

UNIVERSITÉ DU QUÉBEC À CHICOUTIMI

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

Par

WALID ETER

**SYSTÈME DE SUIVI DES TEMPÊTES DE  
VERGLAS EN TEMPS RÉEL**

Analysis of real time icing events

September 2003



### **Mise en garde/Advice**

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

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

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

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

## Abstract

The study of atmospheric ice accumulation on power networks is a major concern for cold climate regions like Quebec. This includes the understanding of several complex phenomena in order to reduce the risk of damage to the existing and future electrical overhead networks. In this perspective, it is important to understand these phenomena by the analysis of the anterior atmospheric icing events and to model them.

The present research was carried out within the framework of the NSERC/HYDRO-QUÉBEC/UQAC Industrial Chair on Atmospheric Icing of Power Network Equipment (CIGELE), aiming to analyze the data collected on icing events through the SYGIVRE measurement network of Hydro-Québec, and to establish a model of prediction in real-time for tracking the icing storm evolution by detecting the mechanical charges resulting from the ice accumulation in different regions of Quebec. The available meteorological information of the SYGIVRE database covers 6 years for 28 stations, beginning in 1992, and is based on an improved icing rate meter (IRM). A neural network technique was proposed as the basis for the modeling work.

The exploratory analysis of the icing events contained in the SYGIVRE database constitutes the first step of the study. This analysis consisted in extracting the icing storms from the database while putting the emphasis on each of the database parameters characteristics, and in making the data adapted to neural networks. Four groups of geographically and meteorologically related stations are involved in the prediction, each one being processed separately.

A first prediction model based on neural network techniques has been devised for studying the evolution of icing storms on the basis of the spatial relationships between stations of a group. The variable to be predicted was the *icing storm occurrence*, in a binary format (events occur or not), for each station of a given group at a specific time. The same model has been improved by the integration of meteorological quantitative variables. It has

shown significant improvements over the previous models described in the literature, based on the logistic regression method.

Finally, it has been determined that the addition of a greater number of years of observations to the two models could improve the prediction performance, since this would increase the quantity of information taken into account. This also constitutes an essential requirement for neural network predictions.

## Résumé

L'étude des événements de givrage atmosphérique sur les réseaux de transport de l'énergie électrique constitue une préoccupation importante des habitants des régions de climat froid, telles que la province de Québec. Ceci inclut la compréhension de plusieurs phénomènes complexes en vue de réduire les risques de dommages aux réseaux de transmission existants et à ceux qui seront être implantés dans le futur. Dans cette perspective, il est important de comprendre ces phénomènes par l'analyse des événements de givrage atmosphériques antérieurs et de les modéliser.

La présente recherche, effectuée dans le cadre des travaux de la Chaire industrielle NSERC/HYDRO-QUÉBEC/UQAC sur le givrage atmosphérique des équipements des réseaux électriques (CIGELE), avait pour but d'analyser les événements passés tels qu'enregistrés par le réseau de mesure SYGIVRE d'Hydro-Québec, et de créer un modèle de prédiction en temps réel de l'évolution des tempêtes de verglas pour différentes régions du Québec. Les informations météorologiques disponibles dans la base de données SYGIVRE, mesurées à l'aide d'un givromètre amélioré (*ice rate meter*), couvrent 6 ans, à partir de 1992, ceci pour un total de 28 stations. Une technique par réseaux de neurones a été choisie comme base pour le travail de modélisation.

L'analyse exploratoire des événements de verglas contenus dans la base de données SYGIVRE constitue la première étape de l'étude. Cette étude a consisté à extraire de la base de données les tempêtes de verglas, à faire ressortir les caractéristiques de chaque paramètre de la base de données et à rendre les données adaptées au traitement par réseaux de neurones. Quatre groupes de stations ayant un rapport géographique et météorologique sont impliqués dans la prédiction, chacun étant traité séparément.

Un premier modèle prédictif basé sur la technique des réseaux de neurones a été conçu en vue d'étudier l'évolution des tempêtes en se basant sur la relation spatiale qui existe entre les stations d'un groupe. La variable à prédire à chaque point pour une station est une variable dichotomique qui prend la valeur 1 s'il y a un événement de givre et 0 sinon. Ce modèle

constitue une grande amélioration comparativement à des modèles antérieurs décrits dans la littérature, basés sur la méthode de régression logistique.

Afin d'obtenir un modèle plus réaliste, un second modèle a été créé pour prédire en temps réel le poids de la glace accumulée sur la structure. Le modèle utilise les informations binaires avec des variables météorologiques, et la prédiction est ajustée pour évaluer différents temps futurs. Les résultats ont montré que la meilleure performance possible du modèle peut être réalisée en ajoutant au modèle de prédiction les variables de température et du poids de la glace. Toutefois, le modèle détecte encore difficilement les événements de verglas extrême.

En conclusion, il est apparu qu'inclure un plus grand nombre d'années d'observations aux deux modèles devrait améliorer la performance de la prédiction, puisque ceci augmenterait la quantité des informations à étudier. Il s'agit d'ailleurs d'une condition essentielle aux prédictions par réseaux de neurones.

## **Acknowledgements**

This study was carried out at the University of Quebec at Chicoutimi within the framework of the NSERC/HYDRO-QUEBEC/UQAC Industrial Chair on Atmospheric Icing of Power Network Equipment (CIGELE).

I take this opportunity to express my thanks to all who have helped me, either for technical or moral support, to finish this thesis.

First and foremost, I would like to thank my director and my co-director, Professors Louis Houde and Masoud Farzaneh, for their support and supervision during all the course of my M.Sc. thesis work. I would also like to thank CIGELE for the financial support during my M.SC. study.

I also would like to thank my friends and colleagues, who gave so generously of their time and knowledge, corrected my mistakes, and created an environment in which ideas developed rapidly.

Finally, I would like to thank all my family, specifically my brother, Khaled, and his wife, Isabelle, for their encouragement and tolerance during the months I spent finishing this research.

# Contents

<b>ABSTRACT .....</b>	<b>II</b>
<b>RÉSUMÉ .....</b>	<b>IV</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>VI</b>
<b>CONTENTS .....</b>	<b>VII</b>
<b>LIST OF FIGURES.....</b>	<b>X</b>
<b>LIST OF TABLES.....</b>	<b>XIV</b>
<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 BACKGROUND.....	2
1.1.1 Problematic.....	2
1.1.2 Research Overview .....	3
1.1.3 Survey of the Literature .....	5
1.2 OBJECTIVES.....	10
1.3 METHODOLOGY.....	11
1.4 PERTINENT LITERATURE.....	13
1.4.1 Spatial analysis.....	14
1.4.2 Prediction of ice accretion.....	15
1.4.3 Insertion of the present work .....	15
<b>CHAPTER 2 ARTIFICIAL NEURAL NETWORK.....</b>	<b>18</b>
2.1 ARTIFICIAL NEURAL NETWORK DEFINITION (NN).....	19
2.1.1 History of Neural Network technology .....	19
2.1.2 Neural Networks capability .....	21
2.1.3 Basic principles of a Neural Networks .....	22
2.1.4 The Artificial Neuron .....	24
2.1.5 The Activation function (transfer function).....	25
2.1.6 Single layer and Multilayer Neural Networks .....	28
2.1.7 Neural Networks Architectures.....	31
2.1.8 Kinds of Neural Networks.....	32
2.1.9 Training of Artificial Neural Networks (learning).....	33
2.1.10 Error function .....	36
2.1.11 Validation of Artificial Neural Networks .....	37
2.2 BACKPROPAGATION .....	38

2.2.1	<i>Generalization</i>	42
2.2.2	<i>Underfitting, Overfitting and Noise</i>	43
2.2.3	<i>Determination number of hidden units (neurons) and layers</i>	44
2.3	RECURRENT NETWORKS	46
2.3.1	<i>Elman Network</i>	46
2.4	PERCEPTRONS (PCM)	50
2.5	LINEAR FILTER	51
2.6	CHOICE OF NEURAL NETWORKS	52
2.7	PREPROCESSING	54
<b>CHAPTER 3</b>	<b>SYGIVRE DESCRIPTION</b>	<b>56</b>
3.1	INTRODUCTION	57
3.2	GEOGRAPHICAL AND TEMPORAL DISTRIBUTION	58
3.3	SYGIVRE DATABASE	63
3.3.1	<i>Ice type</i>	63
3.3.2	<i>IRM signal</i>	66
3.3.3	<i>Ice accumulation weight</i>	67
3.3.4	<i>Temperature</i>	67
3.3.5	<i>Certified events</i>	69
3.3.6	<i>Variables characterisation</i>	70
3.4	DATABASE FOR THE ICING STORM STUDY	71
3.4.1	<i>Icing events</i>	72
3.4.2	<i>Storm events</i>	74
3.4.3	<i>Temperature missing</i>	89
3.5	DATA AVAILABLE FOR NEURAL NETWORK	91
3.6	DATA ANALYSIS	95
<b>CHAPTER 4</b>	<b>ICING STORMS PROPAGATION</b>	<b>103</b>
4.1	INTRODUCTION	104
4.2	VARIABLES	105
4.3	BINARY NETWORK MODEL	108
4.3.1	<i>Performance function and validation</i>	109
4.3.2	<i>Model setting</i>	114
4.3.3	<i>Discussion</i>	125
4.4	BINARY NETWORK MODEL WITH SUPPLEMENTARY VARIABLES	127
4.4.1	<i>Model characterization</i>	128
4.4.2	<i>Results</i>	129
4.4.3	<i>Discussion</i>	137

4.5	CONCLUSIONS .....	138
<b>CHAPTER 5 PREDICTION OF ICE ACCUMULATION WEIGHT .....</b>		<b>140</b>
5.1	INTRODUCTION.....	141
5.2	AVAILABLE VARIABLES.....	142
5.3	NEURAL NETWORK MODEL.....	145
5.4	CRITERIA.....	150
5.5	MODEL ANALYSIS .....	152
5.5.1	<i>One hour ahead .....</i>	<i>152</i>
5.5.2	<i>Analysis of results when using different hours ahead with different previous hours .....</i>	<i>163</i>
5.6	DISCUSSION.....	171
5.7	CONCLUSIONS .....	172
<b>CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS .....</b>		<b>175</b>
6.1	GENERAL CONCLUSIONS.....	176
6.2	RECOMMENDATIONS .....	181
<b>REFERENCES .....</b>		<b>182</b>

## List of Figures

Figure 2-1 Neural Network .....	23
Figure 2-2 Artificial Neuron .....	24
Figure 2-3 Logistic Transfer Function .....	27
Figure 2-4 Artificial Neuron with Activation Function .....	28
Figure 2-5 Single-Layer Neural Network .....	30
Figure 2-6 Two-Layer Neural Network .....	31
Figure 2-7 Recurrent Architecture .....	32
Figure 2-8 Two-layer Backpropagation Network .....	39
Figure 2-9 Overfitting of noisy sine function .....	44
Figure 2-10 A simple recurrent network .....	47
Figure 3-1 Station locations for the SYGIVRE network .....	59
Figure 3-2 The principle station groups .....	62
Figure 3-3 General certified event sample .....	70
Figure 3-4 Dissociation of a certified event .....	75
Figure 3-5 Overlapped and propagated icing events .....	79
Figure 3-6 Idle times less than 3 hours for the Côte Nord group .....	81
Figure 3-7 Storm event propagation between stations of the Côte Nord group .....	82
Figure 3-8 Two different storms due to the difference of accretion types .....	83
Figure 3-9 Example to discuss the mergence of the icing events at station 1 .....	84
Figure 3-10 A big idle time with variable accretion type .....	86
Figure 3-11 The four groups of stations .....	93

Figure 3-12 Histogram of icing events and storm events duration (h) for all stations of the four groups .....	98
Figure 3-13 Observations of ice accumulation weight for all stations of group 1 .....	99
Figure 3-14 Observations of ice accumulation weight for all stations of group 2 .....	100
Figure 3-15 Observations of ice accumulation weight for all stations of group 3 .....	101
Figure 3-16 Observations of ice accumulation weight for all stations of group 4 .....	102
Figure 4-1 Simple sample of prediction for the group 1 .....	106
Figure 4-2 Spatial propagation between group 1's stations for a sample storm event ...	107
Figure 4-3 General example for one, two and three previous hours for a given storm event at one station .....	108
Figure 4-4 Neural network sample applied to the group1 in the case of binary network model .....	109
Figure 4-5 Efficiency measure for group 1 in the case of total validation (binary model) .....	117
Figure 4-6 Efficiency measure for group 2 in the case of total validation (binary model) .....	117
Figure 4-7 Efficiency measure for group 3 in the case of total validation (binary model) .....	117
Figure 4-8 Efficiency measure for group 4 in the case of total validation (binary model) .....	117
Figure 4-9 Efficiency measure for group 1 in the case of validation by separation (binary model) .....	118

Figure 4-10 Efficiency measure for group 2 in the case of validation by separation (binary model) .....	118
Figure 4-11 Efficiency measure for group 3 in the case of validation by separation (binary model).....	118
Figure 4-12 Efficiency measure for group 4 in the case of validation by separation (binary model) .....	118
Figure 4-13 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 1 .....	121
Figure 4-14 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 2 .....	122
Figure 4-15 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 3 .....	123
Figure 4-16 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 4 .....	124
Figure 4-17 Efficiency measure for group 1 in the case of total validation for 1 to 6 previous hours .....	132
Figure 4-18 Efficiency measure for group 2 in the case of total validation for 1 to 6 previous hours .....	133
Figure 4-19 Efficiency measure for group 3 in the case of total validation for 1 to 6 previous hours .....	134
Figure 4-20 Efficiency measure for group 4 in the case of total validation for 1 to 6 previous hours .....	135
Figure 4-21 Efficiency measure for group 1 in the case of validation by separation ....	136

Figure 4-22 Efficiency measure for group 2 in the case of validation by separation ....	136
Figure 4-23 Efficiency measure for group 4 in the case of validation by separation ....	136
Figure 5-1 General example for 1 and 2 hours ahead with 1 and 2 previous hours applied to some ice weight values at station 4.....	145
Figure 5-2 Neural network sample applied to group 1 in the case of ice accumulation weight prediction .....	146
Figure 5-3 Effect of number of neurons in the hidden layer for one hour ahead .....	148
Figure 5-4 Signal of observed and resultant predicted values for station 11 in group 1	153
Figure 5-5 Correlation between the observed and the predicted values for the optimum state of each group and the corresponding R-square value.....	155
Figure 5-6 Sample of failed prediction for station 16 in group 2 .....	157
Figure 5-7 Sample of failed prediction for station 24 in group 4 .....	158
Figure 5-8 Errors between the predicted and observed values versus temperature or ice weight (one sample from each group) .....	160

## List of Tables

Table 3-1 State of measurement seasons for all stations of the SYGIVRE network.....	60
Table 3-2 Number of occurrence at each temperature level for all stations .....	68
Table 3-3 Icing events distribution over six seasons .....	73
Table 3-4 Idle time frequency in terms of time ranges .....	76
Table 3-5 Number of icing events inside certified events .....	77
Table 3-6 Storm events distribution between the measurement seasons .....	87
Table 3-7 List of all stations of the SYGIVRE database with the corresponding number of certified events, icing events and storm events.....	88
Table 3-8 List of all stations with the corresponding number of icing events and storm events with and without missing temperature readings .....	90
Table 3-9 Stations with multiple instruments .....	94
Table 4-1 Sample storm event in 1996 at group 1 .....	107
Table 4-2 Observed versus predicted items truth table.....	110
Table 4-3 Total validation criteria in the case of six previous hours for binary network model.....	115
Table 4-4 Validation by separation criteria for binary network model .....	116
Table 4-5 Rate of the extreme ice accretion rate values for the positive and negative predictions.....	120
Table 4-6 Total validation criteria when using six previous hours and duration variable as network input .....	131

Table 4-7 Validation by separation criteria of binary network model with supplementary variables .....	131
Table 5-1 Average of prediction errors for the extreme observations and the least ones for all stations .....	162
Table 5-2 R-square values to different hours ahead and previous hours for group 1 .....	164
Table 5-3 R-square values to different hours ahead and previous hours for group 2 .....	164
Table 5-4 R-square values to different hours ahead and previous hours for group 3 .....	165
Table 5-5 R-square values to different hours ahead and previous hours for group 4 .....	165
Table 5-6 R-square values in the case of 2 hours ahead for group 1, trial 1 .....	167
Table 5-7 Average R-square values for different hours ahead and previous hours in group 1 .....	168
Table 5-8 R -square values for different hours ahead and previous hours at station 16 in group 2 .....	168
Table 5-9 Average R-square values for different hours ahead and previous hours in group 2 .....	169
Table 5-10 R -square values for different hours ahead and previous hours at station 17 in group 3 .....	169
Table 5-11 Average R-square values to different hours ahead and previous hours for group 3 .....	170
Table 5-12 Average R-square values for different hours ahead and previous hours in group 4 .....	170

# **CHAPTER 1**

## **INTRODUCTION**

## **1.1 Background**

### **1.1.1 Problematic**

Icing precipitations are meteorological phenomena that occur frequently in cold regions, such as Canada and several other cold-climate countries. These atmospheric phenomena affect the overhead electrical power networks, and cause mechanical damages to equipments. From time to time, electricity failures are observed, a problem that causes considerable economic losses. For instance, the damage cost of a single ice storm, like the one of Quebec, in 1998, can be estimated to as high as five hundred million dollars.

Ice accretion on transmission lines causes physical damages to the line structure. In spite of the strong structure of the transmission lines, a risk always exists due to the damage provoked by the charge of atmospheric icing, and further by the existence of the wind. In order to reduce the risk, practical ways were proposed, such as monitoring the charges on the transport lines in real-time, in order to prevent ice adhesion on the conductors. This is done by controlling the power loss in the lines, which depends directly on the change of electrical intensity of current in the conductors.

Therefore, it is very important that a system be devised that can predict the disasters resulting from ice and snow accretions, thus preventing much damage. In the event of an imminent disaster, such a system could enable the preparation of the necessary equipments

and personnel beforehand according to advance estimations, and also help restore any damage promptly after the disaster.

The following study is based on the historical data brought from the SYGIVRE network of Hydro-Quebec. This network has been installed in 1992 in order to monitor specifically icing events in Quebec. It is based on an improved Icing Rate Meter (IRM), which is a measuring instrument used to obtain an estimate of the ice accumulation weight. For every hour, the instrument gives at all times updated meteorological conditions, and, during icing events, precise measurements of ice accumulation as well [5]. The network is composed of 28 stations connected together, and distributed throughout the Province of Quebec. Collecting icing data makes possible the understanding of the climatology of icing events happening in Quebec.

### **1.1.2 Research Overview**

The present work is part of the research done at the NSERC/HYDRO-QUÉBEC/UQAC Industrial Chair on Atmospheric Icing of Power Network Equipment (CIGELE), which is devoted to studying several phenomena associated with atmospheric icing, in particular its effects on power network equipment. The CIGELE research program includes mathematical modeling studies as well as physical and numerical simulation of a number of fundamental phenomena related to the process of icing and ice shedding of cables and conductors. In addition, it includes the development of electrical discharge in air

gaps and on the ice surface. The program also includes probabilistic studies of icing, and aims at improving the models, allowing better equipment design under icing conditions.

This work concerns the study of atmospheric events that happened in the Province of Quebec in order to create a system that can follow up icing storms, and is adapted to the SYGIVRE network of Hydro-Quebec. The purpose of such a system is to predict icing events for different times ahead by adjusting the predictions in real-time.

The procedure that has been followed for carrying out this research consisted in the modelling of icing events using the historical data of the SYGIVRE network, in order to devise a prediction tool.

The present analysis is based on neural network techniques, and its purpose is to find a better way to follow up the events, and to develop a basic knowledge that could help making predictions. Two models are thus proposed for predicting the occurrence of icing events and the ice accumulation weight in real-time.

In previous works, two major studies have aimed at understanding and modeling the icing events related to the SYGIVRE database. In the first study, several kinds of networks have been used to predict the ice load on transmission lines using the data of one particular station [1]. The other one has been conducted in order to study the spatial relationship

between stations throughout the Province of Quebec [2]. These studies will be explained later in some detail.

### **1.1.3 Survey of the Literature**

Understanding the behaviour of icing atmospheric phenomena is necessary in order to be able to avoid disasters. The evolution of such phenomena has not yet been represented by theoretical models. It is thus required to conduct an analysis of the past atmospheric icing events in order to obtain the maximum amount of information possible, which will be used in the design and verification of such models. For this reason, Hydro-Québec has developed, over the past three decades, two icing measurement networks (PIM and SYGIVRE) for collecting measures of events and storing them into databases.

The passive ice meters network (PIM), which comprises approximately 180 stations, is deployed throughout the Province of Quebec in the vicinity of electric transmission networks, and has been collecting data since 1974. Another automatic monitoring system, SYGIVRE, has been developed recently. It includes 28 stations distributed in the Province of Quebec that record real-time meteorological data, and is available to users via satellite. The data available for the present study has been collected since 1992, during 6 seasons (each of which lasted from September to June during consecutive years). At present, more stations have been added, and the data for more than 10 measurement years could be available for future studies. The two databases, PIM and

SYGIVRE, are useful for the study of icing events because their measurements are at large scale, and their data is of good precision, and have been recorded over a long period of time.

An icing database used by the Iceland State Electricity (RARIK) power company to store and manipulate information on icing of overhead lines has been prepared and constructed [3]. The database comprise both icing and damage data, where 3 100 records were included.

In 1986, a preliminary study outlined the potential of the PIM icing database (10 years of recorded data were included), and showed a possibility for conducting a statistical analysis [4].

Some other studies have been carried out, and some are still in progress, aiming to validate the measuring instruments used to evaluate the ice accretion on structures [5]. A comparative study of icing rate meter and load cell measurements based on the analysis of data recorded by SYGIVRE is used for estimating the ice accretion loads from the ice meter, using the sites' load cell readings [6]. Savadjiev et al. [7] estimated the ice accretion weight by converting the measuring tension force of transmission cables into linear ice mass for two of the ice stations studied (Mt. Belair). Other studies have been carried out in order to investigate the correlation between the number and the duration of signals emitted

per hour by the ice rate meter (IRM), and the actual ice accretion. Two different techniques were used: laboratory testing [8], and mathematical modelling [9].

By using the historical meteorological data of the available databases, it is possible to conduct studies in several directions. One of the major problems that can be addressed is the estimation of the return period for catastrophic icing events, such as the return period of the annual and extreme values of ice accretion, and the return period for wind speeds associated with icing events. It is also possible of creating statistical models capable of estimating the ice accretion, and the associated wind speed, and others that can estimate the combined wind-on-ice loads and the creation of distributions of glaze. Another work that can be achieved is the exploration of the databases, the grouping of stations by determining the relations between each other according to the historical events, and the creation of a risk region map according to the intensity of icing events [14]. In the case of prediction of icing events at future times, some topics can be studied: the prediction of annual and extreme values of ice accretion, the creation of a system that can prevent catastrophic icing events by predicting and detecting their existence in real time, and the prediction the ice accretion, also in real time.

The analysis of ice observations of historical data was used to derive estimations on the return period of extreme values. For example, Laflamme [10] used the pooling of data between three neighbouring stations to define the sample of annual maximum glaze ice accumulation, and reported a significant improvement in the goodness-of-fit for the

probability distribution function of ice accumulation. Savadjiev et al. [11] showed that the return period for wind speed associated with icing events can only be specified correctly on the basis of statistical analysis of field data for the annual recurrence and persistence of icing events. Farzaneh et al. [12] determined the distribution of *annual number of icing event recurrence*, *ice residency period*, and *total annual ice residency period* using the data of the PIM network.

Many statistical models were set to estimate icing events based on historical data. Druez et al. [13] calculated the values of the main icing parameters of the two conductors of the Mt. Valin test site, and they compared these values for freezing rain and in-cloud icing. Savadjiev et al [11] created two probabilistic models concerning the combined wind-on-ice loads, and the specification of radial ice thickness versus linear ice weight for 150 icing measurement sites in Quebec. Chouinard et al. [14] estimated an event-based model for the distribution of glaze ice, and of the associated wind speed for the data of the PIM network. Laflamme et al. [15] performed the preliminary analysis of 20 years of ice accretion measurements, and improved the study of icing storms trajectories.

Many models were set to predict ice accretion from historical records. Lu et al [16] created a model to estimate ice loads occurring in Winnipeg, Canada, from the available 126 years of weather data, and to predict its future extreme values. Felin [17] examined the correlation existing between observed glaze ice accretions with estimated values obtained from a proposed model for 150 observations.

The icing events distribution has been the subject of previous study at CIGELE [18, 2]. One part of the study consisted in exploring the two databases of Hydro-Quebec by characterizing the temporal distribution of events, and by studying the spatial and temporal patterns between stations. The other part consisted in determining the geographic grouping of the SYGIVRE network stations according to the events that have occurred at these stations, and in describing the relations between them, and, further, to explicate the grouping with supplementary variables by using the logistic regression method.

The recent advances in neural network methodology for modeling nonlinear, dynamical phenomena [19, 20, 21] along with the impressive successes in a wide range of applications, encouraged people to investigate the applications of neural networks for the prediction of atmospheric phenomena. Ohta, et al. [22] developed an on-line disaster warning system for detecting snow accretion on power lines by using neural networks, and his system was confirmed being highly practicable. It has been demonstrated also that neural networks could be used for predicting ice accretion loads on transmission line structures [1]. In an other study, a model was created for estimating the ice accretion loads of one site, and it was verified on data collected at other icing sites [23]. A neural network for the prediction of tornados was described [24], and this network was implemented, in order to detect the existence/non-existence of tornados based on ground observations. The network is also employed elsewhere to predict damaging wind [25].

Finally, the study of historical meteorological data, especially for the SYGIVRE database has not yet covered the aspect concerning the prediction of icing events in real time based on the spatial relationship between multiple stations. Most real-time predictions for icing events using neural network techniques were carried out only for one station. It is thus important to monitor the behaviour of icing events by considering multiple stations, and to exploit the spatial relationship between them, in order to obtain a robust and stable system with a satisfactory degree of precision.

## **1.2 Objectives**

The main goal of this study consists in formulating spatial models that could serve to develop a better understanding of the recorded icing events of the SYGIVRE network. These models should be adapted for analyzing, and tracking the evolution of the available icing events. Such models have therefore been created, that have the capability of predicting the icing storm events in real-time before it begins, based on the spatial propagation of these events.

The purpose of this study, was to create two models for real-time prediction. Both these models have the function of detecting the occurrence of icing events as well as the ice accumulation weight. Thus, these models could behave as a warning system capable of predicting beforehand the eventuality of a disaster according to the occurrence of icing

events or to the snow ice accretion, by emitting a signal pointing to the threatened stations. The design of these models is based on neural network techniques.

As part of CIGELE's research activities, a first approach using logistic regression modelling has already been tested [2]. The approach proposed by the present study for solving the problem is based on the artificial neural network method, which is a powerful tool for performing predictions. The resulting neural network has been tuned in order to achieve optimal performance.

### **1.3 Methodology**

The methodology upon which the present work is based consists of four steps.

The first step concerns the analysis and the description of the information supplied by Hydro-Quebec. This is the data set provided by the SYGIVRE database, which was the most appropriate database available.

The second step, required before the actual modeling work could begin, was to gain a good understanding of this data set, in order to explicate the actual icing storms in terms of the icing events of the database, for modeling purposes.

Third, came the modeling work itself, which led to the creation of a spatial model, for dealing with the prediction of icing storms evolution.

In the fourth step, a second model has been created by improving the first one, in order as to be able to predict, in real time as well, the ice accumulation weight on the power lines.

For the purpose of information processing, a neural network approach has been chosen. Neural networks are one of the techniques of artificial intelligence used to develop expert systems. Biologically inspired neural networks are finding an increased number of applications in the modelling of complex dynamical systems. They have the capability of learning given samples and abstract essential characteristics from inputs containing irrelevant data, and have been applied to several fields successfully. Neural networks also caused, in recent research, much interest for the prediction of spatial-characteristic phenomena. They thus appear suitable for the analysis of atmospheric icing events. Many computer codes are presently available for neural network design. MATLAB is general-purpose mathematical software supporting the programming of neural network using the necessary code supplied in one of its toolboxes. Due to its powerful programming language, it makes it possible to quickly generate neural networks, and manipulate them easily.

For validating the two models created, several tests have been performed, aiming to measure the degree of success in predicting icing events, and to test the capability of the model at detecting the most serious ones.

## **1.4 Pertinent literature**

Based on the information coming from the SYGIVRE database, several works have been carried out aiming to explore the database, and create models serving to help understanding some aspects from the huge amount of implicit information found in the database. The present work intends to be an extension based on two previous relevant studies: *Exploration of static and time dependent neural network technique for the prediction of ice accretion on overhead line conductors* [1], and *Analysis of Spatial Patterns for Icing Events in Quebec* [2]. Because these studies have several subjects in common, they will be examined below in some details, in order to emphasize the degree of the current work's contribution.

Several basic meteorological variables from the SYGIVRE database are included in the analysis. These are: *total ice accumulation weight at a specific time*, *air temperature*, *ice accretion rate*, *wind velocity*, and *icing event duration*, in addition to the signals of the icing rate meter IRM [5].

### 1.4.1 Spatial analysis

The work of Guesdon, et al. [2] was principally based on the SYGIVRE database, where the data of 28 stations spanning 6 seasons, from 1992 to 1998, was included, using only the following variables: *air temperature*, *ice accumulation weight*, *accretion rate* and *duration*.

Three spatially related groupings of the database SYGIVRE have been concluded. The first is located on the area of Montreal to Quebec, and comprises 8 stations: 4, 8, 9, 11, 14, 16, 22 and 27. The second is located on the area of Quebec and of Saguenay / Lac St. Jean, and comprises 3 stations: 5, 10 and 17. The last group is located on the Côte-Nord, and comprises 8 stations 1, 2, 3, 7, 13, 21, 24 and 28.

A logistic regression model has been constructed, which served to determine the relative importance of meteorological variables, and to predict the relationships between stations. The analysis showed that the *duration* variable is the most important one, because it gives the better correlation coefficient among the different variables.

The prediction rate of the model from the whole data is about 80 %. It should be noted that the model gives a high rate of error of false negatives (or *sensitivity*) of about 19 %. This means that this model too often predicts an absence of bond between two stations when there are indeed common icing events.

### 1.4.2 Prediction of ice accretion

The study of Larouche, et al. [1] sought to predict the ice accretion on the transmission line structure based on data coming from the SYGIVRE database. Only the data for the period of February, 1998, to April, 1999, of the Mont Belair station, located near Quebec City, has been used. Five artificial neural networks (ANN) architectures were explored and compared. Two static networks: *multilayer perceptron* and *radial-basis function* were used. In addition, two time-dependent networks have been employed: *finite impulse response* and *Elman*, which were compared with a multiple linear regression-type network (ADALINE).

The neural networks in this study make use of the following input variables: *temperature*, *normal wind*, *speed* and *IRM signals*. It has been observed that the FIR architecture is the most promising for making predictions, since time play an important role in the processing.

### 1.4.3 Insertion of the present work

There is, in fact, a direct relationship between the cited studies and the current study, because of some sharing of topics. At the root there is the fact that all these studies

used the SYGIVRE database, and they have all sought to explore it in order to extract information about atmospheric icing.

The work done by Guesdon, et al. [2] used exactly the same data set as the present work. Two related concerns are useful in our study. The first consists in retaining the stations groupings already defined, which represents regions groupings having different spatial and temporal characteristics. This will help explicate each region's situation separately. The second concern consists in performing comparisons between the prediction results obtained in the regression logistic model and our proposed model.

The study of Larouche, et al. [1] is similar to the present one, with the difference that its focus was on one particular station, and that it used the data of only 1 year and 3 months for the same station. Therefore, this work can be considered as a preliminary work regarding the prediction of atmospheric icing events, and it can be generalized in order to cover all stations by using the overall existing data.

Exploring the neural network methods used in Larouche et al. has seemed interesting enough as to use them in the present work. Of those methods, only one is chosen for addressing our problem: the *Elman* method. The predictions made in the current study are mainly based on the oldest observations in time. This characteristic is not supported implicitly in the static-architecture method of 'backpropagation' like it is the case with FIR.

However, in the case of the time-dependent architecture method of Elman, this characteristic is supported.

At this point, it seems appropriate to mention that the involvement of multiple stations in the prediction somewhat improves the model, since the prediction results for a station then depends on the data coming from the neighbouring stations rather than on the notion of implicit time dependent of the network architecture.

The work of Guesdon et al. [2] showed that a great precision is a required condition in the modelling of icing events. Guesdon already worked on the modelling of icing events with two databases, namely PIM and SYGIVRE, and he found that the latter gave better results, due to its great precision. The measures provided by the SYGIVRE database are very precise indeed, and it is really the sole database that proved usable in a study of the present type.

The work of Savadjiev et al. [6, 7] is about the estimation of ice accretion weight on power transportation lines, of which the present work intends to be a complement. Considering that both works share complementary aspects, some results could contribute to the improvement of the degree of precision achieved by one or another. For example, the degree of precision achieved in the cited work has a direct impact on the precision of the prediction model realized in the present work.

## **CHAPITRE 2**

# **ARTIFICIAL NEURAL NETWORK**

## **2.1 Artificial Neural network definition (NN)**

### **2.1.1 History of Neural Network technology**

Interest in artificial neural networks has grown rapidly over the past few years. Professionals from such diverse fields as engineering, philosophy, physiology, and psychology are intrigued by the potential offered by this technology, and are seeking applications within their disciplines.

This resurgence of interest has been fired by both theoretical and application success. Suddenly, it appears possible to apply computation to realms previously restricted to human intelligence; to make machines that learn and remember in ways that bear a striking resemblance to human mental processes; and to give a new and significant meaning to the much-abused term *artificial intelligence*.

Artificial neural networks are biologically inspired; that is, they are composed of elements that perform in a manner that is analogous to the most elementary functions of the biological neuron. These elements are then organized in a way that may (or may not) be related to the anatomy of the brain. Despite this superficial resemblance, artificial neural networks exhibit a surprising number of the brain's characteristics. For example, they learn from examples the new ones, and experience, generalize from previous examples to new ones, and abstract essential characteristics from inputs containing irrelevant data.

The improved understanding of the functioning of the neuron and the pattern of its interconnections has allowed researchers to produce mathematical models to test their theories. Experiments can now be conducted on digital computers without involving human or animal subjects, thereby solving many practical and ethical problems. From early work it became apparent that these models not only mimicked functions of the brain, but that they were capable of performing useful functions in their own right.

Along with the progress in neuroanatomy and neurophysiology, psychologists were developing models of human learning. One such model, which has proved most fruitful, was that of D. O. Hebb, who in 1949 proposed a learning law that became the starting point for artificial neural network training algorithms. Augmented today by many other methods, it showed scientists of that area how a network of neurons could exhibit learning behaviour.

In the 1950s and 1960s, a group of researchers combined these biological and psychological insights to produce the first artificial neural networks. Initially implemented as electronic circuits, they were later converted to the more flexible medium of computer simulation, the most common realization today. Early successes produced a burst of activity and optimism. Marvin Minsky, Frank Rosenblatt, Bernard Widrow, and others developed networks consisting of a single layer of artificial neurons. Often called perceptrons, they were applied to such diverse problems as weather prediction, electrocardiogram analysis, and artificial vision.

Often underfunded and unappreciated, some researchers had difficulty finding publishers and continuing the efforts; hence, research published during the 1970s and early 1980s is found scattered among a wide variety of journals, some of which are rather obscure. Gradually, a theoretical foundation emerged, upon which the more powerful multilayer networks of today are being constructed

In the past few years, theory has been translated into application, and new corporations dedicated to the commercialization of the technology have appeared. There has been an explosive increase in the amount of research activity, especially in 1987 in the field of artificial neural networks.

### **2.1.2 Neural Networks capability**

Imaginative researchers are devising new uses for artificial neural networks daily. Some of the more traditional applications include

#### **Classification**

In recent works [26], an electrical circuit has been presented that uses neural network to produce n-bit analog-to-digital converter. Neural network used also to identify diseases of the heart from electrocardiograms [27]. Any task that can be done by traditional

discriminant analysis can be done at least as well (and almost always much better) by a neural network.

### **Noise reduction**

An artificial neural network can be trained to recognize a number of patterns [28]. These patterns may be parts of times-series, image, etc. If a version of one of these patterns, corrupted by noise, is presented to a properly trained network, the network can provide the original pattern on which it was trained. This technique has been used with great success in some image restoration problems.

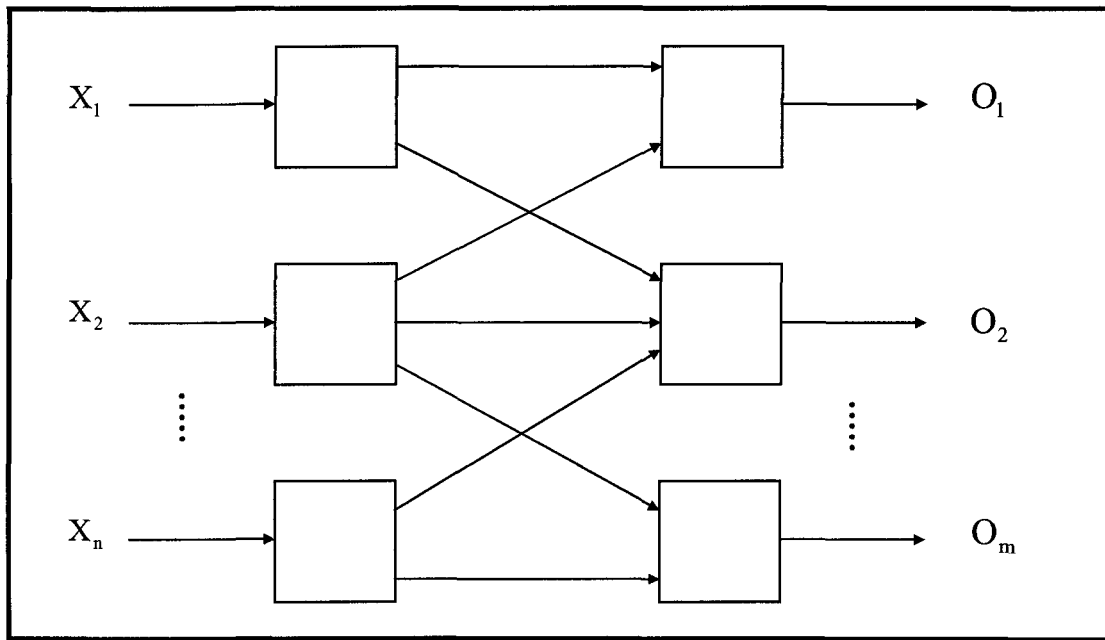
### **Prediction**

A very common problem is that of prediction the value of a variable given historic values of itself (and perhaps other variables). Economic [29] and meteorological [30] models spring to mind. Neural networks have frequently been shown to outperform traditional technique like ARIMA and frequency domain analysis [31].

