d'inhibition des cellules correspondantes à ces dernières.

Les résultats des tests effectués dans ce travail ont montré que les modèles de prédiction sont fonctionnels et estiment correctement les poids d'accumulation de givre et de verglas. Selon la validation des modèles de prédiction avec les données simulées et avec des données manquantes, nous avons remarqué que la somme minimale des erreurs quadratiques des poids d'accumulation prédits est légèrement supérieure à celle obtenue par les tests sans les données manquantes. En effet, elle est passée de 22.72 g^2/m^2 à 23.62

g^2/m^2. De plus, les courbes de prédiction des meilleurs modèles suivent fidèlement les courbes des événements à prédire avec ou sans données manquantes. Cela signifie d'un part que la présence de saisons manquantes a perturbé légèrement la performance du modèle. D'autre part, le modèle de prédiction est valide en présence de valeurs manquantes puisque les modèles testés estiment correctement les poids d'accumulation de givre et de verglas dans les autres stations où toutes les données sont disponibles. Quant aux résultats des tests avec les données des stations composant le groupe de Québec-Saguenay – Lac-Saint-Jean, nous avons constaté, en dépit de la présence d'une saison manquante à la station du Lac Lavoie durant la saison 2001-2002, que le modèle de prédiction proposé dans cette étude est tout à fait apte à prédire les événements de givre et de verglas. La somme des erreurs quadratiques pour les 1327 heures d'observation de ce groupe de stations est égale à 5.44 g^2/m^2. De plus, les allures des courbes de prédiction sont comme prévues. Les événements de givre et de verglas sont correctement prédits pour les deux saisons de givrage même s'il y a une saison manquante. Toutefois, des résultats insatisfaisants sont aussi obtenus lors des expériences réalisées. En validant le modèle de prédiction avec les données simulées et avec celles du groupe du Saguenay – Lac-Saint-Jean, les résultats montrent que les performances des réseaux de neurones ne sont pas toujours bonnes. Les poids d'accumulation de givre et de verglas à toutes les stations ne sont pas prédits correctement par la plupart des modèles testés puisque leurs performances dépendent des paramètres d'apprentissage. Pour un réseau multicouche, ces paramètres sont choisis arbitrairement et l'obtention de bons paramètres est le résultat d'une procédure par essais.

En conclusion, les résultats de cette étude permettent d'affirmer d'une part que l'application des méthodes statistiques d'imputation est une première solution pour les traitements de données manquantes. Cependant, ceci n'est pas suffisant pour exploiter la base de données SYGIVRE à cause de la présence des saisons manquantes. C'est pourquoi nous avons introduit le principe d'inhibition des cellules correspondantes à ces saisons. Cette étude, nous l'espérons, aura apporté certains éléments de réponses aux problèmes causés par la présence de données et de saisons manquantes. De plus, même si beaucoup de points demeurent ambigus, et ainsi témoignent de la complexité du phénomène, nous souhaitons que ce travail contribue à la compréhension des problèmes reliés au suivi en temps réel des événements de givre et de verglas en stimulant, par les interrogations qu'il soulève, le besoin de les étudier.

7.2 Recommandations

Certaines notions devaient être étudiées au cours de la réalisation du présent travail dans le but d'améliorer les résultats obtenus ou pour découvrir leur applicabilité dans cet axe de recherche. Cependant, elles n'ont pas été abordées. Aussi, des remarques et suggestions se sont révélées au fur et à mesure des réalisations des différentes étapes de ce travail. En résumé, les recommandations pour les futurs travaux connexes à ce mémoire sont les suivantes :

- utiliser absolument les autres variables d'études comme la vitesse du vent, la direction du vent, l'humidité relative de l'air afin de tenir compte de tous les conditions d'apparition des accumulations de givre et de verglas;
- développer une collaboration étroite avec les responsables du réseau SYGIVRE d'Hydro-Québec pour éviter les lacunes ou les pertes de données;
- examiner la possibilité d'une collaboration avec Météo-Canada pour obtenir le maximum de relevés météorologiques ou climatologiques des différentes stations de mesure;
- envisager la possibilité d'étudier les données de givrage en temps réel en plus de l'étude en différé. Ceci nous permettra de suivre concrètement leurs évolutions spatiales et temporelles.
- étudier l'applicabilité d'une architecture de réseau de neurones récurrent pour modéliser le fait que les sorties du modèle sont étroitement liées à ses entrées. Dans

ce cas, il faudra rechercher un autre algorithme plus approprié pour intégrer les saisons manquantes lors du processus d'apprentissage des exemples;

