



FACULTE DE GENIE ELECTRIQUE ET DE L'INFORMATIQUE  
DEPARTEMENT D'ELECTRONIQUE

## MEMOIRE DE MAGISTER

Présenté par :

**BENMAHDJOUR Khalida**

En vue de l'obtention du diplôme de Magister en Electronique

Option télédétection

Intitulé :

## Prévision des précipitations par réseaux de neurones

Devant le jury d'examen composé de :

Mr. AMEUR Soltane	Professeur à l'UMMTO	Président
Mme. AMEUR Zohra	Maître de conférences (A) à l'UMMTO	Rapporteur
Mr. HADDAB Salah	Maître de conférences (A) à l'UMMTO	Examineur
Mr. LAGHROUCHE Mourad	Maître de conférences (A) à l'UMMTO	Examineur
Mr. LAHDIR Mourad	Maître de conférences (B) à l'UMMTO	Examineur
Mr. CHIKHI Nourredine	Ingénieur à ONM de T.O	Invité

Soutenu le : .....

## Remerciements

Avant tout je remercie dieu de m'avoir donnée la volonté pour préparer ce modeste travail.

Je tiens à remercier vivement Madame AMEUR Zohra, Maître de Conférences à l'UMMTO, directrice du laboratoire LAMPA, d'avoir assurée la responsabilité d'être ma directrice de mémoire, pour ses précieux conseils qu'elle m'a faits bénéficier tout au long de mon travail, pour sa constante disponibilité, Mais également pour toute la confiance qu'elle a portée en moi.

Je remercie également Mr. AMEUR Soltane, professeur à l'UMMTO, pour m'avoir fait l'honneur de présider mon jury de mémoire.

Que Mr. LAGHROUCHE Mourad, Maître de Conférences à l'UMMTO, trouve ici, l'expression de ma profonde reconnaissance, d'abord pour ses conseils et ses orientations, mais aussi pour avoir accepté de participer à ce jury.

Toute ma gratitude va également à Mr. HADDAB Salah, Maître de Conférences à l'UMMTO, qui m'a fait l'honneur de faire partie de jury de mon mémoire.

Mes sincères remerciements s'adressent aussi à Mr. LAHDIR Mourad, Maître de Conférences à l'UMMTO, d'avoir aimablement accepté de participer au jury de ce mémoire.

J'exprime ma profonde reconnaissance à Mr. ADDANE Abdelhamid, professeur à l'USTHB, pour m'avoir accordée un peu de son temps, et livrer de nombreux conseils et encouragements, et aussi pour l'intérêt qu'il a manifesté à l'égard de mon travail.

Mes remerciements s'adressent aussi à Mr. CHIKHI Nourredine, Chef de la station météorologique de Tizi Ouzou, pour toute sa gentillesse et sa serviabilité. C'est grâce aux données qu'il a mises à notre disposition que nous avons pu appliquer notre modèle sur le site de Tizi-Ouzou. Encore une fois nous lui sommes vraiment reconnaissants.

Je remercie très chaleureusement Mr. TAHANOUT et Mr. LAZRI, enseignants chercheurs à l'UMMTO, pour leur aide très précieuse.

Mes remerciements vont aussi à mes parents, à mes chers frères Sidali, Mohammed et Hamza, pour leur compréhension et leur présence encourageante.

Un grand merci du fond du cœur à Mr. REZGUI, qui m'a encouragée, soutenue, poussée à faire mon mémoire. Merci pour tes encouragements et ta confiance.

Je remercie enfin, tous mes camarades de ma promotion pour leur sympathie et leur amitié.

*A la mémoire de ma grande mère*

## Résumé

Le sujet central de ce travail est l'application du Time Delay Neural Network (TDNN) à la prévision des séries temporelles de précipitations. De manière générale, on cherche à prouver l'efficacité du TDNN pour la prévision des précipitations. Dans un premier temps, nous exposons le fondement mathématique des réseaux de neurones. Nous mettons l'accent sur les réseaux de neurones statiques qui ne prennent pas en considération la notion du temps, nous passerons au cas d'un réseau dynamique où le temps joue un rôle fonctionnel applicable pour la prévision des séries temporelles de précipitations. Dans un deuxième temps nous présentons le TDNN, sa structure, son fonctionnement et enfin, les deux algorithmes d'apprentissages : l'algorithme de descente de gradient et celui de Lavenberg-Marquardt appliqués à des données mensuelles de précipitations. Notons qu'une série temporelle de précipitations contient des caractéristiques locales telles que les pics isolés entre plusieurs valeurs basses par exemple. Ces caractéristiques ne sont pas fixes dans le temps, ce qui rend la prévision de leurs occurrences très difficiles. Le Time Delay Neural Network a été proposé pour identifier les caractéristiques locales des précipitations indépendamment de leur position dans le temps. Nous mettons ensuite en évidence les corrélations entre la sortie désirée et la sortie obtenue par le réseau, ainsi l'erreur quadratique moyenne nous permet de mettre en évidence les limites du TDNN pour la prévision des précipitations. Les différents résultats obtenus prouvent l'efficacité de l'approche abordée. Cependant, d'autres applications peuvent être utilisées afin de prouver la robustesse de notre technique telle que par exemple l'utilisation d'une série d'images satellitaires.

**Mots clés :** précipitation, prévision des précipitations, prévision numérique du temps, ARIMA, réseaux de neurones, Time Delay Neural Network.

# SOMMAIRE

INTRODUCTION.....	1
<b>CHAPITRE I : LES MOYENS D'OBSERVATIONS METEOROLOGIQUES</b>	
I.1. Préambule.....	3
I.2. L'observation météorologique.....	3
I.2.1. Les stations pluviométriques au sol .....	3
I.2.1.1. pluviomètre .....	3
I.2.1.2. Pluviographe à augets.....	4
I.2.1.3. Disdromètre (capteur de gouttelettes) .....	5
I.2.2. Mesure des précipitations par télédétection.....	7
I.2.2.1. Radar météorologique .....	7
I.2.2.1.1. Principe de fonctionnement.....	7
I.2.2.1.2. L'équation radar.....	8
I.2.2.1.3. Radar météorologique et mesures hydrologiques.....	9
I.2.2.1.3.1. Facteur de réflectivité radar.....	9
I.2.2.1.3.2. Intensité des précipitations (Relation Z-R).....	10
I.2.2.1.4. L'effet doppler.....	10
I.2.2.2. Satellite météorologique .....	11
I.2.2.2.1. Catégories de satellites météorologiques.....	12
I.2.2.2.2. Caractéristiques spectrales des satellites météorologiques.....	13
I.2.2.2.3. Le principe de mesures par un satellite météorologique.....	14
I.2.2.2.3.1. La mesure par les satellites géostationnaires.....	15
I.2.2.2.3.2. La mesure par satellites à défilements .....	17
I.3. discussion .....	17

## **CHAPITRE II : LES PHENOMENES METEOROLOGIQUES ET LES APPROCHES DE PREVISION**

<b>II.1. préambule.....</b>	<b>18</b>
<b>II.2. les nuages.....</b>	<b>18</b>
<b>II.2.1. Formation des nuages.....</b>	<b>18</b>
<b>II.2.2. Classification des nuages.....</b>	<b>19</b>
<b>II.3. les précipitations.....</b>	<b>20</b>
<b>II.3.1. Formation des précipitations.....</b>	<b>20</b>
<b>II.3.2. Type des précipitations.....</b>	<b>21</b>
<b>II.3.3. Types de précipitations en fonction du mécanisme qui cause le mouvement vertical .....</b>	<b>21</b>
<b>II.4. Les Phénomènes météorologiques complexes.....</b>	<b>21</b>
<b>II.4.1. Les orages.....</b>	<b>21</b>
<b>II.4.1.1. Mécanisme de formation d'un orage.....</b>	<b>21</b>
<b>II.4.1.2. Cycle de vie d'un orage.....</b>	<b>22</b>
<b>II.4.1.3. Classification des orages.....</b>	<b>23</b>
<b>II.4.2. Les tornades.....</b>	<b>25</b>
<b>II.4.2.1. Condition de formation d'une tornade.....</b>	<b>25</b>
<b>II.4.2.2. Types de tornades.....</b>	<b>26</b>
<b>II.4.3. Les cyclones.....</b>	<b>26</b>
<b>II.4.3.1. L'intensité du cyclone.....</b>	<b>27</b>
<b>II.4.4. Les lignes de grains.....</b>	<b>27</b>
<b>II.4.5. Complexe convectif de méso-échelle.....</b>	<b>28</b>
<b>II.5. Prévision des précipitations.....</b>	<b>28</b>
<b>II.6. Les différentes approches de prévision des précipitations.....</b>	<b>30</b>
<b>II.6.1. La prévision numérique du temps (NWP).....</b>	<b>30</b>

<b>II.6.2. Les approches statistiques.....</b>	<b>31</b>
<b>II.6.2.1. La méthode ARIMA (Auto Regressif Integrated Moving Average).....</b>	<b>32</b>
<b>II.6.2.1.1. Typologie du modèle ARIMA.....</b>	<b>32</b>
<b>II.6.2.1.2. Construction de model ARIMA.....</b>	<b>34</b>
<b>II.6.3. Les approches mixtes.....</b>	<b>35</b>
<b>II.7. Discussion .....</b>	<b>35</b>
 <b>CHAPITRE III : RESEAUX DE NEURONES POUR LA PREVISION DES PRECIPITATIONS</b>	
<b>III.1. Préambule .....</b>	<b>36</b>
<b>III.2. Historique .....</b>	<b>36</b>
<b>III.3. Réseaux de neurones pour la prévision.....</b>	<b>37</b>
<b>III.3.1. Définition d'un neurone biologique.....</b>	<b>37</b>
<b>III.3.2. Définition d'un neurone artificiel .....</b>	<b>38</b>
<b>III.3.3. Principe de fonctionnement.....</b>	<b>39</b>
<b>III.3.4. Réseau de neurone artificiel.....</b>	<b>41</b>
<b>III.4. Propriétés des réseaux de neurones.....</b>	<b>42</b>
<b>III.4.1. l'approximation universelle.....</b>	<b>42</b>
<b>III.4.2. la parcimonie.....</b>	<b>43</b>
<b>III.4.3. La généralisation.....</b>	<b>43</b>
<b>III.5. Apprentissage des réseaux de neurones.....</b>	<b>44</b>
<b>III.6. Algorithmes d'apprentissages.....</b>	<b>47</b>
<b>III.6.1. Algorithmes de premier ordre .....</b>	<b>47</b>
<b>III.6.2. Algorithmes du deuxième ordre .....</b>	<b>47</b>
<b>III.7. Prédiction par réseau de neurones .....</b>	<b>47</b>
<b>III.8. Réseaux de neurones pour la prévision des précipitations.....</b>	<b>48</b>
<b>III.8.1. Les réseaux de neurones statiques (feedforward neural network).....</b>	<b>48</b>

<b>III.8.2. Les réseaux dynamiques .....</b>	<b>49</b>
<b>III.8.2.1. Réseaux dynamiques récurrent.....</b>	<b>50</b>
<b>III.8.2.2. Time delay neural network (TDNN).....</b>	<b>51</b>
<b>III.8.2.2.1. La structure de TDNN .....</b>	<b>52</b>
<b>III.8.2.2.2. Architecture et modélisation du TDNN .....</b>	<b>53</b>
<b>III.8.2.2.3. Le fonctionnement de TDNN.....</b>	<b>54</b>
<b>III.8.2.2.4. Apprentissage de TDNN.....</b>	<b>55</b>
<b>III.8.2.2.4.1. Méthode du gradient récursif (Algorithme de 1<sup>er</sup> ordre).....</b>	<b>56</b>
<b>III.8.2.2.4.2. Algorithme avec poids partagé .....</b>	<b>57</b>
<b>III.8.2.2.4.3. Méthode de Levenberg-Marquardt (Méthode de 2<sup>em</sup> ordre).....</b>	<b>58</b>
<b>III.8.2.3. Time Delay Radial Basis Function (TDRBF).....</b>	<b>59</b>
<b>III.9. Tests statistique pour l'évaluation des prévisions.....</b>	<b>60</b>
<b>III.9.1. L'Erreur Quadratique Moyenne (EQM).....</b>	<b>60</b>
<b>III.9.2. Le coefficient de corrélation linéaire (R).....</b>	<b>61</b>
<b>III.10. Discussion .....</b>	<b>61</b>
 <b>CHAPITRE IV : APPLICATION DU TDNN POUR LA PREVISION DES PRECIPITATIONS</b>	
<b>IV.1. Préambule.....</b>	<b>62</b>
<b>IV.2. Les données expérimentales.....</b>	<b>62</b>
<b>IV.3. TDNN pour la prévision d'une série temporelle de précipitation.....</b>	<b>64</b>
<b>IV.3.1. Application de la méthode d'apprentissage de premier ordre : l'algorithme de gradient récursif.....</b>	<b>65</b>
<b>IV.3.2. Application de la méthode d'apprentissage de deuxième ordre : l'algorithme de Lavemberg-Marquardt.....</b>	<b>72</b>
<b>IV.4.Evaluation des performances des algorithmes proposés.....</b>	<b>81</b>
<b>IV.5. Discussion.....</b>	<b>82</b>
<b>CONCLUSION.....</b>	<b>84</b>

# Introduction générale

# Introduction

La connaissance de l'atmosphère a toujours été et restera la préoccupation de la communauté scientifique surtout à notre époque où d'importants bouleversements sont observés. En outre, beaucoup de domaines comme l'agriculture, la navigation, l'aéronautique, le génie civil et bien d'autres ont besoin des données très précises sur l'état de l'atmosphère. Ce dernier peut s'exprimer par la combinaison de diverses grandeurs physiques, telles que la température, l'humidité de l'air, le vent, la pression, le rayonnement solaire et les précipitations. Ces grandeurs caractérisent tout à fait l'atmosphère. Elles sont mesurées à l'aide d'instruments météorologiques dont on cherche constamment à améliorer les performances.

Aujourd'hui les systèmes de mesures très sophistiqués sont mis au point. Ils permettent l'acquisition à distance et en temps réel. De plus, ces systèmes facilitent l'archivage des données (mesures météorologiques).

Il est important de noter qu'une bonne connaissance du phénomène dépend à la fois d'une observation de qualité et aussi du traitement effectué sur les données acquises.

Beaucoup de méthodes de traitement sont utilisées. Chacune ayant ses avantages et ses limitations. Le cadre de ce travail est l'utilisation des réseaux de neurones pour la modélisation et la prévision des données météorologiques. On s'intéressera à la prévision des précipitations en guise d'application.

Les réseaux de neurones sont une technique de traitement de l'information. Concrètement, ils se traduisent par des algorithmes mettant en jeu des concepts associés à la nature du cerveau pour la notion d'apprentissage.

Diverses applications portant sur l'intelligence artificielle et sur les réseaux de neurones proprement dits, furent développées au cours de ces dernières années : La réalisation des filtres adaptatifs et des applications à la reconnaissance de forme et à la commande des moteurs en robotique.

Compte tenu des propriétés de ces réseaux de neurones, l'idée est donc de construire un modèle neuronale de prévision des précipitations contribuant ainsi à la connaissance du climat dans une région. On a choisi pour notre application le Time Delay Neural Network (TDNN) grâce à sa capacité d'identifier les caractéristiques locales indépendamment de leurs positions dans le temps.

Le travail que nous présentons dans ce mémoire est composé d'une introduction, de quatre chapitres, de quelques annexes et d'une conclusion. L'organisation des chapitres est comme suit :

- Chapitre 1 : Ce chapitre est consacré à la présentation des moyens d'observations météorologiques.
- Chapitre 2 : Dans ce chapitre, nous présentons les différents phénomènes météorologiques complexes ainsi que les différentes approches de prévision des précipitations.
- Chapitre 3 : Ce chapitre est consacré à la présentation des réseaux de neurones pour la prévision des précipitations. Le réseau choisi c'est-à-dire le TDNN et les deux méthodes d'apprentissage utilisées seront également exposés.
- Chapitre 4 : Les différents tests, résultats expérimentaux ainsi que les différentes comparaisons sont donnés dans ce chapitre.

Nous terminons notre travail par une conclusion et quelques perspectives ouvertes par ce thème de recherche

# Chapitre I

## **I.1. préambule**

L'observation météorologique est la base incontournable de la prévision météorologique et de la climatologie. L'existence, la connaissance et l'étude de ces données météorologiques et plus particulièrement celles concernant les précipitations, sont très importantes en hydrologie.

Les météorologues ont, pour ce faire, des moyens très perfectionnés à leur disposition : observatoires terrestres, ballons-sondes dans l'atmosphère, navires et bouées météorologiques, satellites météorologiques et un réseau de transmission rapide et fiable qui est l'œuvre de l'OMM (Organisation Météorologique Mondiale) et de l'U.I.T. (Union Internationale des Télécommunications).

## **I.2. L'observation météorologique**

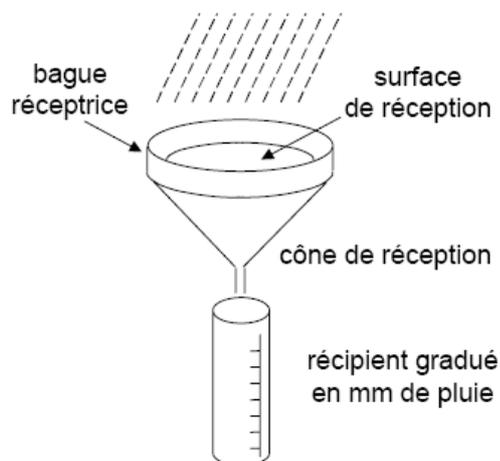
L'observation directe ou indirecte et la mesure des précipitations peuvent se faire à travers différents systèmes ou réseaux avec leur technologie spécifique :

### **I.2.1. Les stations pluviométriques au sol**

Pour un bassin versant donné ou une région donnée, les stations pluviométriques forment un réseau d'observation, et fournissent des mesures ponctuelles. Les données relatives aux stations sont d'une haute importance pour les statistiques climatiques, la planification et la gestion des ressources et les projets de construction. La nature et la densité des réseaux doivent donc tenir compte du phénomène observé, du but des observations, de la précision désirée, de la topographie, de facteurs économiques ou d'autres encore. Parmi les différents instruments permettant la mesure des précipitations au sol, on peut citer entre autre:

#### **I.2.1.1. Le pluviomètre**

Les pluviomètres, dont les modèles sont très divers [Sevruck et Klemm, 1989], indiquent la hauteur d'eau précipitée pendant un intervalle de temps donné, en général 24 heures. L'eau collectée par la surface réceptrice est conservée dans un récipient gradué directement en millimètres d'eau. La lecture du volume permet de connaître la hauteur de pluie précipitée. Après lecture, le récipient est vidé et remis en place.



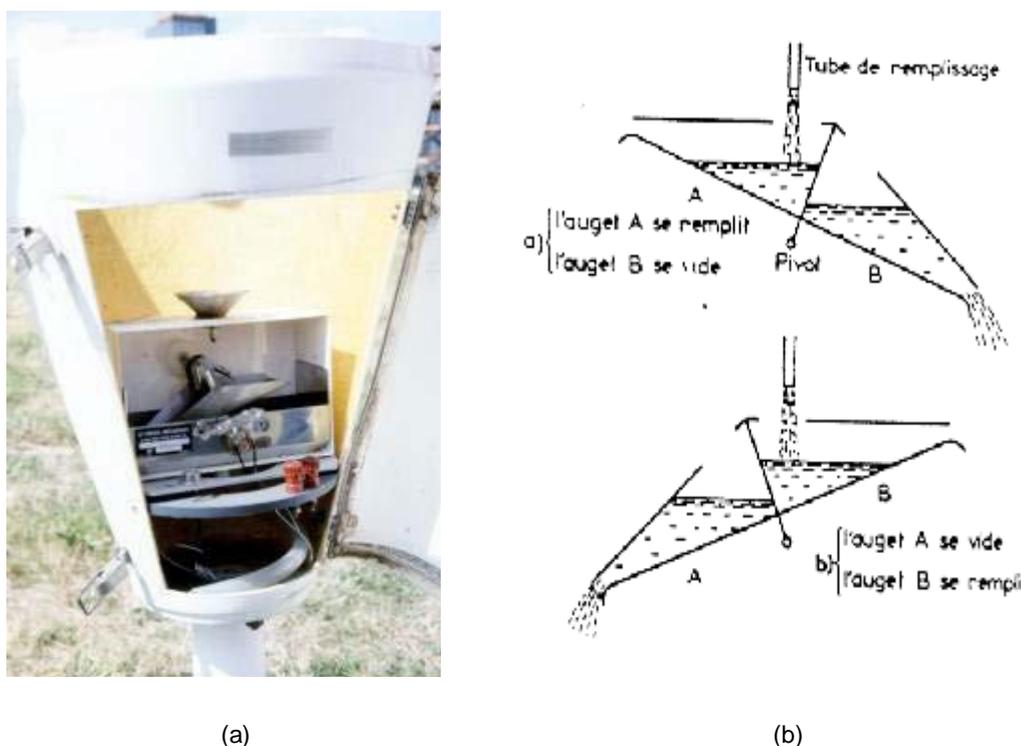
**Figure 1** : Principe du pluviomètre.

### I.2.1.2. Pluviographe à augets

Le pluviographe est un instrument captant la précipitation de la même manière que le pluviomètre mais avec un dispositif permettant de connaître, outre la hauteur d'eau totale, leur répartition dans le temps, autrement dit les intensités.

Il s'agit d'un pluviomètre enregistreur constitué par deux augets de section triangulaire, accolés par l'un de leur côté. Ils peuvent tourner autour d'un axe horizontal. Les augets possèdent deux positions de repos symétriques. Lorsqu'il pleut, l'auget se remplit progressivement et à cause de sa section triangulaire, le centre de gravité de l'ensemble mobile se déplace et entraîne le basculement en position basse et l'auget se vide, remplacé par un autre. Lors de chaque basculement, un dispositif couplé aux augets permet de fermer un contact électrique. Les impulsions électriques sont horodatées et enregistrées sur une centrale d'acquisition de données. Pour mesurer la quantité de pluie tombée, il suffit donc de compter les fermetures de contact et pour connaître l'intensité des précipitations, il faut mesurer le temps séparant deux basculements.

Il existe également des pluviographes à siphon dans lesquels l'eau collectée est stockée jusqu'à un seuil critique où un dispositif de siphonage s'amorce, évacuant l'eau recueillie, et des pluviographes à pesée, le poids du récipient collecteur étant enregistré en continu.



**Figure 2 :** (a) pluviographe à augets basculants, (b) le principe de fonctionnement d'un pluviographe à augets basculants.

### I.2.1.3. Le disdromètre (Capteur de gouttelettes)

Le capteur de gouttelettes ou disdromètre est un instrument utilisé en météorologie pour mesurer la distribution de diamètre des hydrométéores ainsi que leur vitesse de chute. La distribution de diamètre et le nombre de particules de chaque diamètre permet de calculer le taux de précipitation ce qui permet de faire entre autre l'étalonnage des appareils de télédétection comme le radar météorologique. La vitesse de chute est reliée, quant à elle, à la phase de la précipitation. On peut distinguer avec cet appareil la pluie, de la bruine, de la neige, de la grêle ou du grésil sans observateur humain. Cet appareil est utilisé en recherche, dans les stations météorologiques automatiques ou comme capteur de précipitations le long des pistes d'aéroports. Il existe deux types de capteur de gouttelettes : par impact ou optique.

- **Par impact**

C'est un capteur de force qui transforme le choc causé par l'énergie cinétique de l'hydrométéore s'y écrasant en une impulsion électrique. Ce genre de disdromètre est conçu pour la précipitation liquide.

- **Optique (spectropluviomètre optique)**

L'idée générale du spectropluviomètre optique est d'utiliser un faisceau lumineux qui sera coupé par le passage des hydrométéores. On met ainsi un émetteur et un capteur horizontalement avec un espace entre les deux pour le passage de la précipitation. On peut utiliser la lumière visible, l'infrarouge ou même les micro-ondes. Afin de minimiser la diffusion de la lumière par l'air et la précipitation, on utilise maintenant le plus souvent des lasers. Lors du passage d'une goutte au travers du faisceau lumineux, l'énergie lumineuse reçue par la photodiode du capteur décroît. L'amplitude du signal de sortie résultant est proportionnelle au diamètre équivalent de la cible. La vitesse de chute de la goutte au travers du faisceau est déduite du rapport entre la hauteur du faisceau et la durée du signal. Ce genre de capteur peut être utilisé pour mesurer tous les types de précipitations : solide ou liquide.

En disposant deux appareils de ce type à angle droit l'un de l'autre, on obtient une vue à deux dimensions des hydrométéores passant dans leurs faisceaux. En utilisant une caméra au lieu d'un capteur de lumière dans chaque appareil, on peut obtenir des images de chaque cible individuellement. On peut l'utiliser pour étudier la forme des gouttes de pluie selon leur diamètre ou la forme des flocons de neige selon la température extérieure.



(a)



(b)

**Figure 3** : (a) Un module de disdromètre de recherche, (b) Capteur de gouttelettes par impact.

### Ü Les séries chronologiques de pluies

Les séries chronologiques de pluies sont constituées à partir de données pluviographiques en continu sur une durée de plusieurs années ou dizaines d'années. On constitue ainsi une base de données contenant les hyétogrammes et les caractéristiques globales (date, durée, hauteur, intensité, etc.) de tous les événements pluvieux mesurés.

Les séries chronologiques de pluie sont utilisées notamment dans le cadre de la modélisation et la prévision et permettent une étude statistique des résultats obtenus.

## **I.2.2. Mesure des précipitations par télédétection**

La mesure des précipitations par pluviographe est ponctuelle et ne donne que de l'information à une faible distance de la station. Pour connaître les quantités de pluie qui tombent sur une région ou sur un bassin hydrologique, la mesure par télédétection est utilisée.

La télédétection se base sur la mesure du rayonnement électromagnétique émis par les objets dans une large gamme de longueur d'onde. Il est donc essentiel de bien connaître la physique du rayonnement électromagnétique, sa propagation dans l'atmosphère et son interaction avec une interface. Les rayonnements électromagnétiques reçus seront soit directs, c'est-à-dire originaire de la surface, soit indirects, c'est-à-dire réfléchis ou diffusés.

La télédétection désigne la mesure ou l'acquisition d'informations sur un objet ou un phénomène par l'intermédiaire d'un instrument de mesure n'ayant pas de contact avec l'objet étudié. C'est l'utilisation à distance de n'importe quel type d'instrument permettant l'acquisition d'informations sur l'environnement comme par exemple un radar et un satellite.

### **I.2.2.1. Radar météorologique**

Le terme RADAR provient de l'acronyme anglais « RAdio Detection And Ranging », ce signifie approximativement « Détection et Télémétrie Radiométrique », ou plus simplement (détection et estimation de la distance par ondes radio). Le radar a fait son apparition en aéronautique durant la seconde guerre mondiale. Sur les images radar apparaissaient des échos en provenance des particules de précipitations (pluie, neige ou grêle). En aéronautique, ces échos étaient considérés comme des parasites. Les météorologistes y virent au contraire une source précieuse d'information. C'est ainsi que le radar fit son apparition dans le monde de la météorologie. Depuis la seconde guerre mondiale, les systèmes radar se sont sophistiqués et les radars météorologiques occupent désormais une place essentielle dans le dispositif d'observation de l'atmosphère.

#### **I.2.2.1.1. Principe de fonctionnement**

Le fonctionnement du radar est basé sur les propriétés des ondes radio, qui se propagent dans le vide à la vitesse de la lumière. Un émetteur diffuse, au moyen d'une antenne un faisceau d'ondes électromagnétiques concentré dans une direction souhaitée.

Lorsque ces ondes rencontrent un objet, elles sont réfléchies formant ce qu'on appelle écho radar. Cet écho, renvoyé vers le radar est capté par l'antenne qui joue le rôle de récepteur. Le signal réfléchi, après avoir été amplifié, est numérisé et transformé en signal visualisé sur un écran. Puisque le radar a sa propre source d'énergie, une image à deux dimensions de la scène observée est produite de jour et de nuit. L'image radar est contaminée par des échos parasites provenant des obstacles fixes comme par exemple les montagnes ou les constructions au voisinage immédiat de l'antenne.

En analysant le signal réfléchi, il est possible de localiser et d'identifier l'objet responsable de la réflexion, calculer également sa vitesse de déplacement. L'éloignement et la position de la cible observée par rapport à l'émetteur radar peuvent être calculés en se servant du temps écoulé entre l'impulsion de départ et le retour de l'écho avec la relation suivante :

$$Distance = c \frac{\Delta t}{2} \quad (I.1)$$

Où  $C$  : vitesse de la lumière et  $\Delta t$  : temps écoulé entre l'impulsion de départ et le retour de l'écho.

#### I.2.2.1.2. L'équation radar

L'équation radar est la relation fondamentale entre le signal émis par le radar, le signal rétrodiffusé et les propriétés de volume diffusant. Dans le cas où les antennes d'émission et de réception sont confondues (cas monostatique), on a [Sauvageot, 1992] :

$$P_r = P_t \frac{\sigma G_0^2 f^4(\theta, \varnothing) \lambda^2}{(4\pi)^2 r^4} \quad (I.2)$$

Où  $P_r$  : la puissance reçue  
 $P_t$  : la puissance émise  
 $G_0$  : le gain maximum de l'antenne  
 $\sigma$  : la section efficace de rétrodiffusion de la cible  
 $\theta, \varnothing$  : Les angles d'ouvertures en site et en azimut  
 $f(\theta, \varnothing)$  : Fonction caractéristique de l'antenne

En tenant compte de l'atténuation atmosphérique ( $L$ ) et de l'affaiblissement ( $L_r$ ) dans le récepteur, l'équation du radar prendra la forme générale qui est [Sauvageot, 1992] :

$$P_r = \frac{P_t G_0^2 f^4(\theta, \varnothing) \lambda^2 L_r}{(4\pi)^2} L^2 \frac{\sigma}{r^4} \quad (I.3)$$

Pour une cible composée d'un grand nombre d'éléments présents simultanément dans un même volume de résolution  $V$ , avec une répartition supposée homogène (par exemple les gouttes de pluie), la section efficace de rétrodiffusions est égale à  $\sum_i \sigma_i$ , ce terme peut s'écrire sous la forme :

$$\sum_i \sigma_i = V \left( \frac{\sum_i \sigma_i}{V} \right) = V\eta \quad (\text{I.4})$$

Où  $\eta$  : la réflectivité définie comme la section efficace de rétro diffusion moyenne de la cible.