## **2.1.3 Basic principles of a Neural Networks**

Neural network is a network of many simple processors "units", each possibly having a small amount of local memory. The units are connected by communication channels "connections" which usually carry numeric data, encoded by any of various

means. The units operate only on their local data and on the inputs they receive via the connections. It accepts a group of inputs data as Vector  $X$  ( $1 \dots n$ ) and has outputs Vector  $O$  ( $1 \dots m$ ) depending on the network function as illustrated in the Figure 2-1.



**Figure 2-1 Neural Network**

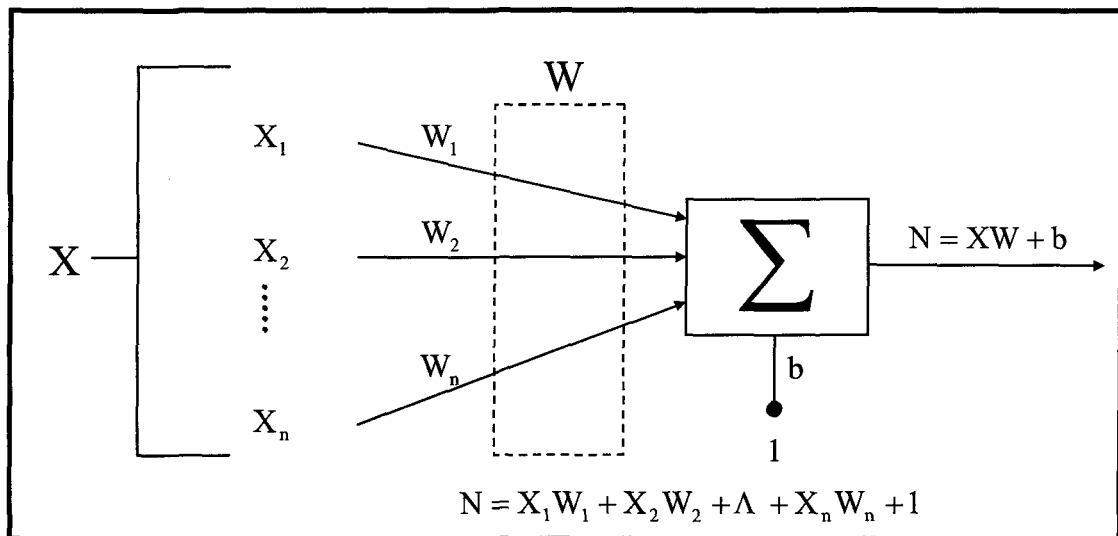
The target vector  $T$  ( $1 \dots m$ ) is defined as, the desired output vector for a given input vector. Ordinarily, this vector is given in many neural network problems and the output has to converge to achieve it. Note that, there are several neural network that do not need targets like Unsupervised networks.

Generally, neural network accept and produce numerical data. This data consists of a collection of many samples called training samples, which serves as exemplars to the

network. Normally, each sample consists of an input and a target vectors. Training samples will be applied to the network successively. Then the network will be trained in such way that, for each applied input, the network produce an output equivalent to the target suggested before the training. Thus the network works as an output predictor.

### 2.1.4 The Artificial Neuron

Each unit of neural network called neuron and each connection between two neurons has weights *connection strengths*, each neuron accepts multiple inputs and gives one output which can be an input to another neuron in the network. Each input is multiplied by a corresponding weight, and all of the weighted inputs are then summed to determine the activation level of the neuron. Figure 2-2 shows a model that implements this idea. A set of inputs labelled  $X_1, X_2, \dots, X_n$  is applied to the artificial neuron.



**Figure 2-2 Artificial Neuron**

These inputs collectively referred to as the vector  $X$ . Each signal is multiplied by an associated weight  $W_1, W_2, \dots, W_n$ , before it is applied to the summation block, labelled  $\sum$ . The set of weights is referred to collectively as the vector  $W$ . The summation block, adds all of the weights inputs algebraically, producing an output that we call  $N$ . This may be compactly stated in vector notation as follows:

$$N = XW + b \quad \text{Eq. 2-1}$$

The neuron has a scalar bias,  $b$ . The bias is simply added to the product  $XW$  as shown by the summing junction. The bias is much like a weight, except that it has a constant input of 1.

### 2.1.5 The Activation function (transfer function)

The  $N$  signal is usually further processed by an activation function  $F$  which is a transfer function to produce the neuron's output signal,  $O$ .

$$O = F(N) \quad \text{Eq. 2-2}$$

The most common used transfer functions in neural network are:

- Threshold function or hard-limit function;
- Squashing function or logistic function "sigmoid", and
- Linear transfer function.

The threshold function has the constant threshold TH value and it works as follow

$$\begin{aligned} O &= 1 \text{ if } N > TH \\ O &= 0 \text{ Otherwise} \end{aligned} \quad \text{Eq. 2-3}$$

The general form of the threshold function is like that



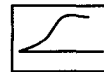
The squashing function (logistic function); it is expressed mathematically as

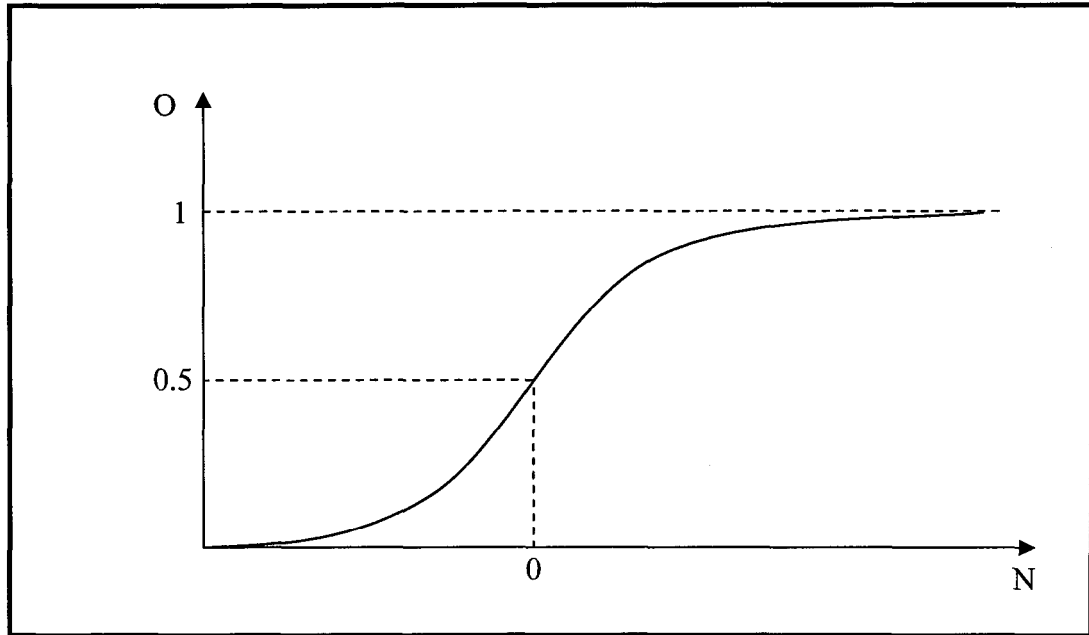
$$F(x) = 1/(1 + e^{-x}) \text{ thus,}$$

$$O = 1/(1 + e^{-N}) \quad \text{Eq. 2-4}$$

This logistic function generates outputs between 0 and 1 as the neuron's network input goes from negative to positive infinity as shown in Figure 2-3, thus it compress the range of its input, so that its output never exceeds the limits [0, 1] regardless of the value of input.

The general form of the logistic function is like that





**Figure 2-3 Logistic Transfer Function**

The linear transfer function produces its input as its output ( $O = N$ ) and its general form

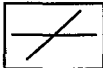
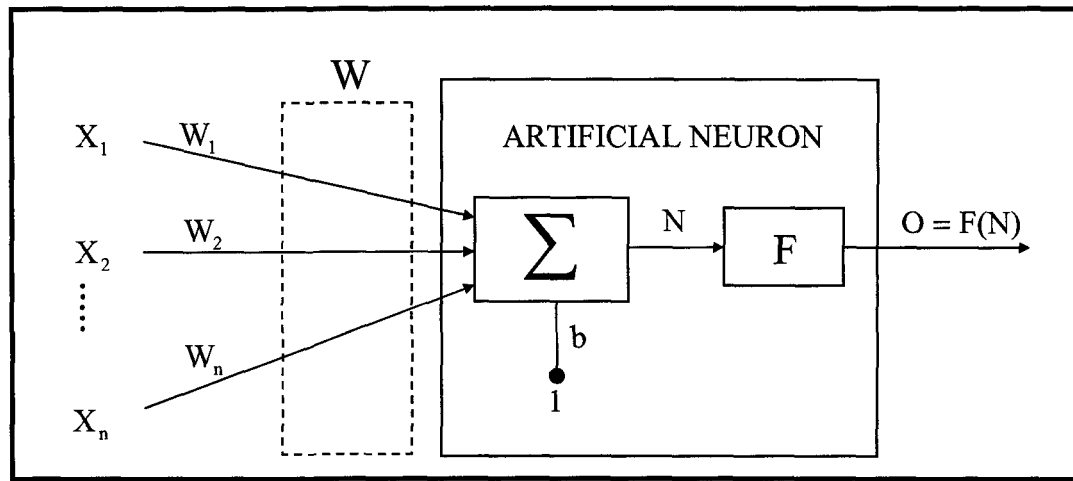
is like that 

Figure 2-4 shows the artificial neuron with activation function; the block labeled F accepts the N output and produces the signal labeled O.



**Figure 2-4 Artificial Neuron with Activation Function**

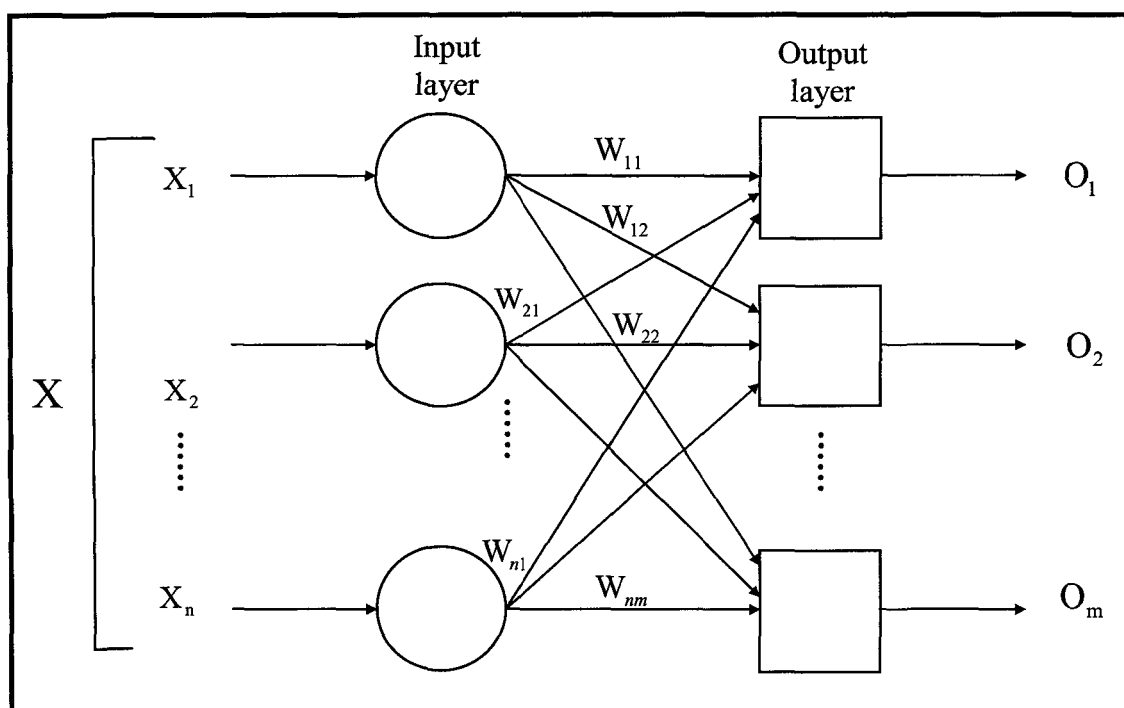
### 2.1.6 Single layer and Multilayer Neural Networks

The single layer network is the simplest network, which is a group of neurons arranged in a layer as shown on the right side of Figure 2-5. Note that, the circular node on the left serve only to distribute the inputs and they are shown as circles to distinguish them from the computing neurons, which are shown as squares. The circular nodes on the left constitute a layer called the input layer and the squares on the right constitute a layer called the output layer which is the active layer. The set of inputs  $X$  has each of its elements connected to each artificial neuron through a separate weight. Each neuron simply output a weighted sum of the inputs to the network.

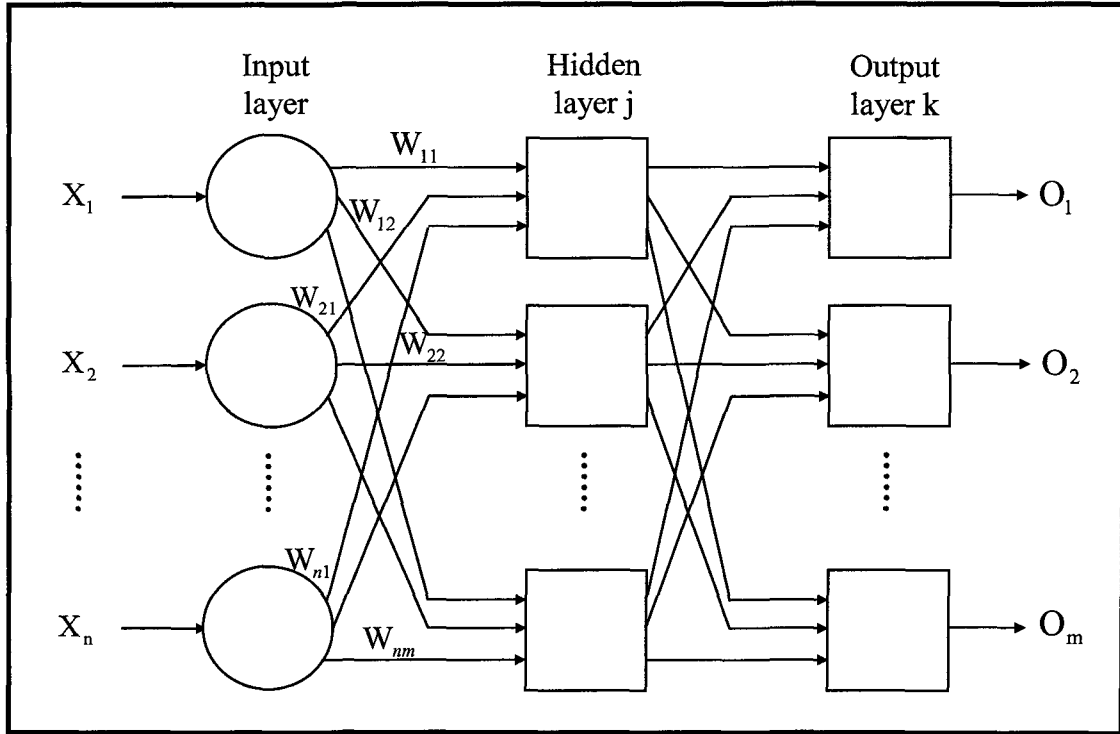
The weight elements are considered as a matrix  $W$ . The dimension of the matrix are  $n$  rows by  $m$  columns, where  $n$  is the number of inputs and  $m$  the number of neurons. For

example, the weight connecting the third input to the second neuron would be  $W_{32}$ . In this way it may be seen that, calculating the set of neuron N outputs, NET for a layer is a simple matrix multiplication. Thus  $NET = XW$ , where NET and X are row vectors.

Multilayer networks are more complex than single layer network but generally offer greater computational capabilities. It may be formed by simply cascading a group of single layers; the output of one layer provides the input to the subsequent layer. The layers located between the input layer and output layer is called hidden layers. Figure 2-6 show a two Layer network with one hidden layer. There are two weighted matrixes  $W^1$  and  $W^2$ , where  $W^1$  represent the weights between the input layer and the hidden layer and  $W^2$  represent the weights between the hidden layer and the output layer.



**Figure 2-5 Single-Layer Neural Network**



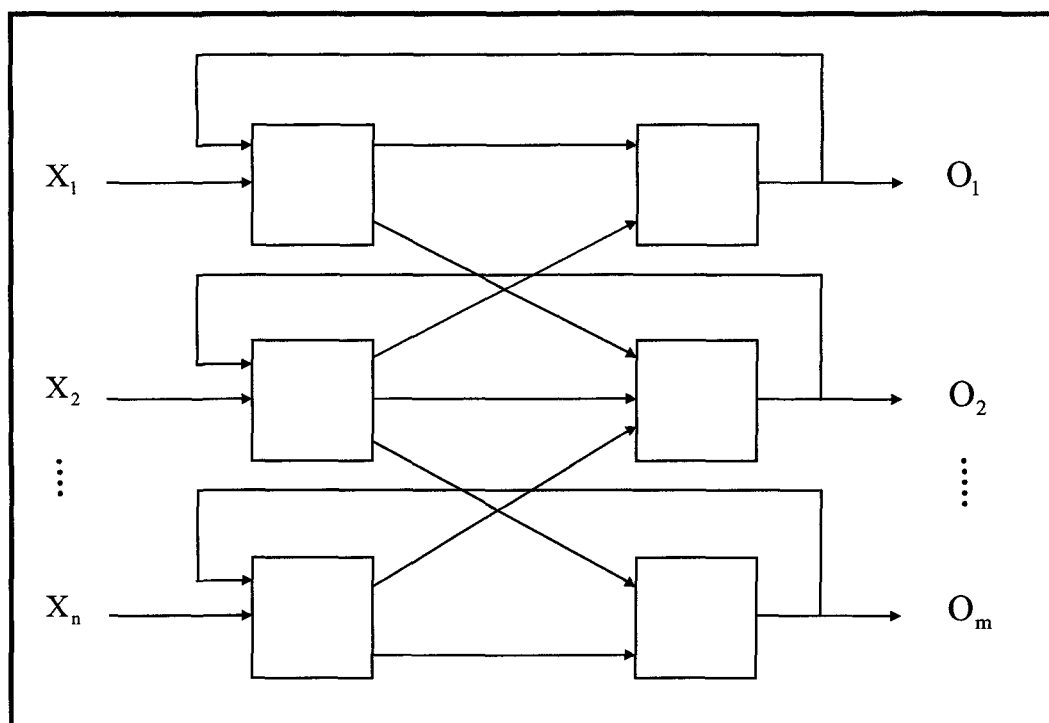
**Figure 2-6 Two-Layer Neural Network**

### 2.1.7 Neural Networks Architectures

Various neural networks architectures are available; considering the following two architectures:

The first one is the standard feedforward architecture which is like the network shown in Figure 2-6 and it is applicable for the methods Backpropagation, Perceptron and linear networks.

The second architecture called recurrent architecture, it differ from the feedforward architecture that it returns previous outputs back to inputs as shown in Figure 2-7; hence, their output is determined both by their current input and their previous outputs. For this reason recurrent networks can exhibit properties very similar to short-term memory in humans in that the state of the network outputs depends in part upon their previous inputs and some applicable methods to this architecture is the Elman network.



**Figure 2-7 Recurrent Architecture**

### 2.1.8 Kinds of Neural Networks

Two major kinds of network topology are feedforward and feedback.

- In a feedforward neural network, the connections between units do not form cycles. Feedforward neural networks usually produce a response to an input quickly.
- In a feedback or recurrent neural network, there are cycles in the connections. In some feedback neural networks, each time an input is presented, the neural network must iterate for a potentially long time before it produces a response. Feedback neural networks are usually more difficult to train than feedforward neural networks.

### **2.1.9 Training of Artificial Neural Networks (learning)**

A network is trained so that application of a set of inputs produces the desired (or at least consistent) set of outputs. Each such input (or output) set is referred to as vector. Training is accomplished by sequentially applying input vectors, while adjusting network weights according to a predetermined procedure. During training, the network weights gradually converge to values such that each input vector produces the desired output vector.

We can categorize the training situations in two distinct sorts supervised and unsupervised:

In supervised training, the correct results (target values or desired outputs) are known and are given to the neural network during training so that the neural network can adjust its weights to try matching its outputs to the target values. After training, the neural

network is tested by giving it only input values, not target values, and seeing how close it comes to outputting the correct target values.

Two kinds of supervised training are found Batch and Incremental. In Batch training, the weights and biases are only updated after all of the inputs and targets are presented, learning proceeds as follows:

```
Initialize the weights.
Repeat the following steps:
    Process all the training data.
    Update the weights.
```

It is used if the network will stay fix, no more inputs and target to be applied to the network in order to change the weights. The training process starts by initializing all weights to small non-zero values. Often these are generated randomly. Then, a subset of the collection of training samples is presented to the network, one at time. A measure of error (like Mean Square Error) incurred by the network is made, and the weights are updated in such a way that the error is reduced. This process is repeated as necessary. One pass through this cycle is called an epoch. The size of the subset (number of training samples used per weight update) is called the epoch size.

In Incremental training, the weights and biases will be updated after each input is presented, inputs and targets will be presented as sequences.

```
Initialize the weights.
Repeat the following steps:
    Process one training case.
    Update the weights.
```

It's used if the network need to be changed, whenever there is new inputs and targets to be added to the network.

Unsupervised learning or Self-organisation in which an (output) unit is trained to respond to clusters of pattern within the input. In this paradigm the system is supposed to discover statistically salient features of the input population. Unlike the supervised learning paradigm, there is no a priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli.

### 2.1.10 Error function

When a neural network is trained, the measure of performance that is optimized is usually the mean square error of the outputs. It is easily computed by summing the squared differences between what a predicted variable should be versus what it actually is, then dividing by the number of components that went into that sum. For any input, the output neurons take on an activation level determined by the input and by the network. For that input, there is a desired set of output activations. Suppose that we are processing case “i” in the training samples. Let the correct (target) activation of output neuron “q” be designated as “ $T_{iq}$ ”. If there are m output neurons, the error for that single presentation is

$$E_i = \frac{1}{m} \sum_{q=0}^{m-1} (T_{iq} - O_{iq})^2 \quad \text{Eq. 2-5}$$

If there are “e” presentations in the epoch, the error for that epoch is

$$E = \frac{1}{e} \sum_{i=0}^{e-1} E_i \quad \text{Eq. 2-6}$$

### 2.1.11 Validation of Artificial Neural Networks

It would be foolhardy to train a network then immediately place it into service. Its competence must be evaluated first. This process called validation. For this reason, three methods for validation and testing neural networks are offered.

The first methods concern the study of the resultant errors of the network which is the difference between the output  $O$  and the target  $T$  of the network. For each specific neuron in the output there is an error  $e_q$ . Thus we obtain a vector “e” called error vector, it has the same size as the output vector. In this vector we might do statistical study in order to determine the efficiency of our network.

Another method can be used to validate the neural networks is to separate the known cases into two disjoint sets. One is the training set, which is used to train the network. The other is the validation set, which is used to test the trained network.

The third method for validation is called cross-validation; we divide the data into  $s$  subsets of (approximately) equal size. We train the network  $s$  times, each time leaving out one of the subsets from training, but using only the omitted subset to compute whatever error criterion. If  $s$  equals the sample size, this is called leave-one-out cross-validation.

## 2.2 Backpropagation

Backpropagation refers to the method for computing the gradient of the case-wise error function with respect to the weights for a feedforward network. By extension, backpropagation is a gradient descent algorithm, in which the network weights are moved along the negative of the gradient of the performance function. The term backpropagation refers to the manner in which the gradient is computed for nonlinear multilayer networks.

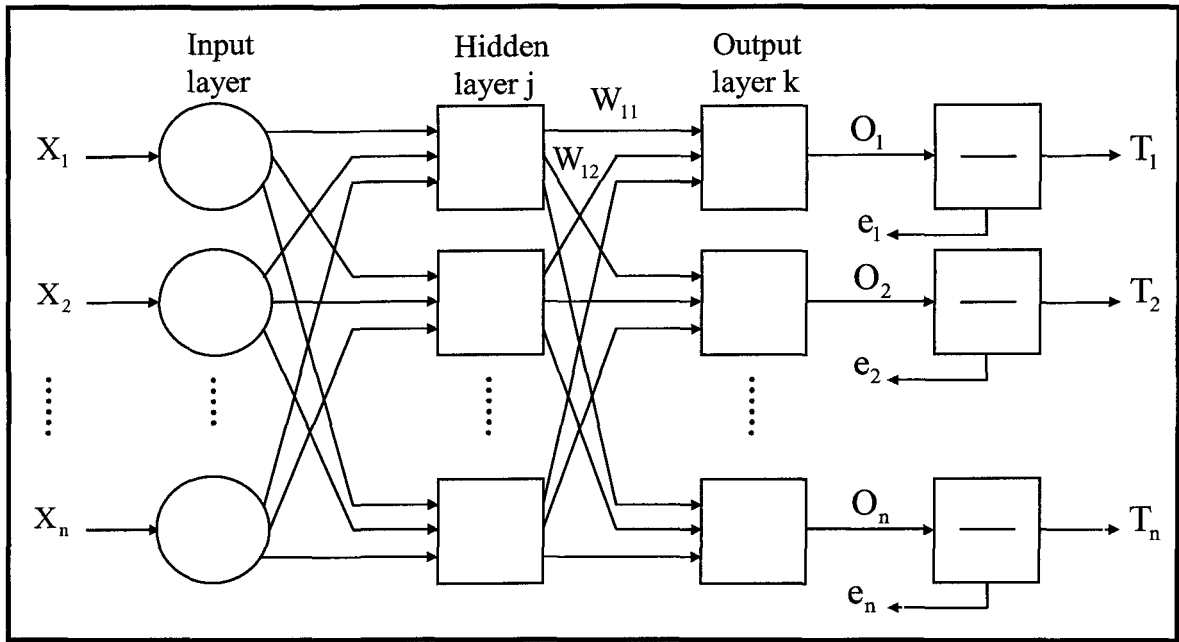
The activation function for the Backpropagation network is the squashing function *sigmoid*,

$$O = 1/(1 + e^{-N}) \quad \text{Eq. 2-7}$$

A fact we use in implementing the Backpropagation algorithm,

$$\frac{\partial O}{\partial N} = O(1 - O) \quad \text{Eq. 2-8}$$

The *sigmoid* compress the range of N so that O lies between zero and one. Figure 2-8 show a two-layer network suitable for training with Backpropagation.



**Figure 2-8 Two-layer Backpropagation Network**

The objective of training the network is to adjust the weights so that application of a set of inputs produces the desired set of outputs. For reasons of brevity, these input-output sets can be referred to as vectors. Training assumes that each input vector is paired with a target vector representing the desired output; together these are called training pairs.

A group of training pairs is called a training set. Training the backpropagation network requires the steps that follow:

1. Select the next training pair from the training set; apply the input vector to the network input
2. Calculate the output of the network

3. Calculate the error between the network output and the desired output (the target vector from the training pair). Automatically calculated from the network.
4. Adjust the weights of the network in a way that minimizes the error
5. Repeat steps 1 through 4 for each vector in the training set until the error for the entire set is acceptably low

At this point, the network is used for recognition and weights are not changed.

Let the vector  $X$  and target vector  $T$  comes from the training set and the output vector  $Y$  and the weight matrix  $W$  between two layers. The calculation in multilayer networks is done layer by layer, starting at the layer nearest to the inputs and finish at the output layer, once the outputs for a layer is found, it serves as input to the next layer and calculated like that:

$$O = F(XW) \quad \text{Eq. 2-9}$$

In order to calculate the new weights for training process we have to define “ $\delta$ ”,

$$\delta = O(1 - O)(T - O) \quad \text{Eq. 2-10}$$

Which is the multiplication of the derivative of the squashing function  $[O(1 - O)]$  and the error “ $e$ ”, we have also training rate coefficient “ $\eta$ ” (serving to adjust the size of the

average weight change) typically 0.01 to 1.01. The following calculation should be done to train the network:

$$\begin{aligned}\Delta w_{pq,k} &= \eta \delta_{q,k} O_{p,j} \\ w_{pq,k}(n+1) &= w_{pq,k}(n) + \Delta w_{pq,k}\end{aligned}\tag{Eq. 2-11}$$

where,

$w_{pq,k}(n)$  = The value of a weight from neuron p in the hidden layer to neuron q in the output layer at step n (before adjustment) note that the subscript k indicates that the weight is associated with its destination layer

$w_{pq,k}(n+1)$  = Value of the weight at step n+1 (after adjustment)

$\delta_{q,k}$  = The value of  $\delta$  for neuron q in the output layer k

$O_{p,j}$  = The value of O for neuron p in the hidden layer j

This calculation must be done to all weights combination between the layers J, k, the entire matrix must be changed each step.

In the case of multilayer we can train the hidden layers by propagating the output error back through the network layer by layer, adjusting weights at each layer.

In order to calculate “ $\delta$ ” for hidden-layer we are using the “ $\delta$ ” and weights for output layer

$$\delta_{p,j} = O_{p,j}(1 - O_{p,j})(\sum \delta_{q,k} w_{pq,k}) \quad \text{Eq. 2-12}$$

Then we can calculate the weights for hidden layer in the same way that output layer weights calculated

$$\begin{aligned} \Delta w_{pq,j} &= \eta \delta_{q,k} O_{p,i} \\ w_{pq,j}(n+1) &= w_{pq,j}(n) + \Delta w_{pq,j} \end{aligned} \quad \text{Eq. 2-13}$$

where, j represents the hidden layer and i represents the layer before the hidden layer.

### 2.2.1 Generalization

During learning, the outputs of a supervised neural network come to approximate the target values given the inputs in the training set. This ability may be useful in itself, but more often the purpose of using a neural network is to generalize--i.e., to have the outputs of the network approximate target values given inputs that are *not* in the training set. The conditions those are typically necessary for good Generalization are:

- The first necessary condition is that the inputs to the network contains sufficient information pertaining to the target, so that there exists a mathematical function

relating correct outputs to inputs with the desired degree of accuracy. It can not be expected a network to learn a nonexistent function

- The second necessary condition is that the function we are trying to learn (that relates inputs to correct outputs) be, in some sense, smooth. In other words, a small change in the inputs should, most of the time, produce a small change in the outputs
- The third necessary condition for good generalization is that the training cases be sufficiently large and representative subsets of the set of all cases that will be generalized

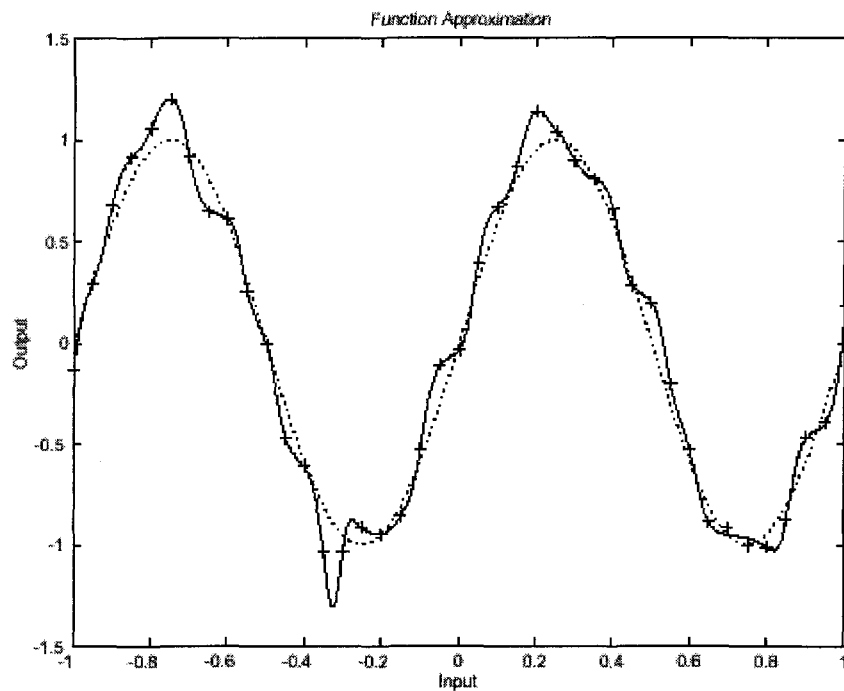
### **2.2.2 Underfitting, Overfitting and Noise**

A network that is not sufficiently complex can fail to detect fully the signal in a complicated data set, leading to underfitting. A network that is too complex may fit the noise, not just the signal, leading to overfitting. Overfitting is especially dangerous because it can easily lead to predictions that are far beyond the range of the training data.

Noise means variation in the target values that is unpredictable from the inputs of a specific network, regardless of the architecture or weights, Noise in the target values increases the danger of overfitting

Figure 2-9 shows the response of a neural network that has been trained to approximate a noisy sine function. The underlying sine function is shown by the dotted

line, the noisy measurements are given by the '+' symbols, and the neural network response is given by the solid line. Clearly this network has overfit the data and will not generalize well.



**Figure 2-9 Overfitting of noisy sine function**

### **2.2.3 Determination number of hidden units (neurons) and layers**

The best number of hidden units depends in a complex way on:

- the numbers of input and output units
- the number of training cases
- the amount of noise in the targets
- the complexity of the function to be learned

- the architecture
- the type of hidden unit activation function
- the training algorithm

In most situations, there is no way to determine the best number of hidden units without training several networks and estimating the generalization error of each. If we have too few hidden units, we will get high training error and high generalization error due to underfitting. If we have too many hidden units, we may get low training error but still have high generalization error due to overfitting.

Maybe we do not need any Hidden Layer in our network, it depends on the complexity of neural network and even if the function we want to learn is mildly nonlinear, we may get better generalization with a simple linear model than with a complicated nonlinear model if there is too little data or too much noise to estimate the nonlinearities accurately.

Multi-layered backpropagation networks are capable of performing just about any linear or nonlinear computation, and can approximate any reasonable function arbitrarily well. They are sensitive to the number of neurons in their hidden layers. Too few neurons can lead to underfitting. Too many neurons can contribute to overfitting, in which all training points are well fit, but the fitting curve takes wild oscillations between these points.

## 2.3 Recurrent Networks

This involves the use of recurrent links in order to provide networks with a dynamic memory. In this approach, hidden unit patterns are feedback to themselves; the internal representations which develop thus reflect task demands in the context of prior internal states. Merely the Elman recurrent will be covered in this part.

### 2.3.1 Elman Network

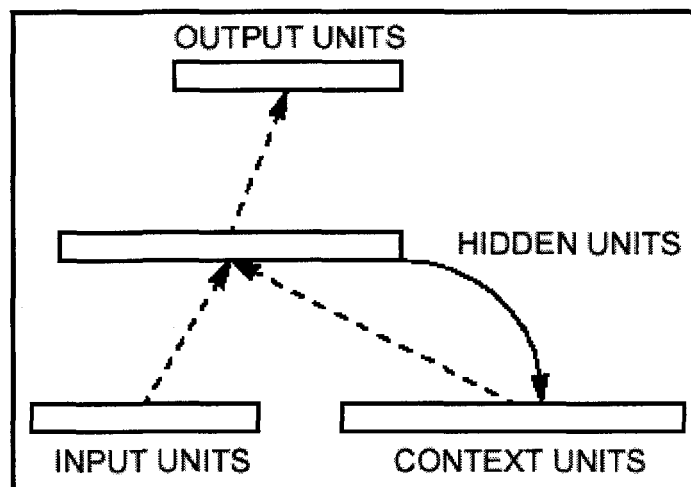
The recurrent connection of the Elman network allows both detecting and generating time-varying patterns. For example, consider the following two vectors.

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

These two vectors appear to be instances of the same basic pattern, but displaced in space (or time, if we give these a temporal interpretation). However, as the geometric interpretation of these vectors makes clear, the two patterns are in fact quite dissimilar and spatially distant.

The aim is to allow time to be represented by the effect it has on processing. This means giving the processing system dynamic properties which are responsive to temporal sequences. In short, the network must be given memory.

The network shown in Figure 2-10 is augmented at the input level by additional units; call these Context Units. These units are also “hidden” in the sense that they interact exclusively with other nodes internal to the network, and not the outside world.



**Figure 2-10 A simple recurrent network**

Consider that there is a sequential input to be processed, and some clock which regulates presentation of the input to the network. Processing would then consist of the following sequence of events. At time “t”, the input units receive the first input in the sequence. Each input might be a single scalar value or a vector, depending on the nature of the problem. The context units are initially set to 0.5 (The activation function used here bounds values between 0.0 and 1.0). Both the input units and context units activate the hidden units; and then the hidden units feed forward to activate the output units. The hidden units also feed back to activate the context units. This constitutes the forward activation.

Depending on the task, there may or may not be a learning phase in this time cycle. If so, the output is compared with a teacher input and backpropagation of error is used to incrementally adjust connection strengths. Recurrent connections are fixed at 1.0 and are not subject to adjustment. At the next time step " $t+1$ " the above sequence is repeated. This time the context units contain values which are exactly the hidden unit values at time " $t$ ". These context units thus provide the network with memory.

In feedforward networks employing hidden units and a learning algorithm, the hidden units develop internal representations for the input patterns which recode those patterns in a way which enables the network to produce the correct output for a given input. In the present architecture, the context units remember the previous internal state. Thus, the hidden units have the task of mapping both an external input and also the previous internal state to some desired output. Because the patterns on the hidden units are what are saved as context, the hidden units must accomplish this mapping and at the same time develop representations which are useful encodings of the temporal properties of the sequential input. Thus, the internal representations that develop are sensitive to temporal context; the effect of time is implicit in these internal states.

Consider now the results of applying this architecture to the Exclusive-OR function which involve processing of inputs which are naturally presented in sequence.

The Exclusive-OR (XOR) function has been of interest because it cannot be learned by a simple two-layer network. Instead, it requires at least three-layers. The XOR is usually presented as a problem involving 2-bit input vectors (00, 11, 01 and 10) yielding 1-bit output vectors (0, 0, 1, 1; respectively).

This problem can be translated into a temporal domain in several ways. One version involves constructing a sequence of 1-bit inputs by presenting the 2-bit inputs one bit at a time (i.e., in two time steps), followed by the 1-bit output; then continuing with another input/output pair chosen at random. A sample input might be:

1 0 1 0 0 0 0 1 1 1 1 0 1 0 1 . . .

Here, the first and second bits are XOR-ed to produce the third; the fourth and fifth are XOR-ed to give the sixth; and so on. The inputs are concatenated and presented as an unbroken sequence.

In the current version of the XOR problem, the input consisted of a sequence of 3000 bits constructed in this manner. This input stream was presented to the network shown in Figure 2-10 (with 1 input unit, 2 hidden units, 1 output unit, and 2 context units), one bit at a time. The task of the network was, at each point in time, to predict the next bit in the sequence. That is, given the input sequence shown, where one bit is presented at a time, the correct output at corresponding points in time is shown below.

```

input:      1 0 1 0 0 0 0 1 1 1 1 0 1 0 1 . . .
output:     0 1 0 0 0 0 1 1 1 1 0 1 0 1 ? . . .

```

Recall that the actual input to the hidden layer consists of the input shown above, as well as a copy of the hidden unit activations from the previous cycle. The prediction is thus based not just on input from the world, but also on the network's previous state (which is continuously passed back to itself on each cycle).

Elman networks, by having an internal feedback loop, are capable of learning to detect and generate temporal patterns. This makes Elman networks useful in such areas as signal processing and prediction where time plays a dominant role. It can be applied also to problems like discovering the syntactic/semantic categories in natural language data.