- appliquer une meilleure méthode d'identification des types d'accumulation;
- rassembler et faire une étude globale des données depuis 1992 en tenant compte des résultats positifs obtenus lors des études antérieures et ceux de la présente étude.

RÉFÉRENCES

RÉFÉRENCES

[1] **Abdi H.**, «Les réseaux de neurones». Presses Universitaires de Grenoble, 1994.

[2] **Allison P.D.**, «Missing Data». Thousand Oaks: Sage, 2001.

[3] **Bouret P., Samuelides M.**. «Réseaux neuronaux». Teknea, 1991.

[4] **Breen R.**, «Regression models censored, sample selected, or truncated data». SAGE University Paper, SN 07-011, 1996.

[5] **Chouinard L.E., Elfashny K.,, Nguyen V. et Laflamme J.**, «Modeling of Icing Events Based on Passive Ice Meter Observations in Quebec». Atmospheric Research 46. Elsevier, pp.169-179, 1998.

[6] **Davalo E., Naïm P.**, «Des réseaux de neurones». Eyrolles, 1990.

[7] **Davidson F.**, «Principles of Statistical Data Handling». SAGE Publications, London, 1996.

[8] **Dempster A. P., Laird N.M., Rubin D.**, «Maximum likelihood from incomplete data via the EM algorithm». Journal of the Royal Statistical Association Society 39. pp. 1-38,1997.

[9] **Draper R., Smith H.**, «Applied Regression analysis». Wiley Series in Probability and Statistics, 4th edition, John Wiley & Sons Inc., 1998.

[10] **Druez J., Bouchard D. et Dufour C.**, «Classification et Analyse Statistique des Événements de Givrage Enregistrés au Mont Valin, Saisons 1986-1998». Université du Québec à Chicoutimi, 2000.

[11] **Druez J., McComber P. et Farzaneh M.**, «Analysis of Atmospheric Icing Events Observed at the Mount Valin Test Site During the 1995-96 Season». Ninth International Offshore and Polar Engineering Conference. Brest, France, pp.574-580, 1999.

[12] **Elfashny K., Chouinard L. E., Laflamme J.**, «Estimation of Combined Wind and Ice Loads on Telecommunication Towers in Québec, Phase 1: Modeling of the Ice

and Wind Observations». Seventh International Workshop on Atmospheric Icing of Structures. Chicoutimi, June 1996.

[13] **Elfashny K.N.G., Chouinard L.E., Nguyen V.T.V. et Laflamme J.**, «Characteristics of Icing Events in Quebec». Eight International Workshop on Atmospheric Icing of Structures. Reykjavik, Islande, pp. 537-542, 1998.

[14] **Eter W., Houde L., Farzaneh M.**, «Système de suivi des tempêtes de verglas en temps réel». Septembre 2003.

[15] **Farzaneh M. et Savadjiev K.**, «Icing Events Occurrence in Quebec : Statistical Analysis of Field Data». Preoceedings of 8th International Offshore and Polar Engineering. Montréal (QC), Canada, 1998.

[16] **Félin B.**, «Remote Sensing of Atmospheric Icing in Québec», Second International Workshop on Atmospheric Icing of Structures». Trondheim Norway, 1984, pp. 39-45.

[17] **Guesdon C., Houde L., Farzaneh M. et Chouinard L.**, «Etudes des répartitions des événements de verglas et de givre à travers le Québec». Automne 2000.

[18] **Haldar A., Pon C., McComber P. et al.**, «Validation of Ice Accretion Models for Freezing Precipitation Using Field Data». Seventh International Workshop on Atmospheric Icing of Structures. Chicoutimi (QC), Canada, 1996.