Pour une cible distribuée, l'équation radar s'écrit [Sauvageot, 1992] :

$$P_r = P_t \frac{G_m^2 \lambda^2 \theta_0 \varphi_0 c \tau L^2 L_r}{1024 \pi^2 L n^2} \cdot \frac{\eta}{r^2} \quad (\text{I.5})$$

Où  $\theta_0$  et  $\varphi_0$  sont des angles d'ouvertures à 3 dB en site et azimut,  $c$  : la vitesse de propagation des ondes et  $\tau$  : la durée de l'impulsion radar.

L'équation précédente peut se mettre sous la forme générale suivante :

$$P_r = A \cdot L^2 \frac{\eta}{r^2} \quad (\text{I.6})$$

Où  $A$  est la constante du radar considéré, obtenue à partir des caractéristiques de ce dernier.

### I.2.2.1.3. Radar météorologique et mesures hydrologiques

Les données radar en météorologie servent à préciser le lieu, l'intensité et le type (pluie, neige, grêle, ... etc.) de la précipitation.

#### I.2.2.1.3.1. Facteur de réflectivité radar

Le facteur de réflectivité dépend de la distribution granulométrique des gouttes. Sous l'hypothèse d'un modèle de la distribution de Rayleigh, il s'écrit [Sauvageot, 1992] :

$$Z = \int_{D_{\min}}^{D_{\max}} N(D) \cdot D^6 \cdot dD \quad (\text{I.7})$$

Où  $D$  est le diamètre des hydrométéores (gouttes, flocons, grêlons), et  $N(D) \cdot dD$  exprime le nombre de particules de dimensions comprises entre  $D$  et  $D+dD$ .

Le facteur  $Z$  est une caractérisation de la répartition moyenne des diffuseurs dans le volume d'impulsion  $V$ . Il est donné par :

$$Z = \frac{1}{V} \sum_i D_i^6 \quad (\text{I.8})$$

Compte tenu de forte amplitude des variations de la réflectivité, une échelle logarithmique définie par comparaison avec un niveau de référence de  $1 \text{ mm}^6/\text{m}^3$  exprime le facteur de réflectivité en *dBZ* [Sauvageot, 1992] :

$$Z(\text{dBZ}) = 10 \log ( Z(\text{mm}^6/\text{m}^3) ) \quad (\text{I.9})$$

#### **I.2.2.1.3.2. Intensité des précipitations (Relation Z-R)**

La seule mesure de la réflectivité par le radar ne permet pas la détermination directe de l'intensité des précipitations. Le fondement physique d'une relation quantitative entre la précipitation dans l'air et l'intensité de l'écho radar reçue réside dans le fait que la réflectivité radar est exprimée en fonction de la distribution des tailles des particules (équation I.7). Cette distribution des tailles détermine aussi la hauteur de pluie ou l'intensité de précipitations *R*.

En posant certaines hypothèses sur le comportement de pluie, son homogénéité, et connaissant les lois physiques qui régissent le comportement des gouttes d'eau [William et al 1996], une relation de type puissance entre la réflectivité radar et l'intensité de pluie a été proposée [Marshall et Palmer, 1948]. Elle est de la forme :

$$Z = a R^b \quad (\text{I.10})$$

Où *a* et *b* sont des coefficients qui dépendent de la nature des précipitations.

#### **I.2.2.1.4. L'effet doppler**

L'effet Doppler a été découvert par le physicien autrichien Christian Doppler et le physicien français Hippolyte Fizeau indépendamment l'un de l'autre. Cet effet se manifeste avec toutes les catégories d'ondes sonores, ultraviolettes et lumineuses, lorsqu'une source émettrice et un récepteur sont en mouvement relatif l'un par rapport à l'autre. Il se traduit par une augmentation ou une diminution de la fréquence selon que la source se rapproche ou s'éloigne du récepteur, conformément à une loi précise qui permet de calculer la vitesse de déplacement de la source au récepteur à partir des mesures de fréquence effectuées par le récepteur. Ainsi pour un radar fixe au sol, la comparaison des fréquences émises et rétrodiffusées permettra non seulement de déterminer la position de la cible mais aussi son mouvement. De ce fait, le météorologue se sert du radar Doppler pour détecter les mouvements à l'intérieur des orages qui annoncent le développement de tornades, les vents dans les grandes tempêtes de neige ou les vents dangereux en provenance des orages.

### **I.2.2.2. Satellite météorologique**

La télédétection spatiale est l'ensemble des connaissances et des techniques utilisées pour déterminer les caractéristiques de l'atmosphère de la terre ou d'une autre planète, par des mesures effectuées à partir d'un engin spatiale évoluant à distance convenable de cette dernière. Ce type de méthode d'acquisition utilise la mesure des rayonnements électromagnétiques émis ou réfléchis des objets étudiés dans un certain domaine de fréquence (infrarouge, visible, vapeur d'eau, micro-onde). Ceci est rendu possible par le fait que les objets étudiés (nuages, précipitations ou masses d'aire par exemple) émettent ou réfléchissent du rayonnement à différentes longueurs d'onde et intensités selon leur état.

Un Satellite météorologique est un satellite artificiel qui a comme mission principale de prendre des données pour la surveillance du climat de la terre. Avec chaque nouvelle génération, les senseurs à bord de ces satellites deviennent plus performant et divisent l'information en plus de canaux de telle sorte qu'on peut les utiliser pour différencier les divers phénomènes météorologiques : nuages, précipitations, vents, brouillard, etc.

Depuis la mise sur orbite du premier satellite d'observation TIROS en 1960, plusieurs satellites météorologiques ont été lancés (figure 4). Les données acquises par ces satellites sont liées aux caractéristiques des capteurs embarqués.

Il existe deux catégories bien distinctes de satellites météorologiques : satellites météorologiques géostationnaires et satellites météorologiques à défilement, souvent nommés improprement satellites météorologiques défilants. Ces moyens de télédétection ont rendu possible l'observation des masses nuageuses qui s'étendent sur de grandes superficies.

Le système opérationnel complet comprend 5 satellites géostationnaires : le satellite européen METEOSAT de première (1-7) et seconde (MSG1, MSG2, MSG3) génération, les deux satellites américains GOES, un satellite japonais GMS et un satellite russe GOMS. Tous ces satellites (figure 4) sont dans le plan équatorial à 36 000 Km de la surface de la terre et disposés de façon à assurer une observation optimale de celle-ci.

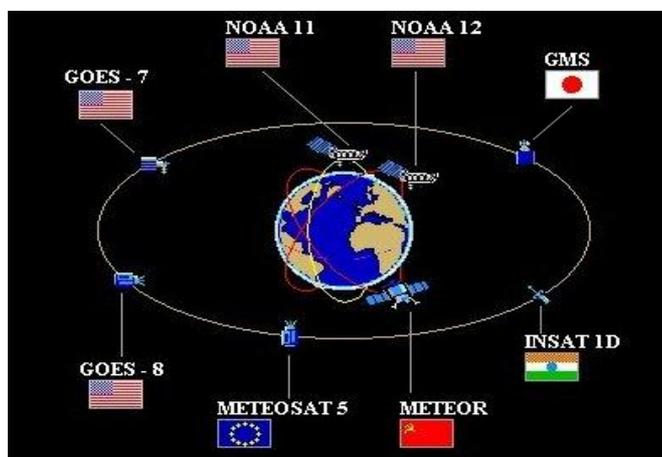


Figure 4 : Satellites météorologiques.

#### 1.2.2.2.1. Catégories de satellites météorologiques

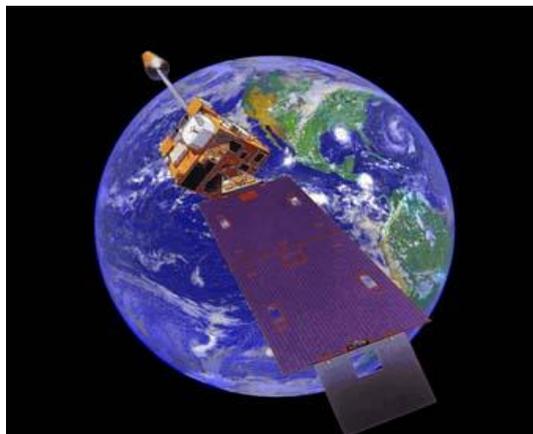
Il existe deux grandes catégories de satellites météorologiques :

##### a) Satellites géostationnaires

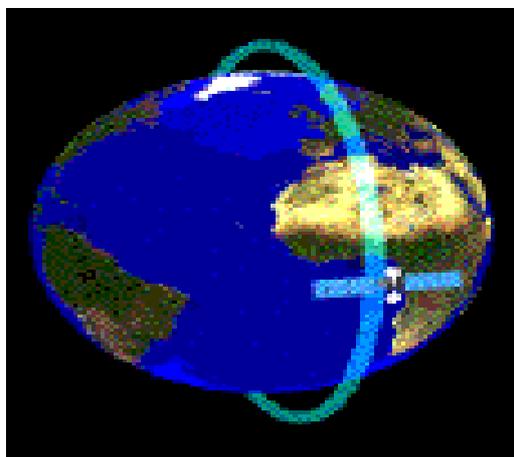
- Ils permettent d'observer toujours le même endroit sur terre ;
- Ils tournent à la même vitesse que la terre et paraissent donc immobiles aux yeux d'un observateur terrestre ;
- Ils sont placés à 36 000 Km au-dessus de la surface terrestre ;
- Ils couvrent environ 1/3 de la planète (surtout les régions peuplées, ils ne représentent pas bien les régions polaires) ;
- Ils tournent sur eux-mêmes à 100 tours par minutes ;
- Ils peuvent fournir de façon continue des informations couvrant une très vaste zone circulaire, centrée sur un point fixe de l'équateur et s'étendant sur environ 42% de la surface de la terre.

##### b) Satellites défilants

- Ils permettent d'observer tous les endroits sur la terre ;
- Ils font le tour de la terre d'un pôle à l'autre ;
- Ils sont placés de 600 à 1000 Km au-dessus de la surface terrestre ;
- Ils sont bons pour les régions polaires ;
- Ils ont comme inconvénient de ne passer que deux fois par jour sur chaque point de la planète.



**Figure 5 :** Orbite géostationnaire [<http://rsd.gsfc.nasa.gov/geos>].



**Figure 6 :** Orbite héliosynchrone [<http://www.met.fsu.edu/explores/IMAGE-IN/image-in.html>].

#### **I.2.2.2.2. Caractéristiques spectrales des satellites météorologiques**

Les satellites météorologiques opèrent sur plusieurs canaux qui correspondent chacun à une bande particulière de longueur d'onde électromagnétique. On peut citer entre autre :

##### **a) Le canal visible**

Les images visibles représentent la quantité de lumière visible rétrodiffusée par les nuages ou la surface de la terre. Plus les nuages sont épais plus la réflexion est forte et plus l'image est blanche. Ce canal n'est utilisable que lorsqu'il fait jour sur la terre.

##### **b) Le canal infrarouge**

Dans ce canal, le radiomètre mesure l'émission thermique de la cible et donc sa température, cela de jour comme de nuit. S'il n'y a pas de nuage sur la ligne de visée, la cible est la surface terrestre, s'il y a des nuages, la cible est leur sommet. Ainsi on en déduit

l'altitude du sommet de nuage : plus il est situé haut dans l'atmosphère plus il est froid et plus l'image est blanche.

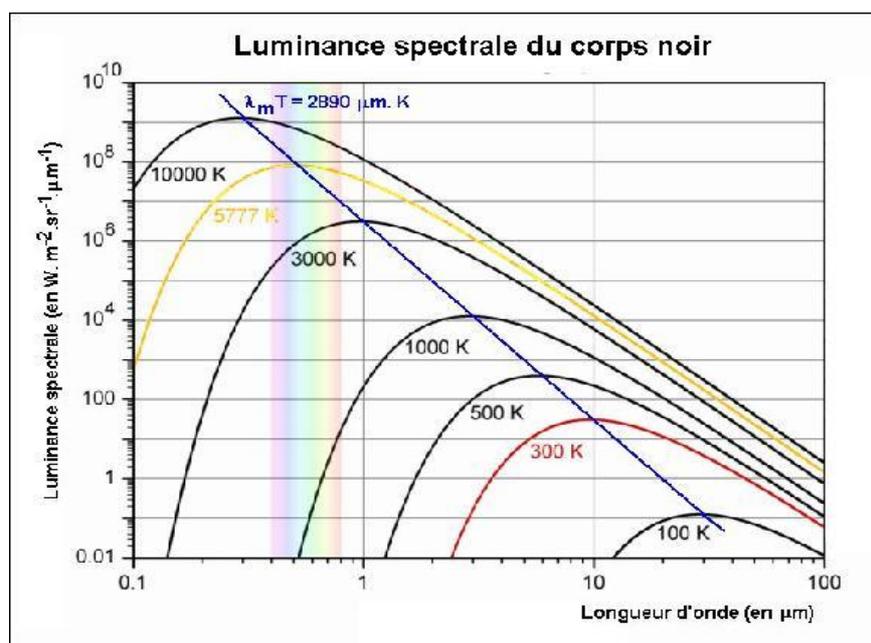
### c) Le canal vapeur d'eau

Dans le canal vapeur d'eau, l'imageur mesure la quantité de vapeur d'eau ou hydrométéores présents dans la moyenne atmosphère. C'est un canal spécial où la vapeur d'eau absorbe très bien le rayonnement. Plus les zones sont humides plus l'image est blanche.

#### I.2.2.2.3. Le principe de mesure par satellite météorologique

Chaque satellite météorologique fournit plusieurs milliers de données très diverses tous les jours. Ces informations sont ensuite utilisées par les météorologues pour suivre en temps réel l'évolution de la météo et en extraire des données dérivées afin de prévoir le temps.

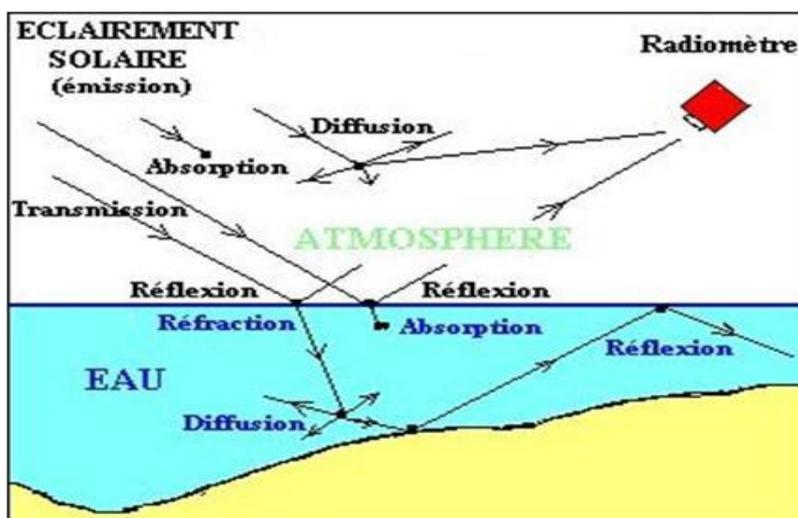
L'avantage du satellite est qu'il peut contrairement à un autre moyen de mesure, prélever les données n'importe où sur terre. Le satellite analyse les ondes électromagnétiques provenant des différents objets. En effet, chaque corps n'a pas le même albédo ce qui permet au satellite de différencier eau et terre par exemple, et suivant le but recherché, les satellites opèrent dans différentes bandes spectrales, dans différents canaux.



**Figure 7** : Distribution du flux d'énergie émis par un corps noir porté à différentes températures.

Pour établir ses mesures à distance, le satellite météorologique utilise des capteurs passifs (le radiomètre), qui se contente de recueillir les ondes émises par le milieu ciblé : c'est la

télédéttection passive. Il peut également utiliser des capteurs actifs (radar, lidar, interféromètres ou diffusiomètre), qui captent parmi les ondes d'autre nature qu'ils ont précédemment émises vers lui, celles qui ont rétrodiffusé : c'est la télédection active.



**Figure 8** : Un schéma expliquant comment un satellite passif reçoit les différentes données.

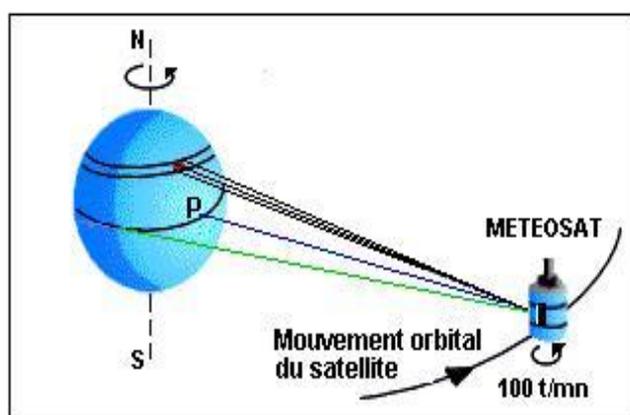
#### I.2.2.3.1. La mesure par satellite géostationnaire

Le principal instrument utilisé sur les satellites géostationnaires est le radiomètre imageur. L'image est obtenue par un balayage combinant la rotation du satellite sur lui-même et l'orientation du radiomètre. Le satellite européen METEOSAT-7 est capable de fournir une image complète du globe toutes les demi-heures. Son successeur MSG (Météosat Seconde Génération) est plus performant et fournit par contre une image toutes les quinze minutes. En effet, l'imageur de MSG est plus perfectionné : il possède douze canaux contre trois pour la première génération, ce qui lui permet de mesurer d'autres paramètres météorologiques. Sur les 12 canaux, on trouve deux canaux visibles, neuf canaux dans le domaine infrarouge et un douzième canal visible HRV à large bande ne couvrant qu'une partie du disque terrestre. La résolution spatiale passe de 2.5 Km à 1 Km pour le canal visible à large bande et de 5 à 3 Km pour tous les autres canaux. Les caractéristiques des 12 canaux de MSG ainsi que leurs objectifs sont résumés dans la table 1.

Canal	Autre appellation	Longueur d'onde centrale ( $\mu\text{m}$ )	Objectif principal
C1	VIS 0.6	0.635	Surface, nuages, champs de vent
C2	VIS 0.8	0.81	Surface, nuages, champs de vent
C3	NIR 1.6	1.64	Surface, phase du nuage
C4	IR 3.9	3.90	Surface, nuages, champs de vent
C5	WV 6.2	6.25	Vapeur d'eau, nuages à haute altitude, instabilité atmosphérique
C6	WV 7.3	7.35	Vapeur d'eau, instabilité atmosphérique
C7	IR 8.7	8.70	Surface, nuages, instabilité atmosphérique
C8	IR 9.7	9.66	Ozone
C9	IR 10.8	10.80	Surface, nuages, champs de vent, instabilité atmosphérique
C10	IR 12.0	12.00	Surface, nuages, instabilité atmosphérique
C11	IR 13.4	13.40	Hauteur des cirrus, instabilité atmosphérique
C12	HRV	0.4-1.1 $\mu\text{m}$	Surface, nuages

**Table 1** : Caractéristiques et objectifs principaux des 12 canaux de MSG.

Cet accroissement des capacités de MSG par rapport à Météosat répond à l'évolution des besoins de la prévision immédiate et de la gestion du risque météorologique à courte échéance et ceci grâce à sa capacité de caractériser les systèmes météorologiques actifs, leurs évolution rapide et leurs probabilité de développement.



**Figure 9** : Le principe de mesure par satellite géostationnaire.

### **I.2.2.3.2. La mesure par satellite à défilement**

Tout comme les satellites géostationnaires, les satellites à défilement sont munis d'un radiomètre à plusieurs canaux qui analyse la terre et son atmosphère. Du fait de leurs mobilité, ils peuvent analyser les régions polaires, contrairement aux géostationnaires. L'image d'une partie du globe est obtenue en combinant le mouvement du satellite et le balayage du radiomètre dans la direction perpendiculaire à son mouvement. Mais le principal atout des satellites à défilement est le radiomètre sondeur qui possède un plus grand nombre de canaux. On commence aussi à voir apparaître un tout nouveau type de sondeur « les interféromètres » qui possèdent un plus grand nombre de canaux.

Ainsi le satellite à défilement Metop est équipé du sondeur IASI (Interféromètre Atmosphérique de Sondage par Infrarouge) muni de plus de 8 000 canaux dans l'infrarouge. Les météorologistes espèrent ainsi pouvoir mesurer les profils atmosphériques avec une bien meilleure résolution verticale (1 à 2 Km) et une meilleure précision (1°C pour la température, 10% pour l'humidité) contrairement au sondeur classique (résolution verticale de quelques kilomètres et une précision de l'ordre de 2°C pour la température et 30% pour l'humidité). IASI peut aussi évaluer la température à la surface d'un océan.

### **I.3. discussion**

Les moyens d'observations présentés dans ce chapitre fournissent des informations sur la scène météorologique observée. La multiplication des moyens d'observations d'une même scène est d'une grande importance sur la compréhension des phénomènes complexes et variables dans le temps. Le prochain chapitre abordera les phénomènes météorologiques et les différentes approches de prévision des précipitations.

# Chapitre II

## **II.1. préambule**

La complexité des processus atmosphériques illustre le défi que constitue leur prévision. De plus, parmi l'ensemble des éléments du temps sensible, les précipitations constituent la variable la plus ardue à prévoir. Il n'est donc pas étonnant que malgré l'existence d'enjeux importants, la prévision des pluies reste un domaine encore imparfaitement résolu par les techniques actuelles.

Ce chapitre est consacré à la présentation des différents phénomènes météorologiques, et les différentes approches de prévision des précipitations.

## **II.2. Les nuages**

Physiquement, un nuage résulte de la condensation de la vapeur d'eau par refroidissement d'une masse d'air humide. Un nuage est donc un ensemble de minuscules gouttelettes d'eau ou de cristaux de glace en suspension dans l'atmosphère. Les nuages entrent dans le cycle de l'eau et donc de l'énergie dans l'atmosphère. Ils sont une composante essentielle du climat. Ainsi, ils réfléchissent une grande partie du rayonnement solaire vers l'espace, mais captent pratiquement tout le rayonnement infrarouge émis par la terre. De ce fait, ils influencent sur la formation de plusieurs phénomènes météorologiques.

### **II.2.1. Formation des nuages**

Les principaux processus de formation des nuages sont :

#### **a. La Convection**

La convection est l'un des processus qui permettent la formation des nuages. Au niveau du sol, lorsque le soleil brille, l'air qui contient de la vapeur d'eau se réchauffe. Il commence à s'élever, et puisqu'il s'élève, il commence à se refroidir. Un nuage se forme lorsque l'air humide se refroidit jusqu'à atteindre une température critique : la vapeur d'eau condense alors sur toutes les particules qui flottent dans l'air, et forme ainsi des nuages.

#### **b. Le relief**

Les nuages peuvent aussi se former grâce à la topographie de la terre, c'est à dire grâce aux collines ou aux montagnes. De tels nuages s'appellent les nuages orographiques. L'air est obligé de passer par dessus une montagne, et lorsqu'il s'élève, il se refroidit. S'il se refroidit jusqu'à son point de condensation, la condensation a lieu et l'eau contenue dans l'air devient liquide : elle se transforme donc en nuage.

### **c. L'effet de Foehn**

Quand l'air s'élève au dessus des montagnes, il refroidit et arrive à saturation en vapeur d'eau. La condensation a alors lieu et l'eau passe de l'état vapeur à l'état liquide. L'eau est alors sous la forme d'un nuage ou de pluie et reste sur place pendant que l'air, lui, continue à s'élever. Quand l'air passe de l'autre côté de la montagne, il s'est débarrassé d'une partie de son eau. Il est plus sec et plus chaud. Cette différence de température et d'humidité entre les deux versants de la montagne est connue sous le nom d'effet de Foehn.

### **d. La rencontre de deux masses d'air**

Il n'y a pas que le relief qui oblige l'air à s'élever : lorsqu'une masse d'air chaud rencontre une masse d'air froid, donc plus lourd, l'air chaud est obligé de monter. La frontière entre l'air chaud et l'air froid est appelée front. Quand l'air chaud monte en altitude, il se refroidit : des nuages peuvent alors se former.

### **e. Les mouvements horizontaux**

Parfois, le vent transporte de l'air chaud et humide à un endroit qui peut être beaucoup plus froid. Quand cet air chaud souffle sur une surface froide, il va se refroidir, l'humidité qu'il contient va condenser et former du brouillard, qui est en fait un nuage.

## **II.2.2. Classification des nuages**

Les nuages peuvent être classifiés en fonction de plusieurs facteurs. Une classification générale et précise de chaque type de nuage permet ainsi à l'observateur d'identifier les nuages dans le ciel.

### **a. Classification selon le genre**

Le genre dans une taxinomie est le regroupement d'espèces. La classification des nuages par le genre repose avant tout sur leur forme générale.

### **b. Classification selon l'espèce**

Il s'agit de la nature propre à plusieurs choses. Cela permet de les considérer comme appartenant à une catégorie distincte. Ainsi, l'espèce d'un nuage est plus précise que son genre. Dans le but d'affiner la classification des nuages, ces derniers sont regroupés selon leur espèce.

### **c. Classification selon la variété**

Une variété est une subdivision de l'espèce. Donc, une variété est encore plus précise qu'une espèce. Les variétés de nuages sont des variations mineures de la forme du nuage.

## **II.3. Les précipitations**

En météorologie, le terme précipitation désigne des cristaux de glace ou des gouttelettes d'eau qui, ayant été soumis à des processus de condensation et d'agrégation à l'intérieur des nuages, sont devenus trop lourds pour demeurer en suspension dans l'atmosphère et tombent au sol ou s'évaporent en virga avant de l'atteindre.

### **II.3.1. Formation des précipitations**

Pour qu'il y ait des précipitations, il faut que les gouttelettes ou les cristaux composant les nuages se transforment en gouttes de pluie. Ce phénomène est lié à l'accroissement de ces éléments dont la masse devient suffisante pour vaincre les forces d'agitation. Ce grossissement peut s'expliquer par les trois processus suivants :

#### **a. La condensation**

Les gouttes commencent à se former dans l'air généralement au-dessus du point de congélation quand l'air soulevé devient légèrement sursaturé par rapport à la température environnante. Pour cela, il faut cependant des noyaux de condensation, poussières ou grains de sel, sur lesquels la vapeur d'eau se dépose. Il y a d'abord formation de très fines gouttes qui donnent le nuage. À mesure que ces gouttes montent, elles passent sous le point de congélation mais resteront surfondues s'il n'y a pas présence de noyaux de congélation.

#### **b. Coalescence**

La coalescence est l'amalgamation de deux ou plusieurs gouttelettes par collision pour en former une plus grosse.

#### **c. L'effet Bergeron**

L'effet Bergeron est le plus efficace des processus de formation des gouttes de pluie ou de neige. Lorsque des cristaux de glace se forment finalement par congélation de gouttelettes, ils ont une pression de saturation moindre que celle des gouttelettes environnantes. Les gouttelettes s'évaporent donc et la vapeur d'eau va se déposer sur les cristaux.

### II.3.2. Type des précipitations

La précipitation peut prendre les formes suivantes :

- **Sous forme liquide** : Pluie, Bruine, Pluie verglaçante, Bruine verglaçante.
- **Sous forme solide**: Neige, Neige en grains, Neige roulée, Grésil, Grêle, Granule de glace, Cristal de glace.

### II.3.3. Types de précipitations en fonction du mécanisme qui cause le mouvement vertical

On distingue deux types de précipitations en fonction du mécanisme qui cause le mouvement vertical :

#### a. Les précipitations stratiformes

Qui viennent du soulèvement lent et à grande échelle de l'humidité qui se condense uniformément.

#### b. Les précipitations convectives

Les précipitations convectives résultent de la brusque élévation de masses d'air chargées d'humidité, par la poussée d'Archimède, à cause de l'instabilité de l'air.

## II.4. Les Phénomènes météorologiques complexes

Dans l'atmosphère on distingue plusieurs phénomènes météorologiques complexes :

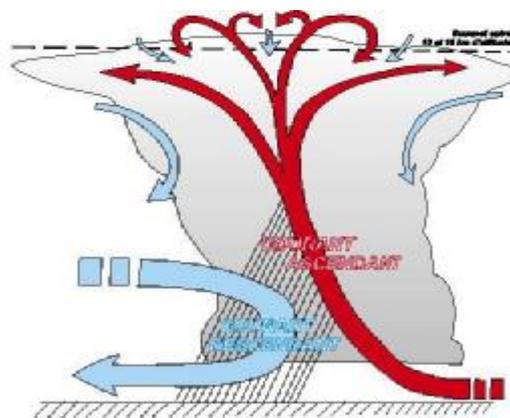
### II.4.1. Les orages

Un orage est une perturbation atmosphérique d'origine convective associée à un type de nuage particulier : le cumulonimbus. Ce dernier est à forte extension verticale, il engendre des pluies fortes à diluviennes, des décharges électriques, de foudre accompagnées de tonnerre. Dans des cas extrêmes, l'orage peut produire des chutes de grêle, des vents très violents, et rarement des tornades.