## **2.4 Perceptrons (PCM)**

Perceptrons is a neural network system, has a neuron use the threshold activation function. The Perceptron neuron produces a 1 if the network input into the transfer function is equal to or greater than a value; otherwise it produces a 0. The threshold activation function gives the Perceptron the ability to classify input vectors by dividing the input space.

The Perceptrons system can contain single layer or multiple layers each layer constructed from a number of neurons. The Perceptron Multilayer network has the same shape as Figure 2-6 with more hidden layers can be added.

Perceptrons network are especially suited for simple problems in pattern classification [20]. They are fast and reliable networks for the problems they can solve. They have several limitations. First, the output values of a perceptron can take on only one of two values (0 or 1) due to the threshold activation function. Second, perceptrons can only classify linearly separable sets of vectors. If a straight line or a plane can be drawn to separate the input vectors into their correct categories, the input vectors are linearly separable.

## **2.5 Linear Filter**

The linear networks are similar to the perceptron, but their transfer function is linear rather than hard-limiting. This allows their outputs to take on any value, whereas the perceptron output is limited to either 0 or 1. Linear networks, like the perceptron, can only solve linearly separable problems.

Linear networks may only learn linear relationships between input and output vectors. Thus, they cannot find solutions to some problems. However, even if a perfect solution does not exist, the linear network will minimize the sum of squared errors.

Single-layer linear networks can perform linear function approximation or pattern association.

The design of a single-layer linear network is constrained completely by the problem to be solved. The number of network inputs and the number of neurons in the layer are determined by the number of inputs and outputs required by the problem.

Nonlinear relationships between inputs and targets cannot be represented exactly by a linear network. It may give the optimal solution for nonlinear problems like regression, which deliver the global minimum (nonlinear problems may contains many local minimums in the error surface) of the MSE.

If the relationship between inputs and targets is linear or a linear approximation is desired, then linear networks are made for the job. Otherwise, backpropagation may be a good alternative.

## ***2.6 Choice of neural networks***

In fact, MATLAB support several kinds of neural networks. Each one of them has its own characteristics, neurons type, input type, output type and the layers number in the case of multilayer networks. Furthermore, the network may support several different training algorithms. Also, it can be used to solve a specific type of problems like classification problems, function approximation problems and error correction problems.

An optimum choice of network should be done in order to create our model. So, it is necessary to investigate the major networks alternatives. Two kinds of data will be used in this study binary and quantitative having a temporal and spatial character. Since the

networks ability to treat any problem depend on their applied data properties. So, the available networks were classified as follow:

**Backpropagation:** The data can be used to test the network by paying attention to underfitting and overfitting of the network.

**Elman:** It seems that it is a good alternative to be tested, since it is capable to detect and generate temporal patterns, which is similar to our task where the time plays a dominant role.

**Perceptron:** Binary data could not be involved due to the hard-limit transfer function. Also, the Perceptron Multilayer is not supported by MATLAB and it is not possible to increase the number of neurons of the network. Since, the available data are very complex, the permitted neurons are not enough to train the network. However, it is not a suitable alternative.

**Linear Filters:** The available data can be used to test the network but the miss of Linear Filters multilayer network in MATLAB, causes a problem like the problem found in the case of Perceptron network.

Consequently, it has been deduced that only the networks kind **Backpropagation** and **Elman** are acceptable alternatives and they are suitable to fit the given data due to the

following two raisons: they are able to accept the data without any restriction and they are able to treat problems like the present problem. Thus, in the remaining of study these networks will be the basis to create our proposed models. Noting that, the **Elman** and **Linear Filter** networks have been used to create relevant models [1], moreover the **Backpropagation** network has been tested [23].

## ***2.7 Preprocessing***

It is highly desirable to perform several Preprocessing steps for the input and target vectors sets before applying them to the neural network, which may take the form of a linear transformation of the data, or sometimes a more complicated method, has a significant role in determining the performance of the final system. Although neural networks systems can perform essential arbitrary non-linear mapping from the raw input data directly onto target values, in practice, such an approach will generally give poor results for several reasons. For instance, the temperature and ice weigh input values differ by several orders of magnitude. So, the typical sizes of the inputs may not reflect their relative importance in determining the targets. On the other hand, we may have very large and small values for data. Therefore, the gradient value in the training algorithm may vary in a large range, and the learning value may also be large at some points and small at other points. Thus, the network may choke completely or learn very slowly if the magnitudes of the inputs are too large. Preprocessing of data can solve these problems; make it easy to fit

the model. Yield a more consistent, increase convergence speed, and reduce the training time.

## **CHAPTER 3**

### **SYGIVRE DESCRIPTION**

### **3.1 Introduction**

Hydro-Quebec is closely involved in the study of the phenomena of atmospheric ice accumulation on its structures. In order to improve the monitoring of its power transmission networks, the company has set up a network of icing instruments called SYGIVRE. The principal measuring instrument used in this network is an autonomous instrument called: Icing Rate Meter (IRM). The function of this system is to record the atmospheric icing information in real-time, and to store it in a computer system.

A good understanding of the historical icing data delivered by Hydro-Quebec is necessary before making any sophisticated analysis or taking on modelling. The preliminary step in this work consists in analyzing and characterizing the available data of the SYGIVRE network.

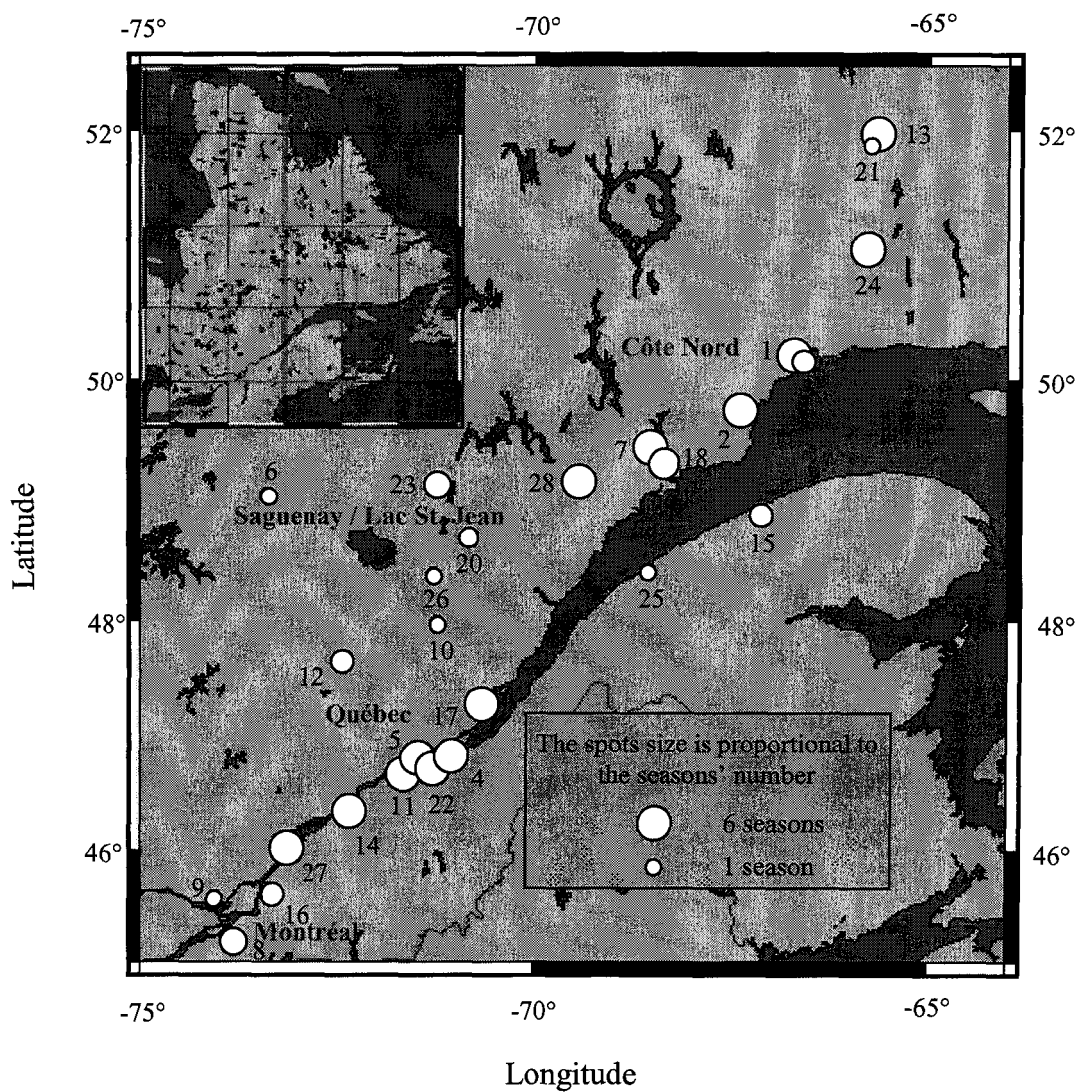
The SYGIVRE network has been setup in 1992 by Hydro-Québec in order to improve the monitoring of its own electrical networks [32,33]. The network consisted of 28 stations connected together, and distributed in the Province of Quebec in a fixed topology. The measures coming from the stations are collected and stored in the SYGIVRE database. This database contains information about icing events.

The data provided by Hydro-Québec covering the period from September, 1992, to April, 1998, were used to predict the behaviour of icing storms with the neural network design.

### ***3.2 Geographical and temporal distribution***

The stations are considered as a fixed mesh, and they are spread out along the St. Lawrence valley in the Province of Quebec, as shown by the map in Figure 3-1. Stations names are shown in Table 3-1. The spatial distribution of stations depends on the proximity of power transportation lines [34].

The SYGIVRE network is in operation since 1992, and is part of an ongoing process. Some measuring stations were installed in 1992, and are still in service, others have been in service for one or two seasons, and some have been put in service for a few seasons, and have been closed after that. An icing season in the SYGIVRE database extends from September of one year to April of the next year. Figure 3-1 gives a detailed measurement period for each station. A complete temporal distribution is given in Table 3-1 where the 'x' refers to the inactivity of the measurement season.



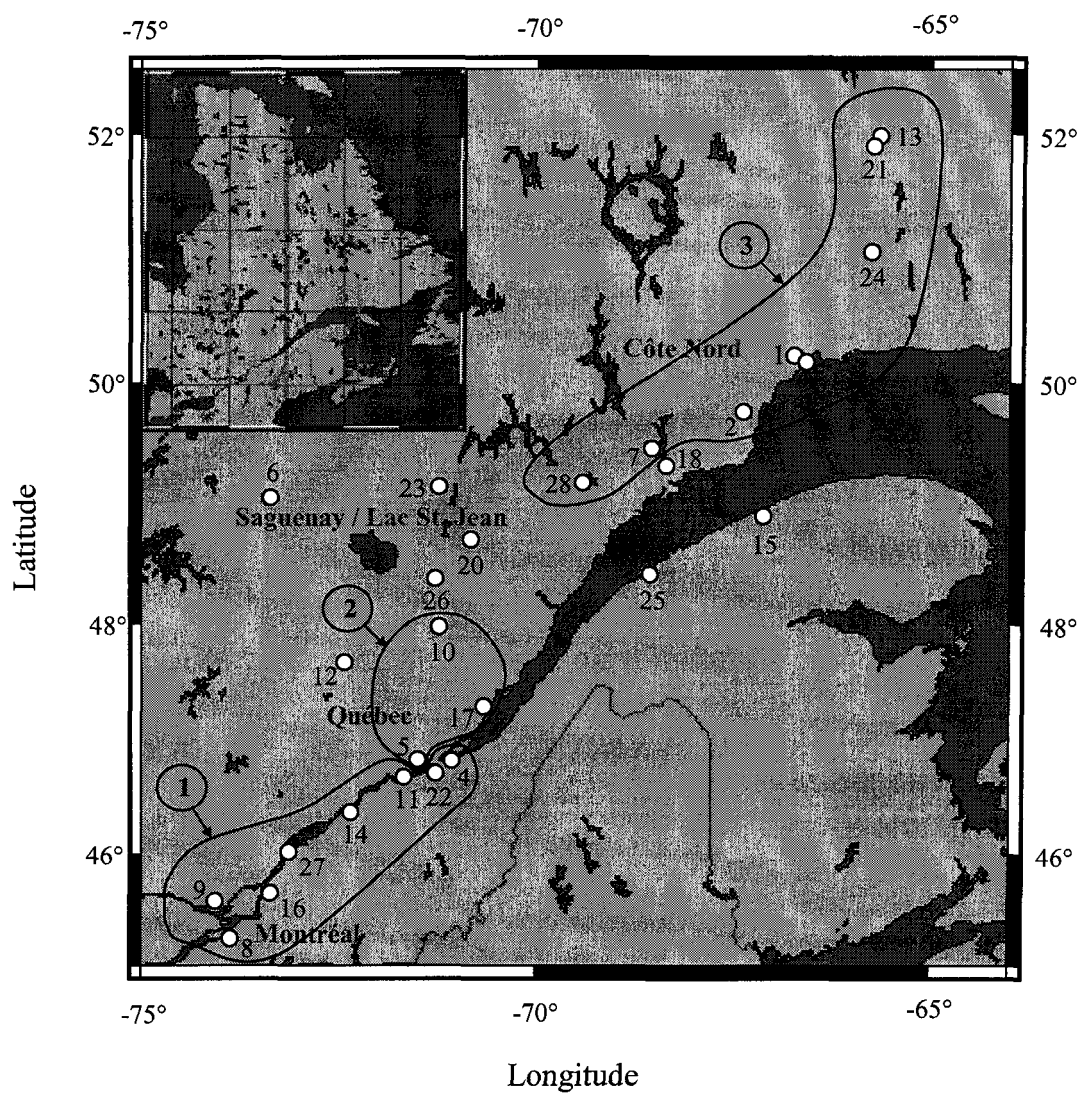
**Figure 3-1 Station locations for the SYGIVRE network**

		Six seasons measurements in the database					
Station Name	Station	1992 /1993	1993 /1994	1994 /1995	1995 /1996	1996 /1997	1997 /1998
STE-MARGUERITE	1	✓	✓	✓	✓	✓	✓
LAC ST-PIERRE	2	✓	✓	✓	✓	✓	✓
POSTE ARNAUD	3	x	x	x	✓	✓	✓
BEAUMONT	4	✓	✓	✓	✓	✓	✓
MONT BELAIR	5	✓	✓	✓	✓	✓	✓
BELEC	6	✓	x	x	x	x	x
LAC CAOUCETTE	7	✓	✓	✓	✓	✓	✓
POSTE CHATEAUGUAY	8	x	x	✓	✓	✓	✓
CHENIER	9	x	x	x	x	x	✓
TOUR DE DARAN	10	x	x	x	x	x	✓
DONNACONA	11	✓	✓	✓	✓	✓	✓
LAC EDOUARD	12	x	x	x	✓	✓	✓
ERIC	13	✓	✓	✓	✓	✓	✓
GENTILLY	14	✓	✓	✓	✓	✓	✓
GROSSE-ROCHE	15	✓	✓	✓	x	x	x
LIGNE EXPERIMENTALE IREQ	16	x	x	x	✓	✓	✓
LAC LAVOIE	17	✓	✓	✓	✓	✓	✓
CENTRALE MANIC 2	18	✓	x	✓	✓	✓	✓
MARVIN	19	✓	✓	x	x	x	x
MONCOUCHE	20	✓	✓	x	x	x	x
POSTE MONTAGNAIS	21	x	x	x	x	x	✓
PONT PIERRE- LAPORTE	22	✓	✓	✓	✓	✓	✓
PONTBRIAND	23	✓	✓	✓	✓	x	x
PREMIO	24	✓	✓	✓	✓	✓	✓
POSTE RIMOUSKI	25	x	x	x	x	x	✓
POSTE SAGUENAY	26	x	x	x	x	x	✓
SOREL	27	✓	✓	✓	✓	✓	✓
VIROT	28	✓	✓	✓	✓	✓	✓

**Table 3-1 State of measurement seasons for all stations of the SYGIVRE network**

Stations 6, 21, and 26 were not considered in the current study because they contain only one season of measurements, which is not enough to allow a reliable analysis.

Based on the work of C. Guesdon et al. [18], that determined spatial and temporal links between stations, we considered three principal groups of stations for analysis. The first group covers the region from Montréal to Quebec City, and includes 8 stations, namely: 4, 8, 9, 11, 14, 16, 22, and 27. The second group covers the regions of Quebec City and Saguenay / Lac St. Jean, and includes 3 stations: 5, 10, and 17. The last one covers the Côte-Nord region, and includes 8 stations: 1, 2, 3, 7, 13, 21, 24, and 28. Figure 3-2 shows the location of each station according to the longitude and latitude coordinates, and shows also the regions covered by each group.



**Figure 3-2 The principle station groups**

### **3.3 SYGIVRE Database**

It appears that a preliminary analysis on the SYGIVRE database should be done in order to understand its major characteristics, and to extract the essential information for the current study.

#### **3.3.1 Ice type**

In order to understand the ice accretion behaviour, a hint about the icing types is necessary. This information helps understanding the icing event formation under complex meteorological conditions. Extra information can be found in G. Fortin's lecture notes [35].

The term ice accretion or icing is employed to describe the process of ice growth on a surface exposed to the atmosphere. The ice growth rate on a surface depends on the impact rate of the ice particles, the airflow characteristics, and the local thermal conditions of the surface.

In general, three types of ice accretion contribute to the formation of ice on surfaces. These are rime, glaze, and frost. The rime and glaze form from water droplets in the atmosphere, which were in supercooled state. A droplet remains in liquid state when the temperature is well below the freezing point at low atmospheric pressures. The droplets

density is the Liquid Water Content (LWC) in  $g/m^3$ . The frost forms from the vapour in the atmosphere by the sublimation process (the transformation of vapour into ice without going through the liquid state). The vapour density depends on the ambient humidity.

## **Rime**

Rime is an ice deposit caused by the impact of supercooled droplets which freeze on a surface instantly by losing their latent heat to the surrounding. This is usually associated with freezing fog with droplet size typically of 10  $\mu m$ . Rime can be formed when the air temperature is well below  $0^\circ C$  (less than  $-5^\circ C$ ) or slightly above  $0^\circ C$ .

When the air temperature is below the freezing point, the supercooled droplets possessing small momentum will freeze instantly on impact, creating air pockets between them. This type of deposit is known as soft rime, and has a low density (less than  $0.1 g/m^3$ ). When the droplets possess greater momentum, or the freezing time is greater, the frozen droplets pack closer together in a dense structure, known as hard rime. The typical density range for hard rime is  $0.1-0.6 g/m^3$ .

When the air temperature is above the freezing point and the wind is calm, air near the substrate remains cold, and the temperature increases as we move away from the surface. Thus this allows the surface temperature to fall down, and become cooler than

0° C. Consequently, water droplets near the surface undergo a supercooled state, and start hitting the surface. However, the temperature around the surface stays slightly above 0°C.

### **Glaze**

Glaze ice will form when the droplet freezing time is sufficiently long for a film of water to cover the accreting surface. Certain water quantity stays unfrozen, and when a second droplet arrives at the same place it adheres to the previous one. The accretion is accomplished at the water solidification temperature, which is 0°C at the atmospheric pressure.

Glaze is usually associated with large values of the LWC and droplet sizes of 500  $\mu\text{m}$  as found in the freezing rain incidents. The latter occurs when there is a layer of below-freezing air near the surface with warmer air aloft. Rain droplets from aloft fall into the cold layer, and transform to supercooled rain. When these hit the surface they freeze immediately into a clear glaze ice.

Glaze ice is compact, smooth, and usually transparent. It is known by its strong adhesion to surfaces. The density of glaze ice approaches that of bubble-free ice (i.e.  $0.917 \text{ g} / \text{m}^3$ ).

## Frost

Frost forms when water vapour in the air sublimates on a cold surface (below  $0^{\circ}\text{C}$ ). It can form also when the ambient temperature is slightly above  $0^{\circ}\text{C}$ . This process looks like the formation of rime at temperature above freezing point, but the difference is that the surface falls below the frost point, and then the water vapour start to sublimate on it. The density of frost varies, and tends to be low (e.g., less than  $0.1 \text{ g} / \text{m}^3$ ). Iceload due to frost is small, and is limited by the degree of humidity of the air.

The SYGIVRE database has included icing types as two categories; the first represents the glaze type which lead to *freezing rain* (V), and the other represents the rime and frost types, and is termed *frost or incloud icing* (G) [36].

### 3.3.2 IRM signal

The IRM is an autonomous system that delivers the increase of weight accumulation every hour. It emits a number of signals during an elapsed hour, and it receives ice deposit on its probe, with a specific fixed predetermined thickness of ice accumulation. As soon as it attains this value of accumulation, it emits a signal, and it takes away the accumulation on its probe. This process is repeated for number of times, which is termed: the number of IRM signals. The ice accretion rate measured in  $\text{g} / \text{m} / \text{h}$  can be estimated using the number

of IRM signals and the accumulation type deposit: either *freezing rain* or *frost*, during the elapsed hour [5].

### 3.3.3 Ice accumulation weight

The IRM estimate the ice accretion rate during an icing event. In order to evaluate the total ice accumulation weight at any time it is enough to aggregate the accretion rates over the passing hours starting from the first time, the icing precipitation starts, until the desired time.

### 3.3.4 Temperature

The ambient temperatures, expressed in degree Celsius, are recorded at almost every hour to all stations. So the SYGIVRE database integrates all those temperature readings.

Several stations contain discrete temperature readings. This is likely because the used temperature measuring instruments have weak precision or error in calibration. Table 3-2 shows certain stations that contain discrete temperature readings. The following readings belong to the class of values: -20, -15, -9, -3.5, -0.5, 0, 0.5, 2.5 and 4.0 °C. The stations that have this characterization are: 1, 2, 5, 8, 7, 11, 13, 17, 18, 19, 20, 23, 24, 27, and 28. The frequency of each discrete temperature for the following stations is displayed in Table 3-2 as well.

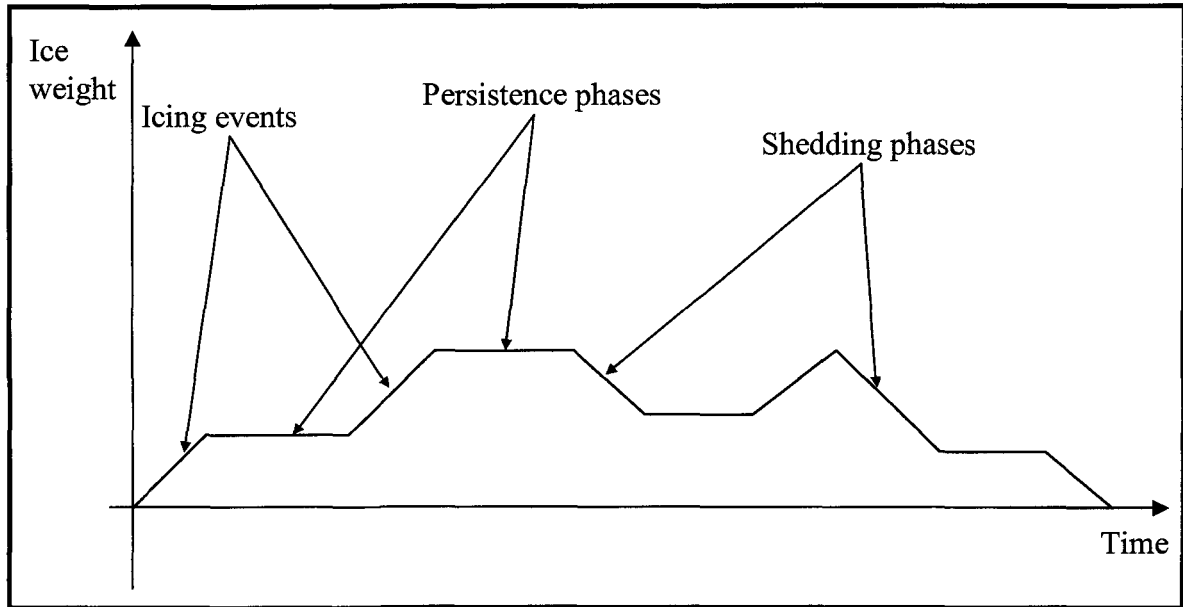
	Station number						
Temperature (°C)	1	2	5	8	7	11	13
-20	582	805	674	99	1 426	309	710
-15	639	187			474	304	2 667
-9	36				445		1 442
-3.5	163				782	21	970
-0.5	335				462		658
0	82				131		132
0.5	51				25		81
2.5	98				106		160
4	187				83		156

	Station number						
Temperature (°C)	17	18	19	20	23	24	27
-20	1 037	410	1 787	1 077	1 752	1 730	484
-15	916	1	2 770	704	1 900	3 596	158
-9	537	5				2 137	
-3.5	1 435	48				2 707	
-0.5	1 440	28	90			994	
0	113	4				71	
0.5	86	3			120	28	
2.5	107	6				164	
4	84	11				318	

**Table 3-2 Number of occurrence at each temperature level for all stations**

### 3.3.5 Certified events

An icing event as defined in the SYGIVRE database is the total time between a first accumulation of ice on the structure and the complete shedding of ice, which is termed as certified event (*CE*) [36]. It consists of a combination of three distinct phases. Each phase has a specific duration. The three phases are the accretion phase (ice accretion increase gradually), the persistence phase (ice stay stable) and the shedding phase (progressively ice loss). So, a certified event from the database has a temporal delimitation which starts by an accretion phase without initial charge followed by a combination of accretion, persistence and shedding phases and ended by a shedding phase after a total charge loss as shown in Figure 3-3 (Figure is only an example of a certified event and other may have quite different forms). The total number of certified events in the database are counted as 1 693 CEs. At the accretion phase time, we can say that an icing event occurs.



**Figure 3-3 General certified event sample**

### 3.3.6 Variables characterisation

Several basic meteorological variables from the SYGIVRE network are present with different degree of importance. Four variables have been included: total ice accumulation weight, air temperature, ice rate accumulation, and icing event duration. These variables are the only values available for the study.

The IRM deliver just the estimate of accumulation during an hour, in our terminology is the *ice rate accretion* variable. The total ice accumulation weight variable is the aggregation of accumulations in the passing hours up to the desired point. The duration variable represents the number of accumulation hours in an icing event. Another essential

measured variable is the air temperature, which is taken at every hour even if there is no accumulation.

This information was recorded by Hydro-Quebec, and was validated at the end of each season. Although the system records all the variables during a season, the database provides the variables values only at the time of certified events. Several variables are not included in the following study since they are not available to all stations. These variables are less important, and include velocity, direction of wind, and relative humidity.

### ***3.4 Database for the icing storm study***

The certified event concept is useful to study icing loads variation with time. Our concern is to detect only the risk at the icing accumulation time or, in other words, at the icing event occurrence when the IRM is attempting to make a reading. An icing storm consists of the icing events belonging to the same meteorological storm. In order to perform analysis for the icing storm, the total time that covers a meteorological storm should be determined, because the non accumulation phases (persistence, shedding) are so long that more than one storm can affect the structure during a certified event. A study should be carried out in order to derive the icing storms.

### 3.4.1 Icing events

An accretion phase (*AP*) consists of a set of consecutive hourly passing accumulations determined by the IRM, starting from the first hour after a non accumulation state until reaching the final one where the accumulation ceases. An icing event belonging to a station is exactly the accretion phase in a certified event.

It is convenient to look at the distribution of icing events depending on each season for all stations. Table 3-3 shows, for all stations in the database, the corresponding number of events in each season, and the corresponding group number for the icing events as well. Often, a certified event contains more than one icing event. That is why the total number of icing events in the database is 6 401.

Six seasons measurements of the database							Total Events Number	Group No
Station	1992 /1993	1993 /1994	1994 /1995	1995 /1996	1996 /1997	1997 /1998		
1	30	25	12	41	52	17	177	3
2	107	124	93	133	106	160	723	3
3	x	x	x	12	4	10	26	3
4	18	2	3	6	24	16	69	1
5	153	163	301	143	258	97	1 115	2
7	80	30	35	39	44	51	279	3
8	x	x	2	21	18	40	81	1
9	x	x	x	x	x	39	39	1
10	x	x	x	x	x	54	54	2
11	44	27	59	19	37	22	208	1
12	x	x	x	34	24	14	72	
13	64	586	32	35	88	70	875	3
14	10	9	9	8	23	28	87	1
15	54	67	20	x	x	x	141	
16	x	x	x	18	22	35	75	1
17	49	91	54	77	70	57	398	2
18	10	x	2	4	2	7	25	
19	105	125	x	x	x	x	230	
20	64	16	x	x	x	x	80	
22	3	5	17	13	25	15	78	1
23	110	68	91	32	x	x	301	
24	93	198	73	121	172	97	754	3
25	x	x	x	x	x	29	29	
27	23	13	27	29	18	25	135	1
28	66	65	45	40	61	55	332	3

**Table 3-3 Icing events distribution over six seasons**

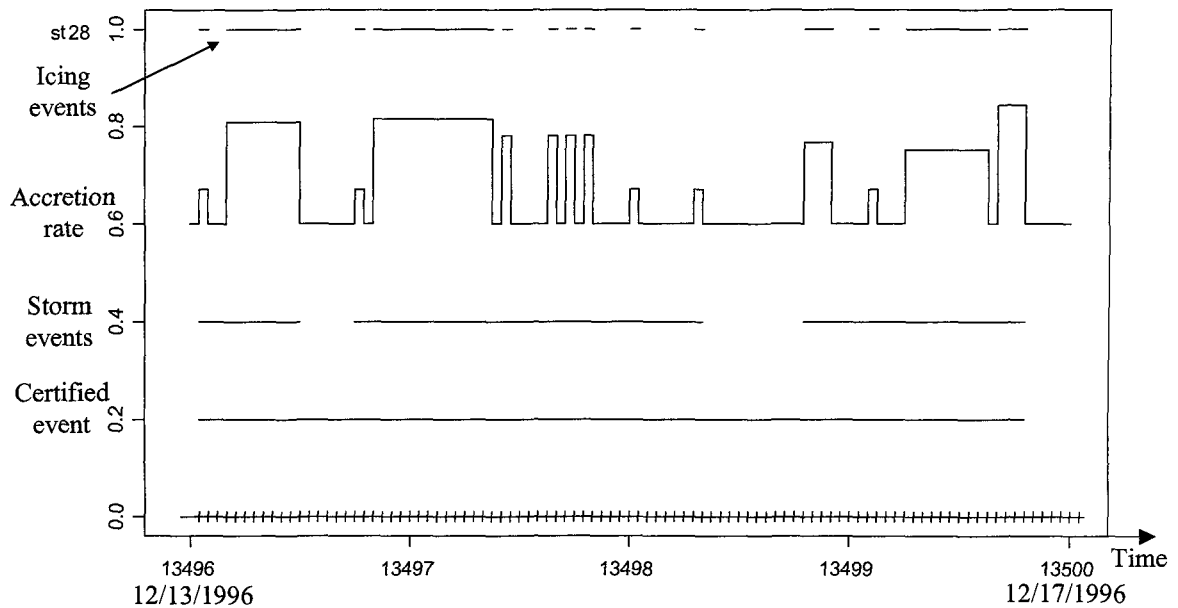
### 3.4.2 Storm events

The notion of icing storm is linked to a geographic location including a group of stations. Any station may be affected when a storm hits the region covered by its group. Here, we need to define the term *storm event (SE)* which represents the icing storm. For a group of stations, a storm event starts with the beginning of accumulation in any station, and finishes when there will be no more accumulation in the stations of the group. The icing events should be used to define the storm events. A certified event includes one or more icing storms.

The effect of a meteorological storm hitting a territory can vary during its occurrence. The deposit precipitations may rise or fall to a trace or even cease temporarily. The IRM is not able to detect the times when there was temporary absence of accumulation or a trace accumulation. Such times should be differentiated from the time of storm terminations in order to define the exact elapsed time of storms for any group.

A sample certified event from the SYGIVRE database is analyzed and shown in Figure 3-4. This certified event belongs to station 28 from the Côte Nord group which lasted 4 days from 12/13/1996 to 12/17/1996. It consists of 14 icing events and 13 non accumulation gaps between them; two of them are storm terminations. Let us define *idle time* as the gap of non accumulation between two icing events. The idle time can be a

separation time between two totally different icing events. It can also be a suspended time for the primary icing event, and the storm should resume after a short time so that the suspended time belong to that storm.



**Figure 3-4 Dissociation of a certified event**

A simple method to define the storm event period is to integrate suspended times. The aim is to estimate the actual storm event duration based on the suspended times integration notion, because it is not possible to directly determine the correct storms durations of the SYGIVRE database because of information loss between icing events.

A visual algorithm is easy to implement as in the example of Figure 3-4. There are 5 idle times of one hour duration, 2 idle times of 4 hours duration, and one with durations 2,

3, 5, 6, 7, and 13 hours. Thus, the aim is to create a robust algorithm capable of defining the storm events durations. This algorithm will be based on the idle times between icing events.

A general analysis of the idle times on the available icing events for the stations of the SYGIVRE database has been carried out. Table 3-4 shows the number of the idle times with durations relying on a specific hourly time range, and its percentage with respect to the number of icing events in the database. It is clear that the idle times durations tied to the two ranges (1, 2) and (2, 3) appear frequently in the database.

Time range (h)		Idle times number	Percentage with respect to the 6 401 icing events
From ≥	To <		
0	1	319	4.98
1	2	962	15.03
2	3	665	10.39
3	4	380	5.94
4	5	167	2.61
5	6	210	3.28
6	7	124	1.94
7	8	85	1.33
8	9	99	1.55
9	10	81	1.27
11	24	607	9.48
			<b>Sum: 57,8 %</b>

**Table 3-4 Idle time frequency in terms of time ranges**

Another factor that affects our algorithm is the number of icing events linked to a certified event. A general analysis for the icing events within the certified events has demonstrated that about 40 % of the certified events contain just one icing event. Thus, most of these certified events will not be included in the algorithm construction, since there are no idle times. As shown in Table 3-5 from 1 693 certified events, 657 cases contain just one icing event and 363 comprise 2 icing events.

<b>Icing event number</b>	1	2	3	4	5	6	7	8	9
<b>Cases number</b>	657	363	200	149	90	49	23	22	20
<b>Percentage</b>	38.81	21.44	11.81	8.80	5.32	2.89	1.36	1.30	1.18

10	11	12	13	14	15	16	17	18
17	17	12	8	7	6	9	6	3
1.00	1.00	0.71	0.47	0.41	0.35	0.53	0.35	0.18

19	20	21	22	23	26	27	28	29
3	1	3	1	3	1	1	1	1
0.18	0.06	0.18	0.06	0.18	0.06	0.06	0.06	0.06

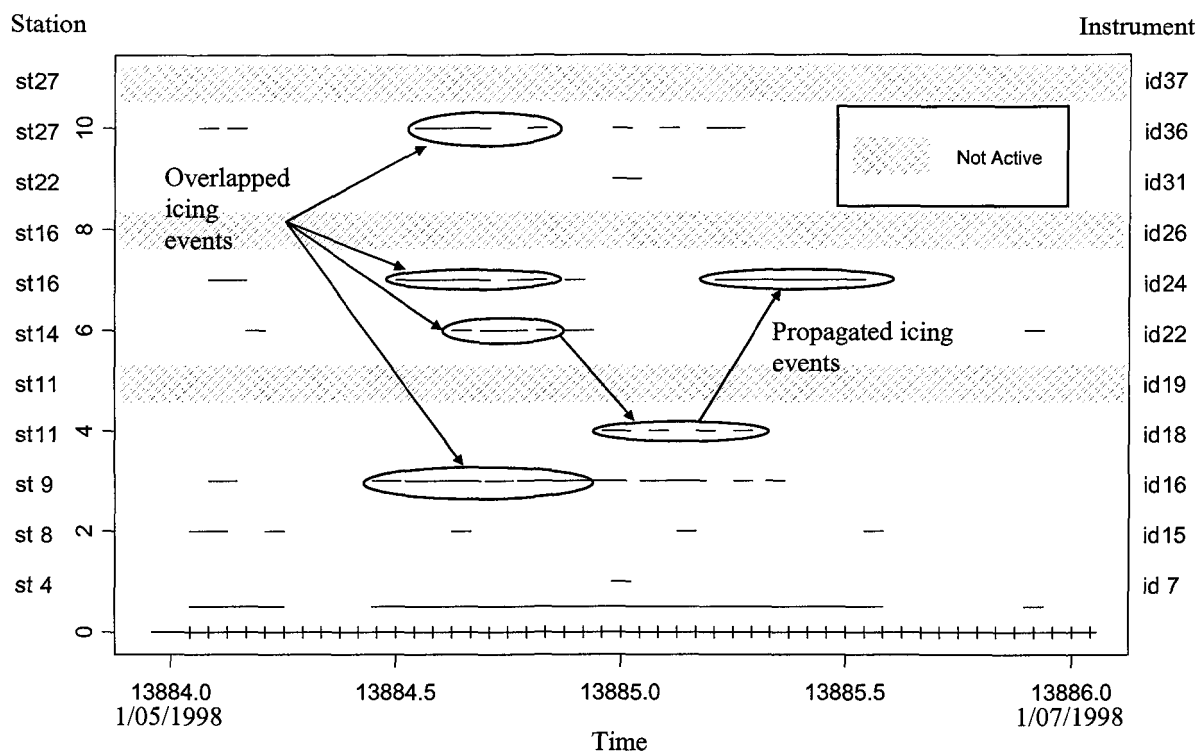
31	32	35	40	41	42	45	48	51
2	3	1	1	1	2	1	1	1
0.12	0.18	0.06	0.06	0.06	0.12	0.06	0.06	0.06

59	82	107	131	193
1	1	1	1	1
0.06	0.06	0.06	0.06	0.06

**Table 3-5 Number of icing events inside certified events**

Since the present study considers multiple stations in the prediction, it is essential to pay attention to the relationship between stations: icing storms may affect all the stations in neighbourhood. For example, Figure 3-5 shows all recorded observations during two days from 1/05/1998 to 1/07/1998 of the Montreal group. As shown in the figure, this group of stations appear on the left, and their corresponding instruments on the right (a station may contains several instruments), and the shaded region near an instrument signifies that this instrument was not active during the case study season. If we consider a station and its neighbouring stations in the same group, an icing event from one station can overlap the icing events of neighbouring stations at the same time. This mean that all overlapped icing events can be considered as part of the storm happening at that time. In other situations, the same icing event of the station can occur at some neighbouring stations, but after a short time, which signifies that the storm was propagated to reach the neighbouring stations.