[19] **Hydro-Québec**, «Givromètre : Fiche technique pour système de mesure de verglas».

[20] **Laflamme J.**, «Space and Time Estimates Ice Loads from two Interactive Networks of Ice Accretion Measurements Sites». International Seminar on Ice Load Measurements. Norway, 1992.

[21] **Laflamme J., Côté Y. et Latour A.**, «Description of Mont-Bélair Ice Load Measurements Site». Seventh International Workshop on Atmospheric Icing of Structures. Chicoutimi, Canada, 1996, p453.

[22] **Laflamme J., Fortin R. et Côté Y.**, «Sygivre, The Hydro-Québec Real Time Icing Events management System». Seventh International Workshop on Atmospheric Icing of Structures. Chicoutimi, Canada, 1996, p459.

[23] **Laflamme J.**, «Nouveaux sites proposés et limites de sous-ensembles géographiquement homogènes; Sites de mesure du givrage atmosphérique». Carte, Hydro-Québec, décembre 1998.

[24] **Larouche E., Rouat J., Bouchard G. et Farzaneh M.**, «Exploration de différentes architectures de réseaux de neurones pour la prédiction de la glace atmosphérique sur les conducteurs des réseaux électriques ». Hiver 2002.

[25] **Little R. and Rubin D.**, «Statistical Analysis with Missing Data, Second Edition». New York: John Wiley and Sons, 2002.

[26] **Little R., Rubin D.**, «Statistical Analysis with Missing Data». John Wiley & Sons, New York, 1987.

[27] **Makkonen L. et Ahti K.**, «Climatic mapping of Ice Loads bases on Airport Weather Observations». Sixth International Workshop on Atmospheric icing of Structures. Budapest, Hongrie, 1993, pp.13-17.

[28] **Maralbashi-Zamini S., Ezzaidi H. et Farzaneh M.**, « Développement de réseaux de neurone pour la prédiction du type et du taux de glace accumulée sur les lignes aériennes de transport d'énergie électrique». Chicoutimi, Canada, 2007.

[29] **McComber P., Druez J., De Lafontaine J., Paradis A. et Laflamme J.**, «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.

[30] **McComber P., De Lafontaine J. et Laflamme J.**, «A Neural System to Estimate Transmission Line Icing». Eighth International Workshop on Atmospheric Icing of Structures. Reykjavik, Islande, 1998, pp. 101-106.

[31] **McComber P., Druez J. et Laflamme J.**, «A comparison of Selected Models for Estimating Cable Icing». Proceedings of Sixth International Workshop on the Atmospheric Icing of Structures». Budapest, Hangary, 1993, pp41-46.

[32] **McComber P., Latour A., Druez J. et Laflamme J.**, «The Icing rate meter, an instrument to evaluate transmission line icing». Seventh International Workshop on Atmospheric Icing of Structures. Chicoutimi (QC), Canada, 1996, pp159-168.

[33] **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, pp149-154.

[34] **Rencher A.**, «Method of multivariate analysis». Wiley Series in Probability and Statistics, 1995.

[35] **Savadjiev K. et Farzaneh M.**, «Analysis and Interpretation of Icing rate Meter and Load Cell Measurements on the Mt. Bélair Icing Site». Proceedings of 9th International Offshore and Polar Engineering. Brest, France, pp.607-611, 1999.

[36] **Savadjiev K. et Farzaneh M.**, «Statistical Analysis of Two Probabilistic Models of Ice Accretion on Overhead Line Conductors». Proceedings of 8th International Offshore and Polar Engineering. Montréal (QC).

[37] **Savadjiev K., Latour A. et Paradis A.**, «Estimation of Ice Accretion Weight from Field Data Obtained on Overhead Transmission Line Cables». Seventh International Workshop on Atmospheric Icing of Structures. Chicoutimi (QC), Canada, 1996, pp.125-130.

[38] **Schafer J. L. et Olsen M.K.**, «Multiple imputation for multivariate missing-data problems: a data analyst's perspective». Behavioral Research, 33, pp. 545-571, 1998.

[39] **Scott J.**, «Regression models for categorical and limited dependent variables", SAGE Publication, London, 1997.