#### II.4.1.1. Mécanisme de formation d'un orage

Pour qu'il y ait orage, il faut qu'il y ait un mouvement vertical rapide de l'air instable et humide vers le haut. Une masse d'air humide s'élève rapidement. En s'élevant, l'air se refroidit et se condense. Cette condensation provoque un dégagement de chaleur latente. Ainsi, l'air saturé devient plus chaud que l'air ambiant et continu donc de s'élever naturellement, on parle de cellules convectives qui donnent forme à des cumulus ou des cumulonimbus. De violents courants d'air verticaux entraînent l'humidité, les fragments de

glace, les grêlons et les gouttelettes d'eau à l'intérieur du nuage. Ces frottements créent une séparation des charges électriques. Les particules les plus fines, chargées d'électricité positive, se dirigent vers le sommet. Les plus lourdes chargées d'électricité négative vont vers la base. À cause des mouvements verticaux très violents et puissants, il y a déséquilibre entre les charges électriques à l'intérieur du nuage ainsi qu'à l'extérieur, entre la base négative du nuage et le sol positif. Quand les charges accumulées deviennent trop importantes, et surtout lorsqu'il y a opposition directe entre ces charges, il y a décharge électrique (éclair) et un orage éclate.



**Figure 10** : Formation d'un orage. <http://www.meteo.fr>.

#### II.4.1.2. Cycle de vie d'un orage

Les orages ont une durée de vie assez faible. Ils durent de 5 minutes à 30 ou 60 minutes en général. Leur vitesse de déplacement se situe entre 30 Km/h et 60 Km/h. La mort d'un orage survient lorsque son carburant vital s'épuise ou ne peut alimenter le nuage. Ceci survient lorsque, d'une part, la pluie qui tombe du nuage vient à ralentir ou renverser le mouvement ascendant de l'air. D'autre part, l'air étant refroidi au contact avec la pluie qui tombe, ceci limite l'apport d'énergie au nuage. Le cycle de vie d'une cellule convective se décompose en trois parties :

##### a. Etape de formation

Marquée par l'intensification des courants ascendants, qui a pour conséquence l'augmentation rapide de la hauteur du sommet du nuage et son extension horizontale [Arnaud. Yand Laloë.F., 1991].

### b. Etape de maturité

Marquée par la présence simultanée de courants ascendants et descendants. L'air est entrainé par les précipitations, et l'évaporation de celles-ci au contact d'un air plus chaud a pour effet de refroidir l'air environnant, augmentant ainsi sa densité et favorisant les courants descendants. L'enclume des cumulonimbus commence alors à s'étaler à la base des couches stables supérieures [Arnaud. Yand Laloë.F., 1991].

### c. Etape de dissipation

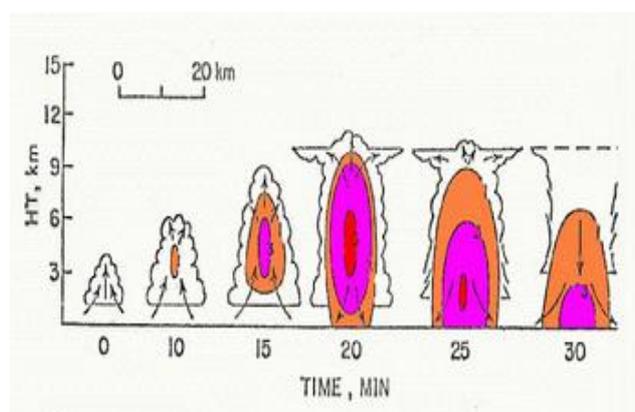
C'est le stade de dissipation des courants ascendants, c'est-à-dire l'épuisement du stock d'énergie latente emmagasinée dans le nuage. L'enclume atteint son développement horizontal maximum et commence à se dissiper [Arnaud. Yand Laloë.F., 1991].

#### II.4.1.3. Classification des orages

Les orages sont classés en plusieurs catégories selon l'énergie potentielle de convection disponible (EPCD) et le cisaillement du vent avec l'altitude :

##### a. L'orage unicellulaire

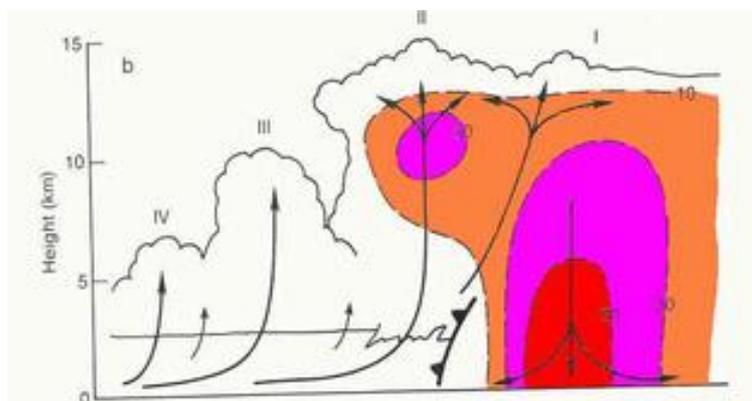
L'orage unicellulaire est caractérisé par une faible énergie (EPCD de 500 à 1000 J/Kg) avec peu ou pas de changement des vents avec l'altitude. Donc le cycle de vie d'environ 30 à 60 minutes de ces orages est caractérisé par un courant ascendant plus ou moins fort et vertical. Ce type d'orage est le plus fréquent. Il peut être associé à une forte averse et des rafales de vent. Les pluies ne sont presque jamais torrentielles et les chutes de gros grêlons sont rarissimes. Dans les régions arides du globe, l'évaporation peut être telle que la pluie n'atteigne pas le sol et forme de la virga sous le cumulonimbus.



**Figure 11** : Cycle de vie d'un orage unicellulaire.

### b. Orage multicellulaire

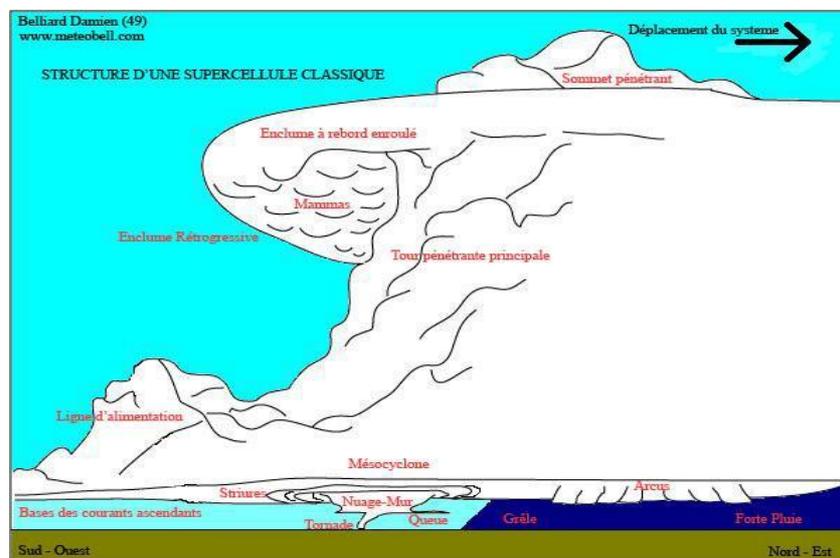
Les orages ne sont pas toujours isolés. Lorsque l'instabilité atmosphérique est marquée (soit de l'air froid en altitude et un changement de direction des vents aux différents étages de l'atmosphère), les orages s'organisent en paquets que l'on appelle orages multicellulaires. En général, l'EPCD est moyen dans ce type d'orage, soit entre 800 et 1 500 J/Kg. Selon l'énergie et l'humidité disponibles, ce type d'orage peut donner des rafales de vents violentes, des pluies diluviennes et très rarement des tornades.



**Figure 12:** Cycle de vie d'un orage multicellulaire.

### c. Orage supercellulaire

C'est certainement l'orage le plus puissant qui puisse exister sur terre. Sa formation est relativement exceptionnelle et est responsable d'un nombre tout à fait remarquable d'orage violent. On l'appelle **supercellulaire** car il n'est formé que d'une unique cellule convective. C'est donc un orage monocellulaire ultra puissant. Elles peuvent produire de la grosse grêle, des vents destructeurs et des pluies torrentielles. De plus, si un cisaillement horizontal du vent en surface est transformé en tourbillon vertical par le courant ascendant et si la rotation est accentuée par le courant descendant, ces supercellules peuvent produire des **tornades**. Un tel cumulonimbus est présenté sur la figure suivante :



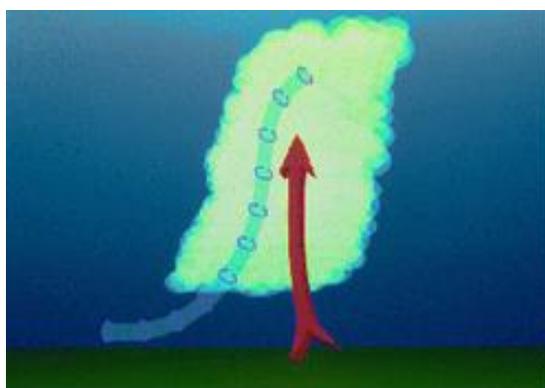
**Figure 13 :** Vue conceptuelle d'un supercellulaire.

## II.4.2. Les tornades

Une tornade est un vortex (tourbillon) de vents extrêmement violents, prenant naissance à la base d'un nuage d'orage (cumulonimbus) lorsque les conditions de cisaillement des vents sont favorables dans la basse atmosphère. De très faibles tornades peuvent également se développer sous des nuages d'averses (cumulus bourgeonnant).

### II.4.2.1. Condition de formation d'une tornade

Une tornade est formée quand un cisaillement des vents dans les premiers kilomètres de l'atmosphère crée une rotation dans l'axe horizontal, cette rotation entre en interaction avec un fort courant ascendant, l'axe horizontal peut basculer et devenir une rotation verticale (figure 14). La rotation sera renforcée ensuite par la circulation de surface.



**Figure 14 :** Basculement du vortex par le courant ascendant.

#### II.4.2.2. Types de tornades

Il existe deux types de tornades :

**Type A** : rassemble les tornades, les plus intenses qui se forment au sein d'un orage violent.

**Type B** : les tourbillons sont générés en l'absence d'orage, la tornade est donc plus difficile à expliquer.

#### II.4.3. Les cyclones

Un cyclone est un terme météorologique qui désigne une grande zone où l'air atmosphérique est en rotation autour d'un centre de basse pression local. On parle également de dépression et de système cyclonique. Par extension, la circulation cyclonique est la direction que prendra le flux d'air autour d'une dépression ou d'un creux barométrique, soit antihoraire dans l'hémisphère nord et horaire dans celui du sud. On distingue principalement trois types de cyclones :

##### a. Cyclone tropical

Le cyclone tropical est une perturbation d'échelle synoptique non accompagnée d'un système frontal, prenant naissance au-dessus des eaux tropicales ou subtropicales et présentant une activité convective organisée ainsi qu'une circulation cyclonique plus intense en surface qu'en altitude. Dans certaines circonstances, lorsque sont réunies un certain nombre de conditions favorables, des formations nuageuses se développent et s'enroulent autour d'un centre de rotation, appelé tourbillon, définissant en surface une circulation fermée. C'est cette perturbation atmosphérique composée de nuages organisés en bandes spiralées qu'on appellera cyclone tropical. Au sein de ce tourbillon, la pression atmosphérique va baisser et une dépression se crée.

##### b. Cyclone extratropicale

Un cyclone extratropical, parfois nommé cyclone des latitudes moyennes est un système météorologique de basse pression, d'échelle synoptique, qui se forme entre la ligne des tropiques et le cercle polaire. Il est associé à des fronts, soit des zones de gradients horizontaux de la température et du point de rosée, que l'on nomme aussi zones baroclines. Les cyclones extratropicaux ont donc des caractéristiques différentes des cyclones tropicaux et des cyclones polaires plus au nord, qui sont alimentés par la convection. Ils sont en fait les dépressions météorologiques qui passent quotidiennement sur la majorité du globe. Avec les anticyclones, ils régissent le temps sur la terre, produisant nuages, pluie, vents et orages.

### c. Cyclone subtropicale

Ils se forment généralement au-delà des tropiques, jusqu'à une latitude de 50° (nord et sud). En effet, on y retrouve une activité orageuse autour de son centre qui tend à lui former un cœur chaud mais on le retrouve dans une zone frontale faible. Avec le temps, la tempête subtropicale peut devenir tropicale.

### d. Cyclone polaire

Un cyclone polaire est un système de basse pression, couvrant habituellement une zone de 1000 à 2000 Kilomètres, identique aux cyclones extratropicaux, dit dépression des latitudes moyennes, mais qui se développent dans les latitudes arctiques et antarctiques. Les cyclones polaires se forment le long du front arctique, une bande de contraste thermique entre l'air venant des pôles et celui venant des latitudes moyennes du globe. L'air circule dans ces dépressions dans le sens contraire des aiguilles d'une montre dans l'hémisphère nord et dans le sens inverse dans l'hémisphère sud sous l'impulsion de la force de Coriolis. Ces systèmes peuvent se produire tout au long de l'année. Cependant, les cyclones d'été tendent à être plus faibles que ceux d'hiver car le contraste thermique est moins marqué. La présence de la banquise et d'eau sans glace côte à côte accentue également ce contraste à certaines périodes de l'année.

#### II.4.3.1. L'intensité du cyclone

L'intensité des cyclones est déterminée par la force du vent maximum qu'il engendre :

- Si le vent est inférieur à 63 Km/h, c'est une dépression tropicale.
- Si le vent est entre 63Km/h et 117 Km/h, c'est une tempête tropicale et le cyclone reçoit alors un nom.
- Si le vent est supérieur à 118 Km/h, c'est un ouragan.

#### II.4.4. Les lignes de grains

Un grain est un événement météorologique au cours duquel la vitesse du vent s'accroît de façon brusque et marquée avec un net changement de direction (45° à 90 °), et qui ne dure que quelques minutes. Ce phénomène, particulièrement redouté des voiliers, est fréquemment accompagné d'averses de pluie, de neige ou d'orages. Selon la définition de l'Organisation Météorologique Mondiale, les rafales de vent doivent être au moins 28 km/h supérieures à la vitesse moyenne du vent pendant une période d'au moins une minute.

#### **II.4.5. Complexe convectif de méso-échelle**

Un Complexe Convectif de Méso-échelle (CCM) est un ensemble orageux se formant généralement en fin de journée à partir d'orages dispersés et qui atteint son apogée durant la nuit alors qu'il s'organise en une large zone circulaire. Après leur formation, ils dérivent dans le flux d'altitude et donnent principalement des précipitations intenses causant des inondations sur de larges régions.

#### **II.5. Prévision des précipitations**

Le choix de la mise en œuvre de l'une ou l'autre des méthodes de prévision de précipitations existantes est naturellement guidé par les besoins de l'utilisateur. Les principales préoccupations de ce dernier sont de trois ordres :

- L'échéance de prévision qui dépend de l'échelle de temps des données.
- Le besoin d'adaptation locale qui dépend de leurs échelles spatiales.
- La forme revêtue par la prévision qui dépend de la nature de ces enjeux, mais aussi des capacités de poste traitement de l'utilisateur.

Le principe d'une prévision météorologique est de prévoir, à partir d'un état mesuré à un instant donné, l'évolution de l'atmosphère pour un instant ultérieur. Une prévision est donc indissociable de l'échéance pour laquelle elle a été émise. Selon un schéma idéal, l'échéance utilisée est définie par les besoins et les contraintes de l'utilisateur. Ce choix impose ensuite le type de prévision à mettre en œuvre. Il est d'usage courant en météorologie de considérer quatre catégories d'échéances de prévision. Chacune correspond à la prépondérance d'un phénomène météorologique particulier et donc à un certain type de méthode de prévision :

##### **a. La prévision immédiate (nowcasting)**

Elle répond aux échéances allant de 0 à 3 heures, ce qui correspond au cycle de vie d'une cellule orageuse. Cette technique repose principalement sur les observations issues de l'imagerie radar et sur l'advection des cellules pluvieuses déjà existantes.

##### **b. La prévision à courte échéance**

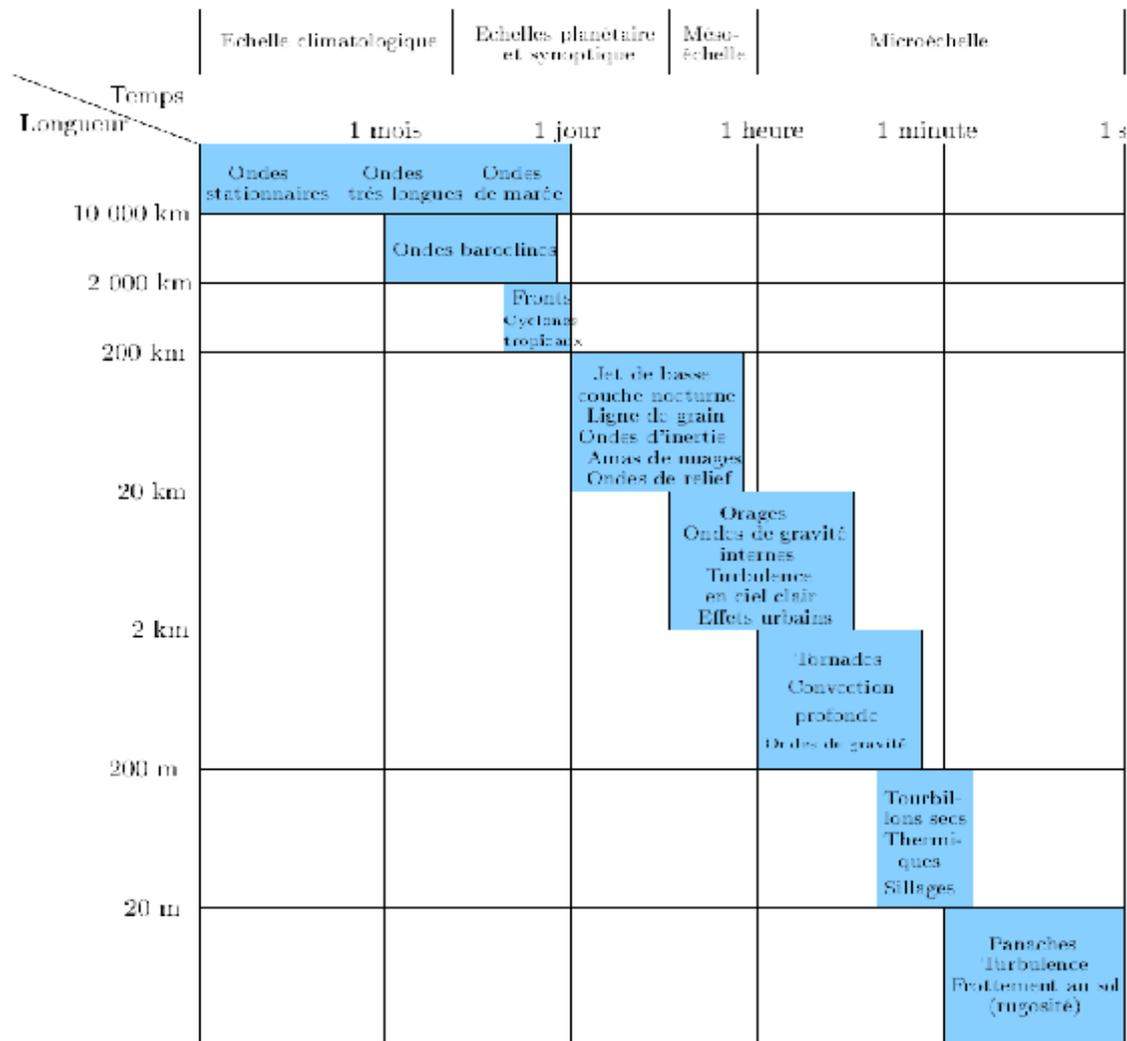
Elle couvre la gamme d'échéance allant de 6 à 72 heures et correspond aux phénomènes rapides de la circulation synoptique, tel que le passage d'un front. C'est actuellement le domaine privilégié des modèles déterministes de prévision numérique.

**c. Prévision à moyenne échéance**

Elle se situe dans des échéances allant de 72 heures (3 jours) à 10 jours voir 15 jours. Elle correspond à la frontogénèse : les fronts qui seront observés ne seront pas encore apparus. Les outils utilisés ici mêlent les approches stochastiques et le modèle numérique déterministe, tel que la prévision d'ensemble.

**d. La prévision saisonnière**

Elle se place au-delà de 15 jours, c'est-à-dire au-delà de la limite théorique de prédictibilité de l'atmosphère. Elle correspond aux phénomènes de très basses fréquences. On ne cherche plus à prévoir le temps, mais plutôt le climat. Dans ce domaine, les meilleurs résultats restent encore ceux obtenus par des approches statistiques.



**Figure 15 :** Schéma des échelles et d'espace associés aux différents phénomènes météorologiques.

## II.6. Les différentes approches de prévision des précipitations

Les différentes méthodes de prévision présentées ci-dessous sont des méthodes objectives sans toute fois perdre de vue que leurs différents résultats doivent être interprétés par un prévisionniste.

### II.6.1. La prévision numérique du temps (NWP)

La modélisation numérique déterministe consiste à résoudre les équations décrivant le comportement de l'atmosphère. Celles-ci forment un système d'équation non linéaire qui ne peut pas être résolu de façon analytique dans le cas général. La résolution se fait donc de façon numérique et c'est pourquoi, le développement de ces méthodes est étroitement lié aux progrès des ordinateurs. On distingue deux étapes dans la construction d'un modèle numérique de prévision du temps [Coiffier 2000]:

- Il faut avant tout établir un système d'équation décrivant l'atmosphère.
- Ensuite, il faut numériser ces équations, c'est-à-dire les remplacer par des équations portant sur des variables discrètes, afin de rendre la résolution numérique possible.

Plus cette discrétisation est fine, plus le nombre de calcul à effectuer est important, et plus la résolution numérique des équations prend du temps. Donc on est obligé de limiter la résolution spatiale de cette discrétisation. Or tous les phénomènes météorologiques n'ont pas la même échelle spatiale. Un grand nombre de ces phénomènes se déroule sur des échelles spatiales inférieures à la résolution de la discrétisation. Ces phénomènes sont dits « sous maille » et nécessitent une paramétrisation. On peut donc dire qu'un modèle numérique de prévision du temps est constitué de deux parties :

- Une partie de résolution numérique des équations des processus atmosphériques dont l'échelle spatiale est supérieure à la résolution du modèle. Ces processus sont en fait liés aux mouvements des masses d'air. C'est la partie dynamique du modèle.
- Une partie de paramétrisation des effets sur les variables du modèle des processus sous maille. Cette partie représente surtout des processus microphysiques et thermodynamiques. C'est la partie physique du modèle.

On distingue en général trois types de modélisation en fonction de la résolution spatiale adoptée :

**a. Les modèles de circulation générales (GCM)**

Ces modèles sont destinés à faire de la prévision à très longue échéance (plusieurs années) sur l'ensemble du globe. On parle plutôt de simulation climatique. Ils sont donc aptes à rendre compte que sur des phénomènes de basses fréquences.

**b. Les modèles de prévisions globaux**

Ces modèles sont destinés à effectuer des prévisions à courts et à moyens termes sur l'ensemble du globe. Ils ont des mailles de l'ordre de 50 à 20 Km de côté. Ces échelles autorisent l'utilisation d'équations simplifiées grâce à l'hypothèse hydrostatique et imposent la paramétrisation de la convection par exemple.

**c. Les modèles à air limités (LAM)**

Ces modèles sont utilisés pour faire des prévisions à courtes échéances sur une région restreinte. Ils utilisent des mailles de 10 à 2 Km de côté, ce qui les oblige généralement à abandonner l'hypothèse hydrostatique et à représenter de façon explicite la convection. Le domaine considéré par les modèles à air limités ne couvrant pas l'ensemble du globe, et leurs conditions aux limites doivent être alimentées par des prévisions issues du modèle globale (forçage du modèle).

En général, les variables liées à la circulation synoptique (partie dynamique du modèle) donnent des prévisions satisfaisantes jusqu'à 3 jours ou plus d'échéance [Atger 2000]. Par contre, en ce qui concerne les variables telles que les précipitations, les performances de prévisions quantitatives chutent fortement pour les échéances supérieures à 24 heures.

**II.6.2. Les approches statistiques**

Les approches purement statistiques telles que les modèles de Markov et les modèles autorégressifs à moyenne mobile (ARIMA) par exemples, cherchent à identifier à partir d'une archive de données, les relations statistiques qui peuvent exister entre certains variables (les prédicteurs) observées à un instant  $t$ , et la variable cible (le prédicteand), observée à un instant  $t+\Delta t$ . Le but ici n'est pas d'explicitier les processus physiques qui se cachent derrière ces liaisons statistiques mais simplement d'exprimer le mieux possible ces liaisons. Une fois ces liaisons établies, on s'en sert pour prévoir la variable cible à l'échéance de  $\Delta t$  à partir de l'observation des prédicteurs au temps  $t$ . Mathématiquement, on peut formuler cette approche de la manière suivante :

$$\begin{cases} \text{Calage : } & f_{\Delta t}(x_t^0) = y_{t+\Delta t}^0 + \varepsilon \\ \text{Prévision : } & y_{t+\Delta t}^s = f_{\Delta t}(x_t^0) \end{cases} \quad (\text{II.1})$$

Où  $f_{\Delta t}$  : est la relation statistique établie pour l'échéance  $\Delta t$

$x_t^0$  : est le prédictor observé à l'instant  $t$

$y_{t+\Delta t}^0$  : est le prédictand observé à l'instant  $t + \Delta t$

$y_{t+\Delta t}^s$  : est la valeur prévue du prédictand pour l'instant  $t + \Delta t$

$\varepsilon$  : est l'erreur de la relation statistique que l'on cherche à minimiser.

### II.6.2.1. La méthode ARIMA (Auto Régressif Integrated Moving Average)

La méthode ARIMA développée par Box et Jenkins en 1976 dont l'objectif essentiel est de permettre une prédiction de l'évolution future d'un phénomène. Elle s'est rapidement généralisée et a été utilisée dans de nombreux domaines, en particulier dans celui de la prévision des précipitations. Elle consiste à sélectionner un modèle de prévision parmi une classe de modèles stochastiques. Voici quelques définitions utiles :

- Les processus autorégressifs supposent que chaque point peut être prédit par la somme pondérée d'un ensemble de points précédents plus un terme aléatoire d'erreur.
- Le processus d'intégration suppose que chaque point présente une différence constante avec le point précédent.
- Les processus de moyenne mobile supposent que chaque point est fonction des erreurs entachant les points précédents plus sa propre erreur.

#### II.6.2.1.1. Typologie du modèle ARIMA

Un modèle ARIMA est étiqueté comme modèle ARIMA (p, d, q), dans lequel :

- p : est le nombre de termes autorégressifs
- d : est le nombre de différences
- q : est le nombre de moyennes mobiles

##### a. Premier critère : la différenciation

L'estimation des modèles ARIMA suppose que l'on travaille sur une série stationnaire. La meilleure méthode pour éliminer toute fluctuation est de différencier. Une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une

version intégrée d'une série stationnaire (d'où le terme Integrated). Une différenciation d'ordre 1 est donnée par l'équation de prédiction suivante :

$$y_t - y_{t-1} = \mu + \varepsilon_t \quad (\text{II.2})$$

Où  $\mu$  est la constante du modèle. Un tel modèle est un ARIMA (0, 1, 0). Il est représenté comme un accroissement linéaire en fonction du temps.

### b. Deuxième critère : l'autorégression

Les modèles autorégressifs supposent que  $y_t$  est une fonction linéaire des valeurs précédentes comme le montre l'équation de prédiction suivante:

$$y_t = \mu + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_n y_{t-n} + \varepsilon_t \quad (\text{II.3})$$

Chaque observation est constituée d'une composante aléatoire  $\varepsilon$  et d'une combinaison linéaire des observations précédentes dans laquelle, les constantes  $\varphi_1$ ,  $\varphi_2$  et  $\varphi_n$  sont les coefficients d'autorégression. Notons que cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA (1, 1, 0) on aura l'équation de prédiction suivante :

$$y_t = \mu + y_{t-1} + \varphi(y_{t-1} - y_{t-2}) + \varepsilon_t \quad (\text{II.4})$$

### c. Troisième critère : la moyenne mobile

Les modèles à moyenne mobile suggèrent que la série présente des fluctuations autour d'une valeur moyenne. Ceci revient en fait à considérer que l'estimation est égale à la moyenne vraie, à laquelle on ajoute une somme pondérée des erreurs ayant entachées les valeurs précédentes comme le montre l'équation de prédiction suivante :

$$y_t = \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_n \varepsilon_{t-n} + \varepsilon_t \quad (\text{II.5})$$

Chaque observation est composée d'une composante d'erreur aléatoire  $\varepsilon$  et d'une combinaison linéaire des erreurs aléatoires passées, dans laquelle les constantes  $\theta_1$ ,  $\theta_2$  et  $\theta_n$  sont les coefficients de moyenne mobile du modèle.

Comme précédemment cette équation porte soit sur les données brutes, soit sur les données différenciées si une différenciation a été nécessaire. Pour un modèle ARIMA (0, 1, 1) on aura l'équation de prédiction suivante:

$$y_t - y_{t-1} = \mu - \theta \varepsilon_{t-1} + \varepsilon_t \quad (\text{II.6})$$

On peut également envisager des modèles mixtes : par exemple un modèle ARIMA (1, 1, 1) aura l'équation de prédiction suivante :

$$y_t = \mu + y_{t-1} + \varphi(y_{t-1} - y_{t-2}) - \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (\text{II.7})$$

### II.6.2.1.2. Construction du modèle ARIMA

#### a. Détermination de l'ordre de différenciation

Une série stationnaire fluctue autour d'une valeur moyenne et sa fonction d'autocorrélation décroît rapidement vers zéro. Donc si une série présente des autocorrélations positives pour un grand nombre de décalages, alors il est nécessaire de la différencier.

#### b. Identification des termes AR (Autorégressifs)

Après que la série ait été rendue stationnaire, l'étape suivante consiste à identifier les termes Autorégressifs nécessaires pour corriger les autocorrélations résiduelles. Cette analyse est basée sur l'examen des fonctions d'autocorrélation partielle (FACP). Rappelons que l'autocorrélation est la corrélation d'une série avec elle-même, selon un décalage défini.