**Figure 3-5 Overlapped and propagated icing events**

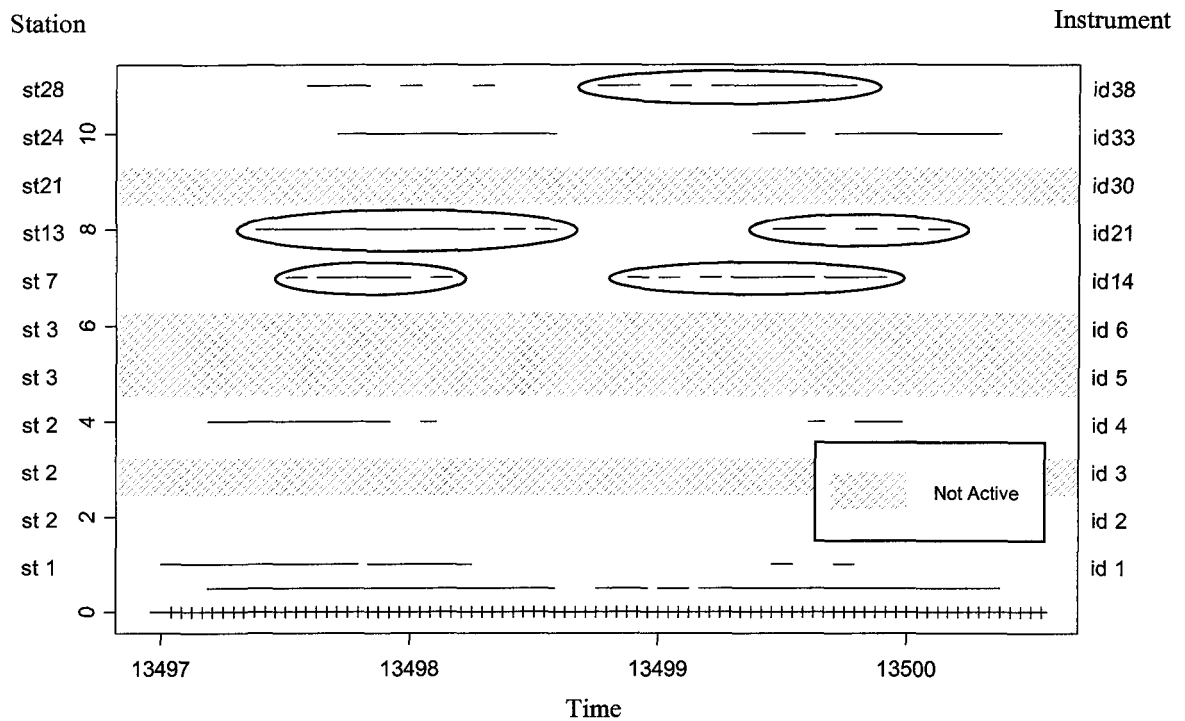
One advantage of grouping is to gather information from a group of stations in order to determine the storm events periods. The relationship between stations will affect the neural network. The network processing and its generalization will be based on this characteristic, and then, the network results at some stations will depend on the data of other stations in the same group.

In order to define correctly the storm events periods, a collection of cases from the database were analyzed in order to characterize them. The cases were selected randomly from the available three groups of stations to cover most of the special cases found in the

database. In each case, the icing events within a certified event belonging to a specific station were monitored attentively along with the simultaneous icing events belonging to the neighbouring stations of the same group. This analysis led to several conceptual decisions proposals, covering the possible situations, which will be exposed below.

Table 3-4 gives the idle times tied to the range from 1 hour to 3 hours; these are the most repeated cases in the database. This signifies that, icing events were stopped many times for short periods, so these idle times can be considered as suspended times. Many examples from the database like the example shown in Figure 3-6 can be located. It is clear that any encircled collection of icing events in Figure 3-6 can be merged and considered as a part of the current storm, since all the idle times inside the circles have periods of less than 3 hours.

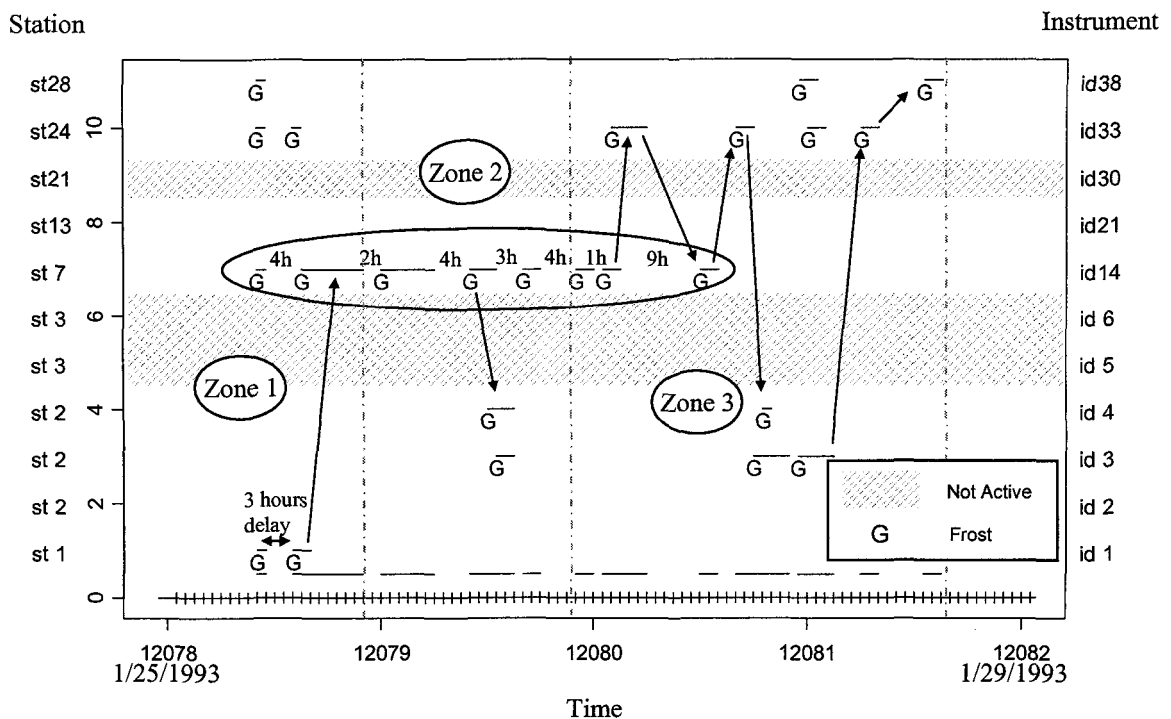
In order to integrate all the suspended times for cases like those mentioned in Figure 3-6, it has been decided that successive icing events with three hours or less suspended times with the same accretion type (Frost [G] or Freezing rain [V]) would be merged together (with some exceptions due to the accretion type variation errors in the database). This decision will be applied on the overall set of icing events.



**Figure 3-6 Idle times less than 3 hours for the Côte Nord group**

It is convenient now to look at several special cases in the database, since the previous decision is not sufficient to detect all suspended times. Some situations are similar to the example shown in Figure 3-7. The icing events of this example are recorded from 1/25/1993 to 1/29/1993, and belong to the Côte-Nord group. It is clear from this example that not all the idle times between icing events for this group are less than 3 hours. However, because of their relatively short duration, the propagation of storm between stations and the icing events have the same accretion type: *frost*, it seems that all the icing events belong to the same meteorological storm.

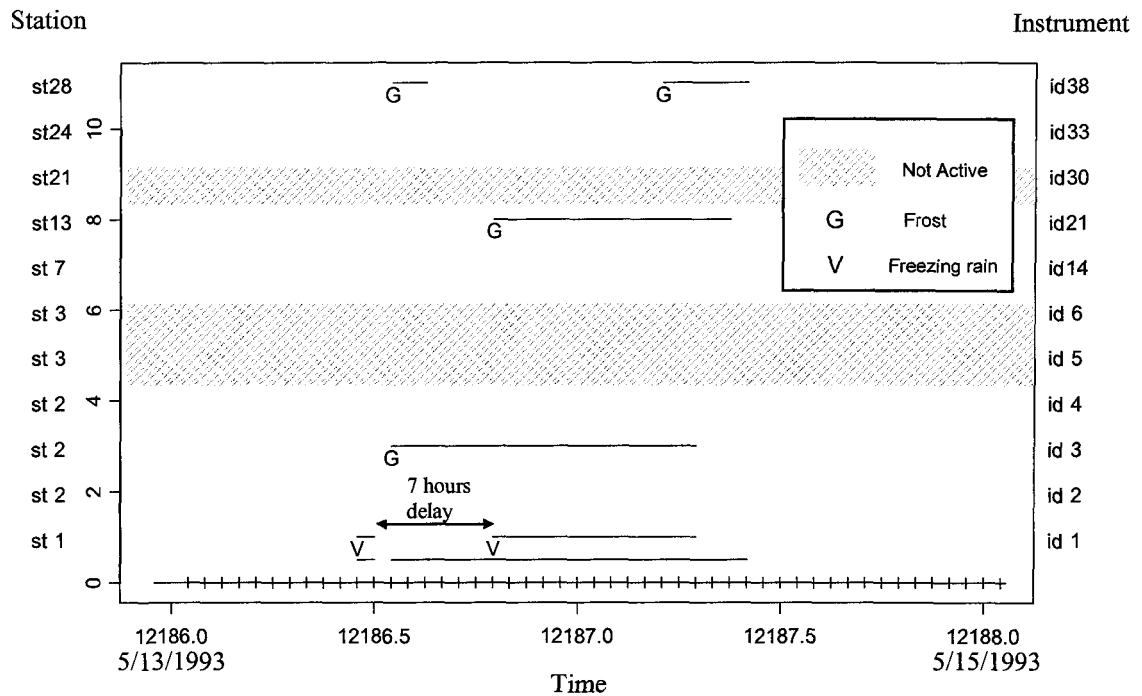
The storm has started in zone 1 in Figure 3-7, and has affected stations 1, 7, 24, and 28. It has continued only at station 7, and has propagated to zone 2 at station 2 for about 3 hours. In zone 3, it has restarted at station 24, and stopped at station 7 for 9 hours. Therefore, the storm has continued to propagate according to the arrows in the figure until it has ended at station 28. Consequently, all the existing idle times between the icing events can be considered as suspended times that can then be merged.



**Figure 3-7 Storm event propagation between stations of the Côte Nord group**

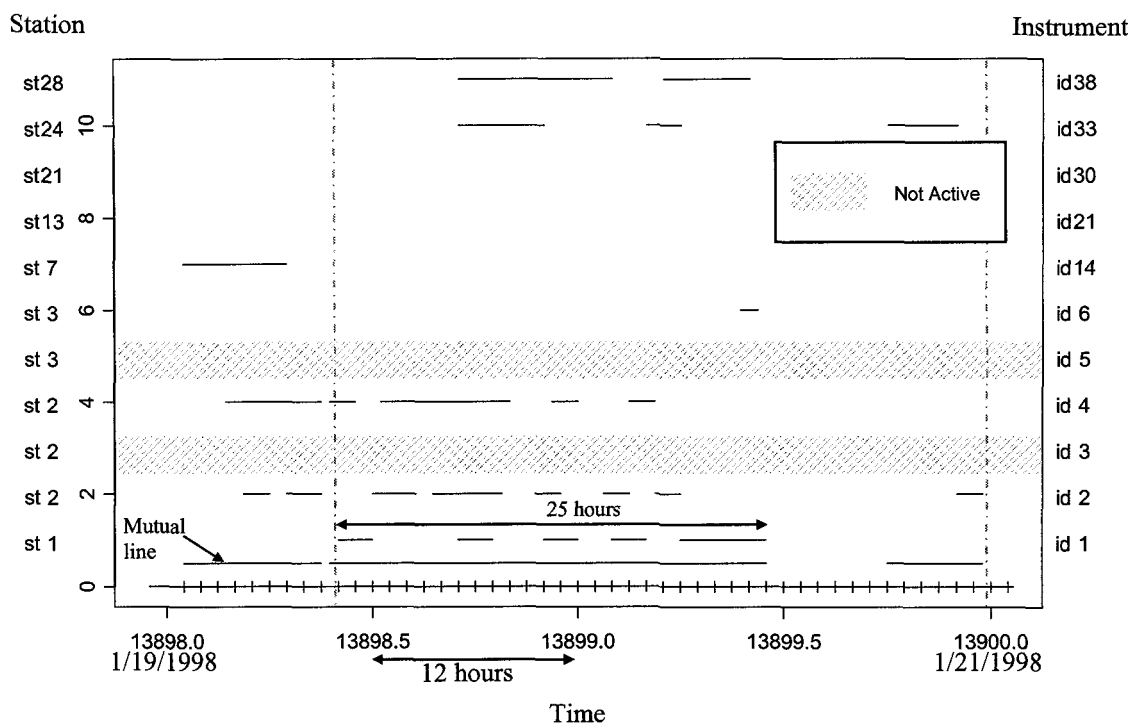
Another example is shown in Figure 3-8. It reveals what happened in the Côte-Nord group from 5/13/1993 to 5/15/1993. It is obvious from the figure that the two icing events at station 1 have *freezing rain* accretion type. This means that these icing events belong to a

freezing rain storm which has affected station 1, and have not propagated to other stations. The suspended time between the two icing events is 7 hours. Since this time is almost long and the icing events have not propagated to other stations, it is unreasonable to consider it as suspended time.



**Figure 3-8 Two different storms due to the difference of accretion types**

Now, let's consider the Côte-Nord group from 1/19/1998 to 1/21/1998, shown in Figure 3-9. The line at the bottom represents the existence of any icing event from the 11 instruments of the group. This is called the mutual line.



**Figure 3-9 Example to discuss the mergence of the icing events at station 1**

The idle times between the icing events at station 1 are presented. It is clear that the storm propagated through stations 1, 2, 3, 24, and 28. At the idle times, some icing events from these stations were active, and this can be verified by the solid mutual line covering the same times. Therefore, the idle times can be considered as suspended times, and the icing events can be merged together to form one bigger storm event.

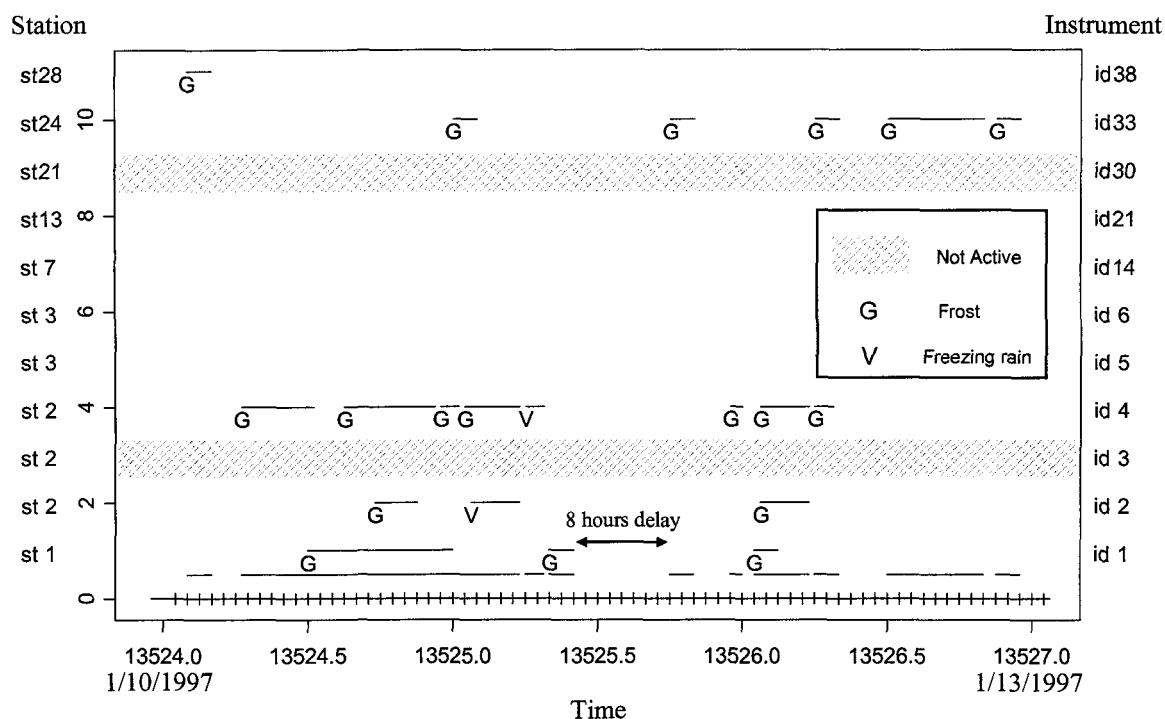
The idle times of a station from a group are determined. In order to know the situation of these idle times, it is enough to look at the icing events in the neighbour stations of the same group that synchronize them. The idle times covered by those icing

events will be considered as suspended times only if the icing events in the neighbouring stations overlap the idle times and the case study icing events have the same type accretion.

It is more apparent to look at the mutual line determined by icing events with the same accretion type. If the mutual line was filled at the moment of idle times, these times will be considered as suspended times.

For a group of stations, we consider an idle time from a station, and the occurring icing events that surround it in the other stations. Those icing events may have variable accretion type, so only a homogeneous type should be selected. The most repeated accretion type is that which is selected, otherwise the more appropriate one could be selected for equal repetition.

Consider the example in Figure 3-10, whereas there are 3 icing events of type *frost* with 2 idle times between them at station 1 from the Côte-Nord group. Based on the previously described rule, the icing events from instruments 2 and 4 are generally of type *frost*. Thus, the first idle time can be considered as suspended time because the mutual line is almost filled at that time. However, at the second idle time, there is a delay of 8 hours at the mutual line (i.e. no icing events at this time).



**Figure 3-10 A big idle time with variable accretion type**

In such cases, if the icing event coming before the idle time is of type *frost*, and the delay is 6 hours or less, then the idle time can be considered as suspended time. On the other hand, if the icing event type is *freezing rain*, the idle time can be considered as suspended time for a delay of 5 hours or less. Thus, based on these decisions the second suspended time in the example of Figure 3-10 is a separation time, because the current delay is 8 hours.

It is convenient to look at the distribution of storm events depending on each season for all stations. Table 3-6 shows, for all stations in the database, the corresponding number of storm events in each season. Table 3-7 gives the number of certified events, icing events

and storm events for all stations; so a total of 2 997 storm events resulted. Those storm events have been determined after applying the previously proposed algorithm to the entire database, whereas each group is processed separately.

Station	Six seasons measurements of the database						Total Events Number	Group No
	1992 /1993	1993 /1994	1994 /1995	1995 /1996	1996 /1997	1997 /1998		
1	18	14	7	14	24	14	91	3
2	50	45	44	44	40	46	269	3
3	x	x	x	7	3	7	17	3
4	12	2	2	3	13	13	45	1
5	92	85	148	53	119	42	539	2
7	38	7	14	16	20	26	121	3
8	x	x	1	17	15	24	57	1
9	x	x	x	x	x	18	18	1
10	x	x	x	x	x	31	31	2
11	23	15	22	12	19	12	103	1
12	x	x	x	15	17	6	38	
13	45	111	9	20	35	30	250	3
14	7	7	5	6	12	14	51	1
15	37	46	11	x	x	x	94	
16	x	x	x	13	10	23	46	1
17	25	50	22	31	35	26	189	2
18	8	x	2	4	2	5	21	
19	78	89	x	x	x	x	167	
20	49	16	x	x	x	x	65	
22	2	4	8	9	12	8	43	1
23	75	46	52	17	x	x	190	
24	47	80	43	48	63	41	322	3
25	x	x	x	x	x	15	15	
27	17	5	10	19	8	9	68	1
28	30	28	17	20	27	15	137	3

**Table 3-6 Storm events distribution between the measurement seasons**

<b>Station</b>	<b>Certified event Number</b>	<b>Icing event Number</b>	<b>Storm event Number</b>
<b>1</b>	74	177	91
<b>2</b>	207	723	269
<b>3</b>	18	26	17
<b>4</b>	37	69	45
<b>5</b>	299	1115	539
<b>7</b>	90	279	121
<b>8</b>	59	81	57
<b>9</b>	23	39	18
<b>10</b>	19	54	31
<b>11</b>	89	208	103
<b>12</b>	35	72	38
<b>13</b>	74	875	250
<b>14</b>	51	87	51
<b>15</b>	31	141	94
<b>16</b>	48	75	46
<b>17</b>	125	398	189
<b>18</b>	18	25	21
<b>19</b>	17	230	167
<b>20</b>	11	80	65
<b>22</b>	42	78	43
<b>23</b>	28	301	190
<b>24</b>	127	754	322
<b>25</b>	14	29	15
<b>27</b>	62	135	68
<b>28</b>	85	332	137

**Table 3-7 List of all stations of the SYGIVRE database with the corresponding number of certified events, icing events and storm events**

Consequently, all discontinuities between icing events have been regulated. Thus, a new duration for the actual or close to the actual icing storms has been deduced. The rest of the study will be based on the resulting storm events.

### 3.4.3 Temperature missing

Several observations in the database, and especially those of type *freezing rain*, have the temperature missing. It is important to look at the temperature values when there is accumulation (i.e. at the time of icing events within the certified events). A total of 666 out of 6 401 icing events do not have temperature readings. Table 3-8 shows the number of icing events and storm events of each station with and without missing temperature readings. Normally, those icing events need be eliminated only when including the temperature variable in the prediction, because the neural network does not accept missing data. Furthermore, the replacement of empty temperature values by approximate values will reduce the prediction precision.

Station	Instrument	Icing events with missing temperature measure	Icing events	Storm events with missing temperature measure	Storm Events
1	1	15	177	11	91
2	3,4	52	723	21	269
3	5,6	0	26	0	17
4	7	0	69	0	45
5	8,9,10,11,12	140	1 115	76	539
7	14	13	279	5	121
8	15	1	81	1	57
9	16	0	39	0	18
10	17	0	54	0	31
11	18,19	40	208	18	103
12	20	0	72	0	38
13	21	34	875	22	250
14	22	0	87	0	51
15	23	0	141	0	94
16	24,26	11	75	8	46
17	25	32	398	15	189
18	27	3	25	2	21
19	28	45	230	31	167
20	29	28	80	16	65
22	31	0	78	0	43
23	32	146	301	86	190
24	33	61	754	23	322
25	34	0	29	0	15
27	36,37	21	135	10	68
28	38	24	332	13	137

**Table 3-8 List of all stations with the corresponding number of icing events and storm events with and without missing temperature readings**

Looking at the periods of persistence and shedding phases, we can notice that there are a few missing temperature readings. An algorithm has been implemented for estimating the values of these missing temperatures. In a group of stations, if one of the stations has a missing temperature at a specific time, this value is estimated as the average of its

surrounding temperature values at that time, coming from either the same station or the neighbouring stations of the group.

### ***3.5 Data available for neural network***

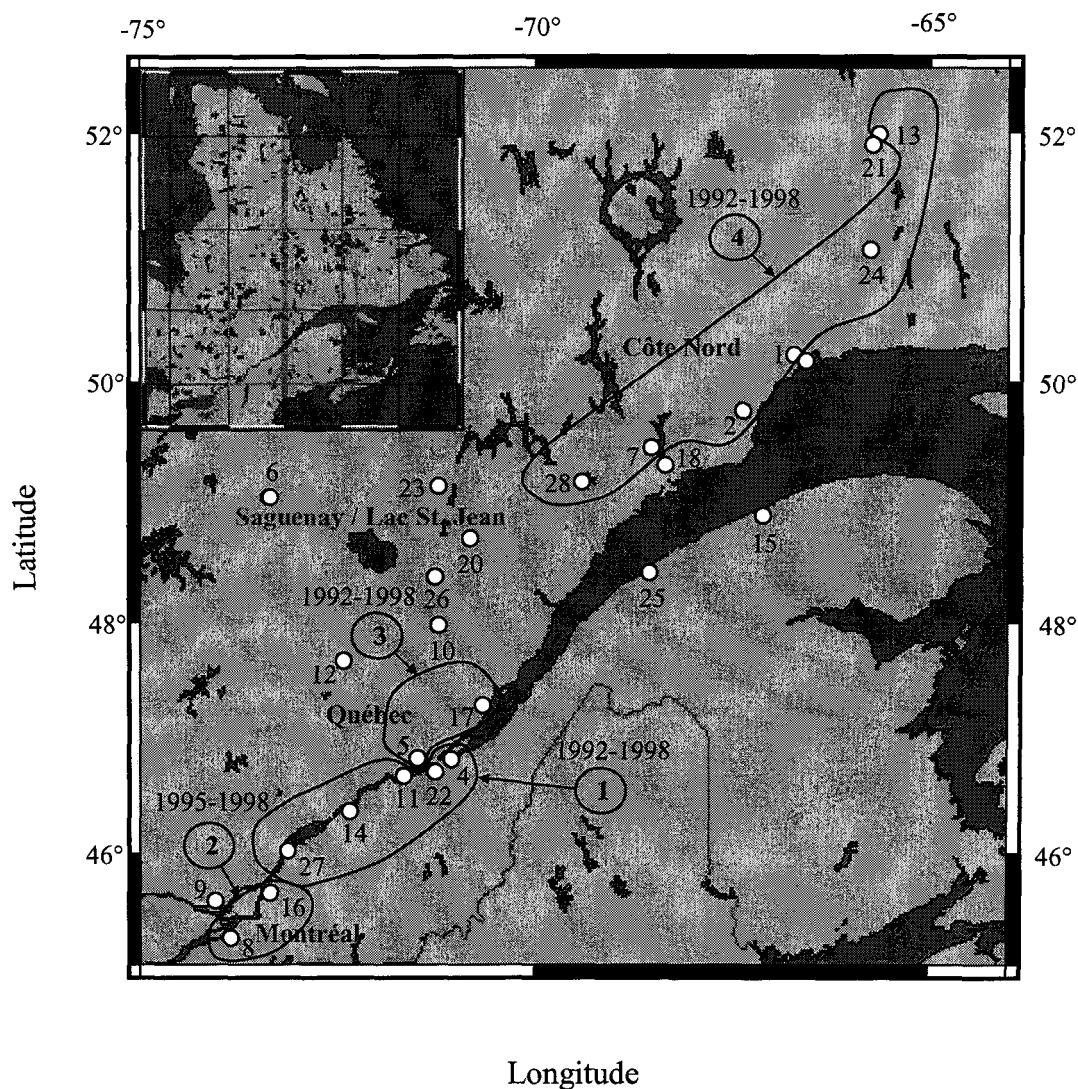
Each group of stations is considered as a separate entity, and is processed independently. The neural network does not allow missing values and, as mentioned before, some stations have missing seasons. So, for each group, only the activated seasons that are common between stations are selected by eliminating the remaining seasons. The groups will be restricted in order to prevent a loss of precision in the prediction.

Referring to Table 3-3, the first group is derived as follow. Stations 4, 11, 14, 22, and 27 that belong to the Montreal group are included, since they support all the seasons of measurements. The second group only accepts the seasons 1995-1996, 1996-1997, and 1997-1998 for the stations 8 and 16 of the Montreal group, since these stations were activated during the above-mentioned seasons. Furthermore, station 8 in 1994-1995 contains two icing events with 1 hour and 3 hours duration, and these were merged as one storm event with 6 hours duration as marked in Table 3-6. Since it is the only event representing station 8 in 1994-1995, it can be discarded in order to be able to define the group.

The third group consists of stations 5 and 17 from the Quebec group, which supports all 6 seasons. It is convenient to note that this group contains many icing events. The fourth and final group consists of stations 1, 2, 7, 13, 24, and 28 from the Côte-Nord group, and all the available seasons are supported. This group contains many more icing events than the other groups as well.

Finally, stations 3, 9, and 10 were eliminated from the current study because of infrequent recorded icing events, as well as the loss of some measuring seasons. Station 3 supports the seasons 1995-1996, 1996-1997, and 1997-1998, with a total of 26 accretion phases. For the stations 9 and 10, only the readings of season 1997-1998 were available. Because considering such stations would minimize the number of deduced groups for that season at a given time, more seasons are needed in order to obtain a better prediction.

The map in Figure 3-11 shows our four selected groups. Group 1 covers the region from Montreal to Quebec, and includes five stations: 4, 11, 14, 22, and 27. Group 2 covers the Montreal region, and includes stations 8 and 16. Group 3 covers the Quebec region, and consist of stations 5 and 17. Finally, Group 4 covers the Côte-Nord region, and includes 6 stations: 1, 27, 13, 24 and 28.



**Figure 3-11 The four groups of stations**

It has been observed that several stations of the SYGIVRE network consist of more than one icing measurement instrument. Table 3-9 shows all stations that contain multiple instruments, these are: 2, 3, 5, 11, 16, and 27. The distribution of storm events according to the instruments for all seasons is displayed so that the aggregation of storm events of all instruments in a station will lead to the total storm events for that station. It is clear that the

spatial proximity of instruments for a station lead to similarity in their events, since all their measurements describe the meteorological condition for the same location. Thus, the need for merging the storm events of each instrument for the same station is self-evident.

<b>Station</b>	<b>Instrument Number</b>	<b>1992 /1993</b>	<b>1993 /1994</b>	<b>1994 /1995</b>	<b>1995 /1996</b>	<b>1996 /1997</b>	<b>1997 /1998</b>
<b>2 (LAC ST-PIERRE)</b>	2	27	56	25	61	41	80
	3	54	4	19	0	0	0
	4	26	64	49	72	65	80
<b>3 (POSTE ARNAUD)</b>	5	5	0	0	0	12	4
	6	6	0	0	0	0	0
<b>5 (MONT BELAIR)</b>	8	0	0	0	0	0	8
	9	0	47	82	52	109	9
	10	0	0	0	0	99	10
	11	108	116	104	91	50	11
	12	45	0	115	0	0	12
<b>11 (DONNACONA)</b>	18	20	14	32	19	37	22
	19	24	13	27	0	0	0
<b>16 (LIGNE EXPERIMENTALE IREQ)</b>	24	0	0	0	7	22	35
	26	0	0	0	11	0	0
<b>27 (SOREL)</b>	36	11	9	10	29	15	25
	37	12	4	17	0	3	0

**Table 3-9 Stations with multiple instruments**

### 3.6 Data analysis

An exploratory analysis on the available icing events and icing storms in each group should be carried out to help in the model creation.

Only the data that belong to the storm event times are maintained for any group. We consider  $k$  stations for the group, and since the unit of measured observations is the hour,  $m$  is the total number of deduced observations related to the storm events for one station. The variable that represents the observations of  $k$  stations at a specific time is termed as the *number of points* which vary from 1 to  $m$ . The total number of observations for any group is  $k \times m$ . The observed values can be the ice accumulation weight, temperature, ice accretion or the icing event duration.

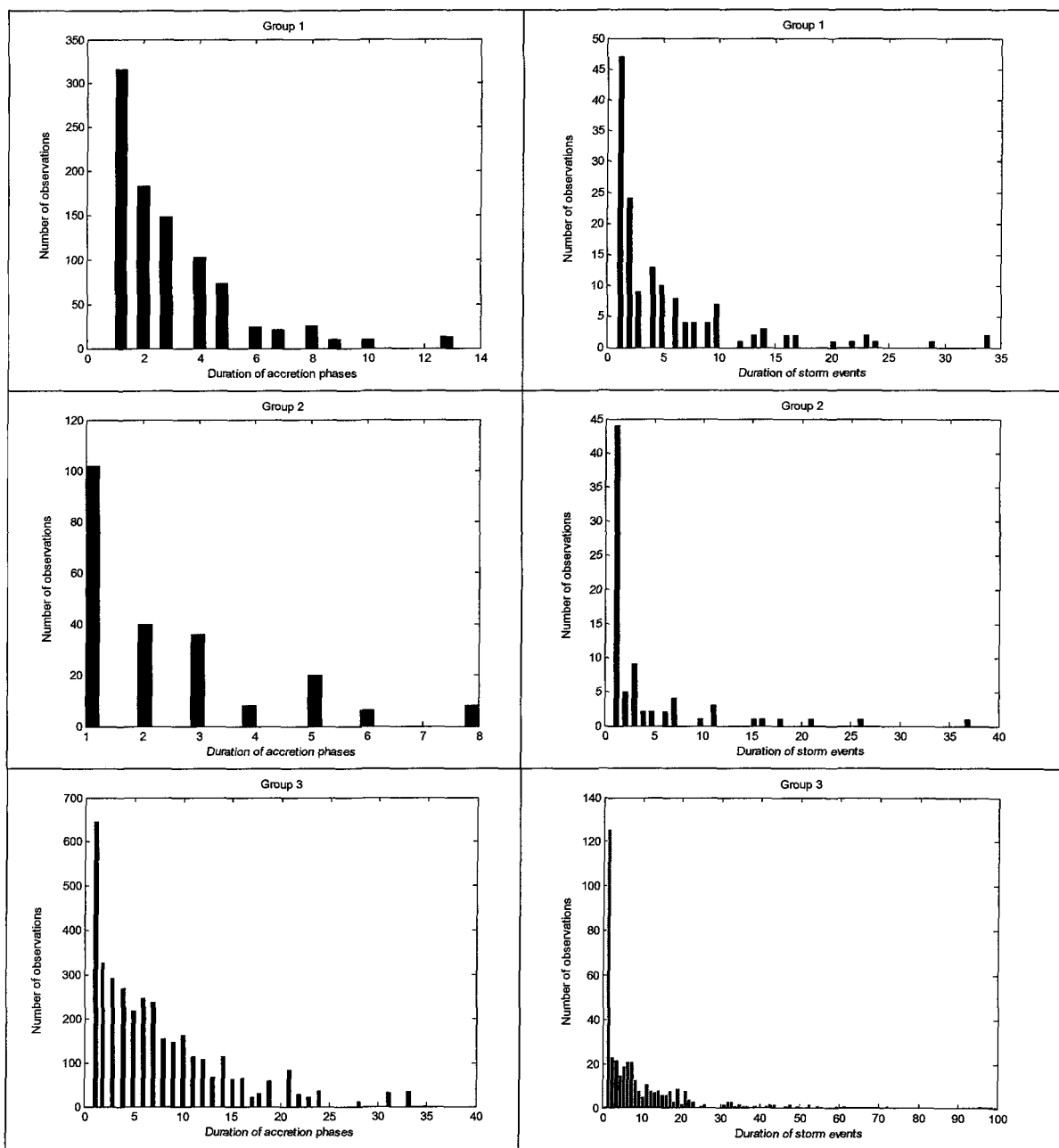
In fact, it has been perceived that three factors affect the maximum number of points for any group of stations: the number of included seasons, the total number of storm events, and the duration of storm events. The last two factors depend on the meteorological situation of the region covered by the specified group.

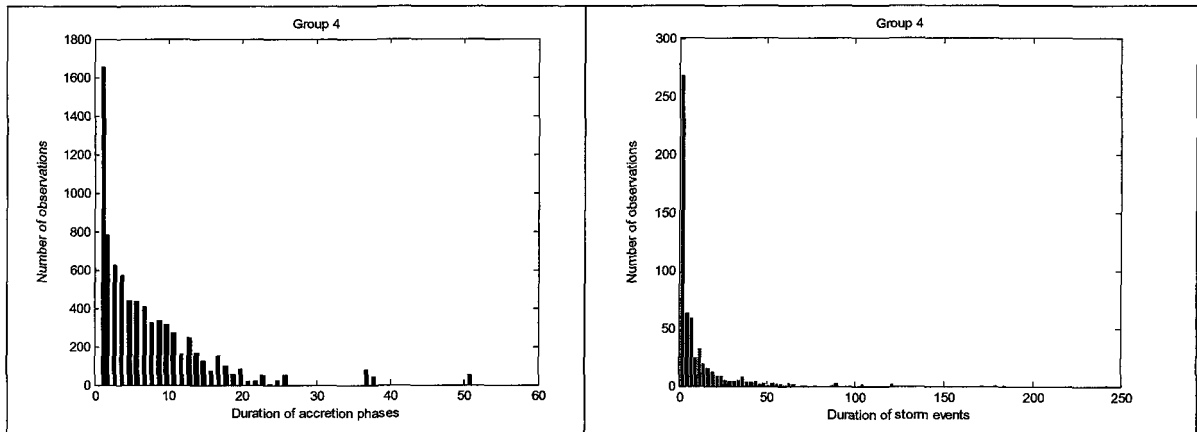
- Group 1 has a relatively small amount of data,  $m = 748$ ;
- Group 2 has little data,  $m = 419$ ;
- Group 3 has a big set of data,  $m = 3\,480$ ; and

- Group 4 has the biggest set of data,  $m = 6\,920$ .

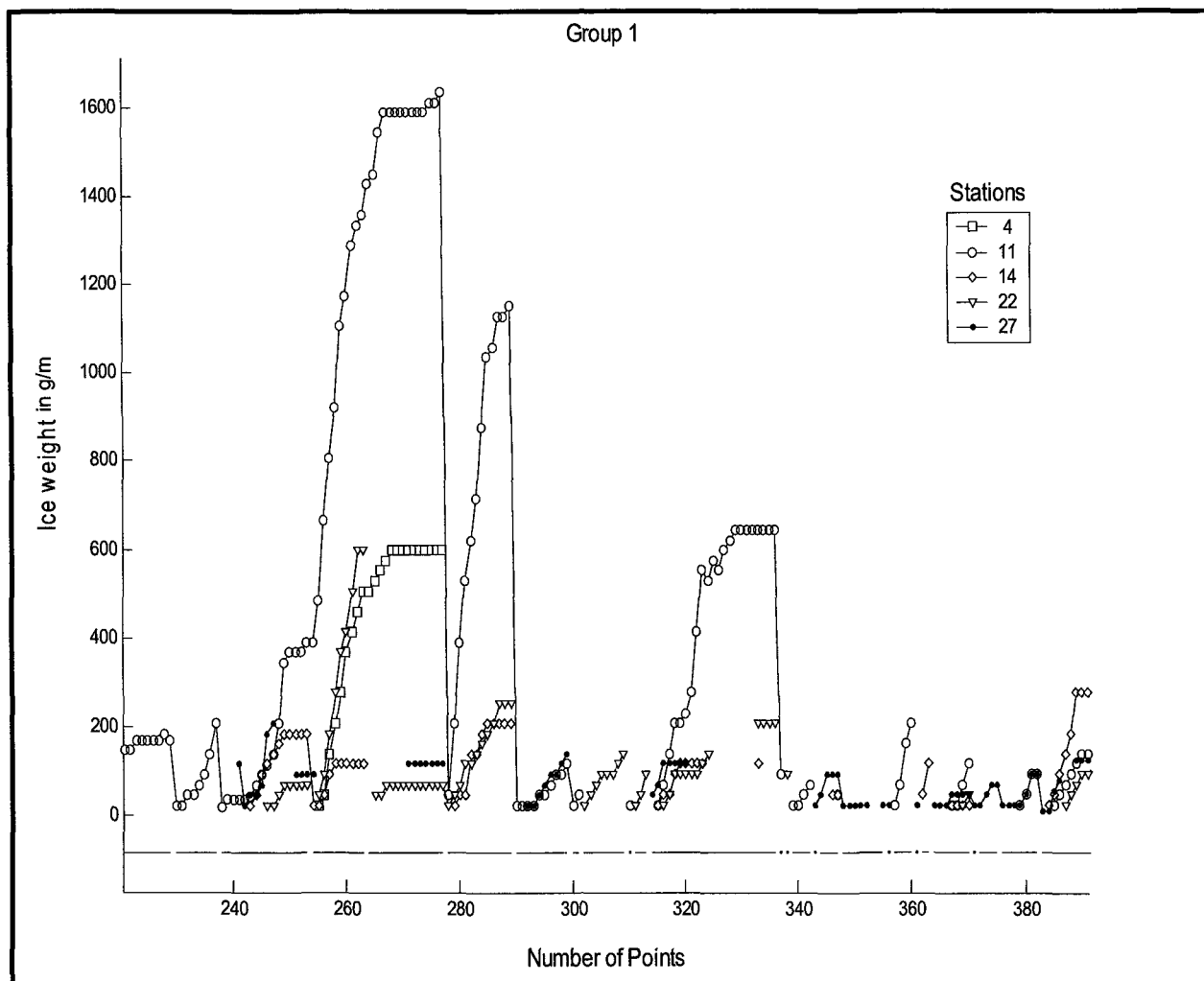
It is evident that as the geographical location of stations moves north from the equator, the events become more serious. Among the groups shown in Figure 3-11, the southernmost ones have fewer events than those in the north.