[40] **TransÉnergie**, «Réseau des sites de mesures SYGIVRE». Carte, Hydro-Québec, novembre 1999.

ANNEXES

ANNEXE A: Détermination des types d'accumulation de givre et de verglas

```
'-------------------------------------------------------------
'Fonction pour déterminer le type d'accumulation
'Temp :     Température de l'air
'TypeRef :  Type d'accumulation durant les dernières heures
'-------------------------------------------------------------
Function DeterminerTypeAcc(Temp As Double, TypeRef As Integer) As Integer

      'Le type d'accumulation ou d'accrétion
      Dim TypeAcc As Integer

      'Selon le dernier type d'accumulation
      Select Case TypeRef

      'Si on ne connait pas le type d'accumulation durant les
      'dernières heures (Case = 0)
              Case 0

                  'Si la température est en dessus de -3 degré C
                  If Temp >= -3 Then
                      TypeAcc = 10        'Verglas
                  Else
                      'En dessus de -6 degré C
                      If Temp >= -6 Then
                          TypeAcc = 12    'Forte probabilité de Verglas
                      Else
                          'En dessus de -9 degré C
                          If Temp >= -9 Then
                              TypeAcc = 21 'Forte probabilité de Givre
                          'En bas de 9 degré C
                          Else
                              TypeAcc = 20 'Givre
                          End If
                      End If
                  End If

      End Select

End Function
```

ANNEXE B: Algorithme 9-13-20-23, estimation des taux d'accrétion de givre et de verglas

```
'-------------------------------------------------------------
't : température de l'air
'G : nombre de signaux du givromètre
'TypeRef : type d'accumulation durant les dernières heures
'-------------------------------------------------------------
Sub EstimerTaux(t As Double, G As Integer, TypeRef As Integer)

    Dim T1      As Double
    Dim TypeAcc As Integer

    'Si Nombre de coups de givromètre supérieur a zéro
    If G > 0 Then
      'Algorithme 9-23 modifié
            TypeAcc = DeterminerTypeAcc(t, TypeRef)
            Select Case TypeAcc
                Case 10
                    T1 = G * 23
                Case 12
                    T1 = G * 20
                Case 21
                    T1 = G * 13
                Case 20
                    T1 = G * 9
                Case Else
                    T1 = G * 16
            End Select
    End If
End Sub
```

ANNEXE C: Détermination du type d'accumulation d'un événement élémentaire

```
'-----------------------------------------------------------------
' Identifier le type d'accumulation d'un événement
'-----------------------------------------------------------------
'En dessus de -3 degré C
If TempMoy >= -3 Then
    If TempMin >= -3 Then
          TypeAcc = 10     'Événement de verglas
    Else
          TypeAcc = 12     'Événement à dominance de verglas
    End If
End If


'En bas de -3 degré C
If TempMoy < -3 Then
    If TempMax < -3 Then
         TypeAcc = 20     'Événement de givre
    Else
         TypeAcc = 21     'Événement à dominance de givre
    End If
End If
```

ANNEXE D: Vérification de la sévérité d'une tempête de givre et de verglas

```
'-------------------------------------------------------------
'Vérifier Présence de forte accumulation
'CNCG : Cumulatif du nombre de coup de givromètre
'TNCG : Taux moyen du nombre de coup de givromètre
'-------------------------------------------------------------
Function ForteAcc(CNCG As Integer, TNCG As Double)

    If TNCG >= 8 And CNCG > 8 Then
        ForteAcc = True
    Else
        ForteAcc = False
    End If

End Function

'-------------------------------------------------------------
'Vérifier Présence de moyenne accumulation
'CNCG : Cumulatif du nombre de coup de givromètre
'TNCG : Taux moyen du nombre de coup de givromètre
'-------------------------------------------------------------
Function MoyenneAcc(CNCG As Integer, TNCG As Double)

    If TNCG >= 4 And CNCG > 4 Then
        MoyenneAcc = True
    Else
        MoyenneAcc = False
    End If

End Function
```