#### c. Identification des termes MA (Moyennes mobiles)

La fonction d'autocorrélation (FAC) joue le même rôle pour les processus de moyenne mobile que la fonction d'autocorrélation partielle (FACP) pour les processus autorégressifs. Si l'autocorrélation est significative au décalage  $k$  mais ne l'est plus au décalage  $k+1$ , alors  $k$  termes de moyenne mobile doivent être ajoutés au modèle.

#### d. Estimation des paramètres

L'estimation des paramètres est obtenue selon la méthode des moindres carrés ordinaire. Cependant, la présence de termes moyenne mobile fait apparaître des non linéarités, ce qui implique l'utilisation de la méthode du maximum de vraisemblance.

#### e. Validation

Cette étape vise essentiellement à examiner les erreurs et à s'assurer que les coefficients  $\varepsilon_t$  suivent un bruit blanc. Pour cela, on effectue une analyse des corrélogrammes (FAC) et corrélogrammes partiels (FACP) des erreurs et on vérifie qu'on ne détecte aucune corrélation significative dans les erreurs.

## **f. Prévision**

A la suite des étapes précédentes nous pouvons alors passer à l'application du modèle et estimer la prévision des variables souhaitées.

Les approches purement statistiques présentent l'avantage d'être beaucoup moins coûteuses à développer et à faire fonctionner que les modèles numériques. Cependant, maintenant elles ont tendances à être abandonnées pour les prévisions à cours et à moyen terme où elles ont été largement supplantées par les modèles numériques. Néanmoins, elles restent encore tout à fait d'actualité pour les prévisions saisonnières.

### **II.6.3. Les approches mixtes**

Bien que paraissant antinomiques, les approches numériques et statistiques s'avèrent finalement complémentaires et gagnent généralement à être fusionnées. La prévision numérique apporte sa forte capacité prédictive des phénomènes dynamiques, alors que les interprétations statistiques permettent de rendre compte à la fois des effets locaux, grâce à l'utilisation de l'information contenue dans les séries de données, mais aussi de l'incertitude liée au chaos dynamique.

## **II.7. Discussion**

Dans ce chapitre nous avons donné un aperçu général sur les phénomènes météorologiques. Ainsi, nous avons mis en évidence les différentes approches utilisées pour la prévision des précipitations.

Du point de vue pratique, pour des prévisions à moyen et à long termes, les composantes non linéaires des systèmes météorologiques et le nombre de variables explicatives prennent plus d'importance. Il est alors préférable d'explorer les modèles non linéaires tels que les réseaux de neurones.

# Chapitre III

### III.1. Préambule

Le cerveau humain est un système d'une extrême complexité capable de résoudre des problèmes très difficiles et complexes. Il est composé de plusieurs éléments différents. L'un des constituants les plus importants est le neurone. Le cerveau contient environ 100 milliards de neurones. Chaque neurone est de forte complexité. En dépit des immenses progrès réalisés au cours des vingt dernières années dans le domaine de la neurobiologie, les spécialistes des neurones biologiques commencent à peine à comprendre quelques uns de leurs mécanismes internes. Des questions aussi fondamentales que celle du codage des informations par le cerveau restent l'objet de nombreuses recherches [Dreyfus 2002] [Touzet 1992]. Les neurones artificiels ressemblent à leurs congénères biologiques dans leur fonctionnement et leur architecture. Aujourd'hui, les réseaux de neurones artificiels (RNA) constituent une technique de traitement de données bien comprise et bien maîtrisée. On les trouve implantés dans divers domaines, tels que les milieux financiers pour évaluer le risque financier, en pharmacutique pour le diagnostic médical, dans le domaine bancaire pour la détection de fraude, sur des cartes de crédit, en aéronautique pour la programmation des pilotes automatiques, dans le domaine de la prévision météorologique,... etc. Les applications sont nombreuses et partagent toutes un point commun essentiel à l'utilité des réseaux de neurones.

Ce chapitre est consacré à la présentation des réseaux de neurones pour la prévision des précipitations. Le réseau choisi pour notre application, c'est-à-dire le TDNN et ses deux méthodes d'apprentissage seront également exposés.

### III.2. Historique

L'origine de l'inspiration des réseaux de neurones artificiels remonte à 1890 où W.James, célèbre psychologue américain, introduisit le concept de mémoire associative. Il proposait ce qui deviendra une loi de fonctionnement pour l'apprentissage des réseaux de neurones, connue plus tard sous le nom de loi de Hebb. Quelques années plus tard, en 1943, les travaux des deux biophysiciens, J.Mc Culloch et W.Pitts montrèrent qu'un réseau de neurone discret, sans contrainte de topologie, pouvait représenter n'importe quelle fonction booléenne et donc émuler un ordinateur. C'est ensuite que D.Hebb, physiologiste américain, présentait en 1949 les propriétés des neurones par le conditionnement chez l'animal. Ainsi, un conditionnement de type pavlovien tel que, nourrir tous les jours à la même heure un chien, entraîne chez cet animal la sécrétion de salive à cette heure précise même en l'absence de

nourriture. La loi de modification des propriétés des connections entre neurones qu'il proposait, expliquait en partie ce type de résultats expérimentaux.

Les premiers succès de cette discipline remontèrent à 1957, lorsque F. Rosenblatt développa le modèle du perceptron. En 1960 Widrow a développé le modèle Adaline (Adaptive Linear Element). Dans sa structure, le modèle ressemblait au perceptron. Cependant, la loi d'apprentissage était différente. Celle-ci était à l'origine de la rétropropagation du gradient très utilisée aujourd'hui avec les Perceptrons Multicouche. En 1969, Minsky et Papert publièrent le livre « Perceptrons » dans lequel ils utilisèrent une solide argumentation mathématique pour démontrer les limitations des réseaux de neurones à une seule couche. Ce livre aura une influence telle que la plupart des chercheurs quitteront le champ de recherche sur les réseaux de neurones.

Quelques années plus tard et plus précisément en 1982, le physicien J.J. Hopfield proposa des réseaux de neurones associatifs et l'intérêt des réseaux de neurones renait chez les scientifiques. C'est ensuite en 1985 que la rétropropagation du gradient apparut. C'est un algorithme d'apprentissage adapté aux perceptrons multicouches. Sa découverte a été réalisée par trois groupes de chercheurs indépendants. Dès cette découverte, nous avons eu la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau, en décomposant cette fonction en une suite d'étapes linéairement séparable. En 1989 Moody et Darken exploitèrent quelques résultats de l'interpolation multi variables pour proposer le Réseau à Fonction de Base Radiale (RFR), connu sous l'appellation anglophone « Radial Basis Function network » (RBF). Enfin, le TDNN qui a été développé à l'origine dans le cadre de la reconnaissance de phonème [Waibel 1989], à la reconnaissance vocale [Bottou 1990] et à la reconnaissance de l'écriture manuscrite [Guyon 1991]. Par la suite, il a été généralisé pour la modélisation et la prévision des données de nature séquentielles.

### **III.3. Réseaux de neurones pour la prévision**

#### **III.3.1. Définition d'un neurone biologique**

L'élément de base du système nerveux central est le neurone. Le neurone est une cellule composée d'un corps cellulaire, de dendrites, d'axone et de synapse.

- Le corps cellulaire contient le noyau du neurone ainsi que la machine biologique nécessaire à la synthèse des enzymes. Sa taille est de quelques microns de diamètre.

- Les dendrites sont de fines extensions tubulaires qui se ramifient autour du neurone et forment une sorte de vaste arborescence. Les signaux envoyés aux neurones sont captés par les dendrites.
- L'axone : c'est le long de l'axone que les signaux partent du neurone. Contrairement aux dendrites qui se ramifient autour du neurone. L'axone est plus long et se ramifie à son extrémité où il se connecte aux dendrites des autres neurones.
- Une synapse est une jonction entre deux neurones, généralement entre l'axone d'un neurone et une dendrite d'un autre neurone.

Le corps cellulaire se ramifie pour former les dendrites. C'est par ces dernières que l'information est acheminée de l'extérieur vers le corps cellulaire où elle est traitée. L'information traitée par le neurone chemine ensuite le long de l'axone (unique) pour être transmise aux autres neurones.

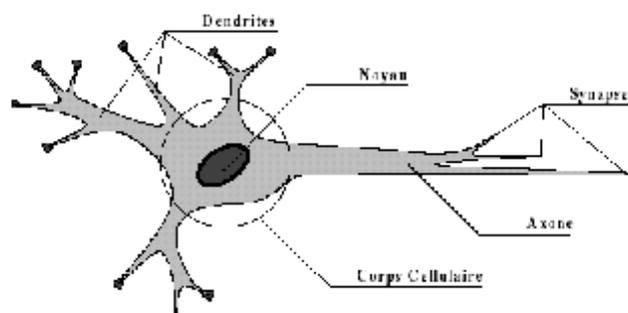


Figure 16 : Le neurone biologique.

### III.3.2. Définition d'un neurone artificiel

Chaque neurone artificiel est un processeur élémentaire. Il reçoit un nombre variable d'entrées en provenance de neurones en amonts. A chacune de ces entrées est associé un poids (paramètre)  $w$ , représentatif de la force de la connexion. Chaque processus élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones avals.

Les réseaux de neurones formels sont issus d'une tentative de modélisation mathématique du cerveau humain. Cette modélisation consiste à mettre en œuvre un système de réseau de neurones artificiels qu'il existe pour chaque élément composant le neurone biologique un élément correspondant dans le neurone artificiel comme le montre la figure 17.

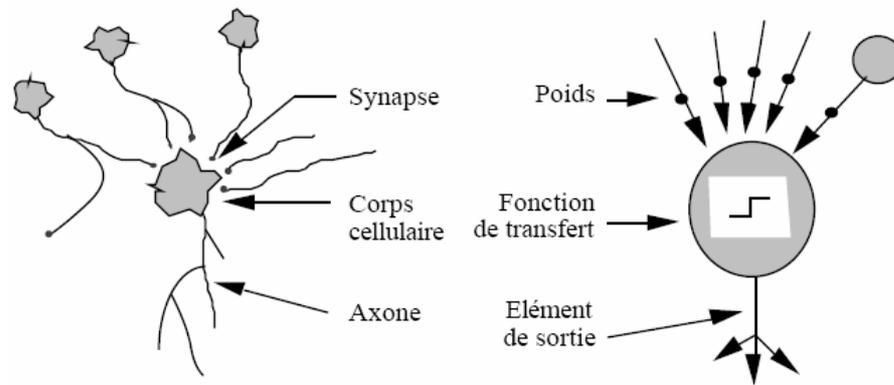


Figure 17 : Mise en correspondance neurone biologique / neurone artificiel.

### III.3.3. Principe de fonctionnement

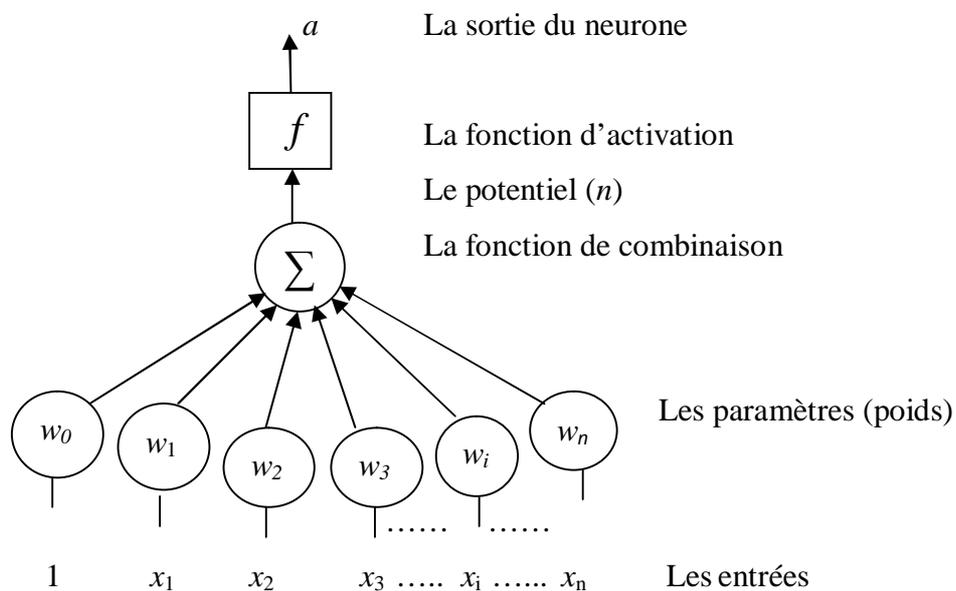


Figure 18 : Neurone mathématique.

Un neurone réalise une fonction non linéaire, paramétrée, à valeurs bornées entre les entrées et la sortie. Autrement dit un neurone réalise une fonction non linéaire d'une combinaison des entrées  $\{x_i\}$  pondérées par les paramètres (ou poids)  $\{w_i\}$ . La combinaison linéaire est appelée potentiel ( $n$ ), à laquelle s'ajoute un terme constant  $w_0$  ou « biais ». Ce principe est illustré sur la figure 18.

Le potentiel d'un neurone est donc :

$$n = w_0 + \sum_{i=1}^{n-1} w_i x_i \quad (\text{III.1})$$

La sortie du neurone sera donc :

$$a = f(n) = f(w_0 + \sum_{i=1}^{n-1} w_i x_i) \quad (\text{III.2})$$

Tel que  $f$  est la fonction d'activation.

### Ø La fonction d'activation

La fonction d'activation (ou fonction de transfert) permet d'introduire un seuil et une saturation dans le réseau. Il existe différentes formes des fonctions d'activation. La plus part sont continues. Les plus courantes sont la sigmoïde log, la sigmoïde tangente hyperbolique, la fonction à base radiale (RBF pour Radial Basis Function) et la fonction linéaire qui se nomment respectivement « logsig », « tansig », « radbas » et « purelin » dans MATLAB. Voici quelques exemples de fonctions d'activation courantes.

- **La fonction sigmoïde**

$$f_1(n) = \frac{1}{1 + e^{-n}}$$

$$f_1'(n) = f_1(n)[1 - f_1(n)]$$

- **La tangente hyperbolique**

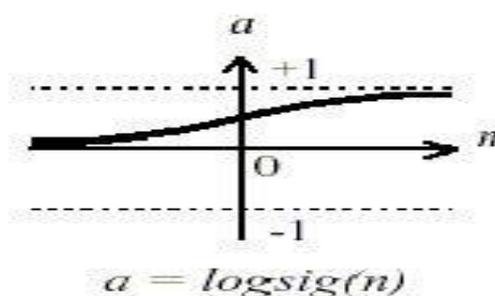
$$f_2(n) = \frac{e^n - e^{-n}}{e^n + e^{-n}}$$

$$f_2'(n) = \frac{4}{(e^n + e^{-n})^2}$$

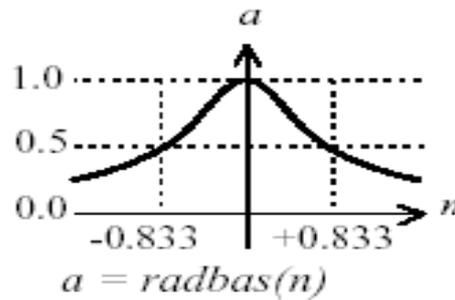
- **Fonction à base radiale**

$$f_3(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

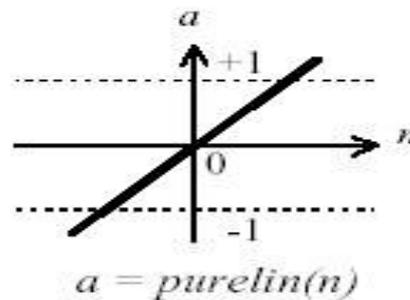
Avec  $x$  le vecteur d'entrée du réseau,  $\mu$  le centre de la fonction et  $\sigma$  la largeur de la fonction.



**Figure 19** : La fonction sigmoïde.



**Figure 20 :** La fonction à base radiale.



**Figure 21 :** La fonction linéaire.

### III.3.4. Réseau de neurone artificiel

Dans le monde biologique, plusieurs structures neuronales existent et chaque structure est dotée d'une fonction particulière. Le neurone reçoit en continu des entrées, ces informations sont interprétées au niveau du corps cellulaire du neurone. La réponse à ces signaux est envoyée à travers l'axone qui fait synapse sur des milliers d'autres neurones [Touzet 1992].

Pour les réseaux de neurones artificiels, Kohonen propose la définition suivante : « les RNA sont des réseaux massivement connectés en parallèle d'éléments simples (habituellement adaptatifs) et d'organisation hiérarchique. Ils sont sensés interagir avec les objets du monde réel de la même manière que les systèmes nerveux biologiques ».

D'une autre manière, un réseau de neurones artificiels (RNA) est un ensemble de neurones formels associés en couches (ou sous groupes) et fonctionnent en parallèle. Chaque neurone dans une couche est connecté généralement à tous les neurones de la couche suivante. Dans un réseau de neurones, l'information se propage de la couche d'entrée à la couche de sortie, en passant par aucune ou plusieurs couches intermédiaires dites couches cachées. Les figures 22 et 23 montrent un RN biologique et un RN artificiel d'une couche d'entrée, d'une couche cachée et d'une couche de sortie.

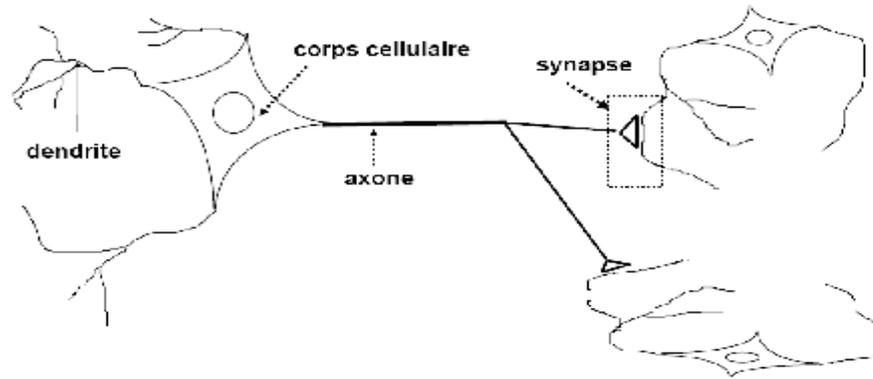


Figure 22 : Réseau de neurone biologique.

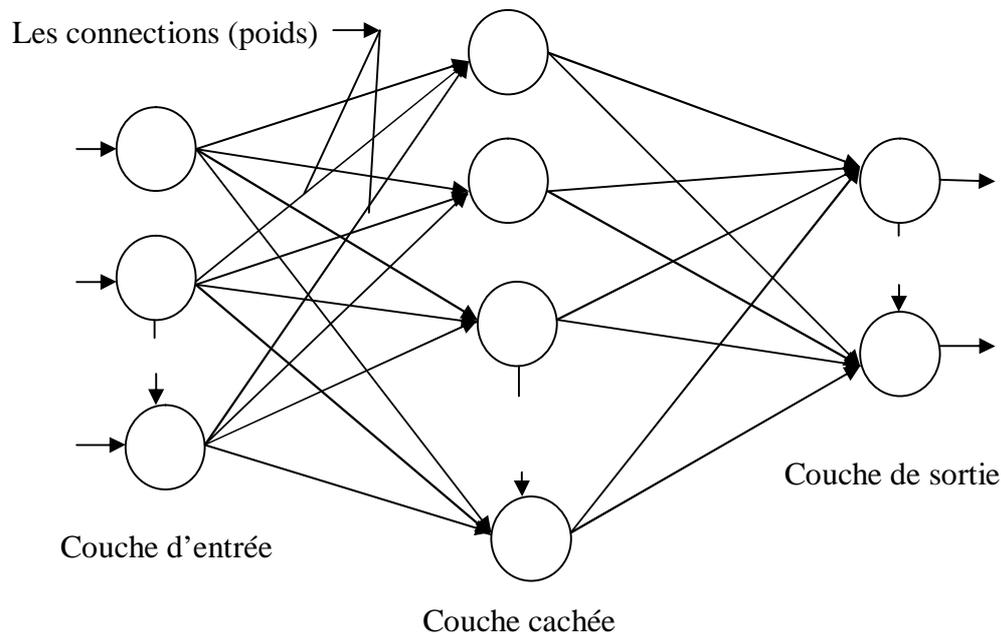


Figure 23 : Réseau de neurone artificiel.

### III.4. Propriétés des réseaux de neurones

Les réseaux de neurones constituent une famille de fonctions non linéaires paramétrées, que l'on met en œuvre pour des tâches de modélisation, de prédiction et de classification. Un réseau de neurones est caractérisé par les propriétés suivantes :

#### III.4.1. L'approximation universelle

L'approximation universelle est une propriété fondamentale des réseaux de neurones bouclés et non bouclés [Thiria 1997].

**Propriété :** *Toute fonction bornée suffisamment régulière peut être approchée uniformément avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un*

*réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation et un neurone de sortie linéaire.*

### **III.4.2. La parcimonie**

Par ailleurs, Baron 1991 montre que si l'approximation dépend des paramètres ajustables de manière non linéaire, elle est plus parcimonieuse que si elle dépend linéairement des paramètres. Plus précisément, on montre que le nombre de paramètres, pour une précision donnée, croît exponentiellement avec le nombre de variables dans le cas des approximateurs linéaires par rapport à leurs paramètres, alors qu'il croît linéairement avec ce nombre pour les approximateurs non linéaires par rapport à leurs paramètres. La parcimonie est donc d'autant plus avantageuse que le nombre de variables du modèle est grand. Donc pour des problèmes à trois variables ou plus, il est généralement avantageux d'utiliser les réseaux de neurones.

### **III.4.3. La généralisation**

L'élaboration d'un bon modèle de réseau de neurones revient au choix de la structure et aussi aux nombres de neurones dans la couche cachée. Concernant les couches d'entrées et de sorties d'un réseau de neurone, elles doivent incorporer autant de neurones que le nombre d'entrées et de sorties du système à modéliser. A l'inverse, le choix du nombre de neurones dans la couche cachée n'est pas imposé par le système à modéliser. Le nombre de neurones à incorporer dans la couche cachée devrait être choisi afin de fournir de bonnes possibilités de généralisation pour l'estimateur neuronale.

La capacité de généralisation est l'une des raisons qui motivent l'étude et le développement des réseaux de neurones artificiels. Elle peut être définie par la capacité d'élargir les connaissances acquises après apprentissage à des données nouvellement rencontrées par le réseau. On parle de sur-apprentissage (over-fitting) quand le réseau a trop parfaitement appris les exemples proposés. Il sera donc incapable de généraliser. Une solution à ce problème consiste à utiliser un critère d'arrêt basé sur une technique dite de validation croisée (en anglais cross-validation). Cette technique consiste à utiliser deux ensembles indépendants de données pour entraîner le réseau : un pour l'apprentissage (l'ajustement des poids) et l'autre pour la validation. Le critère d'arrêt consiste alors à stopper l'apprentissage lorsque l'indice de performance calculé sur les données de validation cesse de s'améliorer pendant plusieurs périodes d'entraînement. La figure 24 illustre le critère de la validation croisée dans le cas d'un indice de performance que l'on cherche à minimiser.

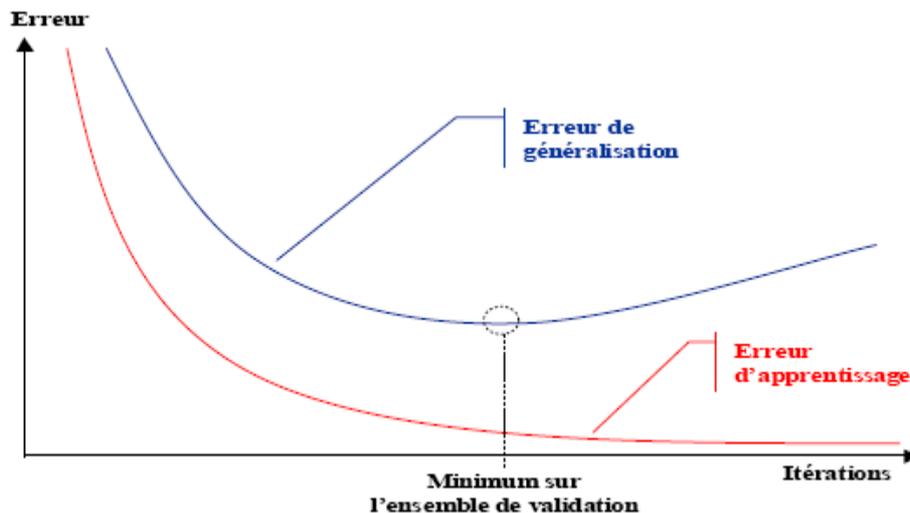


Figure 24 : Illustration de la validation croisée.

### III.5. Apprentissage des réseaux de neurones

L'apprentissage est vraisemblablement la propriété la plus intéressante des réseaux de neurones. C'est la phase de développement d'un réseau de neurone durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.

Il existe de nombreux types de règles d'apprentissage qui peuvent être regroupées en trois catégories [Hassoum, 1995]: l'apprentissage supervisé, non supervisé et renforcé. Mais l'objectif fondamental de l'apprentissage reste le même : soit la classification, l'approximation de fonction ou encore la prévision. Dans l'optique de la prévision, l'apprentissage consiste à extraire des régularités à partir des exemples qui peuvent être transférées à de nouveaux exemples.

Ø ***Apprentissage supervisé*** : L'apprentissage dit supervisé est caractérisé par la présence d'un professeur qui possède une connaissance approfondie de l'environnement dans lequel évolue le réseau de neurones. En pratique, les connaissances de ce professeur prennent la forme d'un ensemble de  $Q$  couples de vecteurs d'entrée et de sortie que nous noterons  $\{(p_1, d_1), (p_2, d_2), \dots, (p_Q, d_Q)\}$ , où  $p_i$  désigne un stimulus et  $d_i$  la cible pour ce stimulus c'est-à-dire les sorties désirées du réseau. Chaque couple  $(p_i, d_i)$  correspond donc à un cas d'espèce de ce que le réseau devrait produire (la cible) pour un stimulus donné. Pour cette raison, l'apprentissage supervisé est aussi qualifié d'apprentissage par des exemples. Pour ce type d'apprentissage, la règle la plus utilisée est celle de Widrow-Hoff.

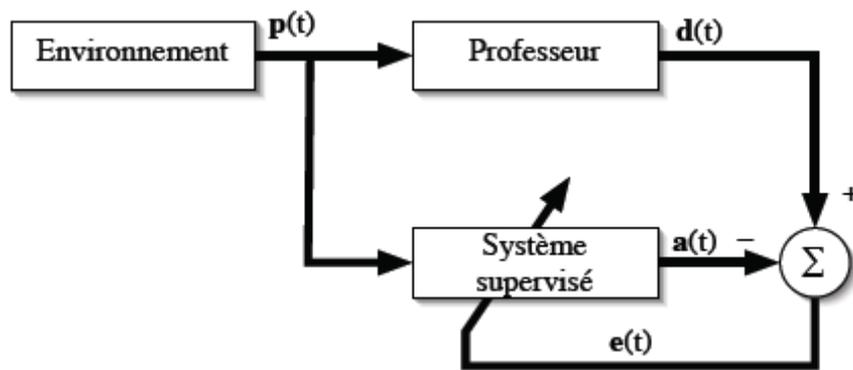


Figure 25 : Schéma bloc de l'apprentissage supervisé.

- Ø **Apprentissage renforcé** : L'apprentissage renforcé est une technique similaire à l'apprentissage supervisé à la différence qu'au lieu de fournir des résultats désirés au réseau, on lui accorde plutôt un grade (ou score) qui est une mesure du degré de performance. Les algorithmes utilisant la procédure d'apprentissage renforcé sont surtout utilisés dans le domaine des systèmes de contrôle.
- Ø **Apprentissage non supervisé** : L'apprentissage non supervisé consiste à ajuster les poids à partir d'un seul ensemble d'apprentissage formé uniquement de données. Aucun résultat désiré n'est fourni au réseau. L'apprentissage non supervisé consiste donc à détecter les similarités et les différences dans l'ensemble d'apprentissage. Les réseaux à apprentissage non supervisé les plus étudiés et utilisés sont les cartes auto-organisatrices de Kohonen.

Le but de l'apprentissage est d'ajuster les poids des neurones, cette tâche nécessite des exemples désignés aussi sous l'appellation d'échantillon d'apprentissage ainsi qu'un algorithme d'apprentissage.

Soit un exemple de  $N$  mesures de la sortie  $\{Y_p^k\}$  pour  $k=1$  à  $N$ , correspondant à  $N$  valeurs du vecteur d'entrée  $\{X^k = [X_1^k \ X_2^k \ \dots \ X_n^k]\}$ , l'ensemble des couples  $\{X^k, Y_p^k\}$  est appelé, ensemble d'apprentissage. Afin de déterminer la meilleure valeur du vecteur paramètre  $W$ , on doit minimiser le coût, qui est une fonction des écarts entre les mesure  $\{Y_p^k\}$  et les valeurs fournies par le modèle  $\{Y^k(W)\}$  (réseau de neurones). On choisit généralement une fonction de coût quadratique  $J(w)$  telle que :

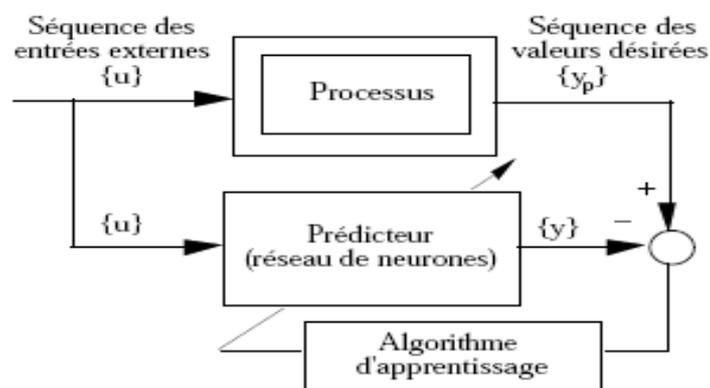
$$J(w) = \sum_{k=1}^N \frac{1}{2} (Y_p^k - Y^k(W))^2 \quad (\text{III.3})$$

L'apprentissage des modèles nécessite des algorithmes de minimisation du coût  $J(w)$  qui permet d'avoir un vecteur paramètre qui garantit une sortie de réseau égale ou proche de la sortie désirée, on distingue deux types d'algorithmes :

- Un algorithme de minimisation itératif qui recherche le minimum du coût en modifiant les paramètres à chaque itération à partir du calcul de coût total. Parmi ces algorithmes on trouve l'algorithme du gradient et l'algorithme de Newton.
- Un algorithme de minimisation récursif qui effectue des modifications des paramètres calculés à l'aide d'un coût portant sur un nombre limité d'exemple (coût partiel). Dans ce cas, une modification des paramètres est effectuée après la présentation de chaque exemple. Parmi ces algorithmes on cite, l'algorithme de Widrow Hoff, la méthode des moindres carrés récursifs. On peut également utiliser la récursion du filtre de kalman étendu.