A general view of the ice accumulation weight values should be explored to help understand the amount of variation for the ice weight. It is interesting to look at the durations of icing events and storm events, since the duration of an event generally reflects its size. Figure 3-12 shows the histogram for the durations of icing events and storm events, where the data for all stations in a group were merged together. It could be noted that most of icing events and storm events have short duration. Figure 3-13, Figure 3-14, Figure 3-15, and Figure 3-16 show a sample of observations where the ice accumulation weight values is traced with respect to the number of point for all stations in the four groups. The discontinued lines below the graphics in the figures represent the durations of storm events.





**Figure 3-12 Histogram of icing events and storm events duration (h) for all stations of the four groups**



**Figure 3-13 Observations of ice accumulation weight for all stations of group 1**

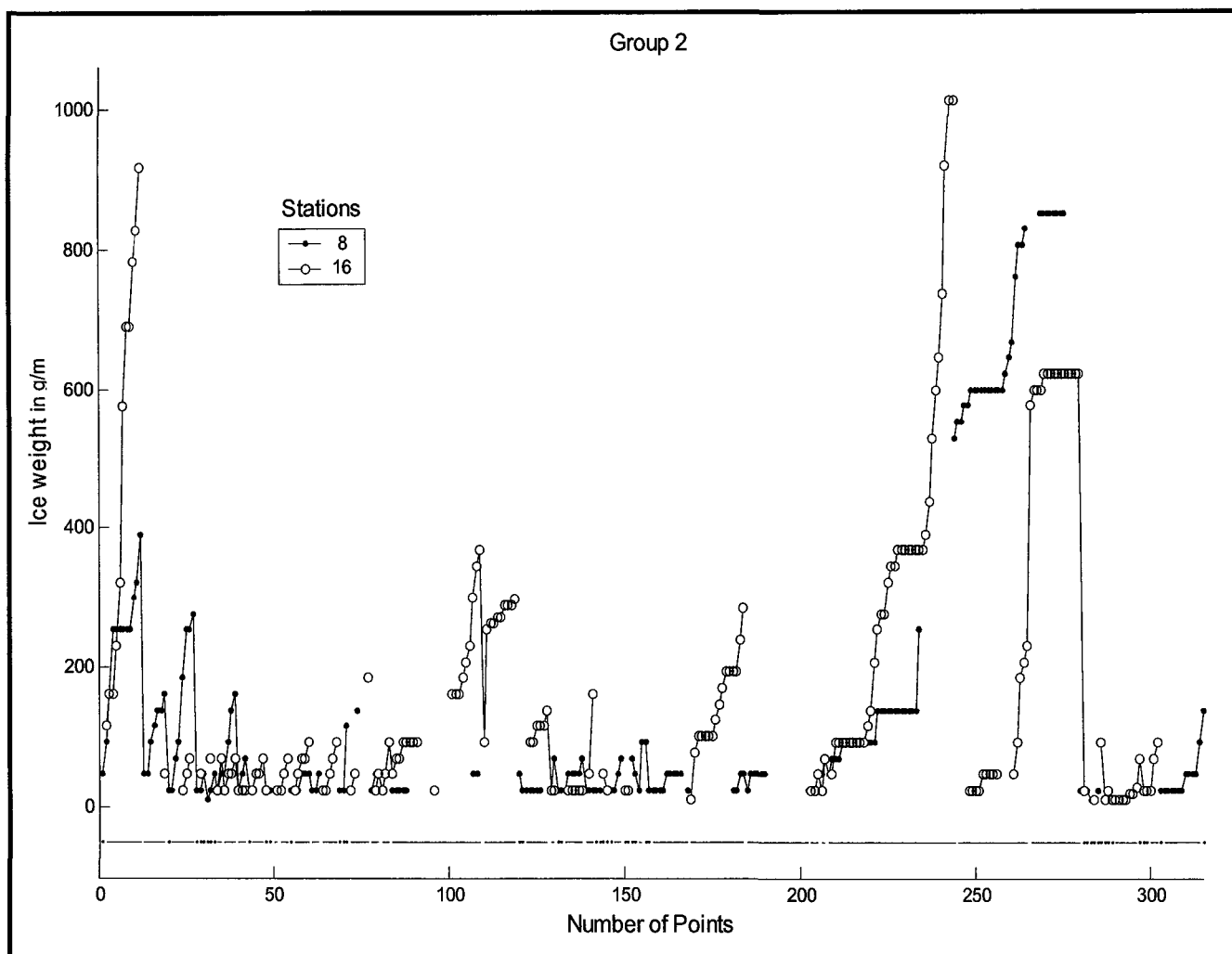


Figure 3-14 Observations of ice accumulation weight for all stations of group 2

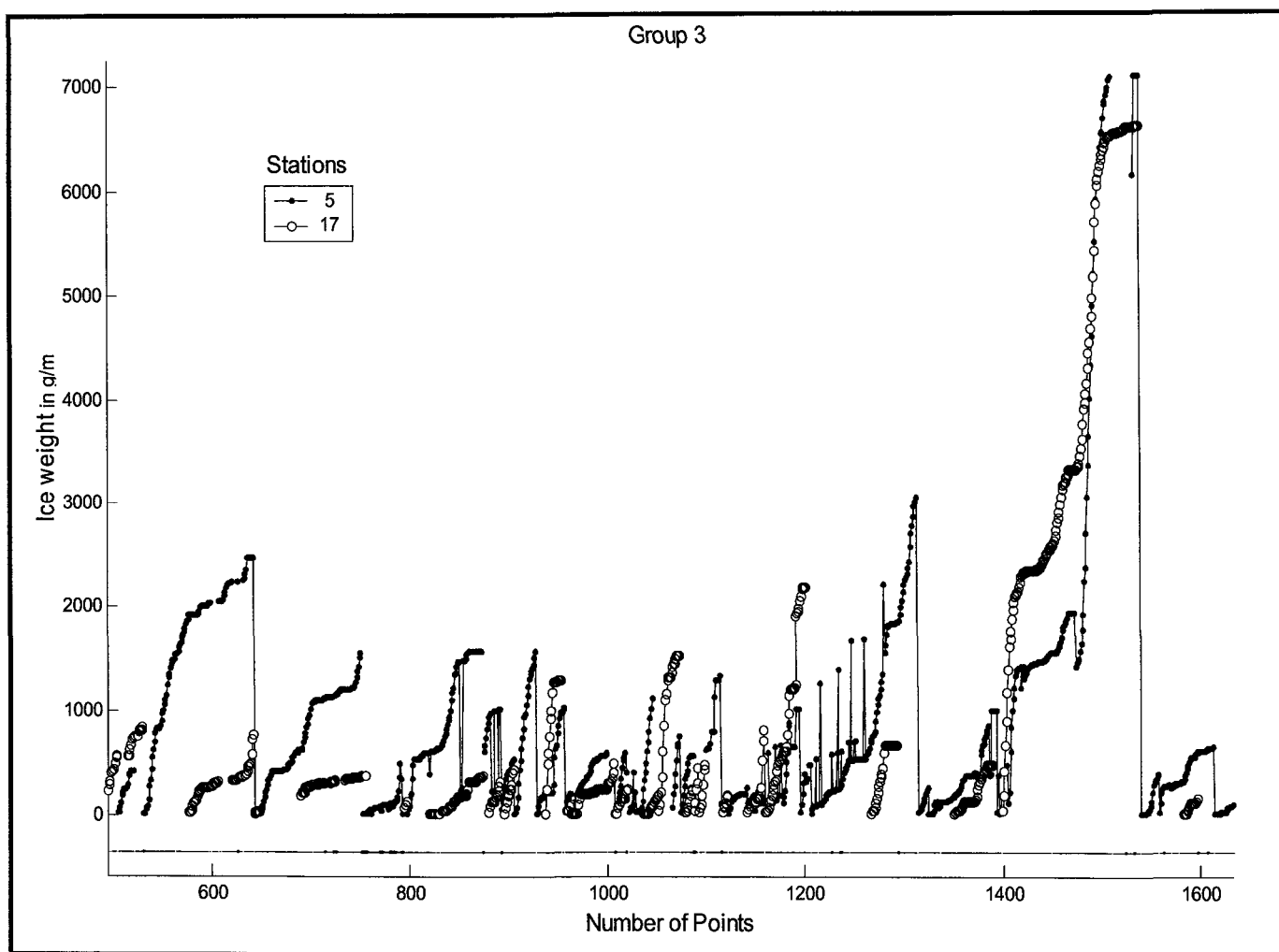


Figure 3-15 Observations of ice accumulation weight for all stations of group 3

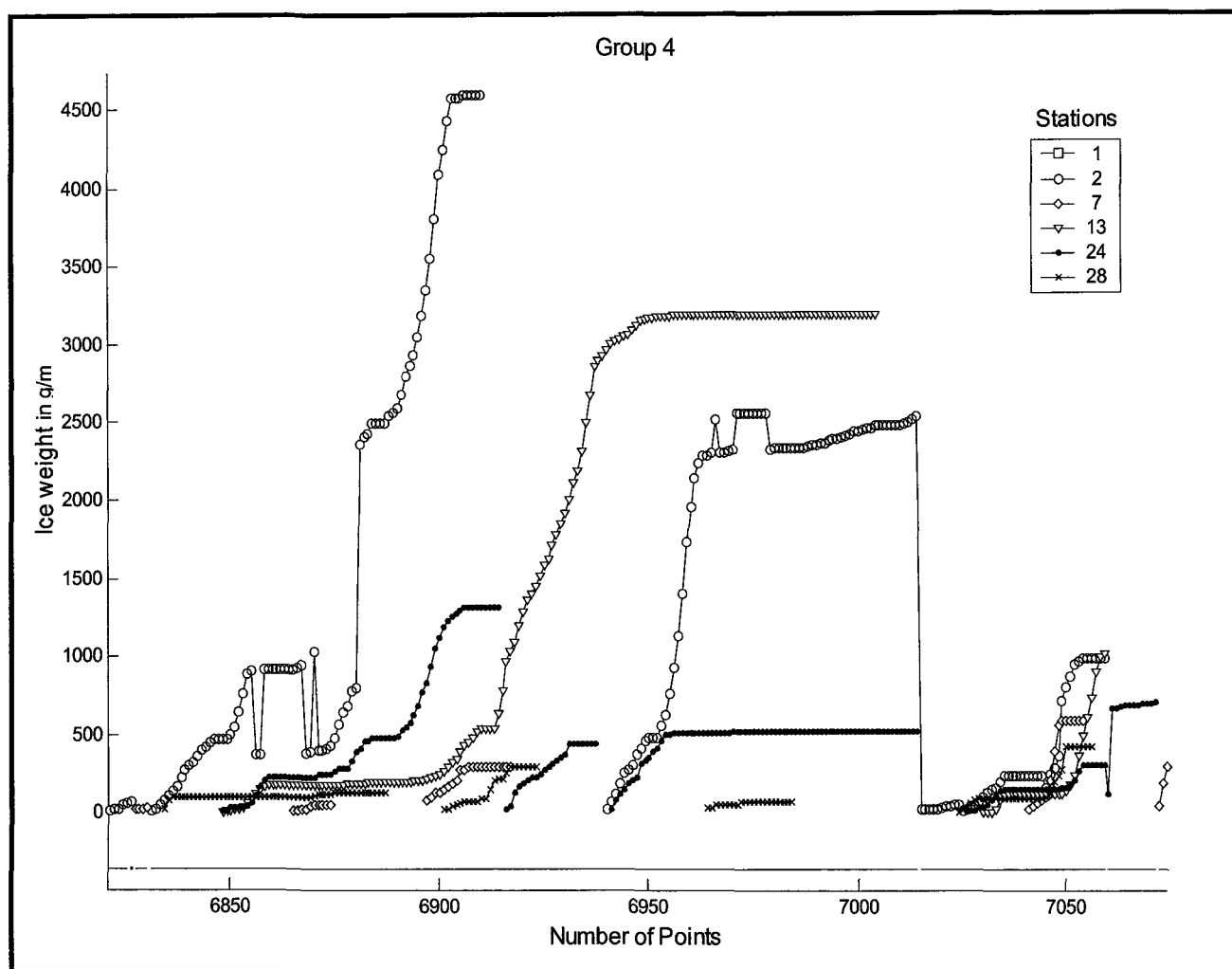


Figure 3-16 Observations of ice accumulation weight for all stations of group 4

## **CHAPTER 4**

### **ICING STORM PROPAGATION**

## **4.1 Introduction**

The data related to icing storms from the SYGIVRE database are very complex. Therefore, it is necessary to establish a prediction model, in order to track the evolution of a storm within a given region. A neural network based model is constructed using information on geographical position of stations. The model is designed to predict the icing storm evolution in real-time, where the prediction is conditioned to the fact that there is an icing event. The state of an icing storm at a specific time is predicted using the available information on the icing storm at previous times.

A preliminary study of the icing event in the SYGIVRE database has been carried out by Guesdon et al. [18], where the spatial relationship between stations was analyzed. A logistic regression model was used to determine the relative importance of meteorological variables, in order to predict the relation between stations. However, the model was not well adapted to the study of the storm evolution since it produces a high rate of false negative, which means that the model almost predicts that there is no icing storm when there is one. A neural network model is considered here to address the same problem.

In a preliminary model we are concerned with the spatial propagation of the icing storm by considering the state variable at each station (1 if the icing storm is affecting the station and 0 otherwise). Two models are proposed: the first is a complete binary network

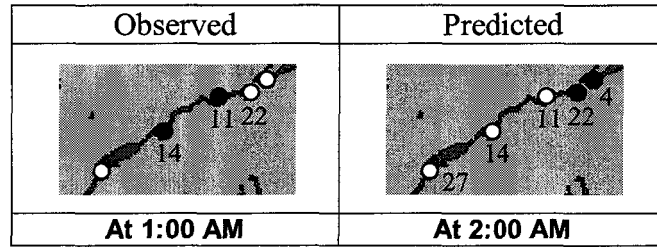
model (only binary data as input/output are used) while the second one integrates quantitative variables, mainly meteorological, to achieve better prediction.

The data used for building the model come from the historical information of the SYGIVRE database, which covers 6 seasons of measurements starting from 1992. The model has been set up for the four groups of stations defined in CHAPTER 3.

## 4.2 Variables

Binary information coming from the SYGIVRE database is used to create the prediction model. The variable to be predicted is the state of each station of a group at a specific time. For time  $t$ , the state variable for station  $i$ ,  $X_i(t)$ , takes the value 1 if there is an ice accumulation during this time, and 0 otherwise. For a group of  $k$  stations, the spatial modelling for the icing storm under consideration is equivalent to predicting the vector  $X(t) = (X_1(t), X_2(t), \dots, X_k(t))$ . The variables used as predictors are the same variables, but from previous times for the complete binary model,  $X(t-1)$ ,  $X(t-2)$ , etc. A more sophisticated model uses some meteorological variables to achieve a better prediction.

Figure 4-1 illustrates the observed state for the stations 4, 11, 14, 22 and 27 of the group 1 at 1:00 AM and the predicted at 2:00 AM. A dark point indicates that the station is affected by an icing storm. It is clear that a storm event has affected stations 14 and 11 at 1:00 AM and has continued to stations 4 and 22 over the next hour.



**Figure 4-1 Simple sample of prediction for the group 1**

The icing storm is described in terms of binary data, so it should be modelled based on this data. The neural network is implemented via a mathematical model in the following form:

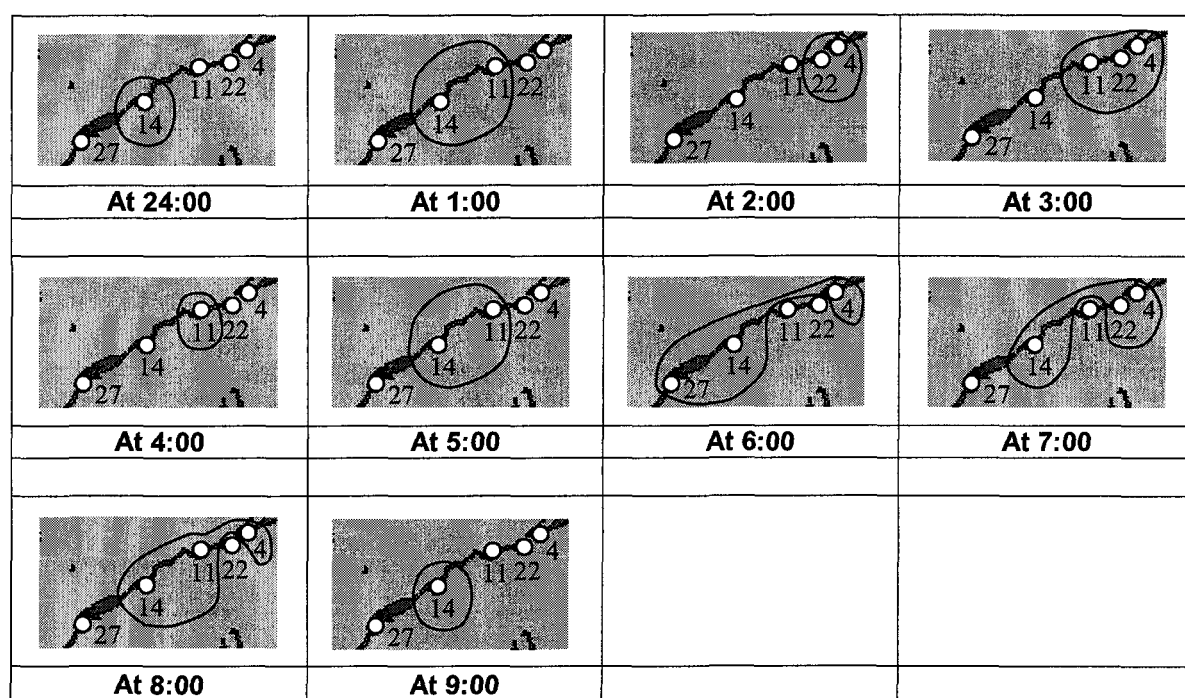
$$\begin{array}{ccc}
 \text{Predicted} & & \text{Observed} \\
 \begin{bmatrix} X_1(t+1) \\ X_2(t+1) \\ \vdots \\ X_k(t+1) \end{bmatrix} & = & F \begin{bmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_k(t) \end{bmatrix}
 \end{array} \quad \text{Eq. 4-1}$$

where  $X_i(t)$  is a binary random variable that represents the state of station  $i$  at the time  $t$ .

Table 4-1 illustrates the available data for the icing storm modelling at stations 4, 11, 14, 22, and 27 from 24:00 2/8/1996 to 9:00 2/9/1996. Figure 4-2 displays the spatial evolution of this icing storm.

Station	2/8/96 24:00	2/9/96 1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00
4	0	0	1	1	0	0	1	1	1	0
11	0	1	0	1	1	1	0	0	1	0
14	1	1	0	0	0	1	1	1	1	1
22	0	0	1	1	0	0	0	1	0	0
27	0	0	0	0	0	0	1	0	0	0

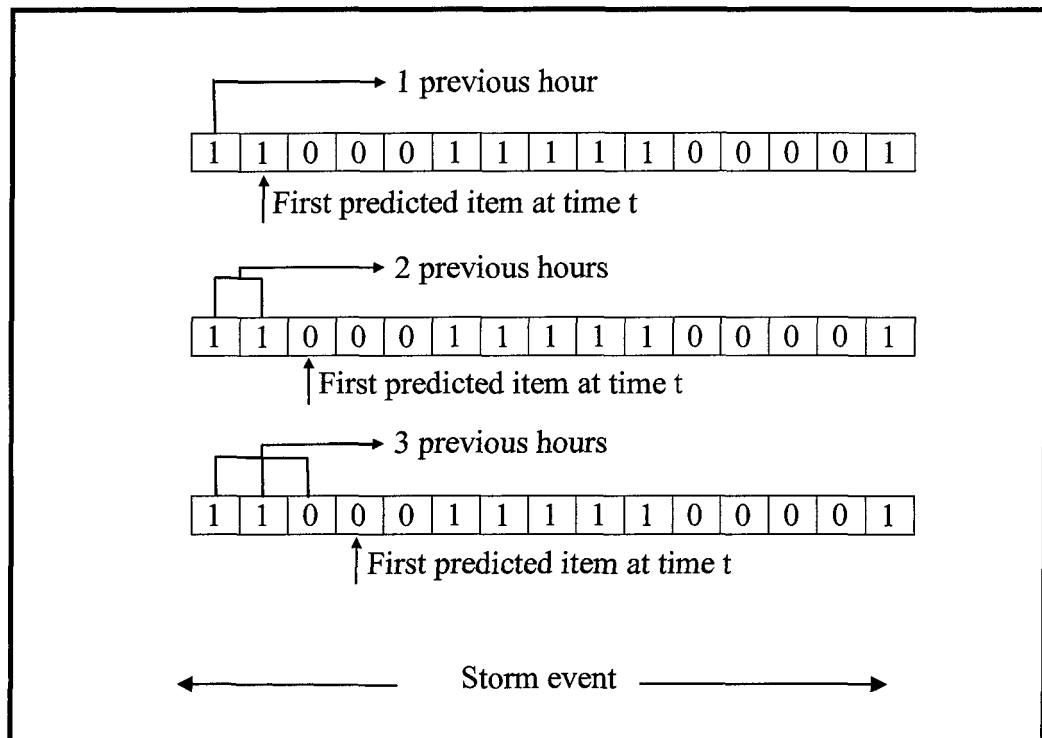
**Table 4-1 Sample storm event in 1996 at group 1**



**Figure 4-2 Spatial propagation between group 1's stations for a sample storm event**

The idea is to use the information of the previous hours for predicting the icing storm state at a specific time. A simple model predicts the icing storm using only the information available one hour before time  $t$  to predict, represented as  $X(t-1)$ , but it is possible to take into account more than one previous hours of information  $X(t-1)$ ,  $X(t-2)$ , etc, as shown in Figure 4-3, since it may give a better chance to make a

good prediction. This has been limited in the current study to a maximum of 6 previous hours, and can be extended in forthcoming studies, when adding more measurement years. It should be noted that, as the number of previous hours increases, there will be more data lost, but more data is also available for a predicted item.

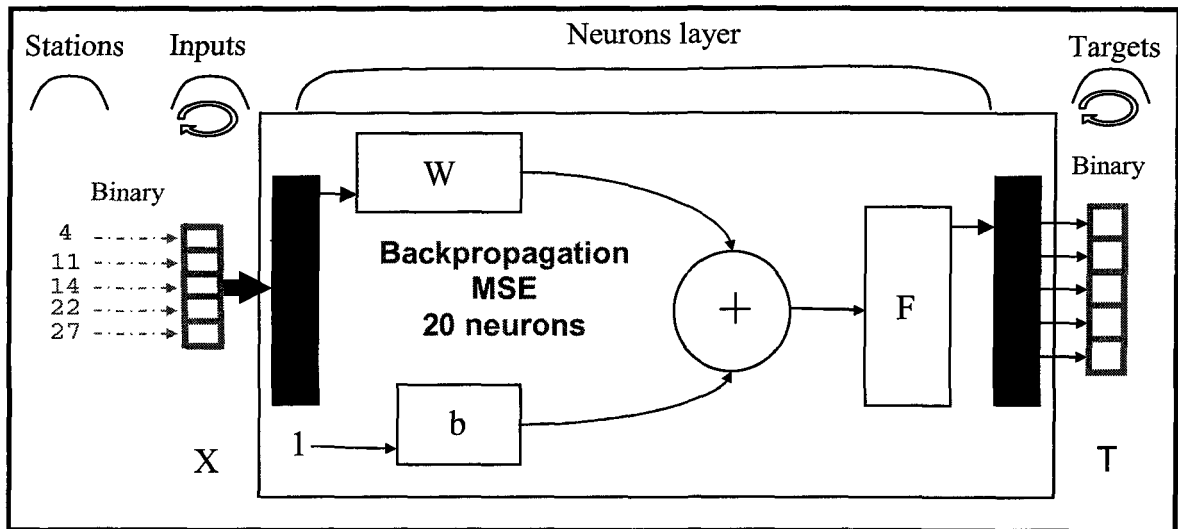


**Figure 4-3 General example for one, two and three previous hours for a given storm event at one station**

### 4.3 Binary network model

The binary network model is a neural network that supports just binary data. It accepts and produces binary data that represents event occurrence for multiple stations.

Figure 4-4 shows a neural network adapted to the binary model, where the data of Group 1 is involved. Each network entry and corresponding output represents a station from the group.



**Figure 4-4 Neural network sample applied to the group1 in the case of binary network model**

### 4.3.1 Performance function and validation

Two functions are considered in order to measure the performance of the binary model. The first one is the classic Mean Square Error (MSE), which computes the average squared error between the network outputs and the observed outputs. This function and its mathematical properties are well-known for its pertinence for continuous variables, but for binary variables it is not always appropriate. The other performance function included in the processing is called *Kappa statistics*. The Kappa was originally devised by Cohen [37]

for evaluating in their classifications inter-observer agreement on a nominal scale. If we consider a set of  $n$  binary observations, and hence  $n$  binary predictions, the set can be represented by a truth table like Table 4-2. At any point in the set, two binary items are available, the observed item that belongs to the network target, and the predicted one that belongs to the network output. There exist four possible states between observed and predicted items: 00, 01, 10, and 11. The number of repetitions for each state of the set of data is represented by:  $n_{00}$ ,  $n_{01}$ ,  $n_{10}$ , and  $n_{11}$ , respectively, where  $n_{11}$  imply the positive fitted observations,  $n_{00}$  the negative fitted observations,  $n_{01}$  the negative failed observations, and  $n_{10}$  the positive failed observations.

		Predicted		
		0	1	
Observed	0	$n_{00}$	$n_{01}$	$n_{0.}$
	1	$n_{10}$	$n_{11}$	$n_{1.}$
		$n_{.0}$	$n_{.1}$	$n$

**Table 4-2 Observed versus predicted items truth table**

The Kappa is defined as follows:

$$K_p = \frac{\frac{n_{00} + n_{11}}{n} - \sum_{i=0}^1 \frac{n_{i.} n_{.i}}{n^2}}{1 - \sum_{i=0}^1 \frac{n_{i.} n_{.i}}{n^2}} \quad \text{Eq. 4-2}$$

where,  $n_{.0} = n_{00} + n_{10}$ ,  $n_{0.} = n_{00} + n_{01}$ ,  $n_{.1} = n_{01} + n_{11}$  and  $n_{1.} = n_{10} + n_{11}$

The main diagonal represents the agreement between predictions and observations while the values outside diagonal are errors. The percentage of agreement can be misleading because it is always positive even without any predictive power. If all the values are at random in the table, the expected value for the diagonal is  $\sum_{i=0}^1 (n_{i.} n_{.i}) / n^2$ , so it is in agreement by chance. The Kappa is then the percentage of agreement corrected for agreement by chance.

The Kappa function is a good measure for evaluating the model but not a performance function in the neural network model. The Kappa is sensitive to small variations in the distribution of values in the truth table, leading to a bad criterion for a convergence function. On the other hand, the MSE is very stable for small variations in the truth table. We therefore consider the MSE as the performance function for the convergence criteria, and the Kappa for the validation criteria.

For the model validation, two validation types have been proposed, total validation and validation by separation. In total validation, the training data is used as is for

validation. In the case of validation by separation, the data is separated into two parts: the first is used for training (2/3), and the other for validation (1/3) [38].

Different criteria have been used to validate our model. One of them was explained previously: it is the goodness of fit Kappa criterion. The next one is called *failed predictions*. It expresses the failed predictions rate for a group of stations.

Other interest validation criteria are required, which concern the determination of efficiency measure of the model. They are: sensitivity, specificity, predictive positive value, and predictive negative value criteria. These criteria are defined based on Table 4-2 as follow:

- The sensitivity indicates the capacity of the model of predicting an icing storm when it is indeed the case. It can be evaluated as the probability of predicted succeeding 1's given the observed 1's :

$$\text{SENS} = \text{Pr}(\text{positive predicted} \mid \text{positive observed}) = n_{11} / (n_{11} + n_{10})$$

- The specificity is a measure of the model behaviour in terms of false alert when nothing is going on. It can be evaluated as the probability of predicted succeeding 0's given the observed 0's:

$$\text{SPEC} = \text{Pr}(\text{negative predicted} \mid \text{negative observed}) = n_{00} / (n_{00} + n_{01})$$

- The predictive positive value is a measure of the value of the prediction when the model is effectively predicting an icing storm. It can be evaluated as the probability of observed succeeding 1's given the predicted 1's:

$$PPV = \Pr(\text{positive observed} \mid \text{positive predicted}) = n_{11} / (n_{11} + n_{01})$$

- The predictive negative value is a measure of the confidence one can have for a prediction of no risk. It can be evaluated as the probability of observed succeeding 0's given the predicted 0's:

$$NPV = \Pr(\text{negative observed} \mid \text{negative predicted}) = n_{00} / (n_{00} + n_{10})$$

An optimum efficiency measure can be obtained while approaching to 1.

A technical interpretation of this is that for a small sensitivity most icing storms can not be detected by the system, so at most times the personnel will not be alerted of the danger. For a small specificity, each time the system provokes an erroneous alert about the existence of an icing storm, it cost the personnel an initiation in order to avoid disaster. As long as the system indicates good weather, the risk of disaster is large, this is the effect of a small predictive positive value, which is different from the effect of predictive negative value, where as long as there is bad weather the chance of improving is large, therefore, the predictive negative value is not annoying.

### 4.3.2 Model setting

The current model has been carried out based on the MSE performance function. The training function, *trainoss*, is used to train our network, since it provides much more convergence than other training functions existing in MATLAB, it can also be applied to the two proposed neural network methods of backpropagation and Elman.

In most trials, the Elman network gives similar results as the backpropagation network. However, the results are a little bit better when using the Elman network in the case of total validation. The Kappa values for the 4 groups (1, 2, 3, 4) in the case of 6 previous hours are (0.94, 0.76, 0.73, 0.74) when the Elman network is used, and (0.83, 0.77, 0.64, 0.66) when the backpropagation network is used. In contrast, the prediction results are more effective when using the backpropagation network in the case of validation by separation, whereas the Kappa values for 1 previous hour are (0.4, 0.13, 0.66, 0.66) versus (0.38, 0.13, 0.66, 0.63) for the Elman network. Therefore, our model only uses the backpropagation method, since the validation by separation is the most vital test for the model.

As the number of neurons increases, the network tends to overestimate the data, so the prediction based on total validation enhances. It has been observed that when performing the training with 50 neurons, an improvement of the prediction takes place and

starts progressing slowly. The Kappa values for group 1 in the case of 1 previous hour for different neurons (20, 30, 40, 50, 60, 80 neurons) are (0.511, 0.51, 0.509, 0.514, 0.513, 0.516), and for group 4 they are (0.651, 0.651, 0.652, 0.652, 0.652, 0.652). In the case of validation by separation, the optimum prediction can be obtained when using a number of neurons lying between 20 and 30. The obtained Kappa results for group 1 are (0.643, 0.632, 0.631, 0.629, 0.629, 0.62), and (0.412, 0.395, 0.397, 0.386, 0.389, 0.386) for group 4. Simply, the involved number of neurons was chosen to be 20 neurons.

### **Validation**

The following tests have been carried out based on the two validation types, total validation and validation by separation for all groups.

The results of Kappa and failed predictions based on total validation type for the four groups are given in Table 4-3. Only the case of six previous hours is shown, because it is the best result obtained for the total validation.

<b>Group</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>Kappa</b>	0.83694	0.77282	0.64543	0.66554
<b>Failed predictions</b>	0.05204	0.05952	0.17719	0.09904

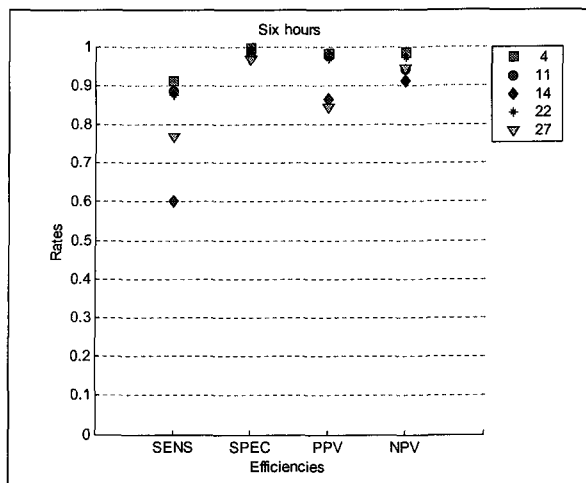
**Table 4-3 Total validation criteria in the case of six previous hours for binary network model**

The results of Kappa and failed predictions based on the validation by separation type for the four groups are given in Table 4-4, at one to six previous hours.

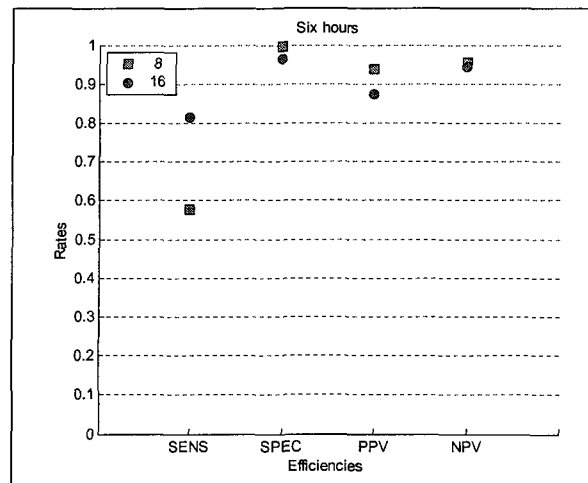
Group		Previous hours					
		1	2	3	4	5	6
1	Kappa	0.4074	0.3373	0.2720	0.2641	0.2697	0.2654
	Failed predictions	0.1582	0.1650	0.1882	0.2024	0.1959	0.1908
2	Kappa	0.1332	0.3268	0.2141	0.1904	0.2337	0.1792
	Failed predictions	0.1786	0.1706	0.2217	0.2286	0.2660	0.2440
3	Kappa	0.6606	0.6578	0.6550	0.6400	0.6272	0.6287
	Failed predictions	0.1685	0.1699	0.1717	0.1791	0.1859	0.1849
4	Kappa	0.6670	0.6694	0.6812	0.6796	0.6775	0.6645
	Failed predictions	0.1050	0.1029	0.1047	0.1068	0.1075	0.1130

**Table 4-4 Validation by separation criteria for binary network model**

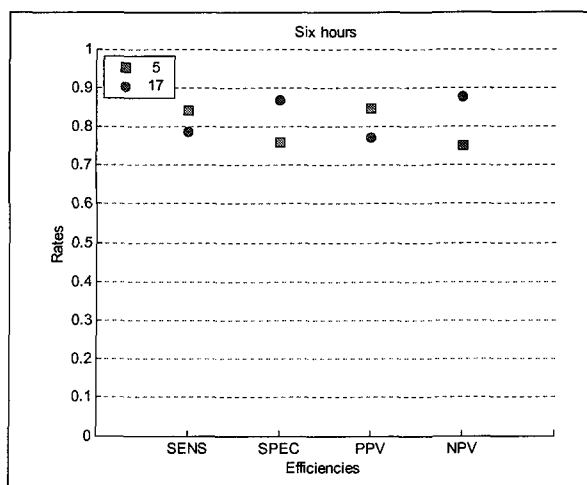
The results of efficiency measures on the stations of each group, at 6 previous hours, are presented in the Figure 4-5, Figure 4-6, Figure 4-7, and Figure 4-8, in the case of total validation. These show the case of six previous hours.



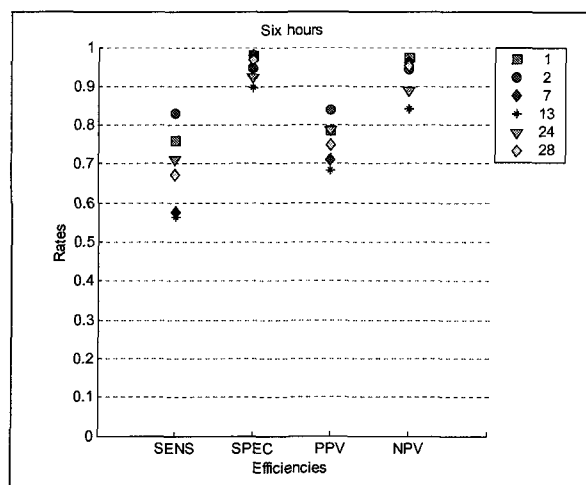
**Figure 4-5 Efficiency measure for group 1 in the case of total validation (binary model)**



**Figure 4-6 Efficiency measure for group 2 in the case of total validation (binary model)**



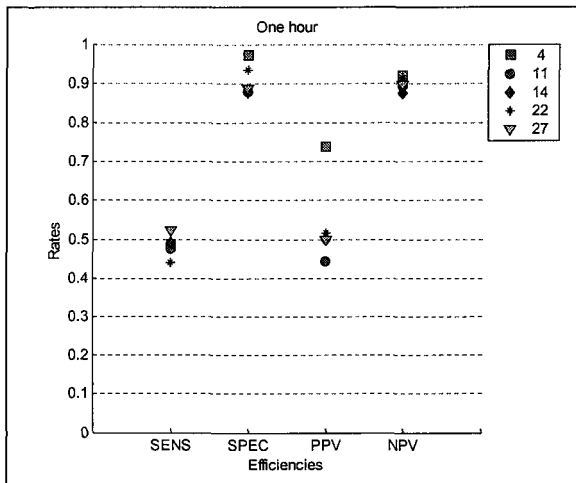
**Figure 4-7 Efficiency measure for group 3 in the case of total validation (binary model)**



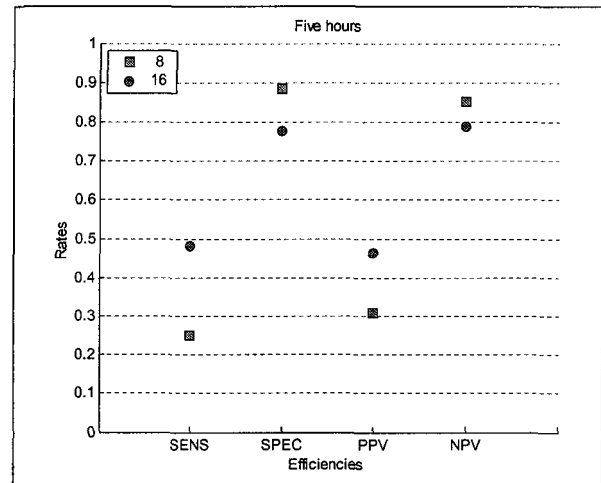
**Figure 4-8 Efficiency measure for group 4 in the case of total validation (binary model)**

Figure 4-9, Figure 4-10, Figure 4-11, and Figure 4-12 display the efficiency measure in the case of validation by separation for the stations of each group, where only the better cases in terms of previous hours have been chosen. This can be distinguished

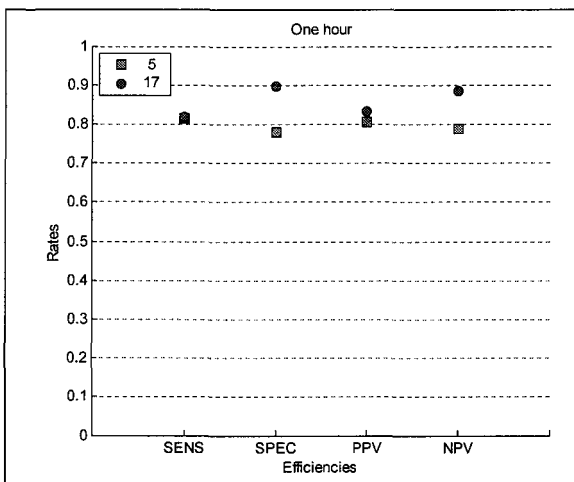
based on the sensitivity, specificity, predictive positive value and predictive negative value, which should tend to be unity.