ANNEXE E: Algorithme d'identification des tempêtes de givre et de verglas

```
'-------------------------------------------------------------
'Vérifier si un événement peut être associé à son précédent
'-------------------------------------------------------------
Function EvenementOK(IT As Double, TypeAcc As Integer, TypeAccPr As
Integer, CNCG As Integer, TNCG As Double) As Boolean

 'Verifier si Temps de latence <= 2
 If IdleTimeOK(IT, 2) Then
        EvenementOK = True
 Else
        'Verifier la corrélation entre les types d'événements
        If TypeAccOK(TypeAcc, TypeAccPr)
           Or DominanceOK(TypeAcc, TypeAccPr) Then

           If IdleTimeOK(IT, 4) Or (IdleTimeOK(IT, 6)
             And TypeAccPr = 20)
             Or (TNCG >= 0.5 And IdleTimeOK(IT, 10)) Then
                EvenementOK = True
           Else
                'Verifier Presence forte accumulation
                If ForteAcc(CNCG, TNCG) Then
                   If IdleTimeOK(IT, 20)
                        Or (IdleTimeOK(IT, 22) And TypeAccPr = 20)    Then
                        EvenementOK = True
                   Else
                        EvenementOK = False
                   End If
                Else
                    'Verifier Presence Moyenne accumulation
                    If MoyenneAcc(CNCG, TNCG) Then
                        If IdleTimeOK(IT, 8)
                              Or (IdleTimeOK(IT, 10) And TypeAccPr=20)
                              Then
                            EvenementOK = True
                        Else
                            EvenementOK = False
                        End If
                    Else
                        EvenementOK = False
                    End If
                End If
            End If

       'Si ce n'est pas le même type d'accumulation
        Else
           EvenementOK = False
        End If
End If

End Function
```

ANNEXE F: Algorithmes pour la simulation de données virtuelles

```
'-------------------------------------------------------------
'-Définir les décalages de temps et les durées des tempêtes
'-Dt est une distribution normale de moyenne mu_dt(s) et d'écart
' type sig_dt(s)
'-D est une distribution normale de moyenne mu_d(s) et d'écart 'type
sig_d(s)
'-------------------------------------------------------------

For s = 1 To NbStations
   If s > 1 Then
        Dt(s) = NormInv(Rnd(), mu_dt(s), sig_dt(s)) + Dt(s - 1)
    Else
        Dt(s) = NormInv(Rnd(), mu_dt(s), sig_dt(s))
    End If
    D(s) = NormInv(Rnd(), mu_d(s), sig_d(s))
    If D(s) < 2 Then D(s) = 2 'Durée minimale d'une tempête

    'Calculs des dates des fins
    Fin(s) = Dt(s) + D(s)
Next


'-------------------------------------------------------------
'Définir les taux de précipitations moyens au sein des stations
'-------------------------------------------------------------

Prec_Moy(1) = NormInv(Rnd(), 6, 3)
Prec_Moy(2) = Prec_Moy(1)*1.5 - 1.2
Prec_Moy(3) = Prec_Moy(1)*0.8 + Prec_Moy(2)*0.7 - 1.5



'-------------------------------------------------------------
'Calculer le taux d'accrétion correspondant
'aux paramètres (Température Temp, Précipitation Prec)
'-------------------------------------------------------------

Function Calculer_TauxAcc(Temp, Prec) As Double
    Dim TauxAcc As Double
    TauxAcc = -0.9 * Temp + 0.5 * Prec + 1.4
    Calculer_TauxAcc = TauxAcc
End Function
```

ANNEXE G: Statistique des tempêtes de givre et de verglas du groupe 1 et du groupe 3

Station	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Beaumont	6	29	56	5	7	9	19
Poste Chateauguay	8	21	40	3	7	5	9
Chenier	7	33	72	5	12	10	37
Donnacona	11	39	79	4	8	7	21
Gentily	8	24	34	3	8	4	7
Ligne exp IREQ	8	26	49	3	9	6	16
Pont Pierre-Laporte	7	23	46	3	7	7	19
Sorel	2	8	11	4	6	6	7
Total/Moyenne/Max	57	203	387	4	12	7	37

Tableau G-1: Analyse descriptive des tempêtes de givre et de verglas, groupe 1, saison 2001-2002.