L'estimation des paramètres peut avoir lieu hors ligne (système non adaptatif) ou en ligne (système adaptatif) [Personnaz et al, 2003] :

- Dans le cas de l'apprentissage hors ligne, l'expérience qui a fourni les séquences d'apprentissage a été effectuée auparavant, donc l'ensemble d'apprentissage est de taille finie. Dans ce cas l'algorithme d'apprentissage peut indifféremment être itératif ou récursif.
- En utilisant l'apprentissage en ligne, les séquences d'apprentissage sont de taille infinie, puisque pour un système adaptatif, les valeurs des paramètres des modules de calcul sont modifiées au cours du temps, pendant l'utilisation du système en fonction de l'environnement. Ainsi, en traitement de signal, il est nécessaire de faire des mises à jour des paramètres des filtres en fonction de l'arrivée des nouvelles mesures. Dans ce cas, il est nécessaire d'utiliser un algorithme récursif.



**Figure 26** : Système d'apprentissage pour la modélisation d'un processus.

### III.6. Algorithmes d'apprentissage

L'algorithme d'apprentissage est la méthode mathématique qui va modifier les poids de connexions afin de converger vers une solution qui permettra au réseau d'accomplir la tâche désirée. On distingue des algorithmes de premier et de deuxième ordre :

#### III.6.1. Algorithmes de premier ordre

Les méthodes dites du premier ordre sont basées sur le calcul du gradient de la fonction coût et le minimum de cette fonction est atteint si son gradient (autrement dit sa dérivée) est nul. On compte parmi ces méthodes, la méthode du gradient simple, la méthode du gradient stochastique.

#### III.6.2. Algorithmes du deuxième ordre

Les méthodes du deuxième ordre sont ainsi appelées parce qu'elles prennent en considération la dérivée seconde de la fonction coût. Parmi ces méthodes, on compte la méthode de Newton et la méthode de Lavenberg-Marquardt.

### III.7. Prédiction par réseau de neurones

La notion de prédiction est l'une des plus fondamentales en apprentissage. Il s'agit d'un problème de traitement temporel de signal. En supposant que nous possédons  $k$  échantillons passés d'un signal,  $x(t-1)$ ,  $x(t-2)$ , . . . ,  $x(t-k)$ , échantillonnés à intervalle de temps fixe. La tâche consiste à prédire la valeur de  $x$  au temps  $t$ . Ce problème de prédiction peut être résolu grâce à un apprentissage par correction des erreurs, mais d'une manière non supervisée, étant donné que les valeurs de sortie désirée peuvent être inférées directement de la série chronologique. Plus précisément, l'échantillon de  $x(t)$  peut servir de valeur désirée et le signal d'erreur pour l'adaptation des poids se calcule simplement par l'équation suivante :

$$e(t) = x(t) - g(x(t-1), x(t-2), \dots, x(t-k)) \quad (\text{III.4})$$

Où  $x(t)$  désigne la sortie désirée et  $x(t-1)$ ,  $x(t-2)$ , . . . ,  $x(t-k)$  représentent la sortie observée du réseau étant donné les  $k$  échantillons précédents et  $g(\ )$  est une fonction de transfert non linéaire. La prédiction s'apparente à la construction d'un modèle physique de la série chronologique. D'après l'équation (III.4), le développement d'un réseau de neurones pour la prévision nécessite :

- ∅ Le choix de réseau de neurones approprié qui représente au mieux le système dynamique à modéliser.

- Ø La collecte des données nécessaires à l'apprentissage et à l'évaluation des performances du réseau de neurones.
- Ø L'estimation de la valeur de  $k$ , c.-à-d. le nombre de valeurs passées dans le vecteur d'entrée contribuant à la prévision à un instant futur.
- Ø La détermination de la complexité optimale de réseau de neurone, c.-à-d. le nombre de couches cachées ainsi que le nombre de neurones dans ses couches.
- Ø L'estimation des valeurs des paramètres correspondant à un minimum de la fonction de coût, c'est-à-dire d'effectuer un apprentissage.
- Ø La prévision pour un horizon donné à partir de réseau de neurones préalablement calibré durant la phase d'apprentissage.

### III.8. Réseaux de neurones pour la prévision des précipitations

On distingue essentiellement deux types de réseaux de neurones formels utilisés dans le domaine de la prévision des précipitations, et cela en fonction du graphe de leurs connexions : Les réseaux de neurones statiques (feedforward neural network) et Les réseaux de neurones dynamiques :

#### III.8.1. Les réseaux de neurones statiques (feedforward neural network)

Dans un tel réseau, le flux d'information circule des entrées vers les sorties sans retour en arrière et le temps ne joue aucun rôle fonctionnel.

Un réseau de neurones statique réalise donc une fonction non linéaire  $\psi$  de ses entrées paramétrées par les coefficients  $w$  du réseau :

$$O(k) = \psi (I(k) ; w) \quad (\text{III.5})$$

Où  $O(k)$  et le vecteur de sortie à l'instant  $k$ ,  $I(k)$  est le vecteur des entrées et  $\psi$  est la fonction non linéaire réalisée par les neurones du réseau.

#### ü Le perceptron multi couche

Le réseau statique le plus connu et le plus répandu utilisé pour la prévision de séries temporelles de précipitations qu'on trouve dans la littérature est le réseau à couche ou bien le perceptron multicouche ou « MLP pour Multi Layer Perceptrons ». Dans cette architecture, les neurones cachés sont organisés en couches, les neurones d'une même couche n'étant pas connectés entre eux. De plus, les connexions entre deux couches de neurones non consécutives sont éliminées comme le montre la figure 27.

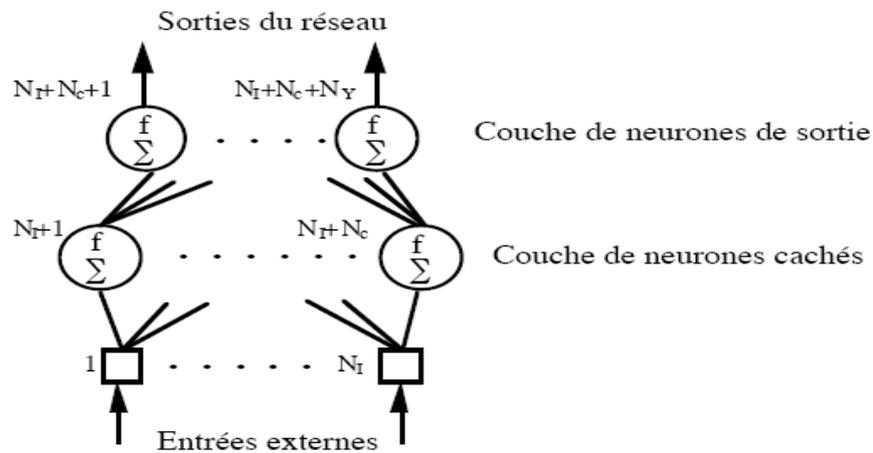


Figure 27 : Réseau à couche.

Dans la figure 28, nous présentons la structure d'un MLP désigné pour la prévision des précipitations. Les neurones de sorties représentent les précipitations à l'instant futur (les valeurs prévues par le réseau), contenant  $N$  éléments, représentant la localisation spatiale des précipitations.

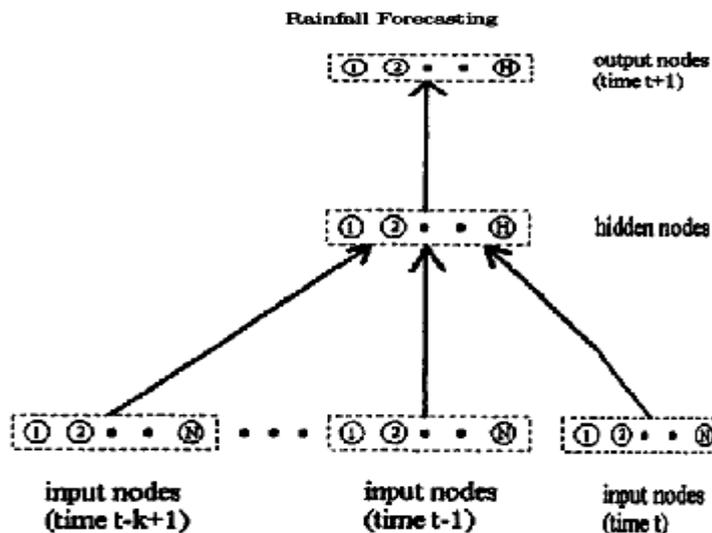


Figure 28 : MLP pour la prévision des précipitations.

### III.8.2. Les réseaux dynamiques

Les réseaux dynamiques peuvent être classés en fonction de la représentation du temps qu'ils utilisent. Cette représentation peut être :

Ø *Interne au modèle* : C'est le cas des réseaux récurrents.

Ø *Externe au modèle* : C'est le cas de TDNN (Time Delay Neural Network) et le TDRBF (Time Delay Radial Basis Function).

### III.8.2.1. Réseaux dynamiques récurrent

Quand un MLP est utilisé comme modèle pour la prévision des séries temporelles de précipitations, on a besoin de déterminer le nombre de retard à appliquer dans la couche d'entrée. Pour remédier à ce problème, un réseau de neurone récurrent est utilisé. Ce type de réseau possède des connexions sous forme de boucles et admet tout type de connexions, c'est-à-dire d'un neurone à n'importe quel autre, y compris lui-même. Il est possible dans ce réseau de trouver au moins un chemin qui revient à son point de départ en suivant le sens de connexion. Un tel chemin est dit cycle. Ces réseaux sont assez intéressants car leur fonctionnement est séquentiel et adopte un comportement dynamique. La sortie du réseau est donnée par l'équation suivante :

$$\hat{y}(k+1) = f(w_x \times x(k) + w_d + d(k-1) + b_1) \times w_y + b_2 \quad (\text{III.6})$$

Tout réseau de neurones récurrent est un système dynamique non linéaire que l'on peut mettre sous forme d'une représentation d'état appelée forme canonique :

$$\begin{cases} x(k+1) = \varphi(x(k), u(k), w) \\ y(k) = \psi(x(k), u(k), w) \end{cases} \quad (\text{III.7})$$

Où  $y(k)$  et le vecteur de sortie à l'instant  $k$ ,  $u(k)$  est le vecteur des entrées et les fonctions  $\varphi$  et  $\psi$  sont des fonctions non linéaires qui peuvent être réalisées par deux réseaux de neurones non bouclés. La Figure 29 représente la forme canonique d'un réseau de neurone récurrent.

Les modèles les plus performants et les plus connus des réseaux récurrents sont, l'architecture de Jordan (Jordan 1986) et celle d'Elman (Elman 1990), qui ont été développées à l'origine pour apprendre les structures linguistiques. Ils ont été plus tard généralisés pour la modélisation et la prévision d'une série temporelle (séries chronologiques). La figure 30 montre la structure de Jordan et celle d'Elman désignées à la prévision des précipitations. Dans l'architecture proposée par Jordan, on trouve la dernière couche rebouclée sur la couche de contexte, et celle d'Elman est légèrement différente de celle de Jordan. Il s'agit d'une rétroaction reliant la couche cachée avec les cellules de la couche de contexte. Notons que ces connexions récurrentes sur les unités de contexte donnent au réseau une dynamique et une mémoire individuelle.

L'algorithme d'apprentissage pour ces deux types de réseaux est la rétropropagation de gradient. Il existe d'autres algorithmes d'apprentissage qui peuvent être appliqués pour les réseaux récurrents tel que : « Real Time Recurrent Learning » et « Time Dependent Recurrent

Back-propagation » [Hertz et al.1991], [Haykin et al.1997], « Back-propagation Through Time » [Williams et al.1990].

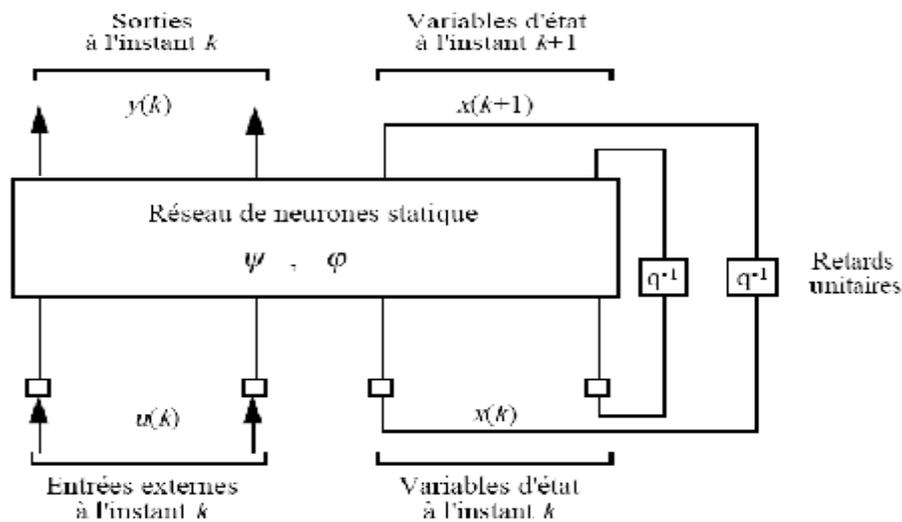


Figure 29 : Forme canonique d'un réseau de neurones récurrent.

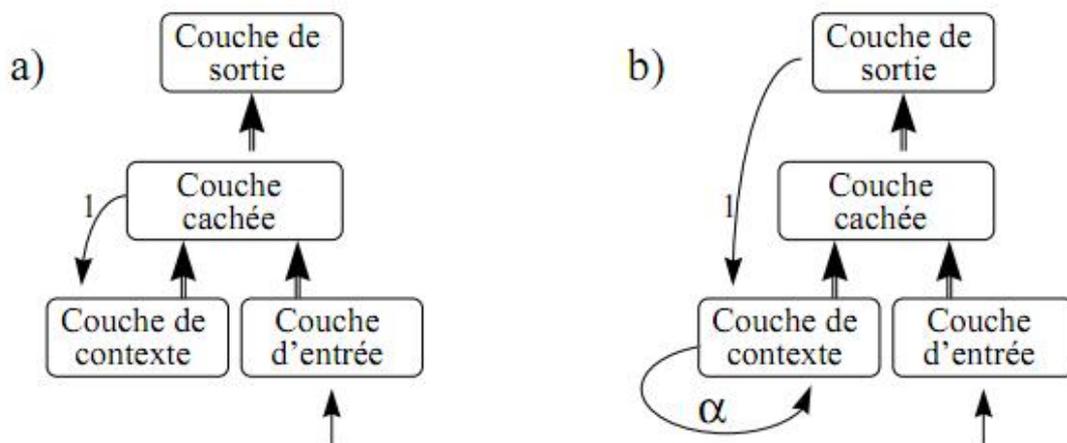


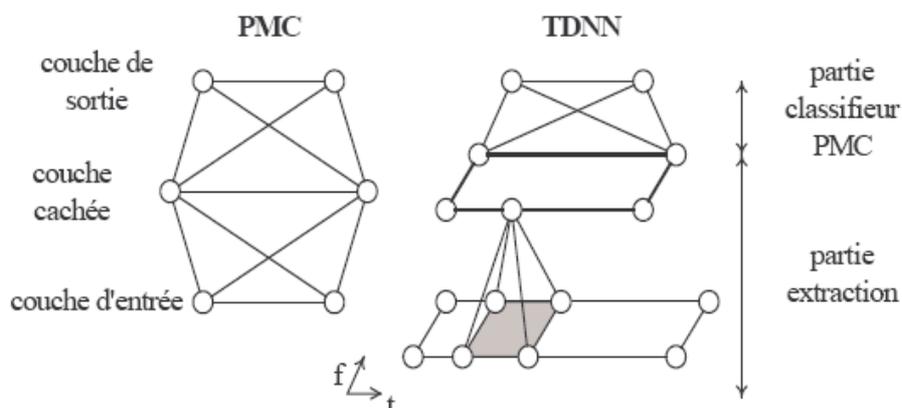
Figure 30 : Réseau récurrent d'Elman (a) et Jordan (b).

### III.8.2.2. Time delay neural network (TDNN)

Les données de précipitations forment un processus aléatoire, dynamique, non linéaire et présentent donc des caractéristiques à variation aléatoire au cours du temps, ce qui rend la prévision de leur occurrence très difficile. Le TDNN a été utilisé dans le domaine de la prévision des précipitations grâce à sa capacité d'identifier les caractéristiques indépendamment de leur position dans le temps.

Les réseaux de neurones de type TDNN sont des réseaux de type convolutif de par leur topologie, ils comportent une fenêtre glissante correspondant à un champ de vision restreint du signal global. C'est un réseau feedforward dynamique où l'information se propage de l'entrée vers la sortie sans retour en arrière et la dynamique se situe dans la couche d'entrée sous forme de retard. Il a été développé à l'origine dans le cadre de la reconnaissance de phonème [Lang et Hinton 1988], [Waibel 1989], à la reconnaissance vocale [Bottou 1990] et à la reconnaissance de l'écriture manuscrite [Guyon 1991]. Il a été par la suite, généralisé pour la modélisation et la prévision des données de nature séquentielles.

Le TDNN se singularise d'un réseau de neurones classique, tel que le réseau multicouche (MLP) par le fait qu'il prend en compte une certaine notion de temps. C'est à dire qu'au lieu de prendre en compte tous les neurones de la couche d'entrée en même temps, il prend une fenêtre du spectre puis effectue un balayage temporel.



**Figure 31** : La différence de structure entre PMC et TDNN.

### III.8.2.2.1. La structure de TDNN

Les TDNN sont constitués comme les Perceptrons multicouches (PMC) d'une couche d'entrée, de couches cachées et d'une couche de sortie, mais ils se différencient de part l'organisation des liaisons inter-couches. Les TDNN introduisent des contraintes qui leurs permettent d'avoir un certain degré d'invariance par décalage temporel et déformation. Celles-ci utilisent trois idées : poids partagés, fenêtre temporelle et délai.

Ø **Poids partagés** : Les poids partagés permettent de réduire le nombre de paramètres du réseau neuronal et induisent ainsi une capacité de généralisation plus importante. Ce concept de poids partagés est le comportement présumé du cerveau humain où des neurones détectent certains traits dans de petites régions de la rétine, essentiellement

de la même manière dans toutes ces régions. On a plusieurs neurones qui calculent la même fonction sur des entrées différentes.

Ø **Fenêtre temporelle** : Le concept de fenêtre temporelle implique que chaque neurone de la couche (l+1) n'est connecté qu'à un sous ensemble de la couche l (nous n'avons plus une connectivité totale). Cette fenêtre temporelle permet que chaque neurone n'ait qu'une vision locale du signal. Il peut être vu comme une unité de détection d'une caractéristique locale du signal.

Ø **Délais** : En plus des deux contraintes précédentes, nous introduisons des délais entre deux fenêtres successives pour une couche donnée.

#### III.8.2.2.2. Architecture et modélisation d'un TDNN

L'architecture retenue comporte deux parties principales. La première, correspondant aux couches basses, implémente les convolutions successives permettant de transformer progressivement les caractéristiques en grandeurs de plus en plus significatives vis-à-vis du problème. La seconde correspond à un MLP (Multi Layer Perceptron) classique, il reçoit en entrée l'ensemble des sorties de la partie TDNN. De plus, un neurone dans la partie extraction est identifié par sa couche l, sa caractéristique f et son emplacement temporel t. Tandis qu'un neurone du classifieur est identifié par sa couche l, et son emplacement t. Le TDNN se caractérise par (figure 32):

- Le nombre de couches : nb\_layers ;
- Le nombre de neurones de chaque couche selon la direction temporelle : window\_t ;
- Le nombre de neurones de chaque couche selon la direction caractéristique : nb\_feat ;
- Le nombre de neurones de la couche l vu par un neurone de la couche l+1 (la taille de la fenêtre temporelle) : field\_t ;
- Le délai temporel entre deux fenêtre successives : delay ;

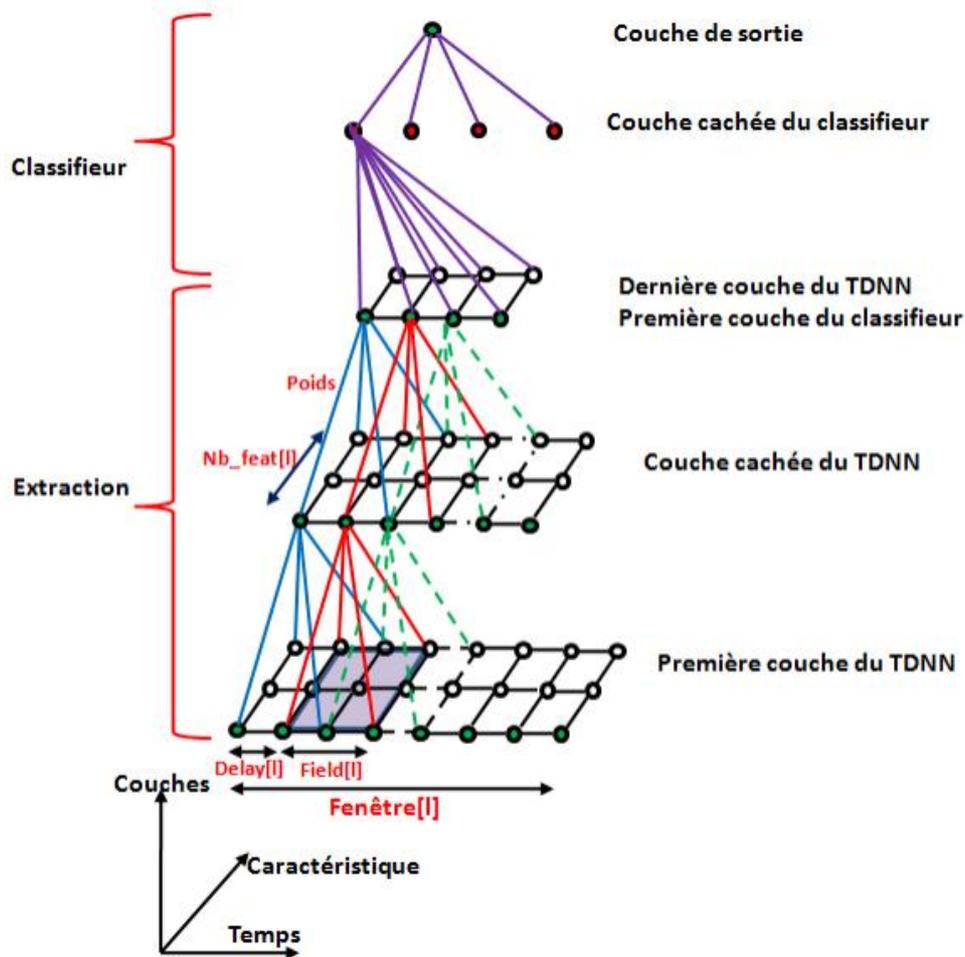


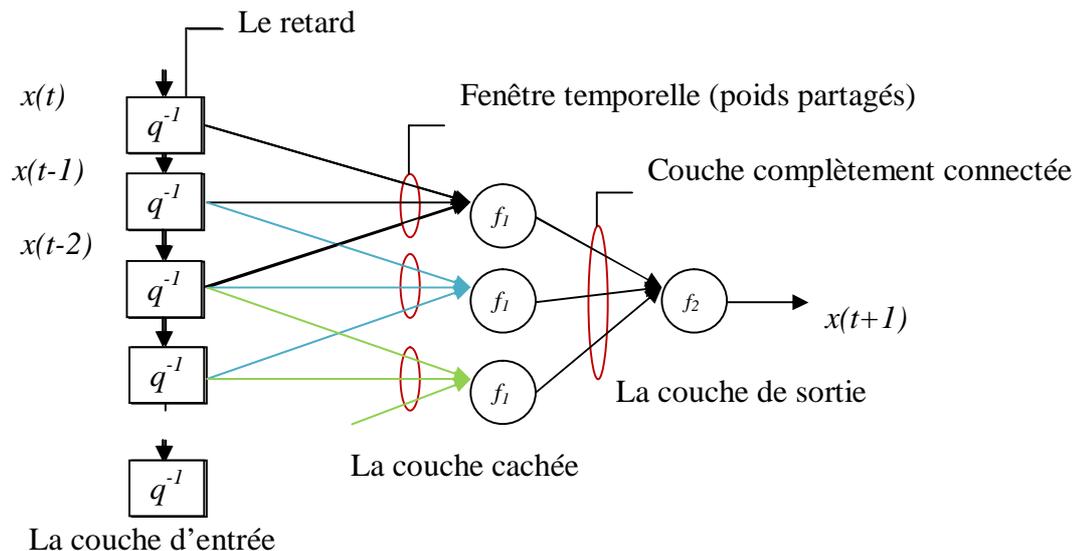
Figure 32 : La structure générale du TDNN.

### III.8.2.2.3. Le fonctionnement de TDNN

Le but du TDNN est non pas d'apprendre basiquement le signal temporel mais d'extraire les caractéristiques de celui-ci. Le champ de vision du neurone est restreint à une fenêtre temporelle limitée. Avec la contrainte des poids partagés, le même neurone est dupliqué dans la direction temps pour détecter la présence ou l'absence de la même caractéristique à différentes places le long de la trajectoire du signal. En utilisant plusieurs neurones ( $nb\_feat$ ) à chaque position temporelle, le réseau de neurones effectue la détection des caractéristiques différentes.

Les opérations réalisées par une couche du TDNN sont de type convolution. Chaque neurone  $k$  de la couche  $l+1$  a un noyau de taille  $w$  (nombre de neurones de la fenêtre temporelle de la couche  $l$ )  $\times$   $f$  (nombre de caractéristiques de la couche  $l$ ). La composante temporelle de la représentation du signal d'origine est éliminée au fur et à mesure en sous

échantillonnant la convolution à chaque couche. Pour compenser cette perte d'informations, le nombre de caractéristiques est multiplié. Nous avons donc une architecture de type bipyramidal. Ce réseau bipyramidal convertit progressivement des informations temporelles en des informations caractéristiques.



**Figure 33 :** TDNN avec une couche cachée de trois neurones, une fenêtre temporelle de taille 3 vue par chaque neurone de la couche cachée et une couche de sortie complètement connectée.

#### III.8.2.2.4. Apprentissage de TDNN

L'apprentissage, c'est-à-dire la détermination des poids neuronaux ou le calcul des paramètres des neurones est une étape fondamentale dans l'utilisation des techniques neuromimétiques. Ces calculs utilisent l'ensemble des données de la série temporelle pour chercher les meilleurs poids possible afin que le réseau de neurones reproduise le comportement du système. Idéalement, ceux-ci doivent permettre de converger rapidement vers le minimum global de la fonction de coût sélectionnée. Le TDNN est un enchaînement de deux sous réseaux, le premier est un réseau dynamique et le deuxième est un réseau statique comme le montre la figure 33. Pour l'apprentissage de ces deux sous réseaux, on peut appliquer la méthode d'apprentissage du premier ordre tel que la méthode de rétropropagation du gradient avec la contrainte de poids partagés. Toutefois, la version utilisée est la version dite du « gradient stochastique » qui permet de converger souvent plus rapidement que le gradient vrai, ou bien les méthodes de deuxième ordre tel que l'algorithme de Lavenberg-Marquardt, l'algorithme Newton, Quasi-Newton, Gauss-Newton, pour différentes largeurs de

la fenêtre temporelle appliquées à l'entrée du TDNN. Nous présentons ci-dessous celles que nous avons mise en œuvre dans ce travail.

#### III.8.2.2.4.1. Méthode du gradient récursif (Algorithme de premier ordre)

L'algorithme récursif du gradient, porte aussi le nom d'algorithme du gradient stochastique, est la méthode la plus utilisée dans la communauté neuronale. Il consiste à modifier les paramètres proportionnellement au gradient de la fonction du coût partiel. Afin d'estimer les paramètres d'un réseau de neurones par l'algorithme du gradient, il faut suivre les étapes suivantes :

1. Choisir la taille du réseau et initialiser les poids et les seuils du réseau (selon le système à modéliser).
2. Calculer les sorties des différentes couches,

La sortie d'un neurone  $i$  est donnée par la relation suivante :

$$y_i = g(n_i) = g\left(\sum_{j=1}^{n_1} w_{ij}x_j\right) \quad (\text{III.8})$$

Où  $g$  est la fonction d'activation du neurone,  $w_{ij}$  représente les poids qui relient le neurone  $i$  aux neurones  $j$  de la couche précédente.

3. Calculer la fonction de coût partiel relatif à l'exemple  $k$ , généralement on choisit une fonction de coût quadratique.

L'erreur quadratique sur la sortie est donnée par la relation suivante :

$$J = \frac{1}{2} \sum_{i=1}^{n_2} (y_i - y_{di})^2 \quad (\text{III.9})$$

Où  $y_i$  est la sortie du réseau de neurones,  $y_{di}$  est la sortie désirée (mesurée).