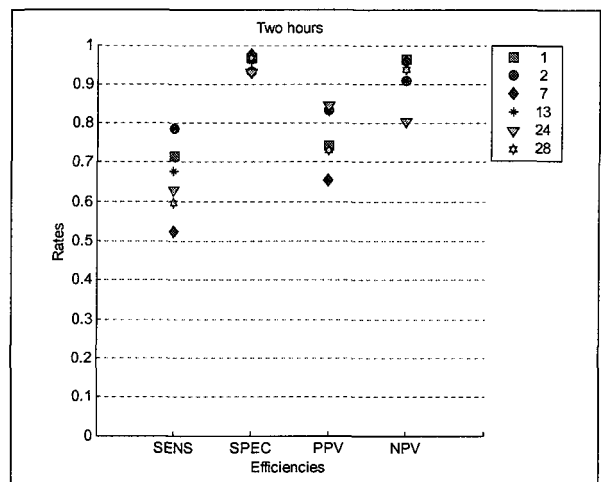
**Figure 4-9 Efficiency measure for group 1 in the case of validation by separation (binary model)**



**Figure 4-10 Efficiency measure for group 2 in the case of validation by separation (binary model)**



**Figure 4-11 Efficiency measure for group 3 in the case of validation by separation (binary model)**



**Figure 4-12 Efficiency measure for group 4 in the case of validation by separation (binary model)**

The number of previous hours for the validation by separation model has been chosen depending on the values of the validation criteria in Table 4-4 and the values of

efficiency measures. For Group 1, the case of one previous hour provides the best results, where a maximum value of Kappa, a minimum value of failed predictions, and a better efficiency measure were obtained. Group 2 has a maximum value of Kappa, and a minimum value of failed predictions at two previous hours, but it has a better efficiency measure as well as good Kappa and failed predictions at five previous hours. Group 3 has the same behaviour as Group 1. Finally, Group 4 has a better efficiency measure as well as good Kappa and failed predictions at two previous hours.

### **Model behaviour for extreme accumulation**

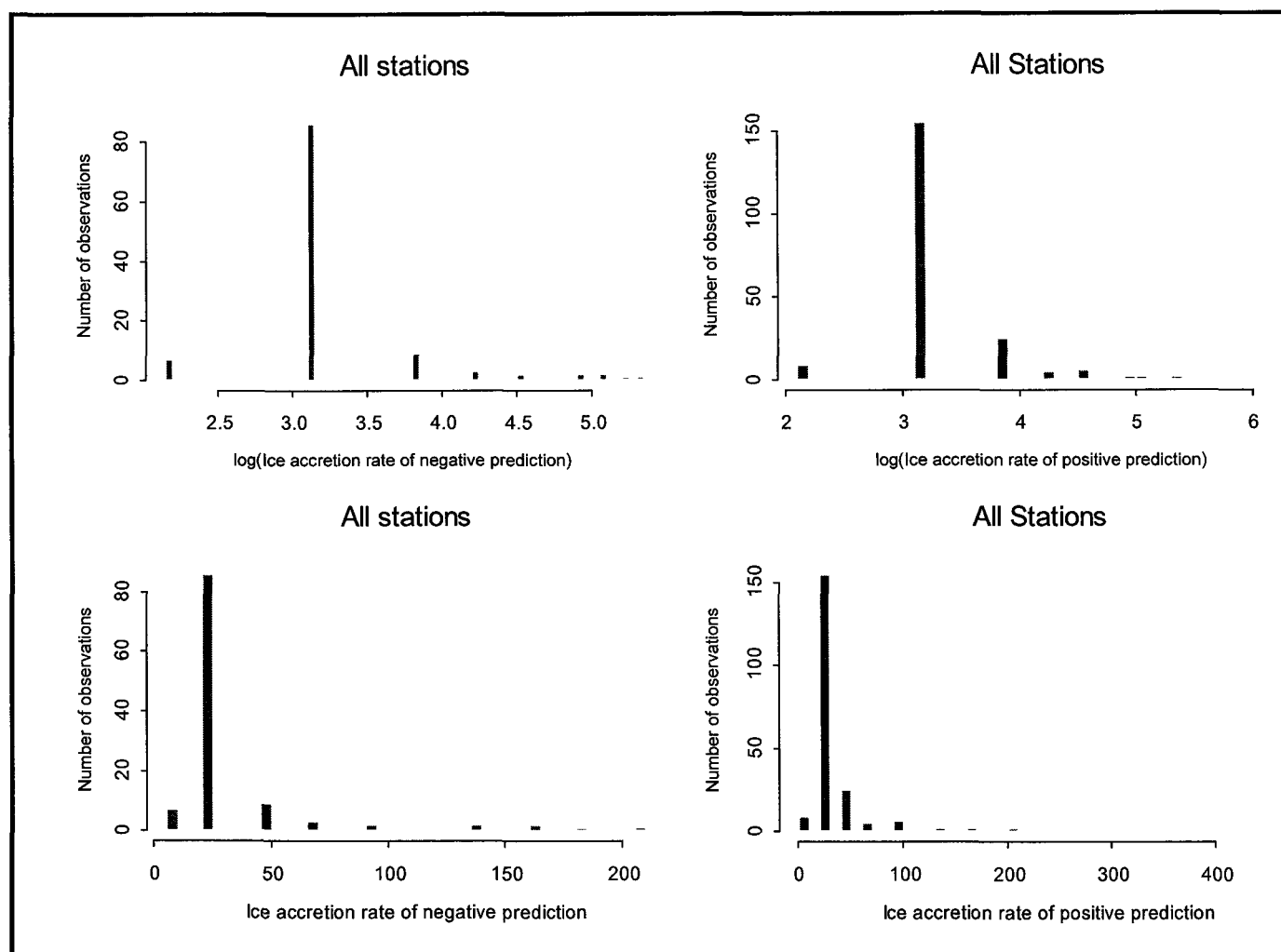
It is important to look at extreme ice accumulations in order to determine if these can be predicted correctly. For a given data set in a group of stations, some predicted points from the set have more serious accumulations than others, so the ice accretion rate helps distinguishing between them. Only the positive observed points (when there was accumulation) are concerned in the analysis. Our main concern is the negative predicted points, in order to explore the frequency of errors for the extreme icing accretion rate points. At the same time, it is important to compare the extreme icing accretion rate values for negative and positive predicted points.

Considering that the prediction results of the extreme accretion rate values resides in a range starting from  $60 \text{ g/m}$ , or 4 in logarithmic scale, for all groups. Table 4-5 shows the rate of extreme ice accretion rate to the total positive values, and similarly for the negative

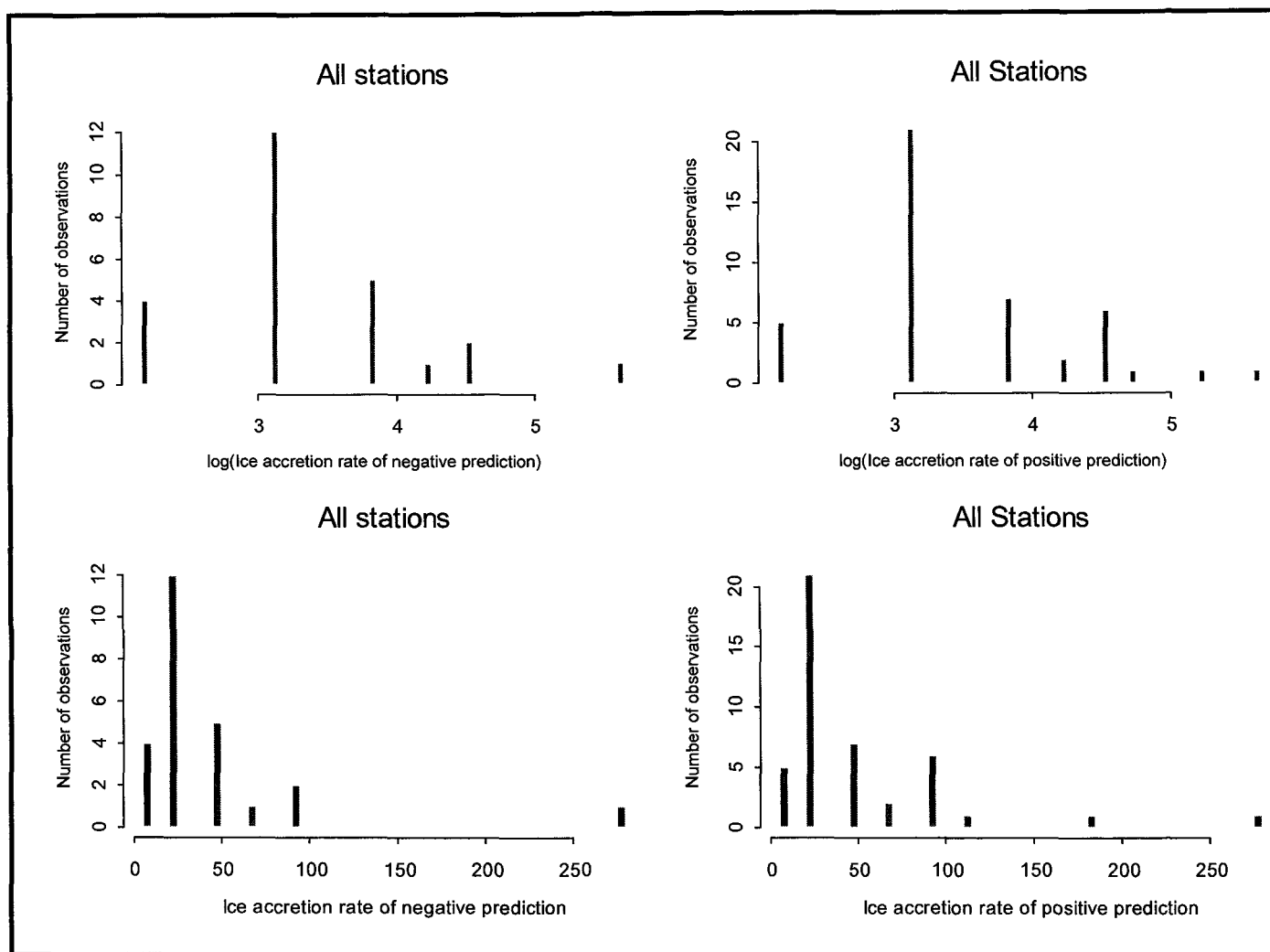
values for all groups. In conclusion, the rate for the positive predicted points is relatively moderate with respect to that negatively predicted for Group 1, but it is higher for the remaining groups. Figure 4-13, Figure 4-14, Figure 4-15, and Figure 4-16 show histograms of the ice accretion rate values for the positive and negative predicted points in the validation by separation model for the four groups. Note that the data related to the stations for each group were merged together.

<b>Group</b>	<b>Negative prediction (%)</b>	<b>Positive prediction (%)</b>
<b>1</b>	7.53	6.98
<b>2</b>	13.51	31.82
<b>3</b>	7.50	24.73
<b>4</b>	4.60	19.26

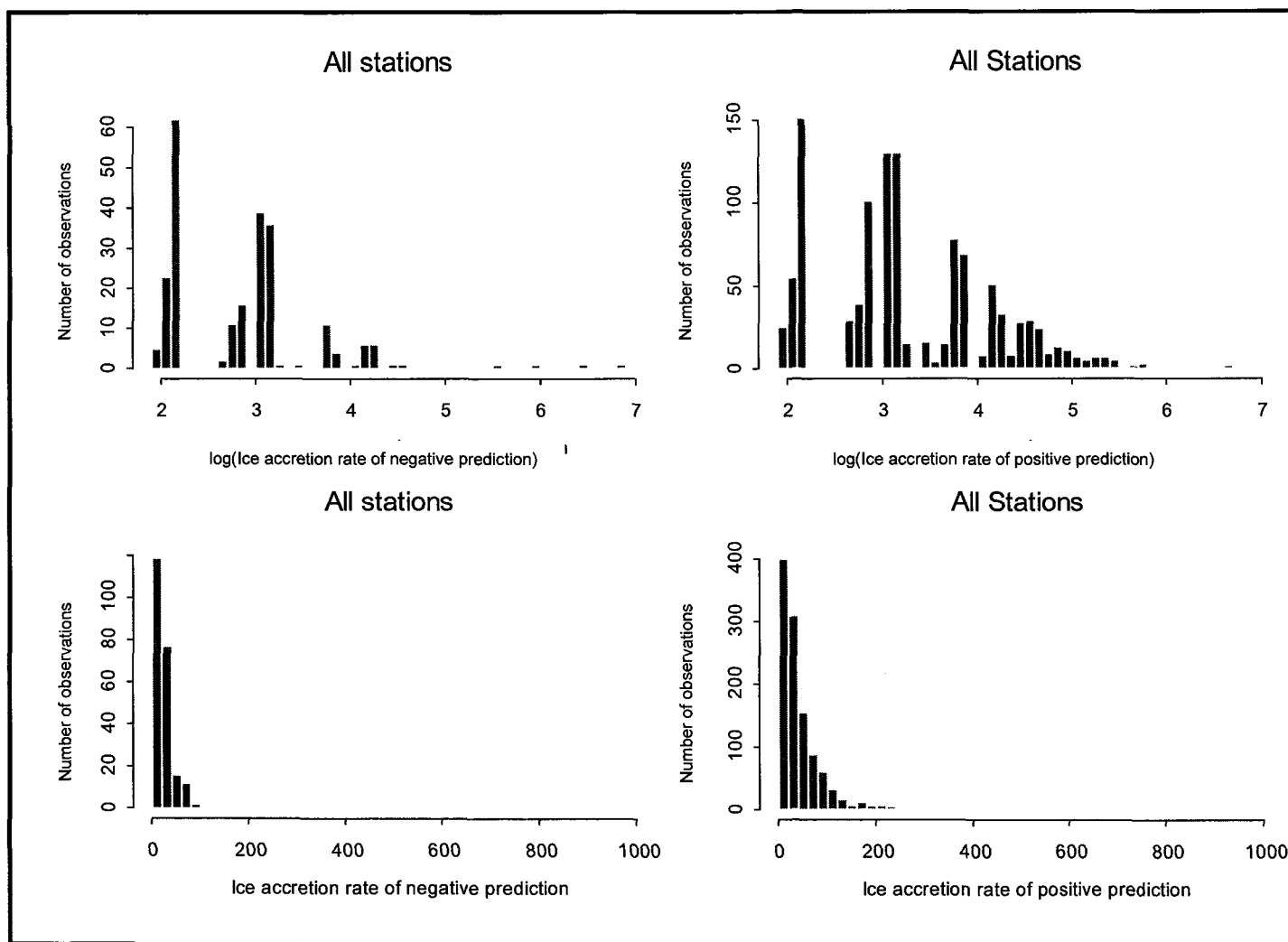
**Table 4-5 Rate of the extreme ice accretion rate values for the positive and negative predictions**



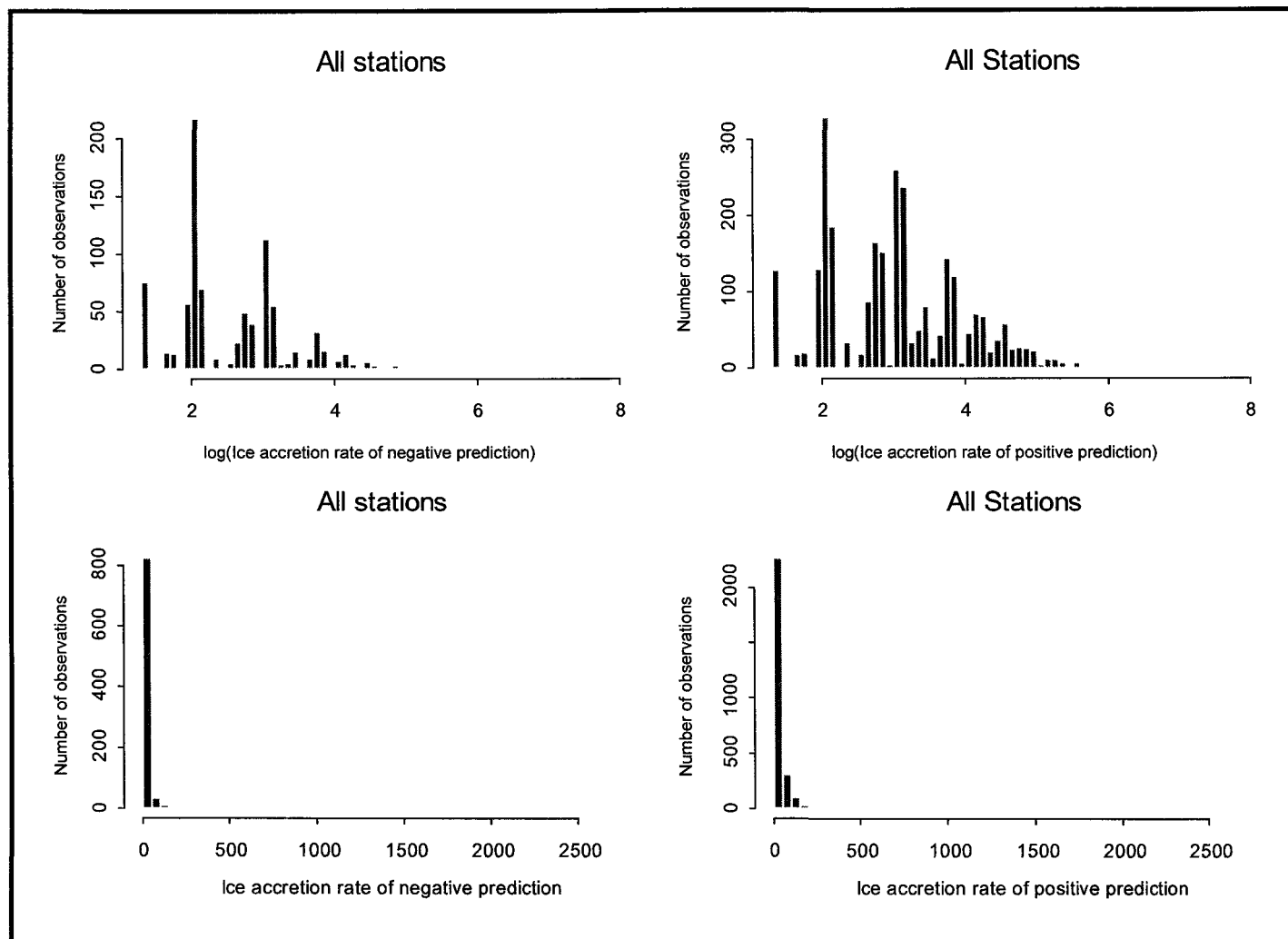
**Figure 4-13 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 1**



**Figure 4-14 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 2**



**Figure 4-15 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 3**



**Figure 4-16 Histograms of the ice accretion rate (g/m/h) of negative and positive predictions for all stations of group 4**

### 4.3.3 Discussion

The purpose of this section is to evaluate the neural network performance for the four groups of stations. For this purpose, a binary model has been established. In fact, our neural network model requires more data in order to enhance its performance.

The spatial propagation of an icing storm does not depend only on the previous state of the station as demonstrated by our model: external factors are important contributors to the behaviour of the icing storm. It is clear from Figure 4-13, Figure 4-14, Figure 4-15, and Figure 4-16 that the model is unable to predict the extreme events better than the insignificant ones. Based on the results of validation, the model tends to provide good prediction, but it encounters difficulties in predicting the extreme events. Referring to Table 4-5, the rate of extreme events positively predicted is almost larger than the rate of extreme events negatively predicted.

It has been observed that, in the case of total validation, the results become better if the number of previous hours taken into account increases, and Group 3 has given a high rate of failed predictions in comparison to other groups.

Probably, the model is not adequate enough for groups 1 and 2 in the case of validation by separation, as demonstrated in Table 4-4, Figure 4-9, and Figure 4-10. In contrast, they perform well in the case of total validation. This is likely due to the loss of data occurring in these groups, since with validation by separation the data diminish. It is not the same case for groups 3 and 4, which support enough data to lead to a better model. This can be verified by Table 4-4, Figure 4-11, and Figure 4-12.

The enhancement of the model in terms of previous hours differs from one group to another. In the case of validation by separation (Table 4-4), Group 1 starts with the best results at one previous hour, and so results continue to display successive degradation until reaching a minimum value at 6 previous hours. The results of Group 2 are greatly random. This is generally because of the data loss. Group 3 behaves like Group 1, and Group 4 also provides random results, but they are almost similar. Generally, as the number of previous hours increase, degradation of results start taking place.

It should be noted that the total validation model has given the best results when applied six previous hours, but this has not been the case for the validation by separation model. For the most part, more data is available for an item, so the chance of data overestimation increases based on the number of previous hours, while the case is not the same for validation by separation, where the model can not make overestimates since the validation data is different from that used for training.

The same model can be carried out by considering only the observations that have extreme ice accumulation, whereas the data applied to the network should represent only those observations. In such case, the importance of the model is remarkable, since the major danger is due to extreme icing events. Unfortunately, the dearth of available data makes it impossible to achieve such a model.

#### 4.4 Binary network model with supplementary variables

In order to improve the prediction of icing storm propagation, we consider the binary model with the addition of meteorological variables. Let  $Y(t)$  be the meteorological variables. The model is then:

$$\begin{array}{c} \text{Predicted} \\ \left[ \begin{array}{c} X_1(t+1) \\ X_2(t+1) \\ \vdots \\ X_k(t+1) \end{array} \right] \end{array} = F \begin{array}{c} \text{Observed} \\ \left( \left[ \begin{array}{c} X_1(t) \\ X_2(t) \\ \vdots \\ X_k(t) \end{array} \right], \left[ \begin{array}{c} Y_1^1(t) \\ Y_2^1(t) \\ \vdots \\ Y_k^1(t) \end{array} \right], \left[ \begin{array}{c} Y_1^2(t) \\ Y_2^2(t) \\ \vdots \\ Y_k^2(t) \end{array} \right], \dots, \left[ \begin{array}{c} Y_1^v(t) \\ Y_2^v(t) \\ \vdots \\ Y_k^v(t) \end{array} \right] \right) \end{array} \quad \text{Eq. 4-3}$$

where  $X_i(t)$  is a binary random variable that represents the state of station  $i$  at time  $t$ , and  $Y_i^j(t)$  represents the  $j^{\text{th}}$  quantitative variables at station  $i$ .

#### **4.4.1 Model characterization**

The current model characteristics are similar to those of the binary network model, except that the model input includes non binary values, although it aims to predict binary data. It is convenient to compare this model with the binary network model in order to observe changes in the new model.

The supplementary variables that could be used for the model are: total ice accumulation weight, air temperature, ice accretion rate, and accretion phase duration. In order to make the problem simpler, the best model performance should be achieved with a minimum number of variables as inputs. It should be noted that the temperature variable leads to loss of observations because of missing temperature readings in the database. It is evident that prediction is less accurate when the number of observations decreases.

The performance function (MSE) was included for the network training. The Kappa was used as a criterion for network validation. In addition, two validation types were involved. In the first one, all data is used for training the network and to validate it (total validation). In the second one (validation by separation), data is separated into two parts, one for training (2/3 of data) and the other one for validation (1/3 of data).

The binary data are our main concern. The same validation criteria as with the binary model were used: the Kappa and the rate of failed predictions. Sensitivity, specificity, predictive positive value, and predictive negative value, which represent the efficiency measure of the model, are used as well.

While it is the binary model that has been built, the backpropagation method was used to train the neural network basing on the training function *trainoss*. It seemed that the same function could be adapted to the new model, due to the similarity between the two models, and therefore it has been kept for the processing.

#### **4.4.2 Results**

Many alternatives exist for selecting the input variables of the network. Several of the trials performed have shown that the use of multiple supplementary variables in input to the network do not enhance its performance. Simply, each variable is applied to the network separately.

A suitable number of network neurons should be determined in order to obtain an optimum model. The number of neurons used in this model differs from the binary model, since more data has been added. For total validation, the number of neurons was chosen to be 50, so there is no need to go further. The Kappa values for Group 1 in the case of 1 previous hour for several training trials with a different number of neurons

(20, 30, 40, 50, 60, 80 neurons) are (0.414, 0.421, 0.42, 0.423, 0.424, 0.424). For Group 4, they are (0.715, 0.714, 0.714, 0.715, 0.716, 0.716). In the case of validation by separation, the number of neurons is determined such as groups 1 and 2 use neural networks with 30 neurons, given that their data is relatively not much abundant. The obtained Kappa results for Group 1 when employing only the temperature variable for training are (0.33, 0.342, 0.309, 0.293, 0.261, 0.258). Groups 3 and 4 use networks with 40 neurons. The Kappa values for Group 4 are (0.543, 0.532, 0.571, 0.498, 0.54, 0.48).

Each variable were applied to the network along with the binary data, and the results were analyzed. It was found that a better performance of the model can be obtained if only the duration variable was included in the network training, where each hour is represented by the duration of the current icing event. The Kappa values for Group 1 in the case of 1 previous hour when each variable (total ice weight, temperature, ice accretion rate, and duration) is used separately for training are (0.26, 0.61, 0.51, 0.64) where the total validation is applied. For Group 4 they are (0.31, 0.61, 0.56, 0.7).

Table 4-6 and Table 4-7 show the validation criteria Kappa along with failed predictions for total validation and validation by separation for all groups. The results indicate that the use of six previous hours for total validation provides better values of Kappa and for failed predictions as obtained in the case of binary model. The results of validation by separation show similarity to those obtained in the binary model.

Group	1	2	3	4
Kappa	0.88639	0.6694	0.73726	0.70674
Failed predictions	0.038926	0.15217	0.13137	0.082355

**Table 4-6 Total validation criteria when using six previous hours and duration variable as network input**

Group		Previous hours					
		1	2	3	4	5	6
1	Kappa	0.3706	0.3505	0.3084	0.2455	0.2527	0.2487
	Failed predictions	0.1931	0.2059	0.2319	0.2585	0.2577	0.2759
2	Kappa	0.3196	0.2441	0.1901	0.1978	0.1103	0.1214
	Failed predictions	0.2553	0.2976	0.3312	0.3732	0.4167	0.4180
3	Kappa	0.6658	0.6722	0.6253	0.6540	0.6318	0.6535
	Failed predictions	0.1667	0.1639	0.1873	0.1731	0.1841	0.1732
4	Kappa	0.7160	0.7039	0.7075	0.6596	0.6745	0.6535
	Failed predictions	0.0833	0.0888	0.0908	0.1069	0.1018	0.1102

**Table 4-7 Validation by separation criteria of binary network model with supplementary variables**

Figure 4-17, Figure 4-18, Figure 4-19 and Figure 4-20 illustrate for the total validation test the four predefined efficiency measure, from 1 to 6 previous hours, when all stations in each group are included.

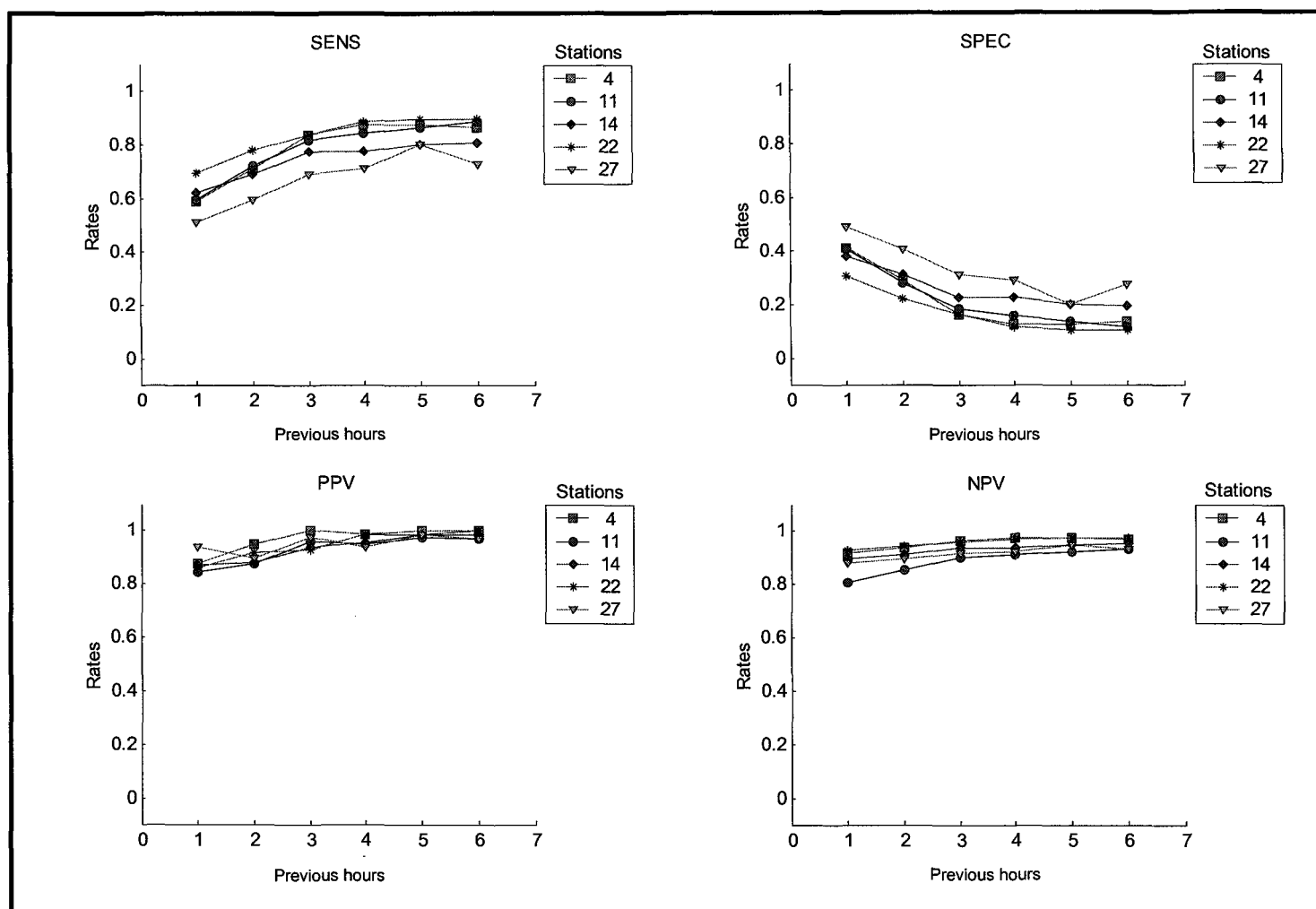


Figure 4-17 Efficiency measure for group 1 in the case of total validation for 1 to 6 previous hours

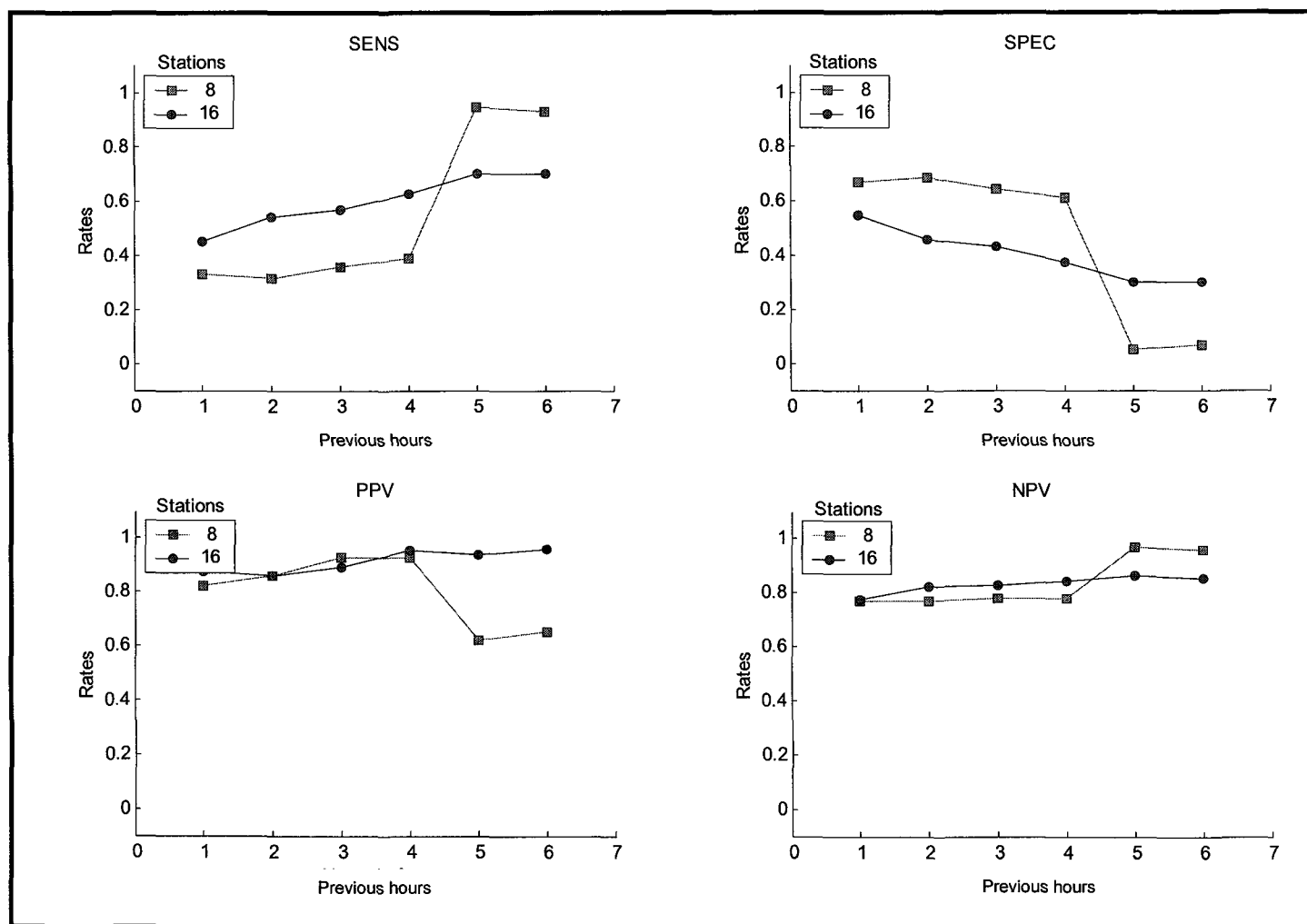


Figure 4-18 Efficiency measure for group 2 in the case of total validation for 1 to 6 previous hours

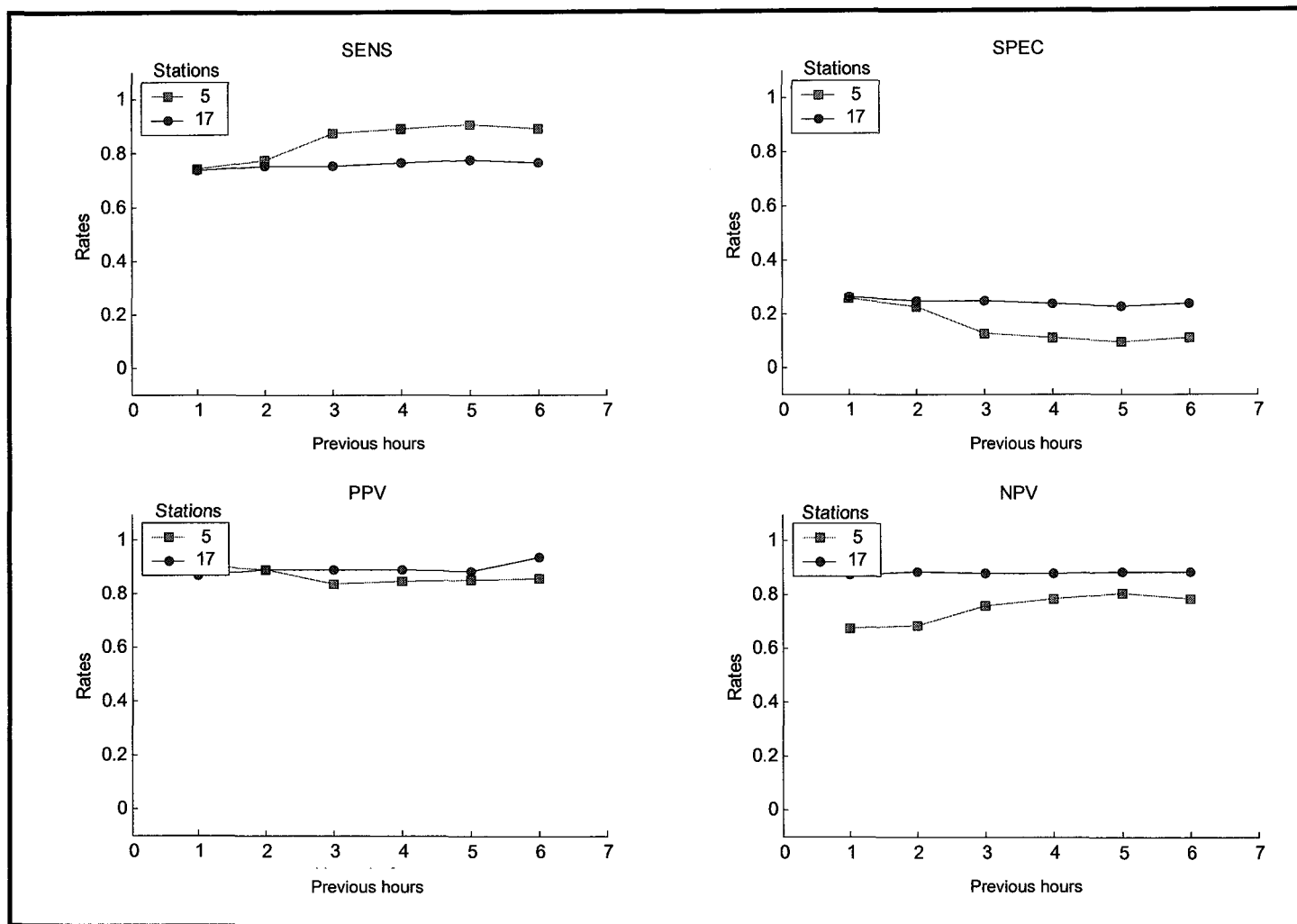


Figure 4-19 Efficiency measure for group 3 in the case of total validation for 1 to 6 previous hours