Station	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Beaumont	3	4	19	1	2	6	15
Poste Chateauguay	6	36	32	6	17	5	10
Chenier	5	35	53	7	15	11	16
Donnacona	*	*	*	*	*	*	*
Gentily	6	27	38	4	13	6	15
Ligne exp. IREQ	4	33	52	8	11	13	18
Pont Pierre-Laporte	9	55	73	6	16	8	29
Sorel	*	*	*	*	*	*	*
Total/Moyenne/Max	33	188	267	5	17	8	29

Tableau G-2: Analyse descriptive des tempêtes de givre et de verglas, groupe 1, saison 2003-2004.

Station	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Ste-Marguerite	23	144	350	6	19	15	57
Lac St-Pierre	*	*	*	*	*	*	*
Poste Arnaud	11	41	125	4	15	11	48
Lac Caouette	19	143	300	8	24	16	48
Eric	16	161	568	10	36	36	115
Poste Montagnais	12	36	73	3	12	6	26
Premio	41	375	1170	9	33	29	153
Virot	21	136	332	6	21	16	66
Total/Moyenne/Max	122	900	2586	7	36	21	153

Tableau G-3: Analyse descriptive des tempêtes de givre et de verglas, groupe 3, saison 2001-2002.

Station	NbSE	DT	CNCGT	DM	DMAX	CNCGM	CNCGMAX
Ste-Marguerite	25	164	540	7	19	22	89
Lac St-Pierre	24	302	713	13	61	30	195
Poste Arnaud	13	50	134	4	11	10	39
Lac Caouette	16	313	298	20	176	19	89
Eric	14	131	333	9	26	24	87
Poste Montagnais	14	35	78	2	6	6	15
Premio	32	252	575	8	30	18	107
Virot	*	*	*	*	*	*	*
Total/Moyenne/Max	138	1246	2671	9	176	19	195

Tableau G-4: Analyse descriptive des tempêtes de givre et de verglas, groupe 3, saison 2003-2004.

ANNEXE H: Statistique des données manquantes du groupe 1 et du groupe 3

Station	2001-2002 (/153 heures)	2003-2004 (146 heures)	Total (/299 heures)	Pourcentage (%)
Beaumont	0	0	0	0
Poste Chateauguay	0	0	0	0
Chenier	11	3	14	5
Donnacona*	0	146	146	49
Gentily	11	15	26	9
Ligne exp IREQ	0	0	0	0
Pont Pierre-Laporte	0	1	1	0
Sorel	0	146	146	49
Total (/Groupe)	22	311	333	14

Tableau H-5: Données de températures manquantes, groupe 1.

Station	2001-2002 (/785 heures)	2003-2004 (/643 heures)	Total (/1428 heures)	Pourcentage (%)
Ste-Marguerite	20	1	21	2
Lac St-Pierre*	785	33	818	6
Poste Arnaud	0	0	0	0
Lac Caouette	20	15	35	3
Eric	2	1	3	0
Poste Montagnais	10	0	10	1
Premio	29	1	30	2
Virot*	1	643	644	45
Total (/Groupe)	867	649	1561	14

Tableau H-6: Données de températures manquantes, groupe 3.

ANNEXE I: Résultats des tests avec les données manquantes

mc|mu	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	100	100	1364	100	100	100	100	100	100
0.1	2259	100	64	100	100	100	100	100	100
0.2	100	3073	1290	100	100	100	100	100	100
0.3	100	100	100	100	100	100	100	100	100
0.4	100	100	100	92	65	100	100	100	100
0.5	1363	100	100	100	100	100	100	100	100
0.6	493	808	100	100	100	100	100	100	100
0.7	100	100	100	100	100	100	100	100	227
0.8	100	100	100	100	100	100	100	3539	1930
0.9	100	100	100	100	100	100	100	100	100

Tableau I-7: Sommes des erreurs quadratiques, J = 5.