4. Calculer le gradient de la fonction de coût par la rétropropagation des erreurs depuis les sorties vers les entrées en utilisant des grandeurs intermédiaires relatives à des neurones qui se trouvent entre ce paramètre et les sorties du réseau:

$$\frac{\partial J^k(w)}{\partial w_{ij}} = \frac{\partial J^k(w)}{\partial A_i} \frac{\partial A_i}{\partial w_{ij}} = \frac{\partial J^k(w)}{\partial A_i} x_j = \delta_i^k x_j \quad (\text{III.10})$$

Où  $A_i$  présente l'activation du neurone  $i$  et  $\delta_i^k$  est la sensibilité de  $J$  aux changements dans le niveau d'activation  $A_i$  de neurone  $i$  de la couche  $k$ .

5. Modifier les paramètres selon la loi suivante :

$$W(k) = W(k-1) - \mu(k) \frac{\partial J}{\partial W} / W(k-1) \quad (\text{III.11})$$

Avec  $\mu(k) \geq 0$  est le pas de progression dans la direction de descente.

Le choix de la valeur de  $\mu(k)$  est très important, on le choisit généralement entre 0 et 1. Il permet d'accélérer la convergence lorsqu'il est proche de 1 ou de la freiner lorsqu'il est proche de 0. Parfois pour accélérer la convergence de l'apprentissage, on utilise un terme qu'on appelle momentum.

Notons que les seuils sont ajustés de la même façon que les poids, et ça en les supposant comme étant poids de connexion reliés à des entrées à valeurs.

6. Répéter à partir de l'étape 2 jusqu'à atteindre une erreur quadratique moyenne prédéfinie, ou jusqu'à atteindre un certain nombre d'itérations prédéfinie.

#### III.8.2.2.4.2. Algorithme avec poids partagés

On parle de poids partagés (connexions locales) dans un réseau, si pour chaque neurone  $i$  d'une couche  $p$ , seulement un groupe réduit de neurones de la couche précédente  $q$ , est connecté à  $i$ . Dans ce cas la minimisation de la fonction de coût  $J(w)$  varie légèrement par rapport à l'équation III.10. En effet, l'algorithme doit assurer une mise à jour des poids afin de donner la même valeur aux poids partagés.

Soit  $w_{ij}$  le poids de connexion entre les cellules  $j$  et  $i$  du réseau. Ce poids étant partagé par d'autres connexions entre cellules, appelons  $N_i$  l'ensemble des couples de cellules qui partagent la connexion  $w_{ij}$ , et d'une façon plus générale  $w_i$  est le poids  $w_{ij}$  qui les relie. On a donc :  $N_i = \{(i_1, j_1), (i_2, j_2), \dots\}$ . Le gradient de la fonction de coût  $J(w)$  par rapport à  $w_i$  est donné par :

$$\frac{\partial J^k(w)}{\partial w_i} = \sum_{(i,j) \in N_i} \frac{\partial J^k(w)}{\partial A_i} \frac{\partial A_i}{\partial w_{ij}} = \sum_{(i,j) \in N_i} \delta_i^k x_j \quad (\text{III.12})$$

De cette façon, on peut modifier l'étape 5 dans l'algorithme de la rétropropagation du gradient stochastique et donner une forme plus générale de la mise au point des poids de connexions. Notons cependant que cette façon de modifier les poids englobe le cas où il n'y a pas de poids partagés et où tous les poids sont différents. Dans ce cas, les différents ensembles  $N_i$  ne contiennent qu'un seul couple  $(i, j)$  indiquant les deux cellules correspondant à la connexion qui a comme poids  $w_i = w_{ij}$ .

### III.8.2.2.4.3. Méthode de Levenberg-Marquardt (Méthode de deuxième ordre)

L'algorithme de Levenberg-Marquardt (LM) est un algorithme itératif, qui consiste à modifier les paramètres proportionnellement au gradient de la fonction du coût total. Il a été appliqué pour l'apprentissage des réseaux de neurones dans plusieurs travaux de recherche. Cet algorithme adopte une direction de déplacement qui n'est plus la direction du gradient, mais une transformation linéaire appropriée du gradient du coût total. Il consiste à modifier les paramètres du réseau selon la formule suivante [Dreyfus et al. 2002]:

$$W(i) = W(i-1) - [H(W(i-1)) + \lambda_i I]^{-1} \nabla J(W(i-1)) \quad (\text{III.13})$$

Où  $\lambda_i$  est un paramètre constant strictement supérieur à zéro,

$H$  est la matrice hessienne ou la dérivée seconde par rapport aux paramètres.

Cet algorithme s'applique à la fonction de coût quadratique parce que son hessien peut être approché par la relation suivante:

$$H = J a^T J a \quad (\text{III.14})$$

Où  $J a$  est la matrice jacobienne ou le jacobien.

Un calcul détaillé de la matrice jacobienne peut être trouvé dans la thèse de doctorat de Mahul [Mahul 2005]. L'algorithme de Levenberg-Marquardt est détaillé par l'algorithme ci-dessous.

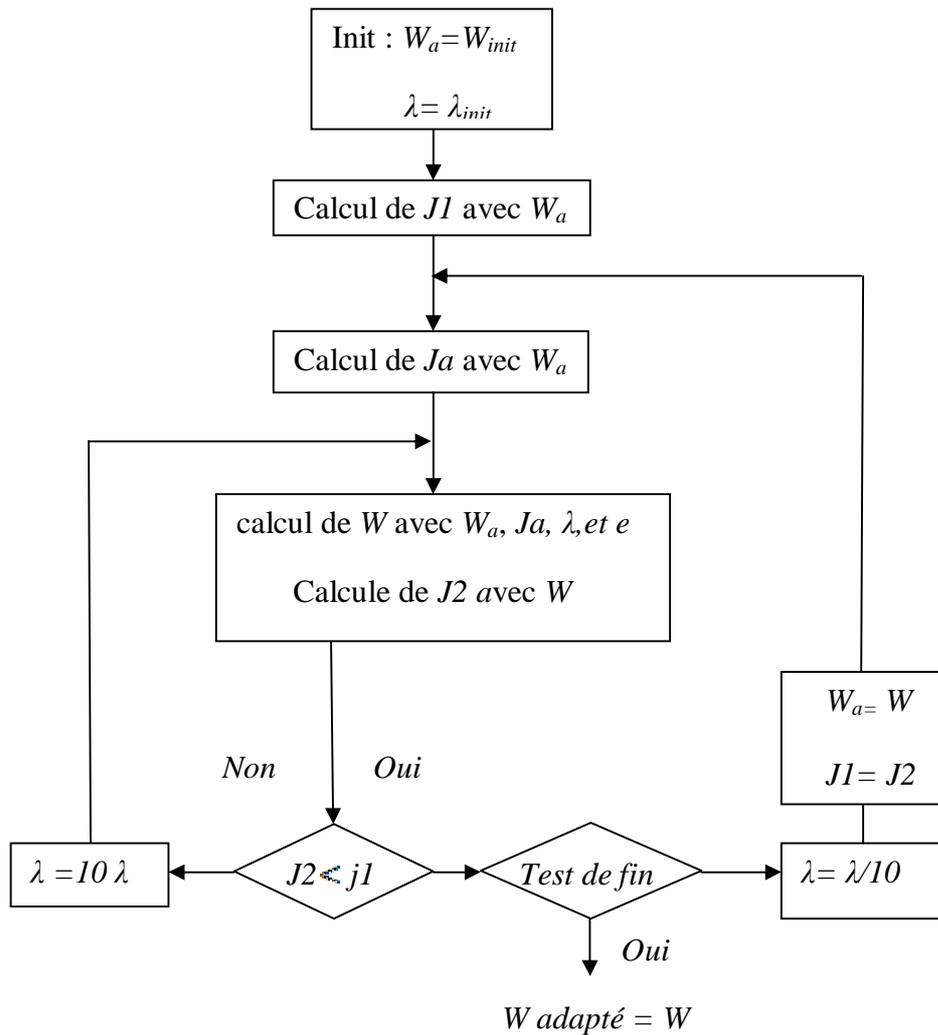


Figure 34 : L'algorithme de Lavenberg-Marquardt.

### III.8.2.3. Time Delay Radial Basis Function (TDRBF)

Les réseaux TDRBF combinent les caractéristiques des TDNN et des RBF. Ce type de réseau de neurones utilise également une fenêtre temporelle à l'entrée du réseau (comme le TDNN) et les fonctions d'activations de type gaussienne (fonctions à base radiales) au niveau de la couche cachée. Plus le vecteur d'entrée est proche du centre d'une Gaussienne, plus la sortie du neurone de la première couche qui lui correspond est élevée. L'appellation « Fonction de Base Radiale » vient du fait que la Gaussienne est symétrique radialement, c'est-à-dire que la valeur de sortie obtenue est la même pour toutes les entrées situées à une même distance du centre de la Gaussienne.

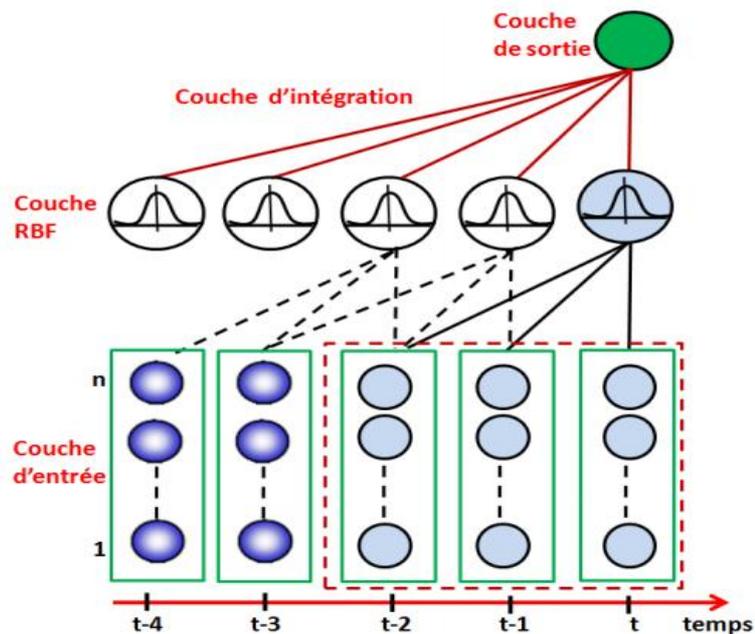


Figure 35 : Topologie du réseau TDRBF.

### III.9. Tests statistiques pour l'évaluation des prévisions

Historiquement, l'évaluation de la qualité des différentes méthodes de prévision du temps est une question centrale qui se pose depuis au moins 1884 [Murphy, 1996]. Différents tests statistiques existent qui permettent à chaque fois de mesurer la relation entre observation (*OB*) et prévision (*PR*) ou ensemble de prévisions. Parmi ces tests, l'Erreur Quadratique Moyenne (EQM), le coefficient de corrélation linéaire *R* et le LEPS (Linear Error in Probability Space) ont été sélectionnés pour des variables continues. Notons que ces tests visent, non seulement, à mesurer la précision des prévisions, mais aussi la performance des modèles. Nous présentons ci-dessous les tests statistiques que nous avons mis en œuvre dans ce travail :

#### III.9.1. L'Erreur Quadratique Moyenne (EQM)

L'erreur quadratique moyenne EQM mesure l'écart entre les grandeurs physiques réelles (observations) et les grandeurs prévues par le modèle. Ce score a été utilisé pour voir si les précipitations étaient surestimées ou sous-estimées par les méthodes de prévisions. L'erreur quadratique moyenne EQM est donnée par la relation suivante :

$$EQM = \frac{1}{n} \sum_{i=1}^n (PR - OB)^2 \quad (\text{III.15})$$

Avec *OB* qui signifie grandeurs physiques réelles (observations), et *PR* les grandeurs prévues par le modèle de prévision.

### III.9.2. Le coefficient de corrélation linéaire ( $R$ )

L'étude des corrélations linéaires est très utilisée en climatologie, ce coefficient ayant pour objet de mesurer l'intensité de la liaison linéaire entre les valeurs prédites par le réseau et les valeurs désirées. Le coefficient de corrélation linéaire  $R$  entre les variables observées ( $OB$ ) et prévues ( $PR$ ) s'obtient par le rapport suivant :

$$R = \frac{cov(PR,OB)}{\sigma_{PR} \times \sigma_{OB}} \quad (III.16)$$

Où  $\sigma_{PR}$ ,  $\sigma_{OB}$  sont les écarts types des valeurs observées ( $OB$ ) et prévues ( $PR$ ) respectivement, et  $cov(PR, OB)$  leur covariance. Notons que le coefficient de corrélation est toujours compris entre -1 et 1.

- Ø Dans le cas où  $R = 1$ , cela signifie qu'il y a une parfaite corrélation entre les valeurs prédites par le réseau et les valeurs désirées,
- Ø Dans le cas où  $R = 0$ , il n'existe pas de corrélation entre les deux variables,
- Ø Dans le cas intermédiaire où  $R$  est proche de 1, on dit qu'il ya un meilleur accord entre la cible et les valeurs prévues.

### III.10. Discussion

Dans ce chapitre, nous avons présenté les définitions et propriétés fondamentales des réseaux de neurones. Leur utilisation est avantageuse en raison de leur propriété d'approximation parcimonieuse de fonction et ils sont appropriés lorsqu'il s'agit de modéliser des processus statiques ou dynamiques.

La présentation des réseaux de neurones nous a amené à expliciter l'apprentissage. Ce dernier se fait sur un ensemble de couples entrées-sorties connus en nombre restreint afin d'estimer les paramètres du réseau en minimisant l'erreur quadratique moyenne. La modélisation de processus dynamiques pour la prévision a été présentée. Nous avons introduit la notion de séries temporelles et d'horizon de prévision. Parmi les applications très nombreuses des réseaux de neurones dans la prévision des précipitations, quelques exemples ont permis de mettre en évidence la pertinence de la présentation précédente. Ainsi, on a mis l'accent sur le TDNN, sa structure, son fonctionnement et son apprentissage. Enfin, on a cité quelques tests statistiques pour évaluer la qualité des différentes méthodes de prévision du temps ainsi que les modèles utilisés. Le chapitre suivant est consacré à l'application du TDNN pour la prévision des données de précipitations.

# Chapitre IV

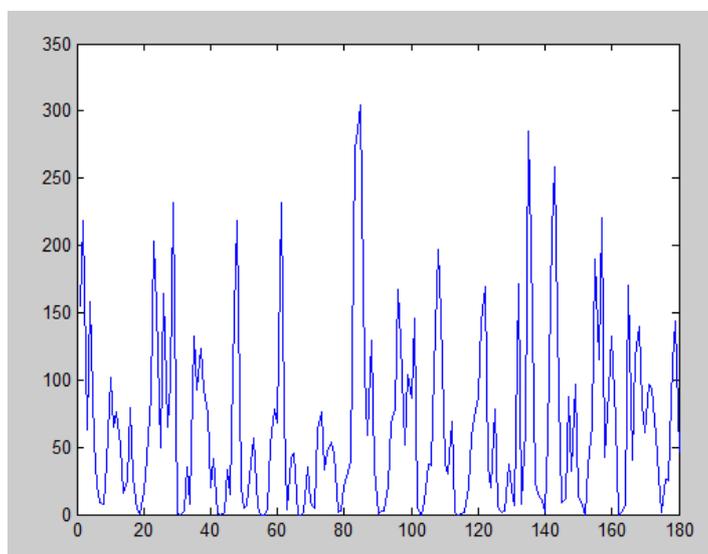
### **IV.1. Préambule**

Les données de précipitations forment un processus aléatoire, dynamique, non linéaire et présentent donc des caractéristiques à variation aléatoire au cours du temps, ce qui rend la prévision de leur occurrence très difficile. Pour toutes ces raisons, le Time Delay Neural Network (TDNN) a été choisi pour cette application. La propriété principale du TDNN est la capacité d'identifier les caractéristiques locales indépendamment de leur position dans le temps.

### **IV.2. Les données expérimentales**

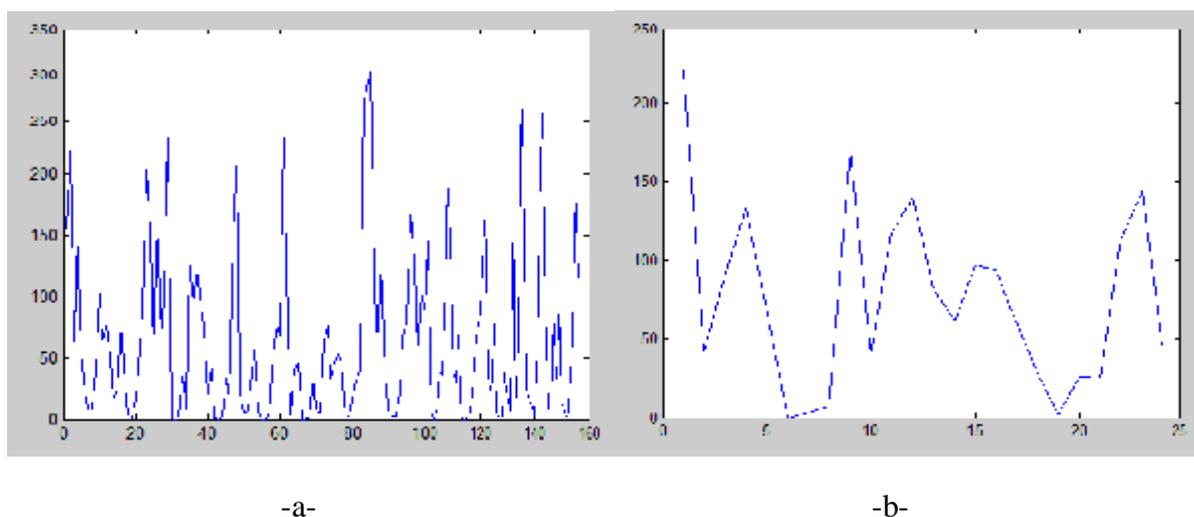
En Algérie, l'office national de la météorologie (O.N.M) dispose de plusieurs stations pour observer l'atmosphère. Dans ce contexte, les stations effectuent des mesures de rayonnement solaire, de nébulosité, de température, d'humidité, de pression atmosphérique, ainsi que des précipitations et des paramètres caractéristiques des nuages. Les données qui en résultent, servent alors à caractériser le temps qu'il fait dans une région donnée. Les observations météorologiques sont effectuées à heures fixes mais elles peuvent être cumulées à l'échelle de la journée et du mois (séries chronologiques). Notons que même si le recueil des données a été fait de façon satisfaisante dans une station météorologique donnée, la garantie de qualité avant l'utilisation, ou en vue de la conservation de ces données, ne peut être acquise qu'au prix d'un minimum d'opérations de contrôle. Une vérification s'impose alors pour la détection des erreurs d'observations et pour corriger la série de données météorologiques mise en jeu.

Les données de précipitations que nous avons utilisées dans notre application, sont des mesures mensuelles relevées à la station météorologique de Tizi-Ouzou, sur une période de 15 ans, allant du mois de janvier 1996 au mois de décembre 2010. Notons que la station météorologique de Tizi-Ouzou est située à la latitude de  $36^{\circ} 42'$  N et à la longitude de  $04^{\circ} 08'$  E. Afin d'illustrer l'évolution des précipitations, nous avons représenté les variations mensuelles de notre base de donnée, par la courbe de la figure 36.



**Figure 36** : Les données mensuelles de précipitations collectées dans la région de Tizi-Ouzou pendant 15 ans (l'intensité de précipitations en fonction du nombre de mois).

Cette base de données a été divisée en deux sous ensembles, le premier sert à effectuer l'apprentissage et le deuxième sous ensemble est utilisé pour tester les performances du TDNN, c'est ce qu'on appelle l'ensemble test. Notons que l'ensemble d'apprentissage contient 160 mois, ce qui fait une période de 13 ans allant du mois de janvier 1996 au mois de décembre 2008 et l'ensemble test contient 24 mois ce qui donne une période de 2 ans allant du mois de janvier 2009 au mois de décembre 2010, comme le montre les courbes de la figure 37 :



**Figure 37** : Les données mensuelles des précipitations : -a- séquence d'apprentissage, -b- séquence de test.

### IV.3. TDNN pour la prévision d'une série temporelle de précipitations

Les précipitations est un phénomène météorologique aléatoire, dynamique et non linéaire et présente donc des caractéristiques très variable au cours du temps. La prévision des occurrences devient alors compliquée. Grâce aux caractéristiques du TDNN, nous l'avons retenu pour le tester sur notre base de données.

La propriété principale du TDNN est la capacité d'identifier les caractéristiques locales indépendamment de leur position dans le temps. Le TDNN est un réseau feedforward dynamique, où l'information se propage de l'entrée vers la sortie sans retour en arrière et la dynamique se situe dans la couche d'entrée, sous forme de retard. Le TDNN se singularise d'un réseau de neurones classique, tel que le réseau multicouche (MLP) par le fait qu'il prenne en compte une certaine notion du temps. C'est à dire qu'au lieu de prendre en compte tous les neurones de la couche d'entrée en même temps, il prend une fenêtre du spectre puis effectue un balayage temporel. Ce qui permet au réseau de prendre en considération les caractéristiques locales de la série temporelle de précipitations. Notons que le TDNN utilise la contrainte de poids partagés ce qui lui permet de réduire le nombre de paramètres du réseau neuronal en induisant ainsi une capacité de généralisation plus importante. Le réseau utilisé contient une couche cachée de trois neurones, de fonction d'activation non linéaire de type sigmoïde et un neurone dans la couche de sortie de fonction d'activation linéaire.

Différentes largeurs de la fenêtre temporelle appliquées à l'entrée du TDNN sont utilisées avec deux algorithmes d'apprentissage : l'algorithme de gradient récursif qui est un algorithme de premier ordre et celui de Lavenberg-Marquardt qui est un algorithme de deuxième ordre, afin de prévoir les valeurs futures des précipitations et de constater leur structure de dépendance dans le temps.

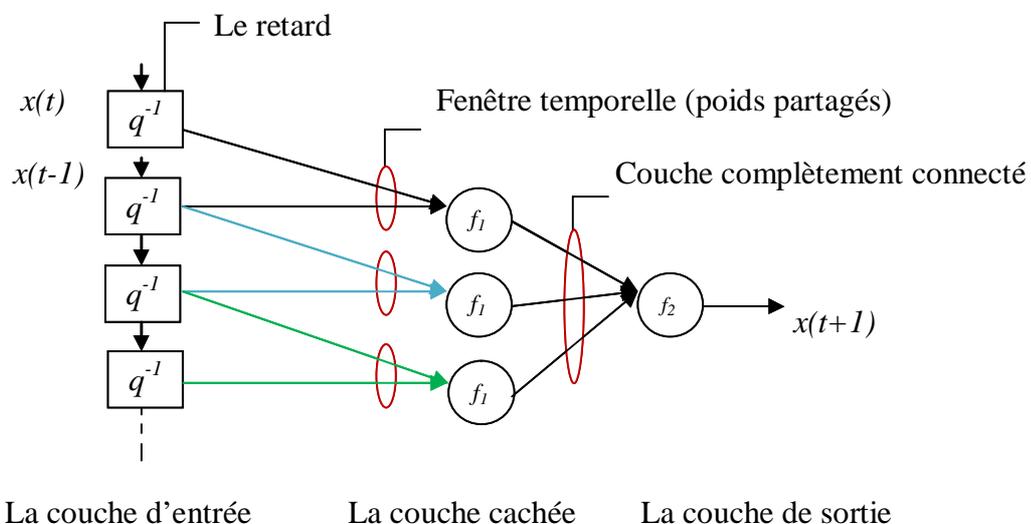
L'apprentissage est stoppé lorsque l'erreur de validation devient minimum, ou bien le nombre d'itération fixé atteint le maximum. Ensuite le réseau est évalué à partir des données différentes de celles utilisées durant l'apprentissage. Par définition ce dernier jeu de données est appelé ensemble test.

L'erreur quadratique moyenne (EQM) et le coefficient de corrélation linéaire  $R$  sont utilisés pour comparer les résultats des différentes largeurs de la fenêtre temporelle appliquées à l'entrée du réseau, ensuite évaluer les performances du TDNN, ainsi que la précision de la prévision.

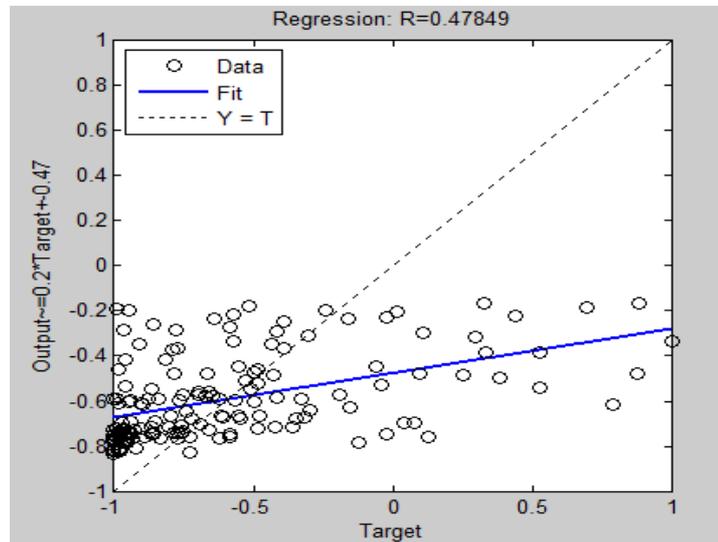
Le pas du gradient est un facteur déterminant dans la vitesse de convergence du réseau de neurones. Le temps d'apprentissage croissant très vite avec la complexité du réseau, il faut donc trouver un pas optimal. Plus le pas est petit, plus le nombre d'itérations de la base d'apprentissage sera important. Tandis que plus le pas est grand, plus ce nombre d'itérations nécessaire sera faible mais le réseau risque de diverger. Dans le cadre de notre expérience le taux d'apprentissage est fixé à 0.01 et un nombre maximum d'itération est fixé à 300 itérations. Durant l'apprentissage, le réseau est validé toutes les 10 itérations pour obtenir les résultats suivants :

### IV.3.1. Application de la méthode d'apprentissage de premier ordre : l'algorithme de gradient récursif

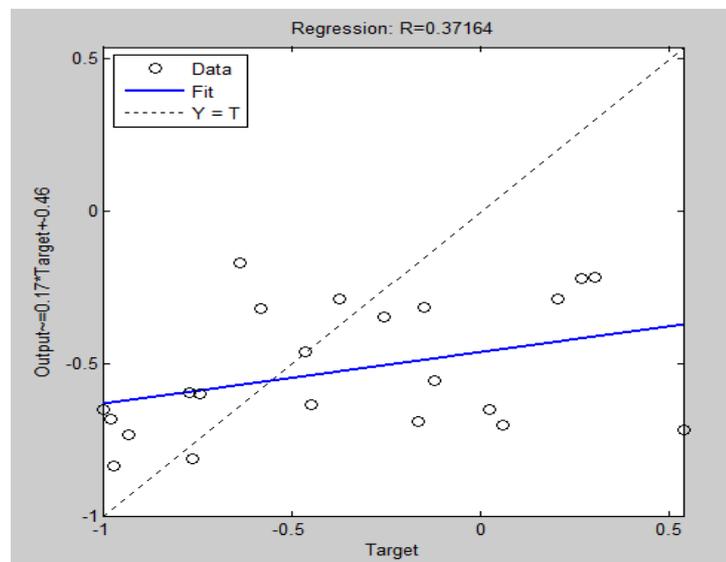
#### a. TDNN avec une fenêtre temporelle de taille 2 reliée à chaque neurone de la couche cachée



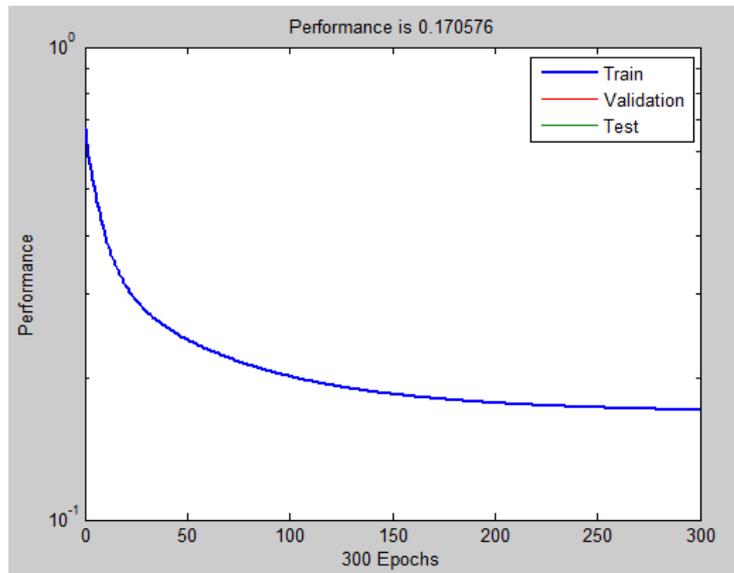
**Figure 38** : TDNN avec une fenêtre temporelle de taille 2 reliée à chaque neurone de la couche cachée.



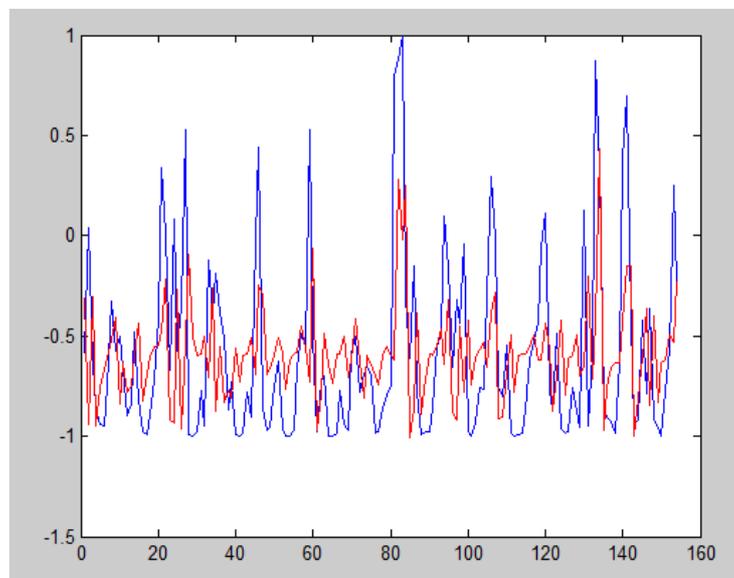
**Figure 39** : Le coefficient de corrélation linéaire R entre les sorties désirées et les sorties prédites par le TDNN (l'étape d'apprentissage).