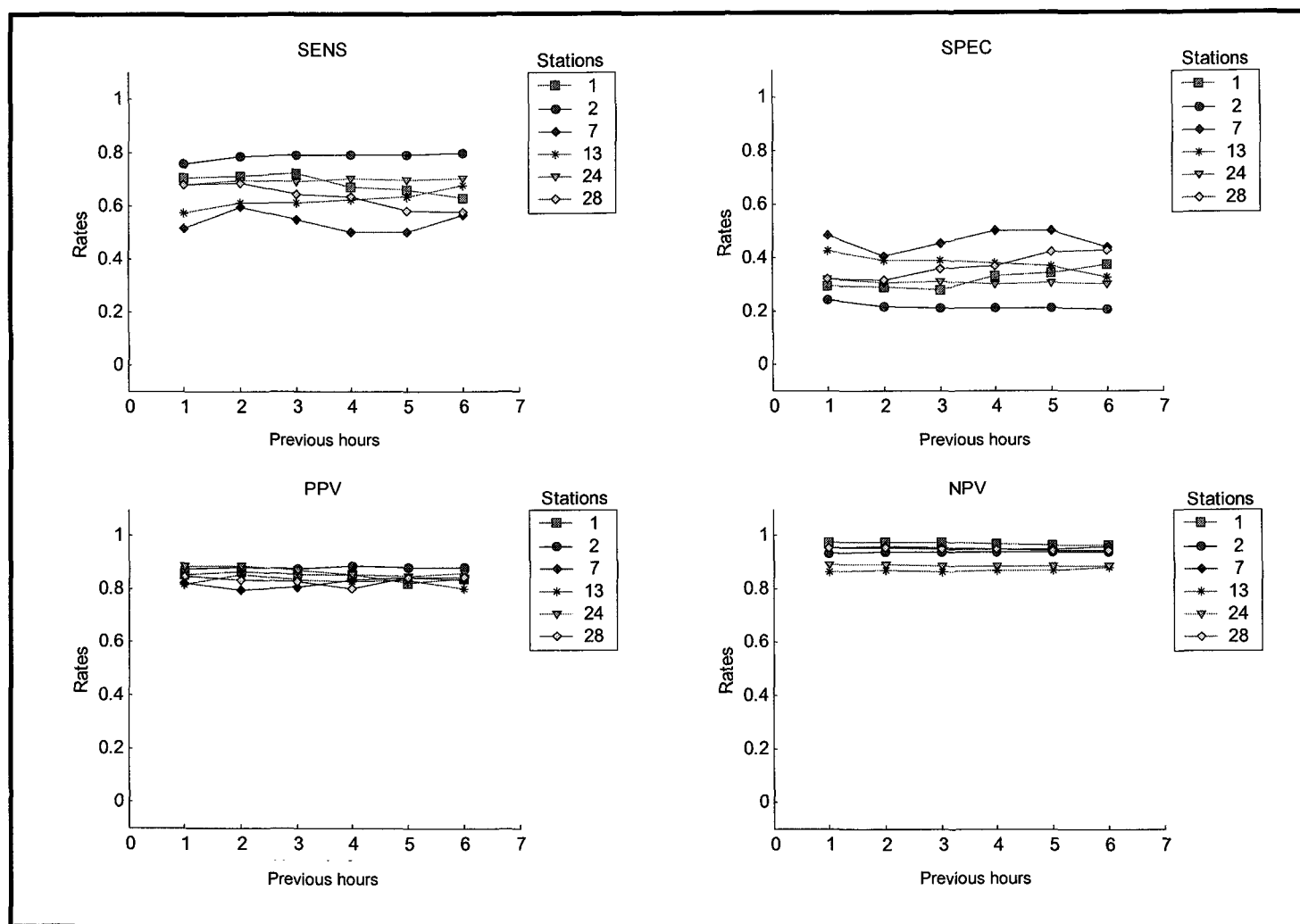
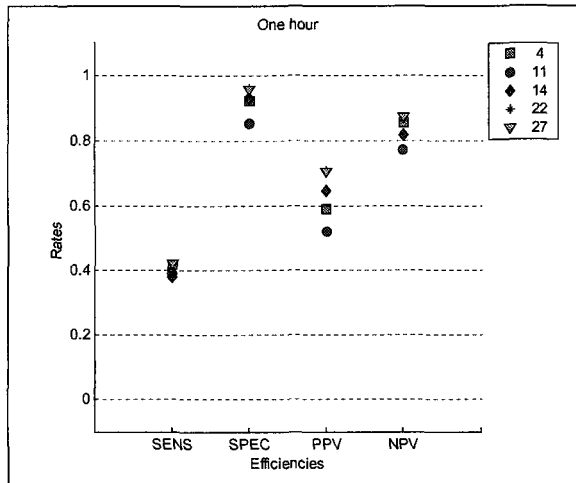
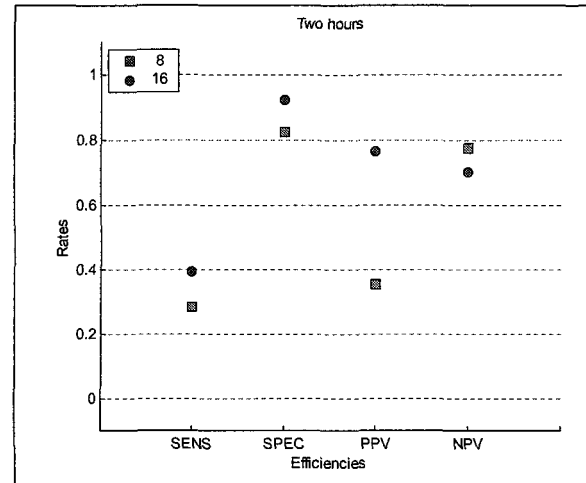


Figure 4-20 Efficiency measure for group 4 in the case of total validation for 1 to 6 previous hours

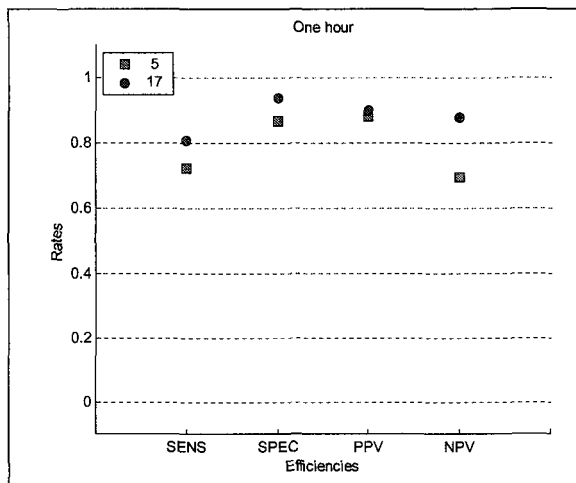
Figure 4-21, Figure 4-22, and Figure 4-23 display the efficiency measure for the four groups in the case of validation by separation. As has been done with the binary model, only the better cases in terms of previous hours have been chosen.



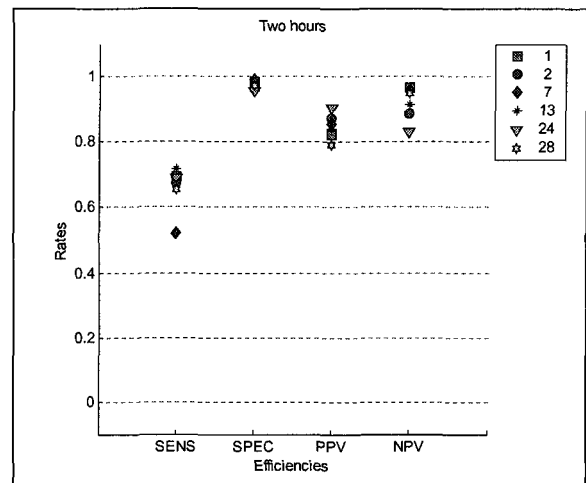
**Figure 4-21 Efficiency measure for group 1 in the case of validation by separation**



**Figure 4-22 Efficiency measure for group 2 in the case of validation by separation**



**Figure 4- Efficiency measure for group 3 in the case of validation by separation**



**Figure 4-23 Efficiency measure for group 4 in the case of validation by separation**

The previous hours in the case of the validation by separation model have been chosen depending on the values of the validation criteria in Table 4-7 and all the computed efficiency measures. Only the previous hours of Group 2 were changed from 5 in the binary model to 2 previous hours in the current model.

#### **4.4.3 Discussion**

The validation by separation model has provided a small improvement in groups 2, 3, and 4, since the values of Kappa and failed predictions are better than those of the binary model. It seems that the graphics of efficiency measure do not show great changes by comparison to those of the binary model, even though it is generally more effective.

In total validation, result enhancement is obvious except in Group 2, where it is less effective. It is worth noting that the specificity in Figure 4-17, Figure 4-18, Figure 4-19, and Figure 4-20 degrades as the number of previous hours increase, especially for Groups 1 and 2. This is likely due to the dearth of data for these groups. It is also worth noting that this was not the case with the binary model, where the specificity was usual.

In most cases, the use of a duration variable provides better results than those obtained in the case of the binary model. This has been recognized after performing several trials, where the training process is repeated. It should be noted that when network training is repeated, result variation is still negligible.

## **4.5 Conclusions**

A preliminary binary network model has been implemented in order to predict icing storm evolution for the predefined four groups based on the spatial relationship between the stations of those groups. A neural network approach has been employed for creating prediction models considering only the occurrence of icing storms in a binary format (events exist or not) at each station of a group at a specific time. These models predict the icing storm state at a specific time, where 1 to 6 previous hours were considered to perform the prediction. The model performs prediction for just one step of time in the future (one hour ahead).

The total validation model works efficiently for the four groups. For validation by separation, a very good model for groups 3 and 4 has been obtained. Group 1 has given a good prediction, but with relatively large errors in sensitivity and predictive positive value. The results of Group 2 were unsatisfactory. Generally, the increase of data may enhance the prediction performance for the groups.

The intervention of the duration variable has enhanced model performance. However, the presence of this variable in the model can not be applied in practice since it is known that the event duration data can not be ready at the real time period before the

passage of the entire event. Therefore, the binary model is the reliable alternative for prediction in real-time.

It has been observed that the neural network strategy generally gives an important reduction of false negatives over the logistic regression model, thus leading to a more appropriate model for the prediction of icing events.

## **CHAPTER 5**

### **PREDICTION OF ICE ACCUMULATION**

## 5.1 Introduction

In CHAPTER 4, a model has been defined in order to study the icing storm evolution in terms of whether it occurs or not in stations of a given region. Occurrence or not means that the icing event is represented only by a logical variable indicating the storm state at a station. In this chapter we are concerned with the severity of the icing storm at each station as represented by the ice accumulation weight.

Like the binary model, the model of ice weight is based on a neural network, the prediction is subject to the fact that there is an icing storm, and the prediction is obtained using previous observations at each station of a group. The major difference is that the predicted variable, *ice weight*, is continuous.

The value of ice weight on the overhead line conductors at a specific time during an icing storm is predicted using the available information at previous times, and the redirection is done for different times ahead. In this model, several quantitative meteorological variables are included as input. The relative importance of these variables should be distinguished in order to optimize the model.

## **5.2 Available variables**

The model that predicts the icing storm evolution could be established based on meteorological variables: total ice accumulation weight, air temperature, ice accretion rate, icing event duration. The values of these exist at every hour during an icing storm. The icing storm occurrence as represented by binary data is included in the study.

Based on the IRM signal, the ice accumulation weight and the ice accumulation rate variables can be estimated. These variables are available for every hour from the SYGIVRE database. The duration variable is represented for each hour by the duration value of the current icing event, where the value is computed as the number of elapsed hours during this icing event. The air temperature variable is given in the database at most times, but there are still missing values. Some storm events have missing temperature readings. These missing temperature values were replaced by approximate values which are estimated from the near actual temperature values as described in the section 3.4.3. Some storm events do not have any temperature reading. Such events have been eliminated when the temperature variable was included in the prediction. This has for consequence a reduction of the number of available data sets.

The variable that the neural network model has to predict is the ice accumulation weight as recorded in the SYGIVRE database. This value of ice accumulation is not the real value, but an estimation from the IRM variable as described in section 3.3.2.

Prediction is performed using the meteorological variables and the binary data. Meteorological variables are included here because using the binary data alone is not sufficient for performing regular predictions, considering that a minimum number of model entries is recommended for simplicity. The ice accumulation weight variable is of foremost importance for effecting the prediction, since it helps the model to remember the variations of total ice accumulation.

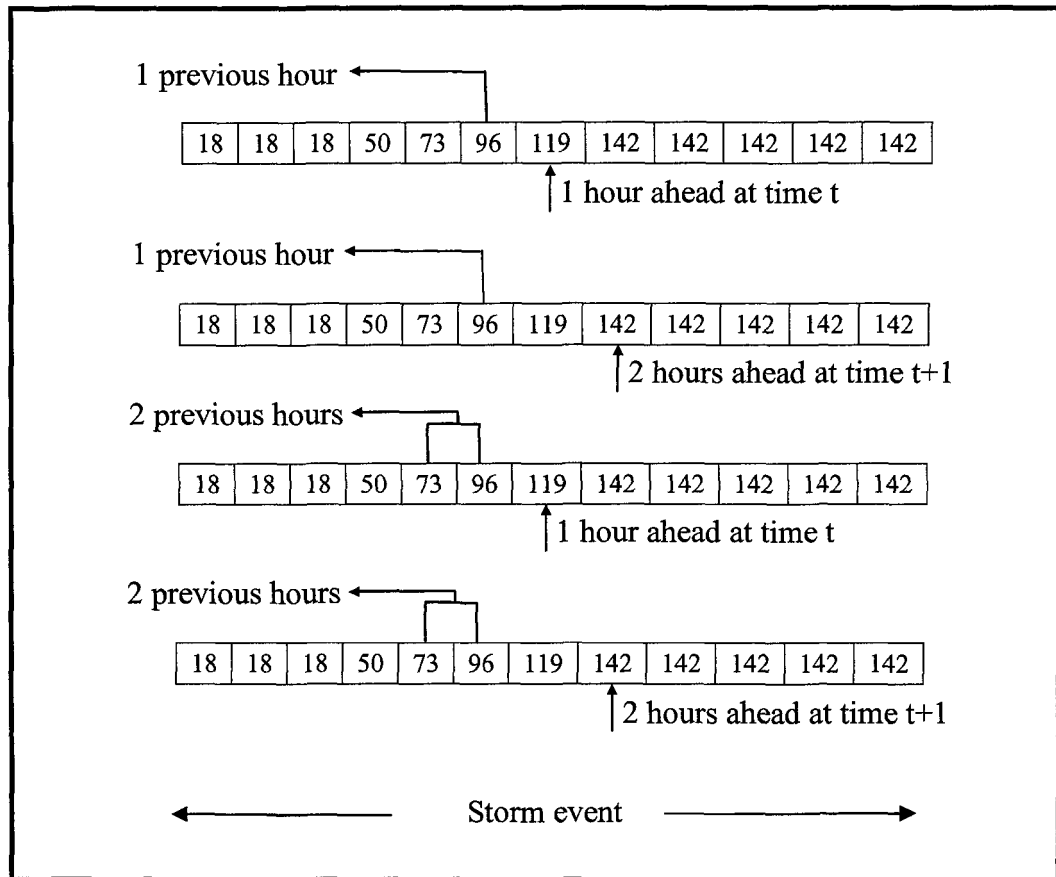
Binary data and quantitative meteorological variables coming from the SYGIVRE database are used for creating the model. Since the unit of time is the hour, a given value for each variable is determined for each hour and all stations. For time  $t$  at station  $i$ , the state variable  $X_i(t)$ , and the quantitative variable  $Y_i^j(t)$  represent the occurrence of an icing storm and the value of the  $j^{\text{th}}$  meteorological variable at this time, respectively. For a group of  $k$  stations, the modeling of ice accumulation is equivalent to predicting the vector  $Y^j(t) = (Y_1^j(t), Y_2^j(t), \dots, Y_k^j(t))$ . When  $j = 1$ , the variable represents the ice accumulation weight. The variables used as predictors are fed from previous times to the model:  $X(t-1), X(t-2), \text{etc}$  for the state variable, and  $Y^j(t-1), Y^j(t-2), \text{etc}$  for the quantitative variable.

The neural network used to predict the ice accumulation weight based on quantitative variables is implemented by means of a mathematical model in the following form:

$$\begin{array}{c} \text{Predicted} \\ \left[ \begin{array}{c} X_1(t+1) \\ X_2(t+1) \\ \vdots \\ X_k(t+1) \end{array} \right] \end{array} = F \left( \begin{array}{c} \text{Observed} \\ \left[ \begin{array}{c} X_1(t) \\ X_2(t) \\ \vdots \\ X_k(t) \end{array} \right], \left[ \begin{array}{c} Y_1^1(t) \\ Y_2^1(t) \\ \vdots \\ Y_k^1(t) \end{array} \right], \left[ \begin{array}{c} Y_1^2(t) \\ Y_2^2(t) \\ \vdots \\ Y_k^2(t) \end{array} \right], \dots, \left[ \begin{array}{c} Y_1^v(t) \\ Y_2^v(t) \\ \vdots \\ Y_k^v(t) \end{array} \right] \end{array} \right) \quad \text{Eq. 5-1}$$

where  $X_i(t)$  represents an ice accumulation weight value for station  $i$  at time  $t$ , and  $Y_i^j(t)$  represents the  $j^{\text{th}}$  quantitative variables at station  $i$ .

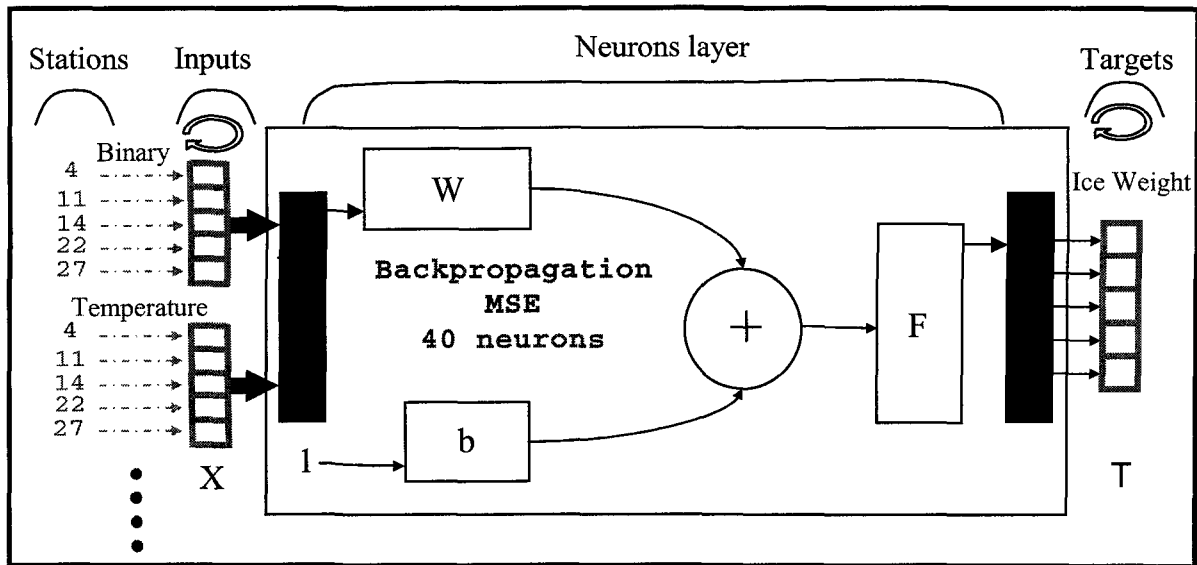
The model uses multiple previous hours, so the information before time  $t$  is used for predicting the ice weight values at a given time, such as:  $\{X(t-1), X(t-2), \text{etc}; Y^j(t-1), Y^j(t-2), \text{etc}\}$ . The model performs prediction for multiple hours after time  $t$  (hours ahead)  $Y^1(t+1), Y^1(t+2), \text{etc}$ . In Figure 5-1 some ice accumulation weight values during an icing storm at Station 4 are considered. The custom model takes into account only one hour ahead at time  $t$ , and this leads to an optimum model compared to the other models that take more hours ahead into account. The present model can predict the ice accumulation weight for 3 hours ahead of recorded data using the data recorded in the previous 3 hours.



**Figure 5-1 General example for 1 and 2 hours ahead with 1 and 2 previous hours applied to some ice weight values at station 4**

### **5.3 Neural network model**

The present neural network accepts binary data and supplementary variables data as input, and produce ice accumulation weight data. The data on the network entry was arranged in a parallel form as shown in Figure 5-2.



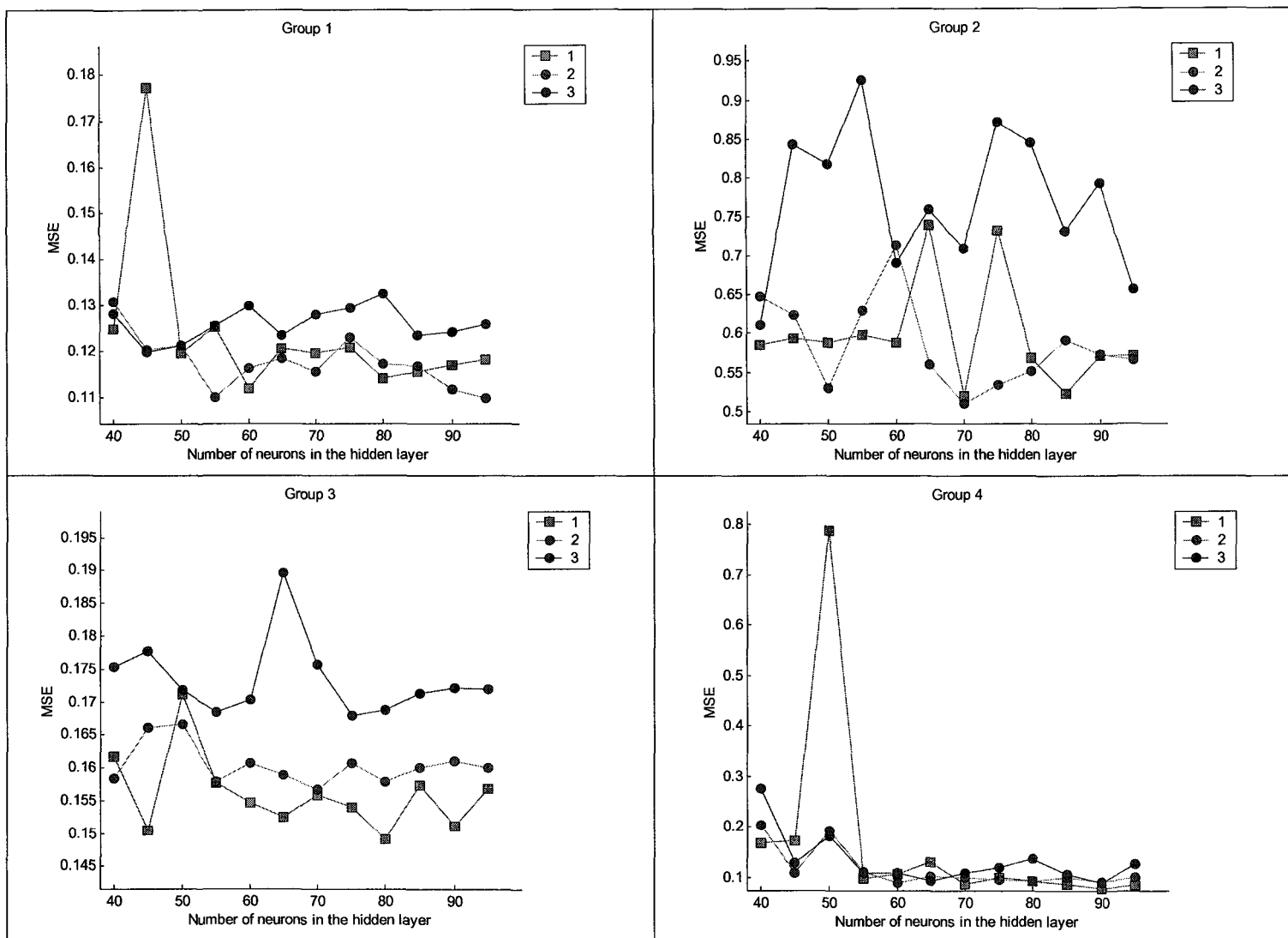
**Figure 5-2 Neural network sample applied to group 1 in the case of ice accumulation weight prediction**

Several basic network characteristics should be considered. These are network architecture, training function, layers, number of neurons, and validation types. These characteristics should be adjusted in order to attain an efficient model. In fact, similar network characteristics used in the binary network model are found to be appropriate in the current model.

For the training part, we use the backpropagation method since it provides acceptable results, and is more appropriate than the Elman method, according to what has been observed with the binary model. Because of its effectiveness, the training function *trainoss* is included in the model's implementation. The standard performance function for

neural networks, MSE, is also involved in the processing, since it is adapted for the processing of continuous data.

The neurons used in the hidden layer have sigmoid transfer functions, and the output layer is formed by neurons having a linear transfer function. The number of neurons in the output layer is equivalent to the size of the network output. In our case, it is the number of stations of the case study group. The number of neurons in the hidden layer could be determined based on Figure 5-3, which shows for all groups the MSE values for different number of neurons, and the influence of the variation in the number of previous hours for the case of one hour ahead. It is worth noting that the MSE results for more hours ahead are similar to those of one hour ahead.



**Figure 5-3 Effect of number of neurons in the hidden layer for one hour ahead**

The minimum values of MSE are obtained with 50 neurons and more. In order to make a simple model with the best possible generalization of the network, a moderate number of neurons with minimum MSE is recommended. It is clear from Figure 5-3 that some cases have smaller MSE values for a small number of neurons, but these cases are rare, and some of them are better suited for a large number of neurons. An optimum model can be made by choosing 60 neurons, while most cases lend minimum values of MSE. Compared to the binary model, this model has showed significant variability of MSE values when repeating the training process.

In order to validate the model, the validation by separation type is used in the process, dividing the data into two parts: the first one used for training (2/3), and the other one for validation (1/3).

Before a training session starts, a preprocessing step as described in the section 2.7 should be done, because the network input contains non binary data. This step is required, since the data must fall within a specified range for a regular prediction. It has been found that the selection of the preprocessing function is difficult since the variation of this function affects the model performance significantly. This is likely due to the inclusion of non binary data in the network's input and output. This problem has not been encountered in the case of the binary model with supplementary variables, where only the duration variable was used in input to the network. Finally, several tests have proved that the most

appropriate function is the logarithmic function. The logarithmic function should scale down the data in the range  $[0, 10]$ . The preprocessing function takes the following form:

$$PX = \log\left(\frac{X + j}{k}\right) \quad \text{Eq. 5-2}$$

where,  $k$  is chosen arbitrarily in order to obtain a better prediction, and  $j$  is chosen as the absolute value of the maximum given temperature values incremented by one, or as only one if the temperature variable is not included.

A simple model can be achieved with a minimum number of supplementary variables in the network entry. An analysis has proved that optimum performance can be achieved by considering the ice weight and temperature together as input variables. This notion is maintained from this point on.

## 5.4 Criteria

Several criteria are needed in order to observe the model's behaviour and its degree of validity. Since the predictive data is composed of quantitative values, the criteria used in the binary model are not applicable.

It is beneficial to examine the correlation between the predicted versus observed values, so that the linear regression method as criteria for validation can be considered [39]. The general purpose of linear regression is to learn more about the relationship between predicted and observed variables. This leads to laying out a straight line that best fits the average of the values.

An important proposed measure is the R-square value, which is an indicator of how well the model fits the data (e.g., an R-square close to 1.0 indicates that we have accounted for almost all of the variability within the specific observed values of the model). The R-square is a function of the performance function MSE.

Another important criterion helps to study the resulting errors between the predicted and observed values. These errors would be traced in terms of the temperature and ice weight variables.

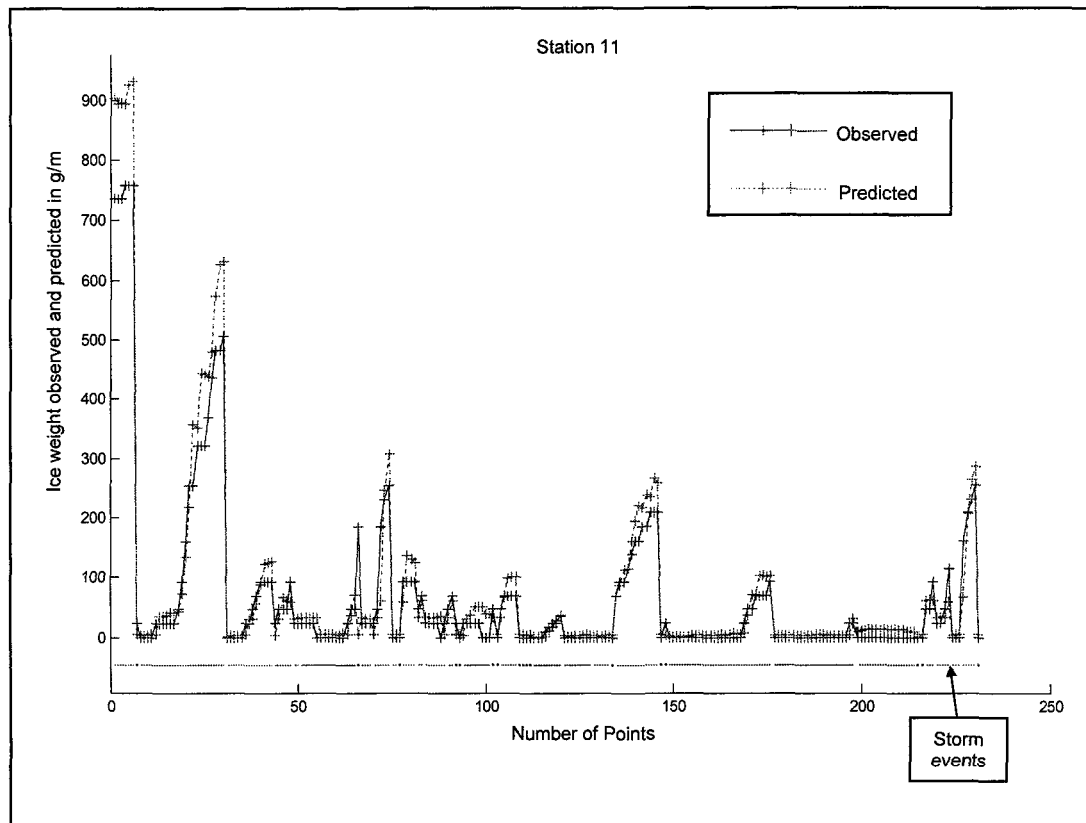
It seems that the neural network generally encounters difficulty in detecting the extreme observations. In order to recognize the prediction detection degree of such observations, an additional criterion was defined. An assumption of critical ice weight point is proposed so all the observations lying above that point are extreme observations, and the ones below that point are least observations. For all extreme and least observations, the average value of the errors between the predicted and observed values is computed.

## **5.5 *Model analysis***

In this section the results are presented in two parts. In the first one, the case of one hour ahead is studied in detail, and in the other one, the case of multiple hours ahead is analyzed.

### **5.5.1 One hour ahead**

An example of the results for Station 11 from Group 1 is illustrated in Figure 5-4. It is clear that the predicted signal (the dotted line) follows very well the observed one, but it is still difficult to detect the extreme observations.

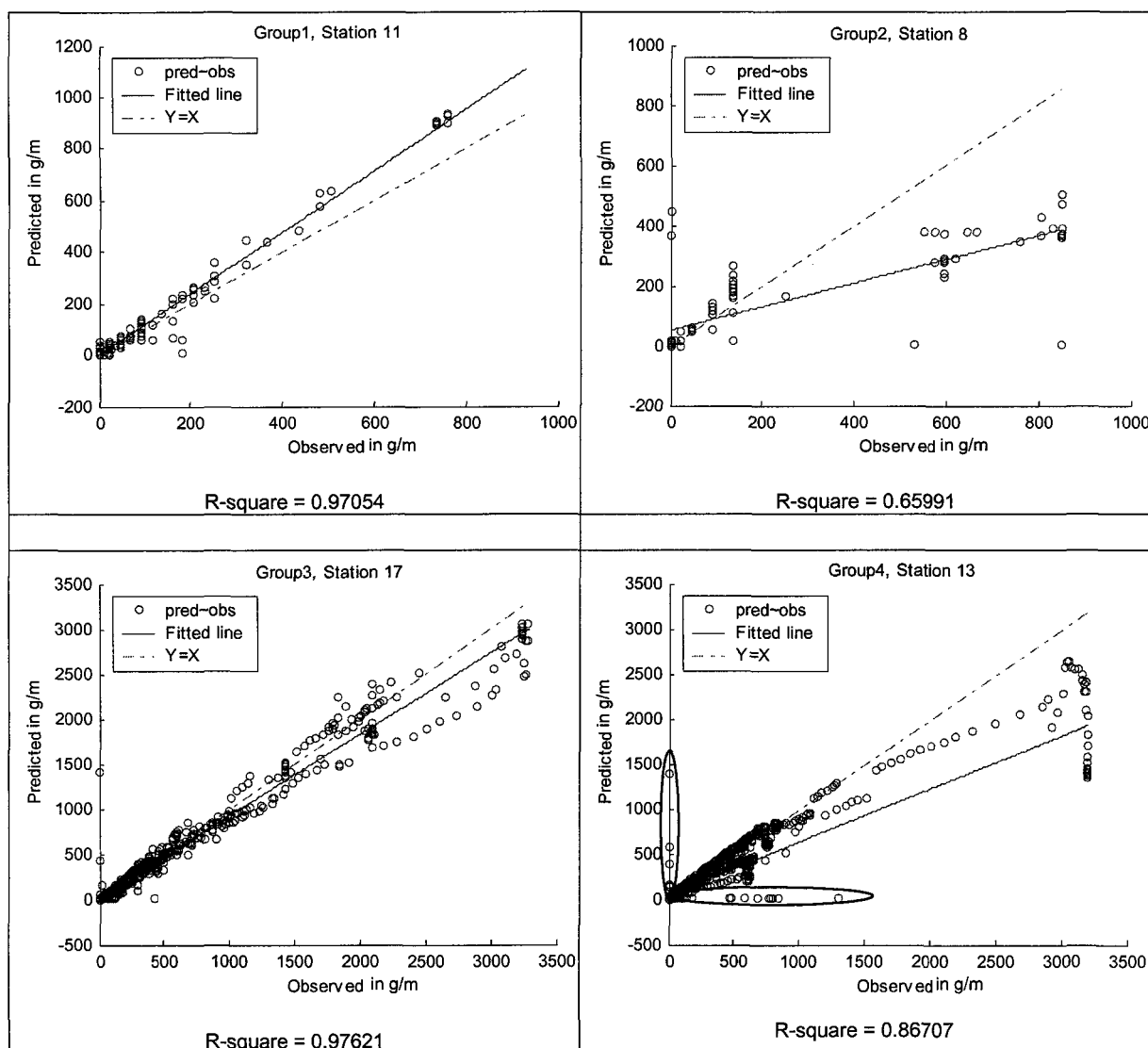


**Figure 5-4 Signal of observed and resultant predicted values for station 11 in group 1**

Figure 5-5 shows the result of the correlation between observed and predicted values for the stations of best performance for each group, and the corresponding R-square value. It can be seen that each group has a distinct behaviour. It is clear that groups 1 and 3 have the better performance, while Group 2 is the worst, usually because of missing data.

It seems that, groups 2 and 4 have many successive observations with equal observed ice weight values. This means that the ice weight stayed stable, and there was no accumulation during that time.

It can be noticed that many observations belonging to Station 8 in Group 2 were not predicted well. It is obvious that the set of observations found far from the dotted line (as the observations approach the line, the model performance increases) take a shape of vertical lines. This is due to the numerous similarities of ice weight observations. In particular, most of the observations that have this characteristic generally encounter a loss of precision in their prediction. This can be noticed at Station 13 in Group 4. At the right side of the figure, many observations have the same value, and their predicted values go far downwards from near the dotted line. In contrast, stations 11 and 17 do not have almost similar observations, but satisfactory predictions.