mc|mu	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	100	100	100	3247	24	100	88	100	100
0.1	100	100	100	87	100	100	73	100	100
0.2	100	100	100	4174	100	100	100	100	100
0.3	1873	100	103	100	100	100	100	61	88
0.4	100	100	100	290	100	100	87	69	100
0.5	2183	1356	87	64	100	100	65	62	92
0.6	100	1871	100	100	95	100	100	100	100
0.7	100	100	100	100	100	100	100	100	100
0.8	1284	100	100	100	100	1930	100	100	100
0.9	100	100	100	100	100	100	1930	100	100

Tableau I-8: Sommes des erreurs quadratiques, J = 6.

mc|mu	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	109	100	87	100	51	100	100	100	100
0.1	100	100	100	100	100	7311	100	100	84
0.2	100	100	100	100	64	5565	100	99	100
0.3	74	100	104	100	100	100	29	65	68
0.4	73	100	100	1930	100	100	100	100	100
0.5	100	100	480	100	100	100	100	100	100
0.6	100	100	32	100	100	100	87	85	100
0.7	100	64	100	100	100	100	100	98	100
0.8	100	87	100	100	100	100	100	100	100
0.9	99	100	100	100	100	100	100	100	100

Tableau I-9: Sommes des erreurs quadratiques, J = 7.

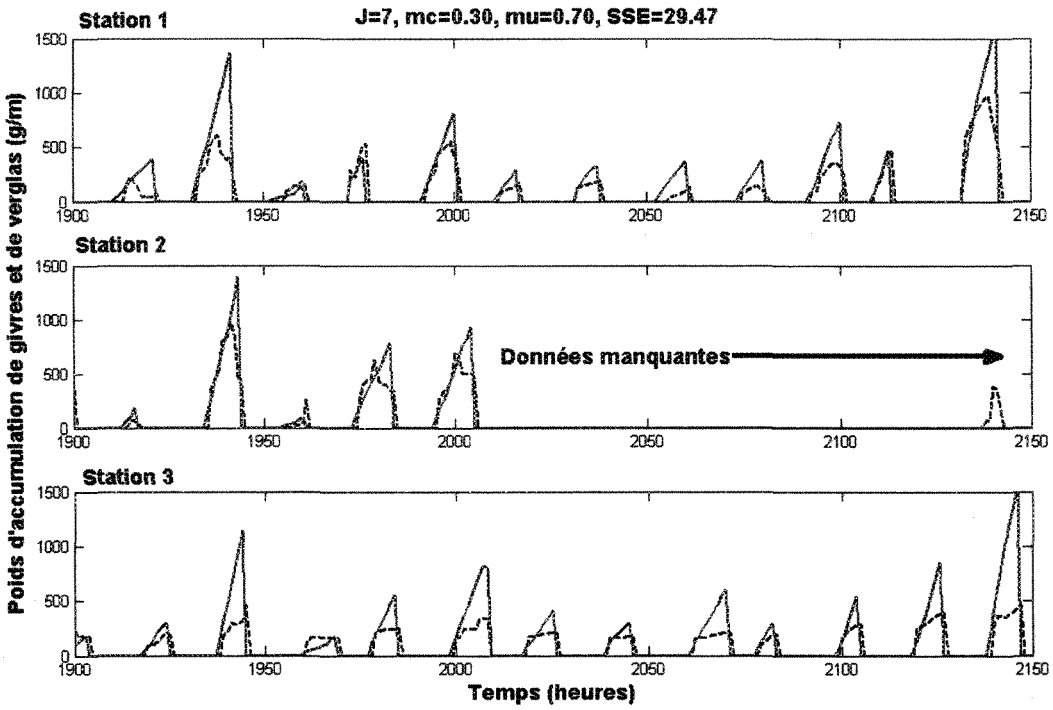

Figure I-1: Prédictions des événements de givre et de verglas avec les données simulées avec données manquantes, J=7.