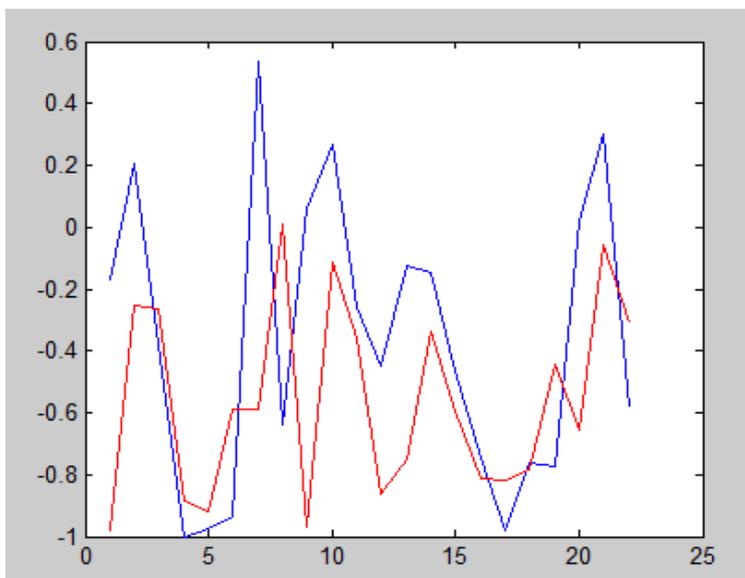
**Figure 40** : Le coefficient de corrélation linéaire R entre les sorties désirées et les sorties prédites par le TDNN (l'étape test).



**Figure 41** : L'erreur quadratique moyenne (EQM) en fonction du nombre d'itération.

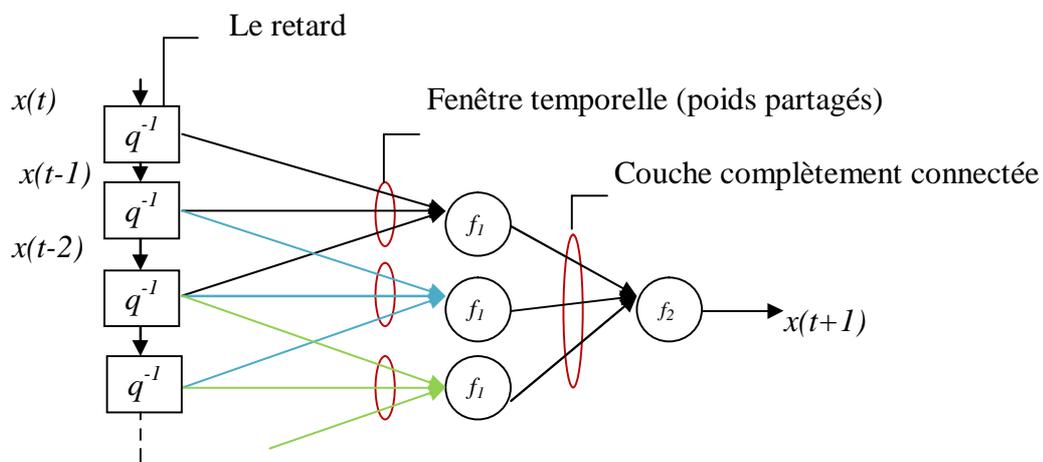


**Figure 42** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape d'apprentissage).

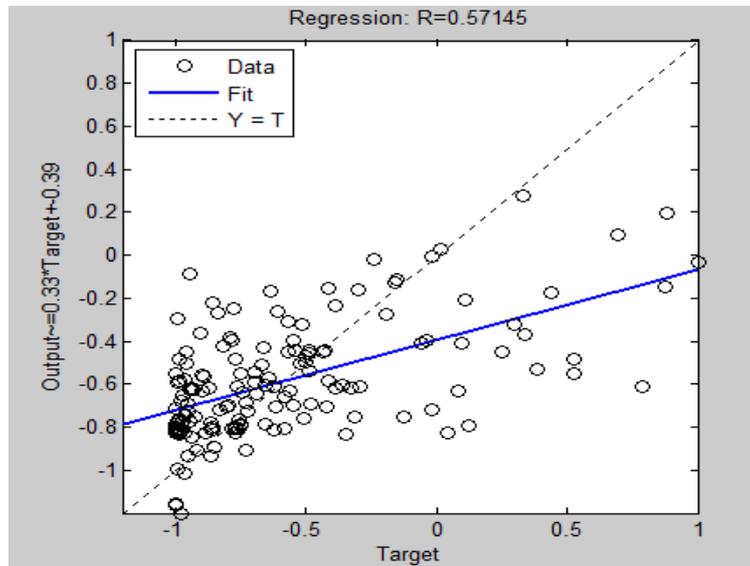


**Figure 43** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape test).

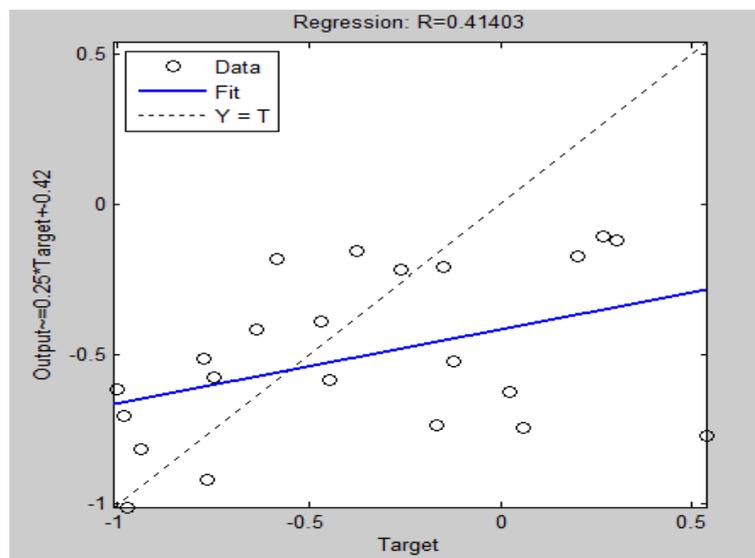
**b. TDNN avec une fenêtre temporelle de taille 3 reliée à chaque neurone de la couche cachée**



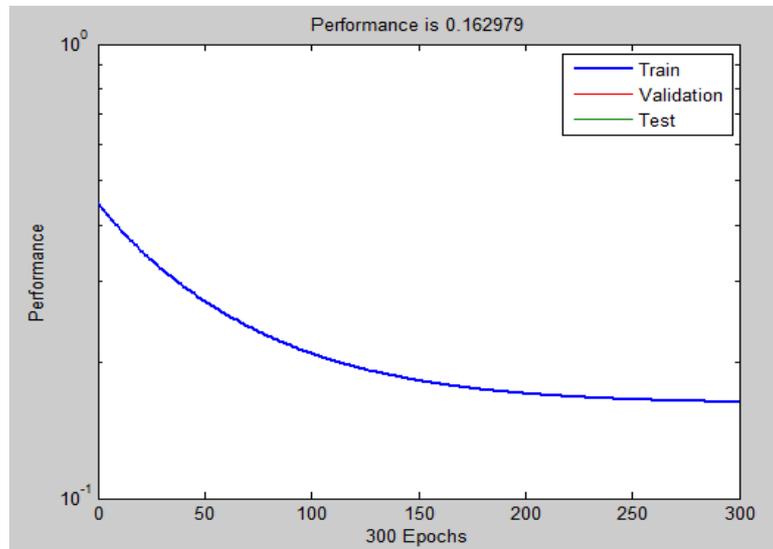
**Figure 44** : TDNN avec une fenêtre temporelle de taille 3 reliée à chaque neurone de la couche cachée.



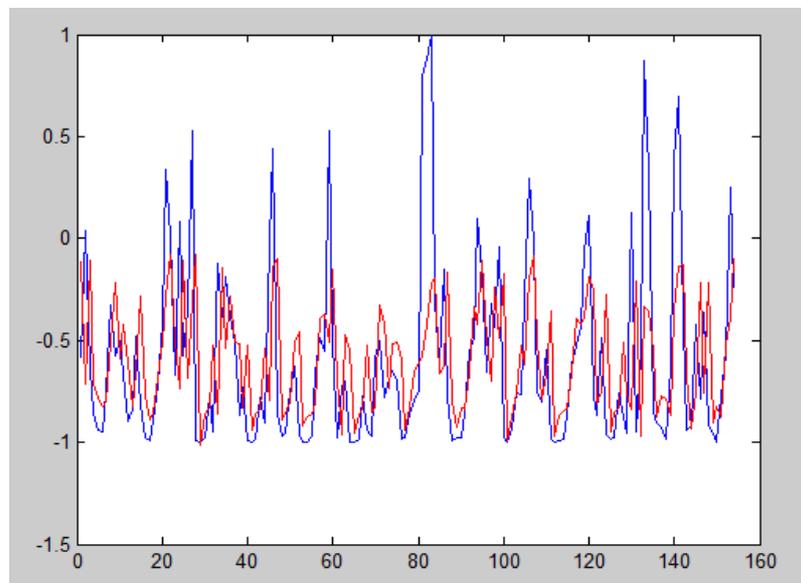
**Figure 45** : Le coefficient de corrélation linéaire  $R$  entre les sorties désirées et les sorties prédites par le TDNN (l'étape d'apprentissage).



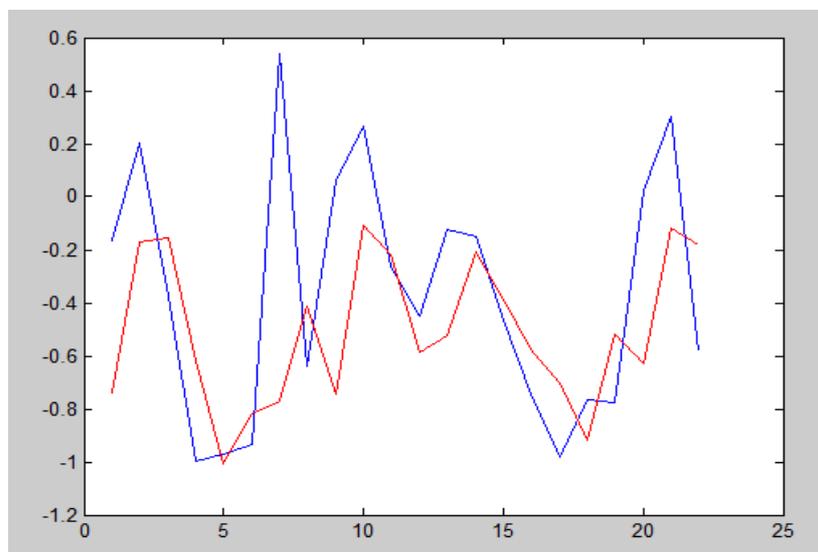
**Figure 46** : Le coefficient de corrélation linéaire  $R$  entre les sorties désirées et les sorties prédites par le TDNN (l'étape test).



**Figure 47** : L'erreur quadratique moyenne (EQM) en fonction de nombre d'itération.



**Figure 48** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape d'apprentissage).



**Figure 49:** Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape test).

#### Interprétation des résultats :

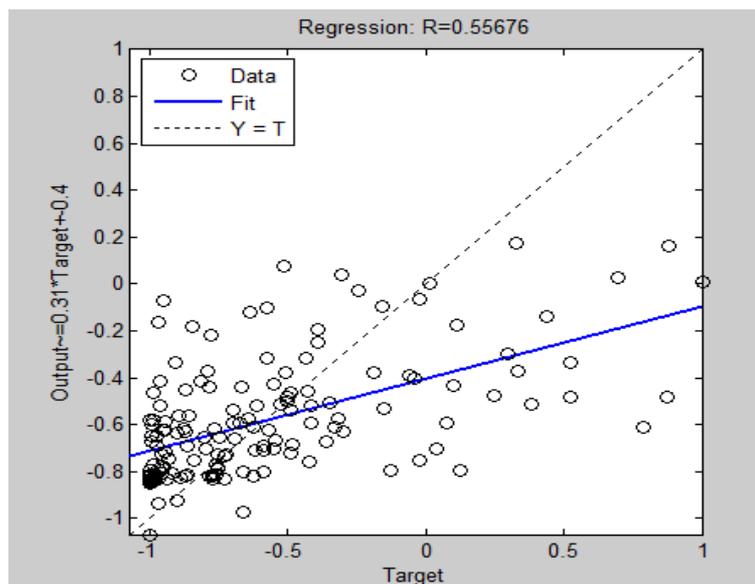
	Apprentissage		Test	
	EQM	R	EQM	R
TDNN avec une fenêtre temporelle de taille 2	0.1705	0.4784	0.2430	0.3716
TDNN avec une fenêtre temporelle de taille 3	0.1629	0.5714	0.2407	0.4140

**Tableau 1 :** Les résultats obtenus après l'étape d'apprentissage et l'étape test.

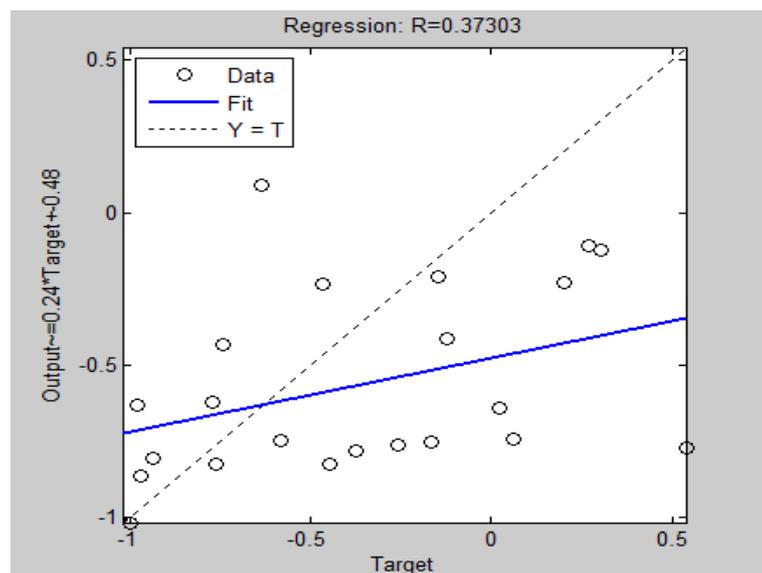
Les résultats obtenus montrent qu'avec l'algorithme de premier ordre (l'algorithme de gradient récursif), l'erreur quadratique moyenne (EQM) diminue en augmentant la largeur de la fenêtre temporelle appliquée à l'entrée du TDNN, et que le coefficient de corrélation linéaire R augmente et se rapproche de 1. Ceci veut dire que la précision de la prévision augmente lorsqu'on augmente la largeur de la fenêtre temporelle. Cependant, quand on dépasse la largeur de taille trois, on trouve que le coefficient de corrélation diminue notablement tandis que l'erreur d'estimation augmente. Ce résultat fixe à trois la taille optimale de la fenêtre temporelle appliquée à l'entrée du TDNN.

### IV.3.2. Application de la méthode d'apprentissage du deuxième ordre : l'algorithme de Lavenberg-Marquardt

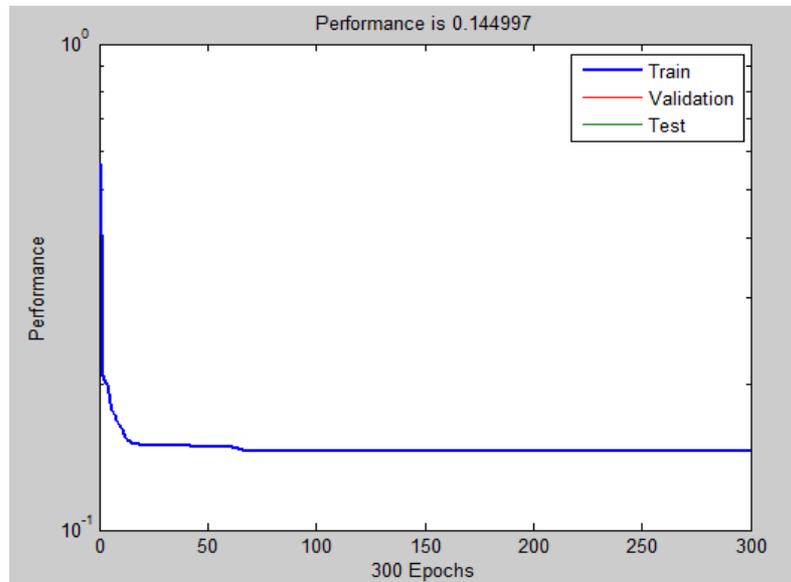
- a. TDNN avec une fenêtre temporelle de taille 2 reliée à chaque neurone de la couche cachée



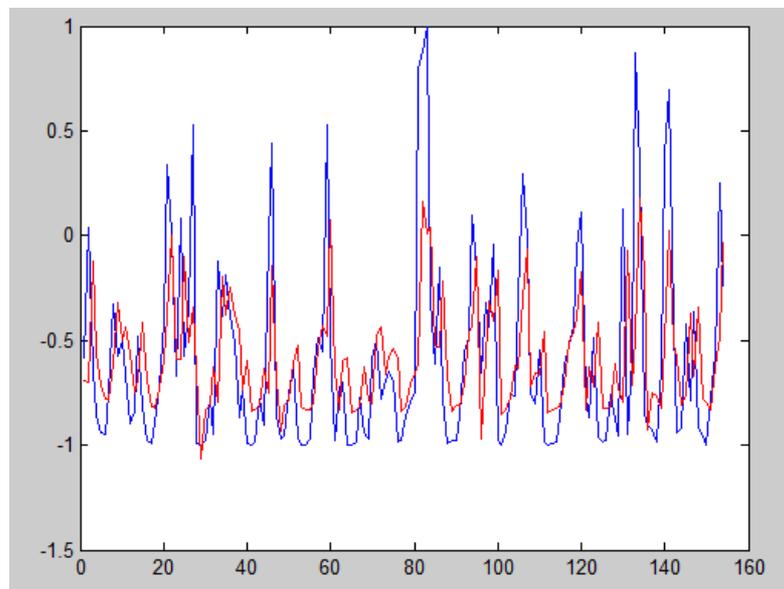
**Figure 50** : Le coefficient de corrélation linéaire  $R$  entre les sorties désirées et les sorties prédites par le TDNN (l'étape d'apprentissage).



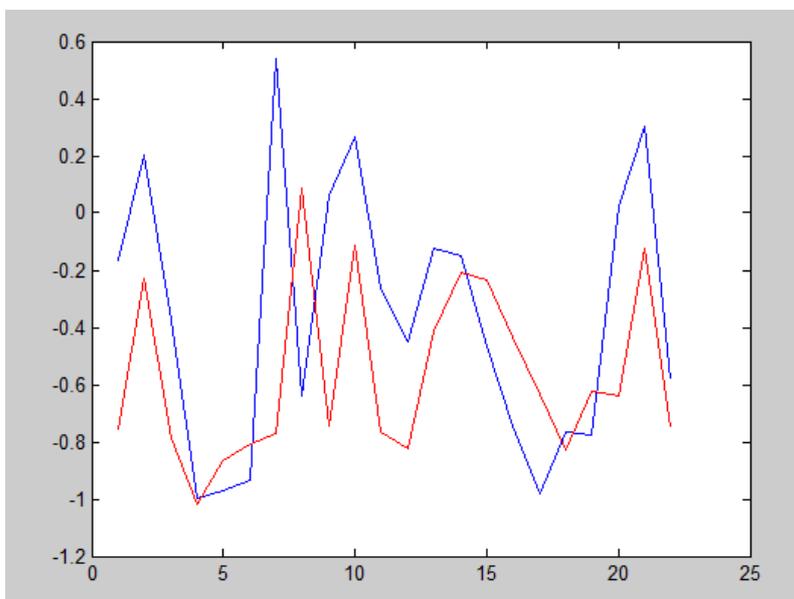
**Figure 51** : Le coefficient de corrélation linéaire entre les sorties désirées et les sorties prédites par le TDNN (l'étape test).



**Figure 52 :** L'erreur quadratique moyenne (EQM) en fonction de nombre d'itération.

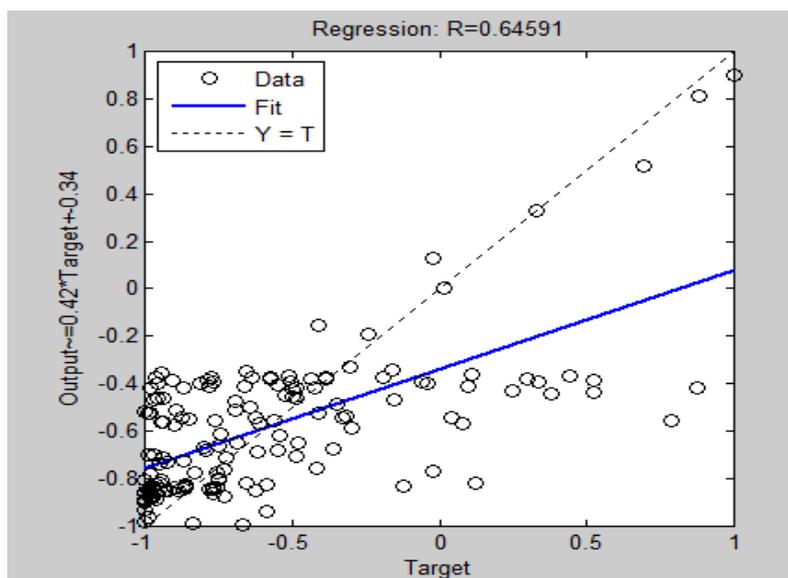


**Figure 53 :** Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape d'apprentissage).

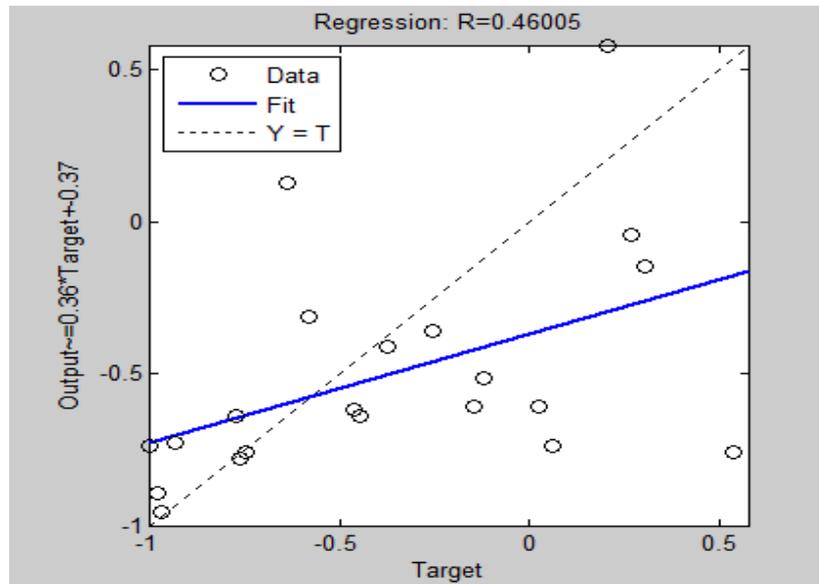


**Figure 54** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape test).

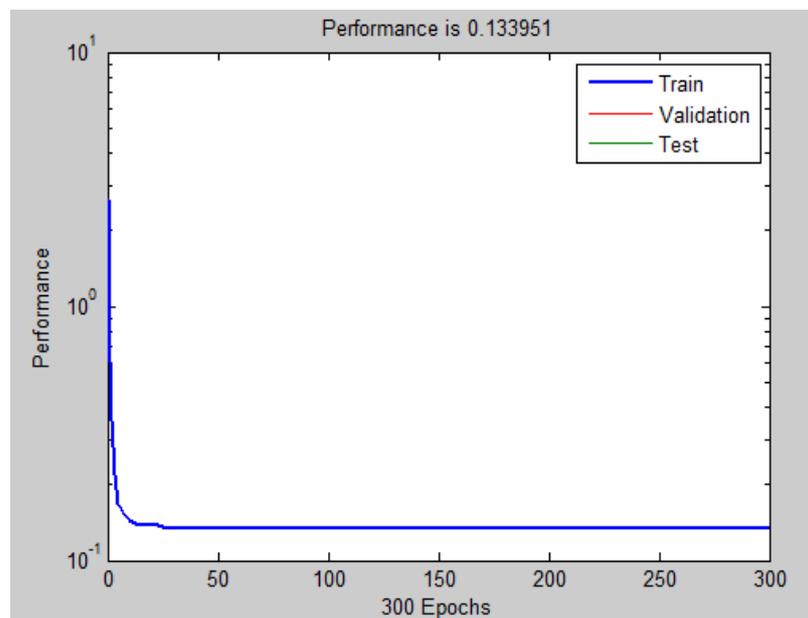
**b. TDNN avec une fenêtre temporelle de taille 3 reliée à chaque neurone de la couche cachée**



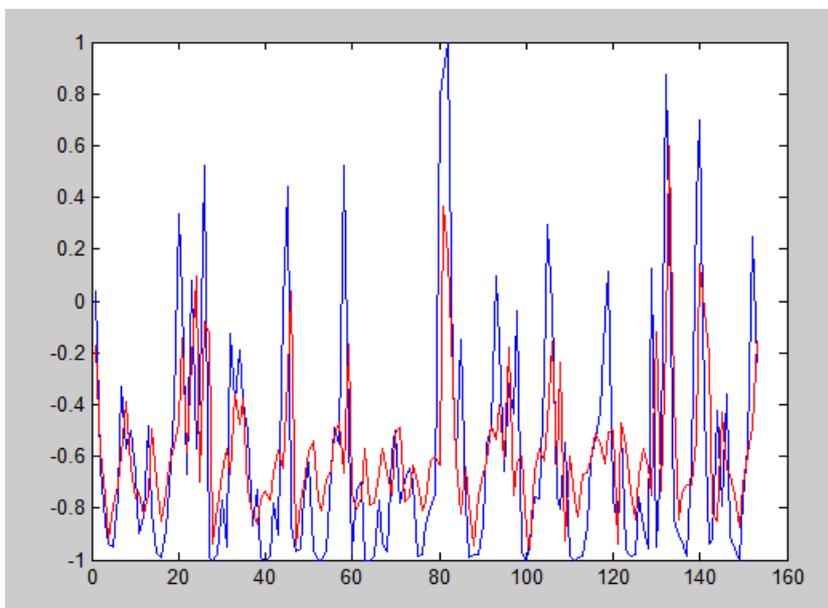
**Figure 55** : Le coefficient de corrélation linéaire R entre les sorties désirées et les sorties prédites par le TDNN (l'étape d'apprentissage).



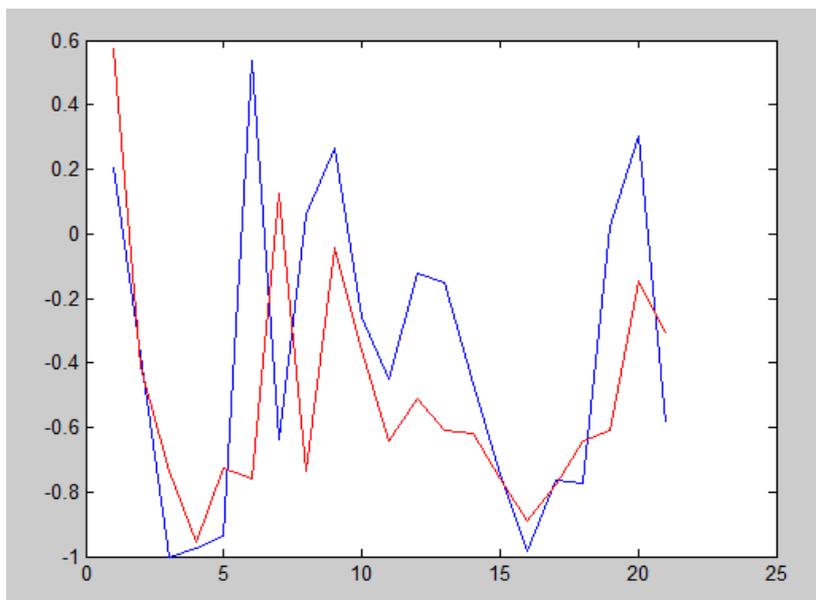
**Figure 56** : Le coefficient de corrélation linéaire  $R$  entre les sorties désirées et les sorties prédites par le TDNN (l'étape test).