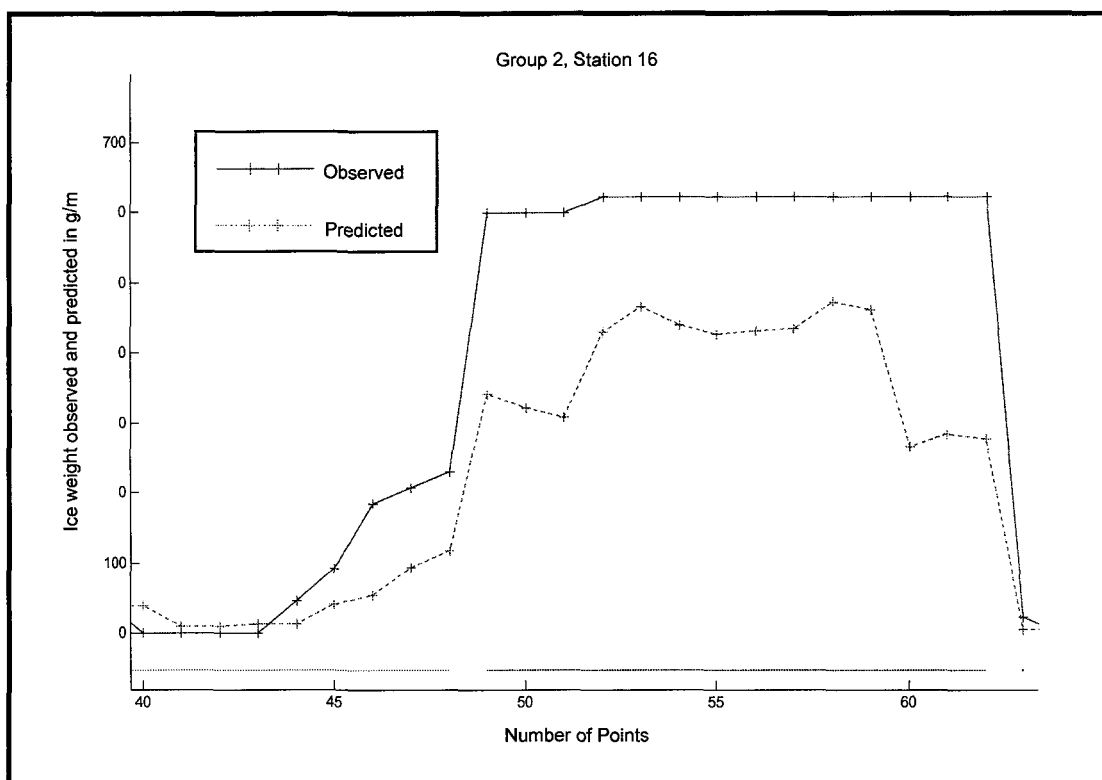
**Figure 5-5 Correlation between the observed and the predicted values for the optimum state of each group and the corresponding R-square value**

Figure 5-6 and Figure 5-7 illustrate the worst samples of predictions. The predicted and observed ice weight values are traced with respect to the number of points. The model

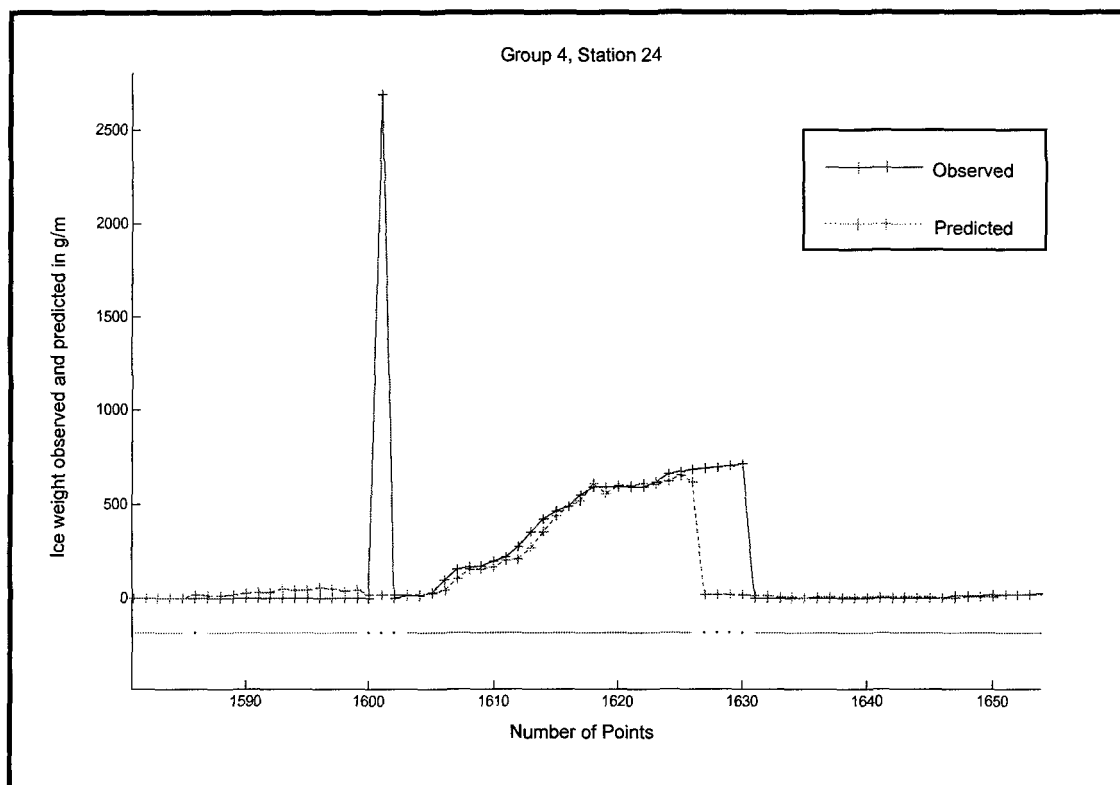
misleadingly predicts various ice weights that are in fact observed to be of the same value (Figure 5-6).

As shown in Figure 5-6, it is always difficult to predict the ice weight values in the terminals of two consecutive storm events, since the transition between the events may baffle the neural network due to the huge difference between the ice weight values. Figure 5-7 shows another prediction problem. When the ice weight changes drastically within the same storm, prediction is perturbed. This kind of problems is often observed.

Consequently, Figure 5-5 shows at the encircled region that several values resulting from the prediction are predicted at almost zero while their corresponding observed values are relatively large, and vice versa.



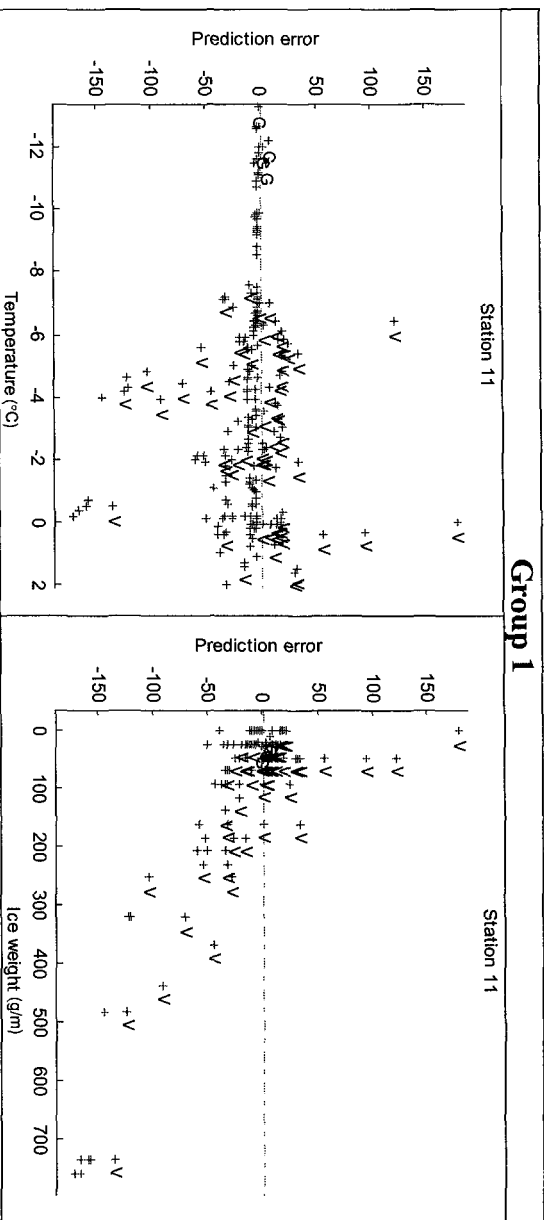
**Figure 5-6 Sample of failed prediction for station 16 in group 2**



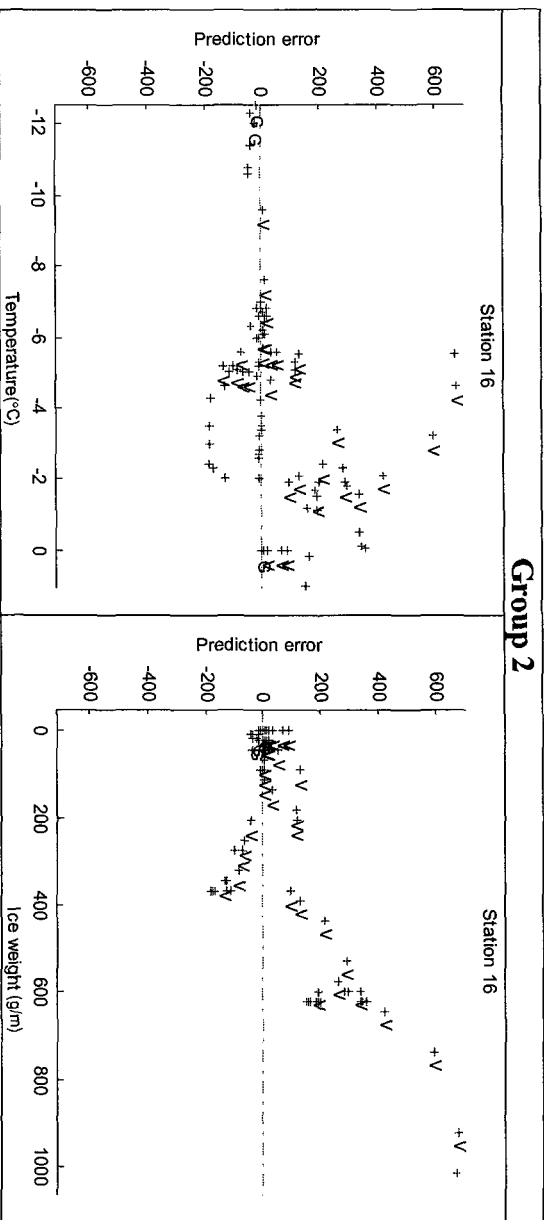
**Figure 5-7 Sample of failed prediction for station 24 in group 4**

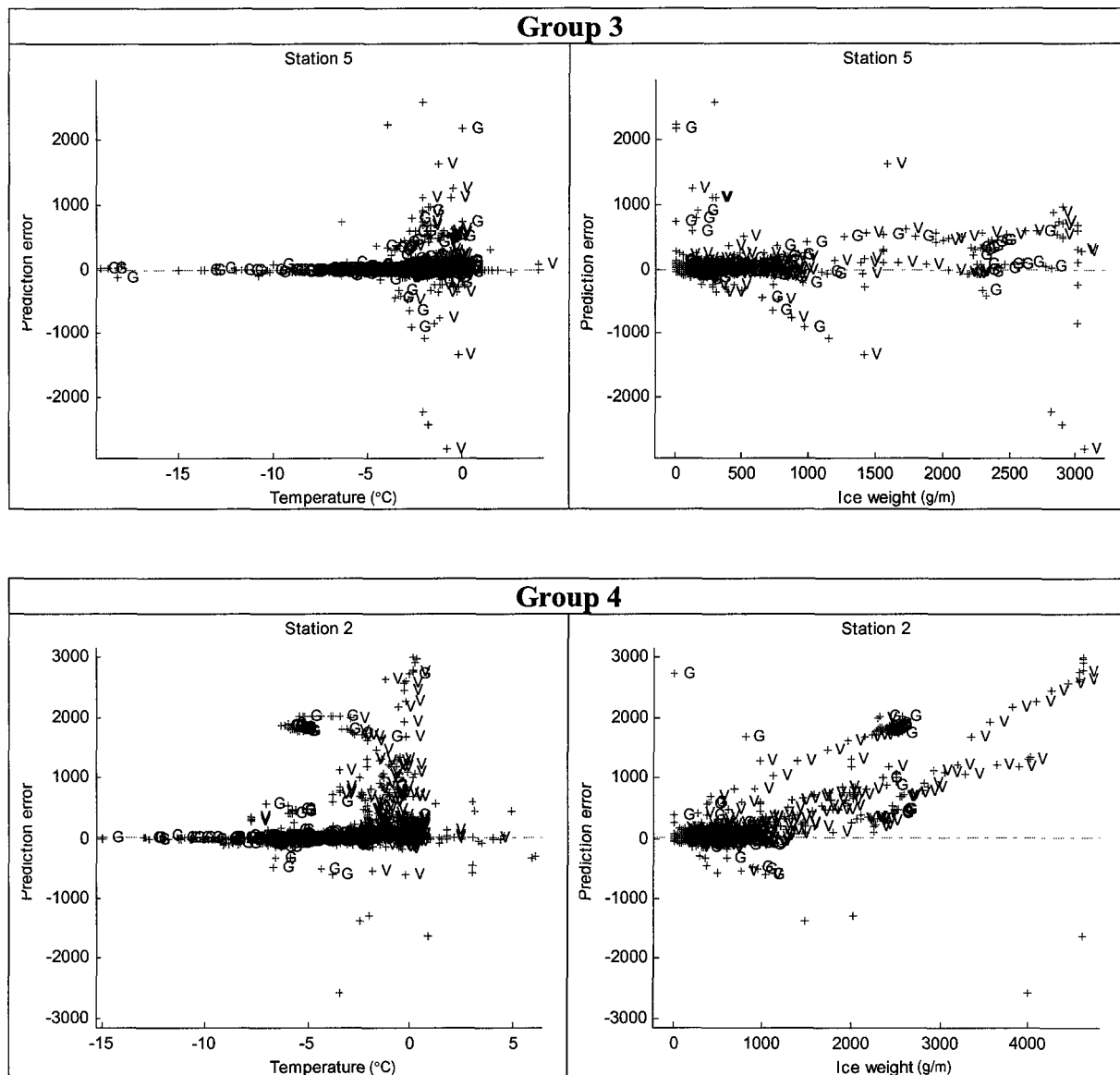
Figure 5-8 illustrates the criterion defined in section 5.4, which presents the prediction error with temperature and ice weight. It is remarked that, greater prediction errors occurred with observations having temperature ranging from  $-5^{\circ}\text{C}$  to  $2^{\circ}\text{C}$ . It is worth noting that most of *freezing rain* accumulation types occurs in that temperature range. We can also mention that the most extreme observations are of *freezing rain* types and they represent the worst cases.

Group 1



Group 2





**Figure 5-8 Errors between the predicted and observed values versus temperature or ice weight (one sample from each group)**

Another criterion was defined in the section 5.4, so the extreme observations should be analysed. Generally, the extreme observations of ice accumulation are infrequent. Table

5-1 shows for all stations considered in this analysis the frequency of observations above a critical point, and the observations below it with the corresponding average of prediction errors. The critical point was specified in each station depending on the size of maximum ice accumulation weight value in the station.

Station	Extreme(>critical point) Least(<critical point)	Frequency	Average prediction error (g/m)
4	>50	34	43.76
	<50	197	6.55
11	>100	35	55.22
	<100	196	10.45
14	>150	38	106.73
	<150	193	13.20
22	>50	37	34.39
	<50	194	5.80
27	>200	34	142.83
	<200	197	15.66
8	>200	30	390.21
	<200	64	41.48
16	>250	37	72.70
	<250	57	24.48
5	>500	186	198.61
	<500	969	44.11
17	>500	210	157.46
	<500	945	16.82
1	>200	236	416.32
	<200	2042	21.29
2	>500	382	453.44
	<500	1896	28.34
7	>150	189	57.05
	<150	2089	7.55
13	>400	462	400.18
	<400	1816	10.63
24	>500	707	345.88
	<500	1571	24.55
28	>200	233	98.93
	<200	2045	9.24

**Table 5-1 Average of prediction errors for the extreme observations and the least ones for all stations**

### **5.5.2 Analysis of results when using different hours ahead with different previous hours**

The result of the prediction is represented by the R-square criterion. Table 5-2, Table 5-3, Table 5-4, and Table 5-5 show the R-square values for all combinations of the number of hours ahead (ahead), and the number of previous hours (previous) taken into account, for the four groups available. For each group of stations, the average of R-square values for all stations is computed.

It is remarked that the variation of previous hours for a fixed number of hours ahead does not affect the model much. The three R-square average values of group 1 for the prediction of one hour head are near 0.8, and for two hours ahead are near 0.7. When looking at other groups, we can observe that the values are almost similar for a fixed number of hours ahead. However, the results coming from Group 2 are quite different, especially for the case of two hours ahead. This is generally due to the stability misses of Group-2 model which suffers from lack of data. Consequently, the results indicate that the use of more than one previous hour may not enhance the model performance.

	Station					
Ahead, Previous	4	11	14	22	27	Average
1,1	0.8429	0.9705	0.8621	0.8457	0.8452	0.8733
1,2	0.8296	0.9598	0.8828	0.9437	0.7694	0.8770
1,3	0.8234	0.8992	0.9439	0.9566	0.7245	0.8695
2,1	0.7240	0.9473	0.5021	0.7727	0.5608	0.7014
2,2	0.6898	0.8368	0.5336	0.7740	0.5084	0.6685
2,3	0.6736	0.8898	0.6102	0.8266	0.4697	0.6940
3,1	0.5775	0.8411	0.2166	0.5788	0.2988	0.5025
3,2	0.6297	0.8396	0.3950	0.5540	0.5030	0.5843
3,3	0.5883	0.8205	0.5584	0.8386	0.3065	0.6225

**Table 5-2 R-square values to different hours ahead and previous hours for group 1**

	Station		
Ahead, Previous	8	16	Average
1,1	0.65991	0.64518	0.65254
1,2	0.66666	0.8141	0.74038
1,3	0.60895	0.73751	0.67323
2,1	0.46397	0.83969	0.65183
2,2	0.070901	0.80988	0.44039
2,3	0.62126	0.47304	0.54715
3,1	0.2779	0.66357	0.47073
3,2	0.32559	0.43363	0.37961
3,3	0.29626	0.42202	0.35914

**Table 5-3 R-square values to different hours ahead and previous hours for group 2**

Ahead, Previous	Station		Average
	5	17	
1,1	0.84449	0.97621	0.91035
1,2	0.85132	0.98531	0.918315
1,3	0.84439	0.98084	0.912615
2,1	0.76207	0.96051	0.86129
2,2	0.75215	0.95595	0.85405
2,3	0.7498	0.88881	0.819305
3,1	0.67507	0.93536	0.805215
3,2	0.68114	0.91604	0.79859
3,3	0.69166	0.90346	0.79756

**Table 5-4 R-square values to different hours ahead and previous hours for group 3**

Ahead, previous	Station						Average
	1	2	7	13	24	28	
1,1	0.72526	0.76186	0.83809	0.86707	0.82813	0.81677	0.80620
1,2	0.92596	0.88028	0.8759	0.94074	0.89346	0.90983	0.90436
1,3	0.78389	0.89683	0.86797	0.86634	0.84276	0.83743	0.84920
2,1	0.79943	0.82064	0.78084	0.88019	0.8168	0.86804	0.82766
2,2	0.79883	0.86397	0.71529	0.91481	0.85811	0.83774	0.83146
2,3	0.75484	0.87221	0.68543	0.86611	0.864	0.67686	0.78658
3,1	0.80798	0.74337	0.73781	0.87018	0.84501	0.83635	0.80678
3,2	0.72351	0.83914	0.74672	0.9254	0.86897	0.72891	0.80544
3,3	0.69287	0.79358	0.69946	0.66998	0.81333	0.78597	0.74253

**Table 5-5 R-square values to different hours ahead and previous hours for group 4**

It is remarkable from these tables that as the number of hours ahead increases, the prediction performance deteriorate. This can be distinguished by looking at the decreasing of R-square average values as the number of hours ahead increases.

For any group of stations, as the number of hours ahead increases, the number of observations related to that group diminishes. This loss of information surely contributes to reduce the prediction performance. Also, since the prediction of the neural network depends on the spatial distribution of icing events for any group of stations, the increasing of the number of hours ahead may weaken the spatial links between the stations.

Some states are similar to that of Table 5-5. For the case of one hour ahead, the average value of R-squared for one previous hour is much smaller than the corresponding value in the case of two previous hours. This is probably due to the random results of the prediction. Usually, the training algorithm takes each time a different track in order to reach its goal, and the output result should be in most times nearly similar. However, it is not the case here. In fact, this exception occurred by hazard, and it does not reflect the habitual results. Therefore, it is essential to repeat the network training under the same conditions, and to observe the results.

Several training trials were performed under the same network conditions, and the R-square values were computed each time. It has been seen that each training process

deliver different R-square values from those obtained from other training processes. However, in all trials, the R-square values are still decreasing as the number of hours ahead increases.

By examining different training trials, we have found that each group had specific behaviour with respect to the variation of R-square values, as follows.

By considering Group 1, we can observe random results, for a fixed hour ahead. Table 5-6 displays the results for 2 hours ahead, as the first training trial. The R-square values at station 14 are very different from each other. On the other hand, as shown in Table 5-7, that presents the average R-square values for different training trials, the values are always almost similar for a fixed number of hours ahead whether repeating the training or changing the number of previous hours.

Ahead, Previous	Station					Average
	4	11	14	22	27	
<b>2,1</b>	0.7095	0.9410	0.4775	0.7784	0.5881	0.6989
<b>2,2</b>	0.6815	0.8642	0.4286	0.8095	0.4674	0.6502
<b>2,3</b>	0.6573	0.8756	0.8073	0.7930	0.4311	0.7129

**Table 5-6 R-square values in the case of 2 hours ahead for group 1, trial 1**

	Ahead, Previous								
	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
<b>Trial 1</b>	0.8608	0.8629	0.8219	0.6989	0.6502	0.7129	0.5651	0.590	0.5334
<b>Trial 2</b>	0.8581	0.8912	0.8688	0.6732	0.6631	0.6161	0.5012	0.5312	0.5762
<b>Trial 3</b>	0.8415	0.8384	0.8607	0.6989	0.7126	0.6707	0.5477	0.5311	0.5896

**Table 5-7 Average R-square values for different hours ahead and previous hours in group 1**

Concerning Group 2, the values are extremely random, and there are many strange values, as shown in Table 5-8, which presents the R-square values for Station 16 in Group 2 with different training trials. However, the average R-square values of the group are not much similar, as shown in Table 5-9.

	Ahead, Previous								
	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
<b>Trial 1</b>	0.9450	0.7325	0.7443	0.9437	0.2424	0.5629	0.8784	0.3038	0.1559
<b>Trial 2</b>	0.9390	0.9505	0.7568	0.7954	0.6809	0.6904	0.7165	0.7159	0.3016
<b>Trial 3</b>	0.7337	0.7425	0.5622	0.9622	0.8156	0.7407	0.8995	0.6014	0.3277

**Table 5-8 R -square values for different hours ahead and previous hours at station 16 in group 2**

	Ahead, Previous								
	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
<b>Trial 1</b>	0.6952	0.6508	0.7106	0.6183	0.2441	0.3265	0.5913	0.2862	0.1138
<b>Trial 2</b>	0.7033	0.8105	0.6746	0.5947	0.5531	0.4701	0.4725	0.4513	0.2706
<b>Trial 3</b>	0.6924	0.7412	0.6098	0.5922	0.6316	0.5431	0.5756	0.3732	0.2809

**Table 5-9 Average R-square values for different hours ahead and previous hours in group 2**

For Group 3, values are not quite random, the average values being nearly similar for a fixed hour ahead and for different training trials results. Table 5-10 and Table 5-11 show the behaviour of this group, where the first table presents the R-square values for Station 17, and the other one presents the average R-square values for the group.

	Ahead, Previous								
	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
<b>Trial 1</b>	0.9857	0.9791	0.9789	0.9442	0.9626	0.9341	0.9231	0.9283	0.8815
<b>Trial 2</b>	0.9685	0.9812	0.9708	0.9373	0.9466	0.9540	0.9011	0.9249	0.9322
<b>Trial 3</b>	0.9838	0.9803	0.9870	0.9243	0.9451	0.8710	0.8854	0.9031	0.9229

**Table 5-10 R -square values for different hours ahead and previous hours at station 17 in group 3**

	Ahead, Previous								
	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
<b>Trial 1</b>	0.9137	0.9167	0.9137	0.8502	0.8541	0.8231	0.8066	0.8126	0.7867
<b>Trial 2</b>	0.9082	0.9115	0.9067	0.8547	0.8532	0.8507	0.7934	0.8031	0.8133
<b>Trial 3</b>	0.9199	0.9143	0.9188	0.8413	0.8563	0.8135	0.7745	0.7957	0.8108

**Table 5-11 Average R-square values to different hours ahead and previous hours for group 3**

Group 4 behaves exactly like Group 1, but several average R-squared values are not much similar for a fixed hour ahead or by varying the training process trials. Table 5-12 shows the average R-square values for that group.

	Ahead, Previous								
	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
<b>Trial 1</b>	0.8978	0.8601	0.8637	0.8332	0.8296	0.8098	0.7756	0.7444	0.7630
<b>Trial 2</b>	0.8842	0.7497	0.8490	0.7986	0.7882	0.8346	0.7832	0.7583	0.6807
<b>Trial 3</b>	0.8551	0.8992	0.8028	0.8274	0.7817	0.7794	0.7826	0.8039	0.7709

**Table 5-12 Average R-square values for different hours ahead and previous hours in group 4**

## **5.6 Discussion**

The prediction results for different previous hours are similar to those obtained with one previous hour. We can also clearly observe the variation of the results when switching to a different number of hours ahead.

The idle times between icing events within storm events, as defined in section 3.4.2, can be interpreted here as consecutive equal ice weight observations. Generally, the model encounters difficulty in predicting those values, because of missing information during idle times.

The analysis of extreme ice weight observations helps evaluating the prediction results related to each station. In Table 5-1, Station 8 from Group 2, and stations 1, 2, 13, and 24 from Group 4 present the worst average prediction errors, whereas the extreme observations have a relatively large average prediction error compared to that of the least observations.

On the other hand, it is remarked that stations 4, 5, 11, 16, and 22 have provided good average prediction errors, because the ratio of average prediction error for the extreme

over the least observations is moderate. This means that the extreme observations may be better detected at these stations than at other stations.

In fact, the R-square variation with the repetition of the training process generally has an influence on stations of groups 1, 2, and 4, and especially on Group 2. It is remarked that, in every trial training processes, the network performance is always distributed over stations of the group under consideration, so the R-square average values stay almost stable. This remark is valid for all groups except for Group 2, since several average R-square values for this group suffer from instability.

It is worth mentioning that the amount of variation of R-square or average R-square values for a fixed hour ahead when changing the number of previous hours or performing different training processes is similar.

## **5.7 Conclusions**

A model has been devised based on the data described in CHAPTER 3, in order to predict the ice accumulation weight by supplying binary data along with other supplementary variables in input to the model. The model has used the four predefined groups, which basically take into account the spatial relationship between stations in each group. Based on the neural network technique, the desired model has been established, and several characteristics were derived from the corresponding binary model. It has been

deduced that the most pertinent supplementary variables used as input to the network were the *ice accumulation weight* and *temperature* variables. In this model, two elements of information are needed in order to make predictions: the number of hours ahead to be predicted, and the number of known previous hours to be considered. In the context of the present study, up to 3 hours ahead have been predicted successfully, using the known data of up to 3 previous hours.

The model has shown a high degree of effectiveness, since the predicted signals follow closely the variation of observed signals. However, each group has a different degree of performance in the order of 3, 1, 4, and 2. The results indicate that various extreme observations with equal values were not predicted well in groups 2 and 4. Furthermore, the model was less predictive in case of Group 2 since there is data missing. The model performance was better with Group 3 than with Group 1, due to the difference in the supplied data. It is notable that Group 1 supports more stations than Group 3.

Also, it can be perceived that the duration variable do not improve the model performance as it was the case with the binary network model using supplementary variables. In contrast, the ice weight and temperature variables play an important role in obtaining acceptable results. The current model display a totally different behaviour compared to the previous models. This can be distinguished by observing the random

results obtained when performing different training processes compared to those obtained with the binary model, which are relatively stable.

The choice of the number of previous hours considered, on the other hand, is not as much sensitive for the prediction. The case when taking into account only one previous hour is almost better than that of more previous hours. Also, as already envisaged, the performance degrades as the number of hours ahead increases.

## **CHAPTER 6**

### **CONCLUSIONS AND RECOMMENDATIONS**

## 6.1 General conclusions

The study of atmospheric icing events for multiple stations spatially related and located in a given region is still at a beginning stage. The complexity of meteorological data of the SYGIVRE database has rendered impractical the application of standard tools of prediction, such as multiple linear regression and other techniques. Recourse to more advanced tools from artificial intelligence was therefore required in order to carry out the present work. The neural network technique has proved very useful for understanding the atmospheric icing phenomena. It has proved its reliability over the regression logistic method, and its ability at detecting the evolution of atmospheric icing precipitation in real-time.

This work has been realized in two steps, both of these considering in their predictions the same given meteorological variables of *ice accumulation weight*, *air temperature*, *ice accretion rate*, and *icing event duration*. Four spatially and geographically distributed groups of stations were included, while the data related to a specific group was studied separately. The storm events have been determined both by considering the fact that a meteorological storm may propagate between the stations of a group, and on the basis of several meteorological characteristics. The data have been organized in order to be compatible with the neural network. The time unit used is the hour, and the prediction is conditioned to the fact that there is an icing storm.

The first step of the work aims at predicting the icing storm evolution in real-time. The prediction has been based on the signal of an icing storm occurrence, which is represented in the database by a binary indication (zeros if there is no icing storm and one if there is one). Two models have been carried out, whereas the icing storm state at a specific time is predicted using the historical information previous to that time. The first one deals with the creation of a binary network model that only includes binary data, while the other integrates meteorological variables with the binary data at its input, aiming to increase the performance of the model.

It has been observed that the integration of the duration variable in the binary model improved the results. However, the *duration* variable has been found not pertinent for effecting a prediction, since the duration of an icing event could not possibly be available in real-time. Therefore, we conclude that the integration of supplementary variables does not bring any advantage to the model.

For purposes of comparison, a prediction model based on logistic regression, and used in the literature for processing the SYGIVRE data, has been studied in order to establish a relationship between the spatial groupings of stations by considering meteorological variables. This model has provided a total of 80% successful predictions, which can be considered good compared to the 94 % achieved with our model for Group 1.

For *sensitivity*, however, a disappointing value of 19% resulted in the case of the model from the literature, whereas our model brought a remarkable improvement of 80 %.

In the case of the logistic regression model, the duration variable explicates well the groupings of the stations for the icing events. This conforms to the results of the binary model with supplementary variables, where it provided the best results using the duration variable.

The second step has been conducted in order to predict the ice accumulation weight in real-time, and for detecting icing storms severity. A model has been implemented by using the binary model as a starting point, because of the similarity between the two models. Many neural network characteristics have been generalized in this model. The ice weight is predicted as a continuous variable rather than as a binary one, and it uses binary data together with other meteorological variables as input. The ice weight value during an icing storm, at a specific time, has been predicted using the information available at previous times, and the prediction was also done for different times ahead.

It has been observed that the model provides the best performance when the ice weight and temperature variables are used together in input to the model. The duration variable did not offer serious effects as that was the case in the preceding model.

It appears that the neural network model is able to track the evolution of icing accumulations. However, we observe that it has some difficulty in detecting extreme icing storms. This is most likely due to the lack of available data.

When considering more than one hour ahead, the prediction based on the spatial relationship between stations has provided promising results, and the model preserves a good performance for up to 3 hours ahead.

It can be concluded that the neural network architectures with a dependent implicit time, like the Elman method, did not bring evident improvements to the model, since the largest effect on the model is due to the spatial character of stations. In fact, there are several conclusions in common between the described models.

Here are the general conclusions of the present work.

The Group 2, that represents the Montreal region, suffers from a lack of observations, maybe because only 3 measurement seasons were included. For this reason, the results obtained for this group were instable and relatively unsatisfactory.

It appears that considering several previous hours for making a prediction it is not quite effective. Generally, in all of the studied models, the consideration of one previous

hour provides almost better prediction results. The inclusion of more data in future studies may render the consideration of previous hours more effective.

The four groups of stations of the SYGIVRE network cover regions having different areas. It seems that the prediction of icing events for any region could be dependent on the size of the area of that region, since the consistency of spatial relationship between stations related to the region diminish when the area increase. On the other hand, the quantity of data related to any group is the major factor affecting the prediction. In fact, the number of observations of the groups is much different. It is thus difficult to observe the effects on the prediction performance in terms of the extension of the group area.

As it is known, the icing events in the SYGIVRE database can be classified in the two types of *freezing rain* or *frost*. Then, two category of icing events are available, each one should be studied separately in order to gain an accurate understanding of its behaviour. Unfortunately, this is difficult to achieve due to data dearth.

## 6.2 Recommendations

In the present study, a model for the prediction of atmospheric icing storms in real-time based on a small database has been devised. Recent measurement years of the SYGIVRE database beyond those used in the current work should eventually be integrated for further analysis. The integration of such data could provide a more consistent system, and give more satisfactory results.

It could be interesting to validate the prediction results using different data, for example, using databases from external sources, such as Environment Canada, which manages a network of meteorological stations distributed in the Province of Quebec.

In order to supply information to the neural network, needed for improving the performance of the model, the use of new meteorological variables is necessary. Such variables could include: wind speed, humidity, station position and station altitude.

The division of data between the two available icing event categories, *freezing rain* and *frost*, is desirable to enable the separate study of each category. Unfortunately, data dearth prevents or renders useless such a division. This problem could be solved by the integration of new data.

## References

- 
1. E. Larouche, J. Rouat, G. Bouchard, M. Farzaneh "Exploration of static and time dependent Neural Network technique for the prediction of ice accretion on overhead line conductors". *Proceedings. Of 9<sup>th</sup> IW AIS2000, Chestert, Royaume-Uni*, session 2.
  2. Guesdon C., Houde L., Farzaneh M. et Chouinard L.. "Analysis of Spatial Patterns for Icing Events in Quebec". *Ninth International Workshop on Atmospheric Icing of Structures. Chester, Royaume-Uni, 2000.*
  3. Isaksson S.P., Eliasson A.J. et Thorsteins E.. "Icing Database, Acquisition and Registration of Data". *Eighth International Workshop on Atmospheric Icing of Structures. Reykjavik, Islande, 1998, pp.235-240.*
  4. Felin B.. "Ten years of standardized field ice accretion measurements in Quebec". *Third International Workshop on Atmospheric Icing of Structures. Vancouver (B.C.), Canada, 1986*
  5. 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.*
  6. 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.*

- 
7. 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.
  8. Laforce J.L., Allaire M.A. et Laflamme J. "Wind Tunnel Evaluation of a Rime Metering using Magnetostrictive Sensor". *Atmosph Research* 36 , 1995, pp287-301
  9. McComber P., De Lafontaine, Druez J. et Laflamme J. and Paradis A. "A comparison of Neural Network and Multiple Regression Transmission Line Icing Models" *Proc 7<sup>th</sup> In Workshop Atm Icing of struct*, 1998, pp175-180
  10. Laflamme J.. "Spatial Variation of Extreme Values in the Case of Freezing Rain Icing". *Sixth International Workshop on the Atmospheric Icing of Structures. Budapest, Hongne, 1993*, pp.19.23.
  11. 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), Canada, 1998*.
  12. Farzaneh M. et Savadjiev K.. "Icing Events Occurrence in Quebec: Statistical Analysis of Field Data". *Proceedings of 8th International Offshore and Polar Engineering. Montréal (QC), Canada, 1998*.
  13. 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, 1999*, pp.574-580.
  14. Chouinard L.E., ElFashny K., Nguyen V.T.V. et Laflamme J.. "Modeling of Icing Events Based on Passive Ice Meter Observations in Quebec". *Atmospheric Research* 46, Elsevier, 1998, pp.169-179.

- 
15. Laflamme J. et Périard G.. "The Climate of freezing Rain over the Province of Quebec in Canada: A Preliminary Analysis". *Seventh International Workshop on Atmospheric Icing of Structures. Chicoutimi (QC), Canada, 1996.*
  16. Lu M.L., Olivier P., Popplewell N. et Shah A.H.. "Predicting Extreme Loads on a Power Line from Freezing Rainstorms" *Ninth International Offshore and Polar Engineering Conference. Brest, France, 1999, pp.594-598.*
  17. Felin B.. "Pluie verglaçante au Québec: observations naturelles comparées aux estimations d'un modèle" *Fourth International Workshop on Atmospheric Icing of Structures. Paris, France, 1988*
  18. C. Guesdon, L. Houde, M. Farzaneh "Études des répartitions des événements de verglas et de givre". *Université du Québec à Chicoutimi.*, 2000, pp173.
  19. Bishop, C. M., 1996: Neural Networks for Pattern Recognition. *Clarendon Press, Oxford*, pp482.
  20. Hertz, J., A. Krogh, and R. G. Palmer, 1991: Introduction to the Theory of Neural Computation. *Addison-Weesley*, pp414.
  21. Masters, T., 1993 : Practical Neural Network Recipes in C++. *Academic Press*, pp493.
  22. Ohta, H., Saitoh, K, Kanemaru, K., Ijichi, Y., Kitagawa, H., Konno, T., 1996, "Application of Disaster Warning System due to Snow Accretion on Power lines Using Neural Networks" *Proc. 7<sup>th</sup> IWAIS, Chicoutimi*, pp149-154.

- 
23. P. McComber(1), J. de Lafontaine, J. Laflamme, J.Druez., A. Paradis, 1996. "Estimation of Transmission Line Icing at Different Sites Using Neural network". *Proc. Of 9<sup>th</sup> ISOPE, Brest, France*, pp. 599-606.
  24. Marzban, C., and G. Stumpf, 1996: A neural network for tornado prediction based on Doppler radar-derived attributes. *Journal of Applied Meteorology*, 35, pp617-626
  25. Marzban, C., and G. Stumpf, 1998: A Neural Network for damaging wind., *Wea. Forecasting*, 13, pp151-163.
  26. Tank, D. W. 1986. Simple "neural" optimization networks: An A/D converter, signal decision circuit, and a linear programming circuit. *IEEE Transactions on Circuits and Systems* CAS33(5):533-41.
  27. Toru Masuzawa, Yoshiyuki Taenaka, Masayuki Kinoshita, Takeshi Nakatani, Haruhiko Akagi, Hisateru Takno and Yasuhiro Fukui, "A Cardiac Output Estimation Model With A Neural Network Toward Total Artificial Heart Control". *IEEE Technology Update Series, Neural Networks Applications*, 1996, pp397-398.
  28. Carpenter, Gail A., and Grossberg, Stephen (1987). "A Massively Parallel Architecture for a Self -Organization Neural Pattern Recognition Machine". *Academic Press, (Computer Vision, Graphics, and Image Processing)* 37:54-115.
  29. Refenes A. P.(ed.): Neural Networks in the Capital Markets. *John Wiley & Sons Ltd*, 1995.
  30. Xiao, R.G. Chandrasekar, V. Liu, H.P. "Development of a neural network based algorithm for radar snowfall estimation". *IEEE transactions on geoscience and remote sensing*, vol.36 (3), 1998, pp716-724.

- 
31. Donald B. Malkoff. "A Neural Network for Real Time Signal Processing". *Neural Computing*, 1991, pp248-255.
  32. Laforce, J.L., Allaire M.A., 1992, "Essais de sondes de type Rosemount à différentes intensités de givrage sec et humide". *Rapt. D'essai HQ-92-02*, GRIEA, Chicoutimi, Québec, 131 pp.
  33. Claffey, J., Jones, K.F., Ryerson, Ch.C., 1993, "Use and calibration of Rosemount ice detectors for meteorological research." *Proc. 6<sup>th</sup> Int. Workshop on Atmospheric Icing of Structures*, pp. 105-109.
  34. Laflamme J.. "Nouveaux sites proposes et limites de sous-ensembles géographiquement homogènes ; Sites de mesure du givrage atmosphérique". *Carte, Hydro-Québec*, décembre 1998.
  35. G. Fortin 'THERMODYNAMIQUE DE LA GLACE ATMOSPHERIQUE'. *Université du Québec à Chicoutimi.*, 2002
  36. Hydro-Québec. "Givromètre: Fiche technique pour système de mesure de verglas".
  37. J. Cohen: "A Coefficient of Agreement for Nominal Scales". In *Educational and Psychological Measurement*, XX(1):37-46, 1960.
  38. H. Demuth, M. Beale "Neural Network Toolbox For Use with MATLAB". *MathWorks*, 2000
  39. StatSoft "Multiple Regression". <http://www.statsoftinc.com/textbook/stmulreg.html>.
  40. MathSoft. "S-PLUS 2000 Guide to statistics, Volume 1". *Data Analysis Products Division. Seattle (WA), Etats-Unis*, 1999.
  41. MathSoft. "S-PLUS 2000 Guide to statistics, Volume 2". *Data Analysis Products Division. Seattle (WA), Etats-Unis*, 1999.

- 
42. MathSoft. "S-PLUS 2000 User's Guide". *Data Analysis Products Division. Seattle (WA), Etats-Unis*, 1999.