ANNEXE J : Résultats des tests avec les données du groupe 2

mc|mu	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	320	34	38	34	34	127	160	34	34
0.1	577	34	31	34	34	27	34	34	34
0.2	34	245	149	34	34	34	34	34	34
0.3	34	245	149	34	34	34	34	34	34
0.4	34	34	150	34	34	34	33	34	49
0.5	28	447	34	34	184	167	34	34	34
0.6	34	34	34	34	55	34	34	34	25
0.7	34	34	34	34	34	34	34	34	34
0.8	34	34	34	34	34	34	34	34	34
0.9	34	34	54	34	34	34	22	34	34

Tableau J-10: Sommes des erreurs quadratiques, J = 5.

mc|mu	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	30	33	34	34	34	34	25	34	28
0.1	34	34	34	34	34	24	22	29	34
0.2	34	34	34	34	34	242	34	34	22
0.3	34	43	25	24	30	34	34	34	34
0.4	34	179	32	34	34	34	34	34	141
0.5	34	295	34	34	34	227	34	34	34
0.6	34	34	34	34	34	34	21	34	34
0.7	34	34	293	34	34	34	34	34	61
0.8	32	34	34	28	22	34	34	21	34
0.9	34	34	16	21	34	34	34	34	34

Tableau J-11: Sommes des erreurs quadratiques, J = 6.

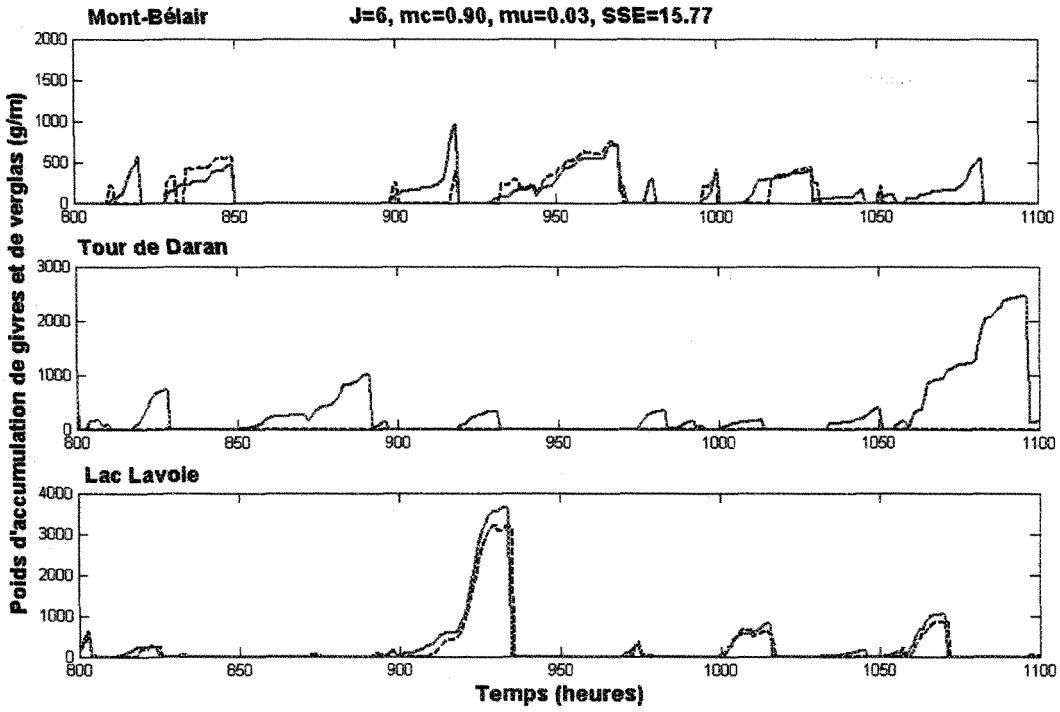

Figure J-2: Prédictions des événements de givre et de verglas, J=6

mc|mu	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	28	33	34	34	34	34	34	34	34
0.1	30	27	12	34	23	34	34	34	34
0.2	34	34	34	34	34	34	34	34	34
0.3	34	34	34	34	34	33	34	29	34
0.4	34	127	34	34	34	11	17	34	34
0.5	34	295	34	34	34	227	34	34	34
0.6	115	34	34	21	34	22	34	34	34
0.7	34	43	5	34	34	27	34	34	32
0.8	34	34	34	22	34	34	34	34	34
0.9	12	34	34	34	16	26	34	34	34

Tableau J-12: Sommes des erreurs quadratiques, J = 7.