**Figure 57** : L'erreur quadratique moyenne (EQM) en fonction de nombre d'itération.

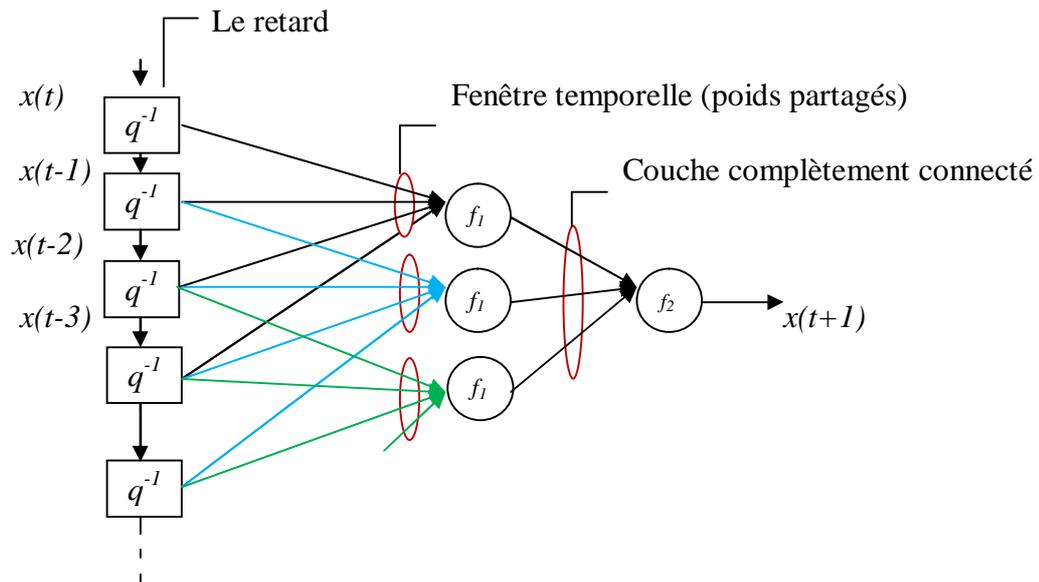


**Figure 58** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape d'apprentissage).

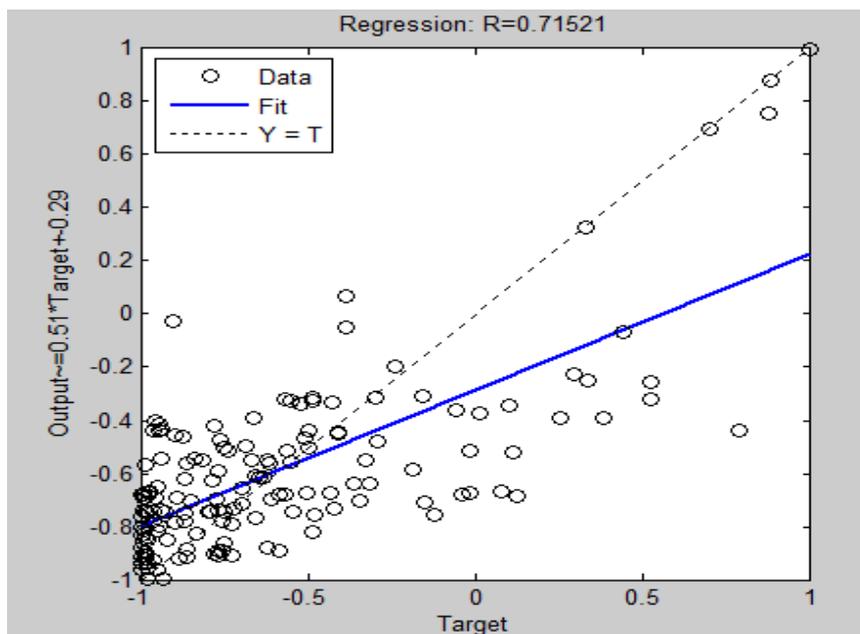


**Figure 59** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape test).

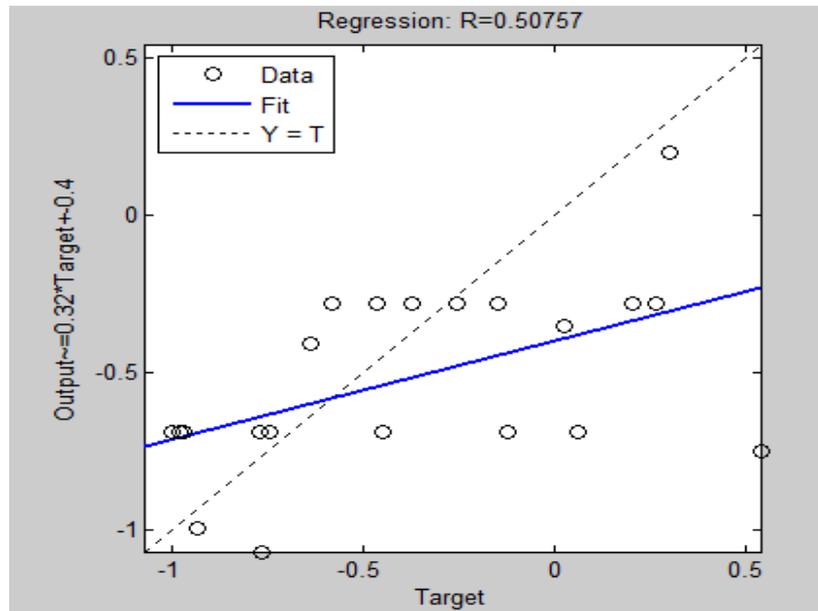
**c. TDNN avec une fenêtre temporelle de taille 4 reliée à chaque neurone de la couche cachée**



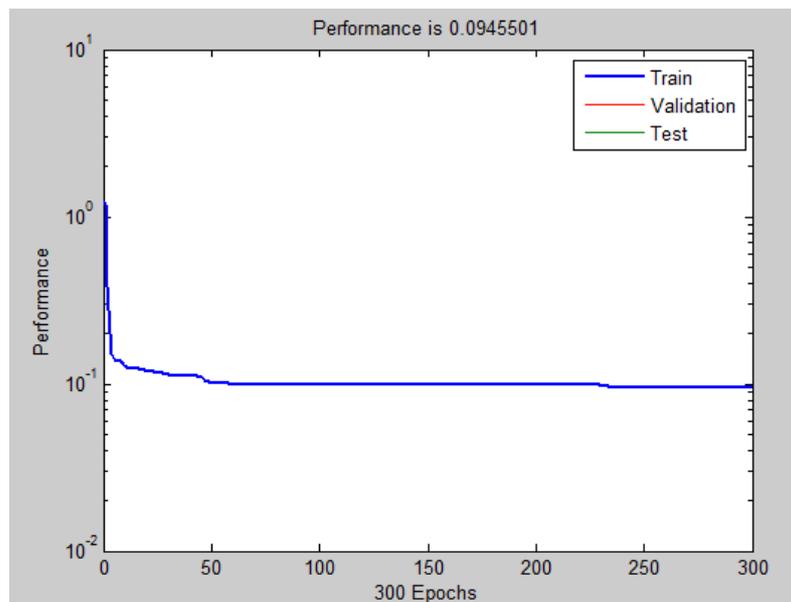
**Figure 60 :** TDNN avec une fenêtre temporelle de taille 4 reliée à chaque neurone de la couche cachée.



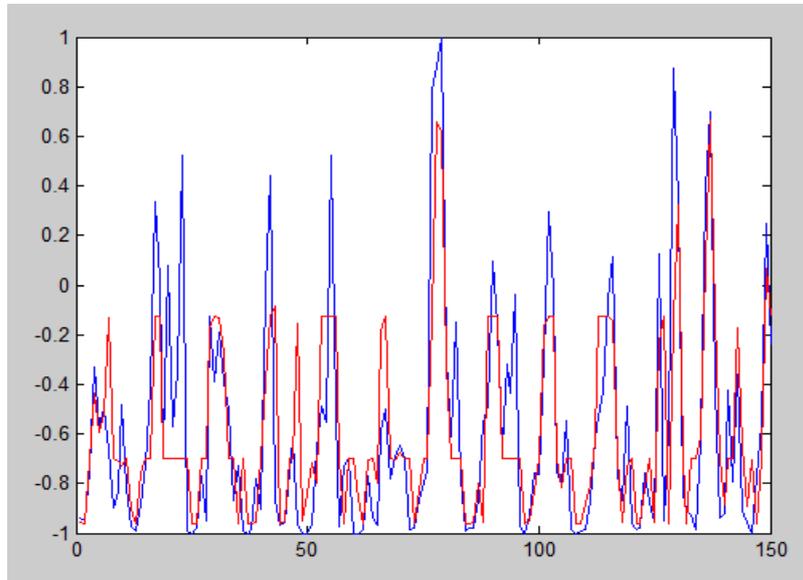
**Figure 61 :** Le coefficient de corrélation linéaire R entre les sorties désirées et les sorties prédites par le TDNN (l'étape d'apprentissage).



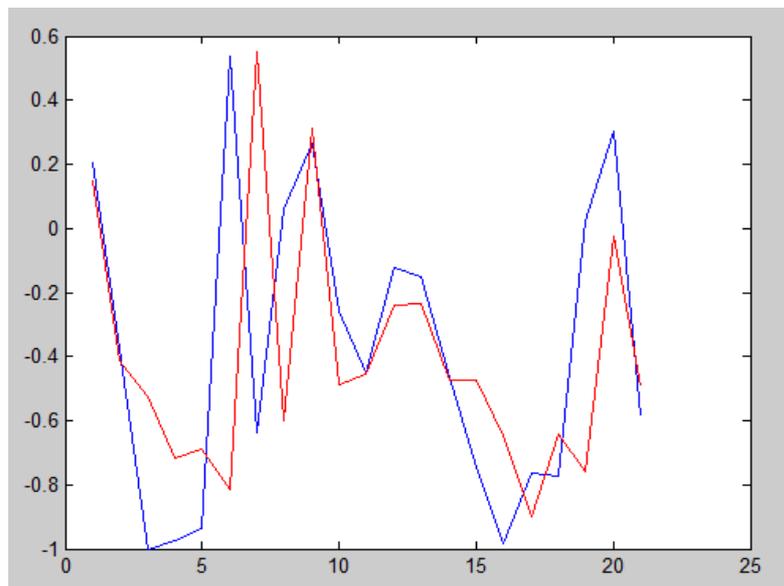
**Figure 62** : Le coefficient de corrélation linéaire R entre les sorties désirées et les sorties prédites par le TDNN (l'étape test).



**Figure 63** : L'erreur quadratique moyenne (EQM) en fonction de nombre d'itération.



**Figure 64** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape d'apprentissage).



**Figure 65** : Les résultats de la prévision, en bleu les sorties désirées et en rouge les sorties prédites par le TDNN (l'étape test).

**Interprétation des résultats :**

	apprentissage		Test	
	EQM	R	EQM	R
TDNN avec une fenêtre temporelle de taille 2	0.1449	0.5568	0.2365	0.3730
TDNN avec une fenêtre temporelle de taille 3	0.1339	0.6459	0.2089	0.4601
TDNN avec une fenêtre temporelle de taille 4	0.0945	0.7152	0.1837	0.5075

**Tableau 2** : Les résultats obtenus après l'étape d'apprentissage et l'étape test.

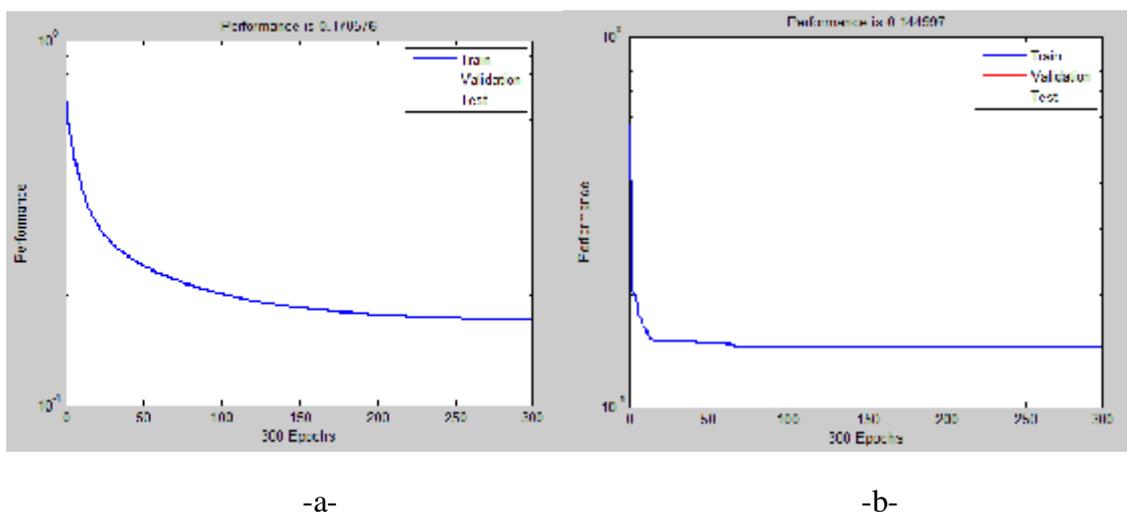
Avec l'algorithme de Lavenberg-Marquardt, On voit que l'erreur quadratique moyenne (EQM) diminue en augmentant la largeur de la fenêtre temporelle appliquée à l'entrée du TDNN et que le coefficient de corrélation linéaire R augmente et se rapproche de 1. Ceci veut dire que la précision de la prévision augmente lorsqu'on augmente la largeur de la fenêtre temporelle. Cependant, quand on dépasse la largeur de taille quatre, on trouve que le coefficient de corrélation diminue notablement tandis que l'erreur d'estimation augmente. Ce résultat qui fixe à quatre (04) la taille optimale de la fenêtre temporelle appliquée à l'entrée du TDNN s'explique par le fait que les données de précipitations étudiées forment une séquence d'un phénomène aléatoire dont la durée de vie est courte.

On constate donc, que par dimensionnement de la taille de la fenêtre et du délai pour chaque couche, on influence la capacité de mémorisation du TDNN. Plus on augmente la taille de la fenêtre, plus le nombre de paramètres libres augmente et plus on augmente la capacité de mémorisation. Cependant une capacité de mémorisation trop importante peut nuire au pouvoir de généralisation du réseau.

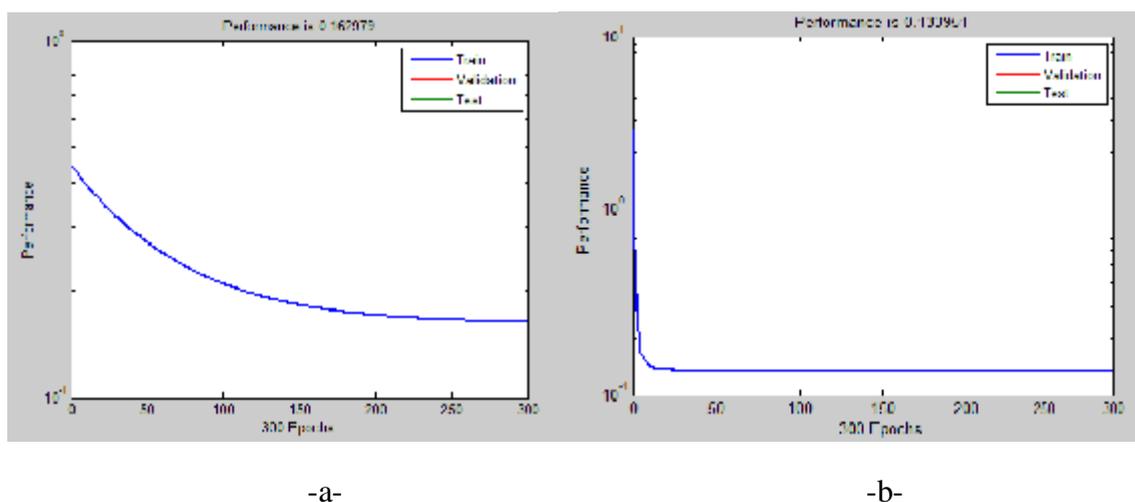
Enfin, on peut conclure que le TDNN donne des résultats raisonnables quand on l'applique pour une tâche de prévision de séries temporelles de précipitations à un pas en avant. Cependant, le réseau a des difficultés de prévoir les valeurs des pics (intensité élevée). Pour remédier à ce problème, il faut présenter au réseau des données plus nombreuses pour l'apprentissage.

#### IV.4. Evaluation des performances des algorithmes proposés

Pour comparer l'algorithme Lavenberg-Marquardt avec celui de gradient récursif, on a entraîné le TDNN avec le même ensemble de données appliqué à l'entrée et avec la même configuration respectivement avec les deux algorithmes d'apprentissage, pour obtenir les résultats suivants :



**Figure 66 :** L'erreur quadratique moyenne (EQM) correspondant à une fenêtre temporelle de taille 2 appliquée à l'entrée du TDNN : -a- avec l'algorithme de gradient récursif, -b- avec l'algorithme de Lavenberg-Marquardt.



**Figure 67:** L'erreur quadratique moyenne (EQM) correspondant à une fenêtre temporelle de taille 3 appliquée à l'entrée du TDNN : -a- avec l'algorithme de gradient récursif, -b- avec l'algorithme de Lavenberg-Marquardt.

Les graphes de l'erreur quadratique moyenne (EQM) (figures 66 et 67) nous montrent clairement que l'algorithme de Lavenberg-Marquardt converge plus rapidement et atteint le minimum après quelques itérations, tandis que, l'algorithme de gradient récursif prend plus du temps pour atteindre le minimum. On voit aussi qu'une erreur quadratique moyenne (EQM) de 0.17 est obtenu avec l'algorithme de gradient récursif, et une EQM de 0.14 avec celui de Lavenberg-Marquardt, pour une fenêtre temporelle de taille 2 appliquée à l'entrée du TDNN. Tandis que pour la fenêtre temporelle de taille 3, on a obtenu une EQM de 0.16 avec l'algorithme de gradient récursif contre 0.13 pour l'algorithme de Lavenberg-Marquardt. Ce qui veut dire que ce dernier est plus précis que l'algorithme de rétropropagation de gradient. On constate donc qu'avec l'algorithme de premier ordre on a une approximation grossière des données par rapport à celle obtenu en utilisant l'algorithme de deuxième ordre.

#### IV.5. Discussion

Ce chapitre illustre la mise en œuvre du TDNN pour la prévision des séries temporelles de précipitations, en utilisant deux méthodes d'apprentissage différentes qui nécessitent un certain savoir faire pour une utilisation efficace. En effet, la convergence des algorithmes n'est pas prouvée et de multiples variables sont à ajuster précisément en fonction du problème traité. Parmi ces variables à fixer, citons par exemple : la taille de la fenêtre temporelle, la contrainte de poids partagé, le gain d'apprentissage, la sélection des exemples pour l'apprentissage et le test, le nombre de neurones dans la couche cachée, la configuration initiale des poids, le nombre d'itérations d'apprentissages.

Les résultats obtenus montrent que la précision de la prévision augmente en augmentant la taille de la fenêtre temporelle appliquée à l'entrée du TDNN. Cependant, quand on dépasse une certaine largeur on trouve que le coefficient de corrélation diminue notablement tandis que l'erreur d'estimation augmente. Ce résultat qui fixe la taille optimale de la fenêtre temporelle à l'entrée du TDNN s'explique par le fait que les données de précipitations étudiées forment une séquence d'un phénomène aléatoire dont la durée de vie est courte.

On a aussi constaté que l'algorithme de deuxième ordre (l'algorithme Lavenberg-Marquardt) est plus performant que l'algorithme de premier ordre ou bien celui de rétropropagation de gradient (gradient récursif) pour la prévision des séries temporelles de précipitations. Ce qui veut dire qu'avec l'algorithme de premier ordre on a une approximation grossière des données par rapport à celle obtenu en utilisant l'algorithme de deuxième ordre.

Enfin, on peut conclure que le TDNN donne des résultats raisonnables quand on l'applique pour une tâche de prévision de séries temporelles de précipitations à un pas en avant. Cependant, le réseau a des difficultés de prévoir les valeurs des pics. Pour remédier à ce problème, il faut présenter au réseau des données plus nombreuses pour l'apprentissage.

# Conclusion générale

## Conclusion générale

L'étude que nous avons menée dans ce mémoire nous a permis de passer en revue les principales méthodes utilisées pour la prévision des précipitations. Etant donné que nous nous sommes intéressés à la méthode de prévision par réseaux de neurones et particulièrement à la prévision par Time Delay Neural Network (TDNN). Les réseaux de neurones sont une nouvelle technique de traitement de l'information. Concrètement ils se traduisent par des algorithmes mettant en jeu des concepts associés à la nature du cerveau pour la notion d'apprentissage. Notons que le TDNN est un réseau feedforward dynamique où l'information se propage de l'entrée vers la sortie sans retour en arrière et la dynamique se situe dans la couche d'entrée sous forme de retard. Le TDNN se singularise d'un réseau de neurones classique, tel que le réseau multicouche (MLP) par le fait qu'il prend en compte une certaine notion du temps. C'est à dire qu'au lieu de prendre en compte tous les neurones de la couche d'entrée en même temps, il prend une fenêtre du spectre puis effectue un balayage temporel.

En résumé, les arguments que nous avons développés au troisième et quatrième chapitre nous ont permis d'opter pour cette approche car une série temporelle de précipitations contient des caractéristiques locales. Ces caractéristiques ne sont pas fixes dans le temps, ce qui rend la prévision de leurs occurrences très difficiles. Le Time Delay Neural Network a été choisi pour identifier ces caractéristiques locales des précipitations indépendamment de leur position dans le temps.

Afin de démontrer l'intérêt de notre choix, on a utilisé des données mensuelles de précipitations relevées à la station météorologique de Tizi-Ouzou, sur une période de 15 ans, allant du mois de janvier 1996 au mois de décembre 2010.

Cette base de données a été divisée en deux sous ensembles, le premier sert à effectuer l'apprentissage, et le deuxième sous ensemble utilisé pour tester les performances du TDNN, c'est ce qu'on appelle l'ensemble test.

On a entraîné le réseau avec un algorithme de premier ordre qui est l'algorithme de gradient récursif, ensuite avec un algorithme de deuxième ordre : celui de Lavenberg-Marquardt. L'erreur quadratique moyenne (EQM) et le coefficient de corrélation linéaire R sont utilisés pour comparer les résultats des différentes largeurs de la fenêtre temporelle appliquées à l'entrée du réseau, ensuite évaluer les performances du TDNN, ainsi que la précision de la prévision. Les résultats obtenus montrent que l'erreur quadratique moyenne diminue en augmentant la largeur de la fenêtre temporelle, et que le coefficient de corrélation augmente et se rapproche de 1. Ceci veut dire que la précision de la prévision augmente en augmentant la largeur de la fenêtre temporelle. Cependant, quand on dépasse une fenêtre temporelle de taille trois pour l'algorithme de premier ordre, et de taille quatre pour l'algorithme de deuxième ordre nous trouvons que le coefficient de corrélation diminue notablement tandis que l'erreur d'estimation augmente. Ce résultat qui fixe la taille optimale de la fenêtre temporelle à l'entrée du TDNN s'explique par le fait que les données de précipitations étudiées forment une séquence d'un phénomène aléatoire dont la durée de vie est courte.

La comparaison des deux algorithmes d'apprentissages nous montre clairement que l'algorithme de Lavenberg-Marquardt converge plus rapidement. Il est aussi plus précis que l'algorithme de gradient récursif. On constate donc qu'avec l'algorithme de premier ordre nous avons une approximation grossière des données par rapport à celle obtenu en utilisant l'algorithme de deuxième ordre.

Enfin, nous pouvons conclure que le TDNN donne des résultats raisonnables quand on l'applique pour une tâche de prévision de séries temporelles de précipitations à un pas en avant. Cependant, le réseau a des difficultés de prévoir les valeurs des pics des précipitations. Pour remédier à ce problème, il faut présenter au réseau des données plus nombreuses pour l'apprentissage.

D'autres applications peuvent être utilisées afin de prouver la robustesse de notre technique telle que par exemple l'utilisation d'une série d'images satellitaires à l'entrée de TDNN.

# Annexes

# Annexe A

## 1. Classification des nuages selon l'altitude

Les nuages se situent dans la troposphère, c'est à dire entre 0km et 10km d'altitude. A 10km, les nuages sont plutôt formés de glace ce qui peut leur donner un aspect fibreux et blanchâtre. La table suivante montre la classification des nuages selon l'altitude :

Groupe	Altitude moyenne de la base	Préfixe du nom du nuage
Étage supérieur	6000 mètres	Cirr...
Étage moyen	2000 mètres	Alto...
Étage inférieur	Du sol à 2000 mètres	Stra...
À développement vertical	500 mètres	Cum...

**Table 1** : Classification des nuages selon l'altitude.

## 2. Classification des nuages selon le genre

Nom	Description
Cirrus (Ci)	Nuages détachés sous forme de délicats filaments blancs composés de bancs ou d'étroites bandes blanches ou en majeure partie blanche. Ces nuages ont un aspect fibreux (chevelu), un éclat soyeux ou les deux.
Cirrostratus (Cs)	Voile nuageux transparent et blanchâtre, d'aspect fibreux (chevelu) ou lisse, couvrant le ciel en totalité ou en partie et donnant le plus généralement lieu à des phénomènes de halo.
Cirrocumulus (Cc)	Banc, nappe ou couche mince de nuages blancs sans ombre propre composés de très petits éléments en forme de granules, de rides, etc., soudés ou non et disposés plus ou moins régulièrement; la plupart des éléments ont une largeur apparente de moins d'un doigt tenu à longueur d'bras.
Altostratus (As)	Nappe ou couche nuageuse grisâtre ou bleuâtre, d'aspect strié, fibreux ou en uniforme couvrant entièrement ou partiellement le ciel et présentant des parties suffisamment minces pour laisser voir le soleil, au moins vaguement, comme au travers d'un verre dépoli. Il ne présente pas de phénomène de halo.

Alto cumulus (Ac)	Banc, nappe ou couche de nuages blancs et gris ayant généralement des ombres propres et composés de lamelles, de galets, de rouleaux, etc., d'aspect parfois partiellement fibreux ou flou, soudés ou non. La plupart des petits éléments ont une largeur apparente comprise entre un et trois doigts tenus à longueur de bras.
Stratus (St)	Couche nuageuse, généralement grise, à base assez uniforme pouvant donner lieu à de la bruine, des cristaux de glace ou de la neige en grains. Lorsque le soleil est visible à travers ces nuages, on distingue facilement son contour. Il se présente parfois en bandes déchiquetées.
Stratocumulus (Sc)	Banc, nappe ou couche de nuages gris ou blanchâtre, ou les deux à la fois, ayant presque toujours des parties foncées, formées de dalles, de galets, de rouleaux, etc., d'aspect non fibreux, soudés ou non; la plupart des petits éléments de forme régulière ont une largeur apparente de plus de trois doigts tenus à longueur de bras.
Cumulus (Cu)	Nuages détachés, normalement denses et aux contours bien délimités, se développant verticalement sous forme de mamelons, de dômes et de tours, dont la partie supérieure bourgeonnante a souvent l'aspect d'un chou-fleur. Les parties du nuage éclairées par le soleil sont d'un blanc éclatant; la base est relativement foncée et horizontale.
Cumulonimbus (Cb)	Nuage dense à extension verticale considérable en forme de montagne ou de tour immense. Sa partie supérieure est presque toujours aplatie; celle-ci prend la forme d'une enclume ou d'un panache. Sous sa base, souvent très foncée, on retrouve fréquemment des nuages bas déchiquetés, soudés ou non avec elle ainsi que des précipitations.
Nimbostratus (Ns)	Couche nuageuse grise, souvent foncée dont l'aspect est rendu flou par des chutes de pluie plus ou moins continues, qui, dans la plupart des cas atteignent le sol. Il masque complètement le soleil sur toute son étendue. Sous sa base on retrouve fréquemment des nuages bas, déchiquetés, soudés ou non avec elle ainsi que des précipitations.

**Table 2** : Classification des nuages selon le genre.

### 3. Classification des nuages selon l'espace

Espèce	Description	Genre associé
Fibratus	Nuages détachés ou en un voile fin consistant en filaments presque rectiligne ou plus ou moins irrégulièrement recourbé ne se terminant pas en crochet ou en touffes.	Cirrus, cirrostratus
Uncinus	Cirrus habituellement en forme de virgule se terminant, dans sa partie supérieure, en un crochet ou une touffe et dont l'extrémité supérieure n'est pas en forme de protubérance arrondi.	Cirrus
Spissatus	Cirrus suffisamment épais pour être gris lorsque vu près du soleil.	Cirrus
Castellanus	Nuages qui présentent dans leur partie supérieure des protubérances en forme de tours et qui leur donnent une apparence crénelées. Les tours sont issues de la même base et arrangées en ligne.	Cirrus, cirrocumulus, altocumulus et stratocumulus.
Flocus	Chaque nuage est une touffe d'apparence cumuliforme dont la base est plus ou moins déchiquetée et qui s'accompagne souvent de virga.	Cirrus, cirrocumulus, altocumulus
Stratiformis	Les nuages s'étendent en une vaste nappe ou couche horizontale.	Stratocumulus, altocumulus et parfois cirrocumulus
Nebulosus	Nuages en forme de voile ou couche, sans détail distinct.	S'applique aux stratus et cirrostratus
Lenticularis	Nuages en forme de lentilles ou d'amandes, souvent allongés et avec un pourtour bien défini montré par une irisation.	Stratocumulus, altocumulus, cirrocumulus
Fractus	Nuages en forme de lambeaux irréguliers ayant une apparence nettement déchiquetée.	Stratus et cumulus
Humilis	Cumulus de faible extension verticale.	Cumulus
Mediocris	Cumulus d'extension verticale moyenne et dont le sommet	Cumulus

	montre de petites protubérances.	
Congestus	Cumulus bourgeonnant de grande extension verticale et dont la partie supérieure ressemble à un chou-fleur	Cumulus
Calvus	Cumulonimbus dont les protubérances du sommet commencent à perdre leur apparence cumuliforme mais dans lequel on ne peut distinguer de parties cirriformes.	Cumulonimbus
Capillatus	Cumulonimbus caractérisé par la présence à son sommet de parties cirriformes fibreuses ou striées, en forme d'enclume, de panache ou d'une vaste masse de cheveux plus ou moins désordonnée. Il est souvent accompagné d'une averse ou d'un orage.	Cumulonimbus

**Table 3** : Classification des nuages selon l'espèce

#### 4. Classification des nuages selon la variété

Variété	Description
Intortus	Cirrus dont les filaments sont irrégulièrement recourbés et, souvent, de façon tortueuse.
Vertebratus	Principalement des cirrus arrangés en forme de vertèbres, de côtes ou ressemblant au squelette d'un poisson.
Undulatus	Nuages en forme de vagues et dont les éléments sont parfois séparés, parfois juxtaposés.
Radiatus	Nuages en bandes parallèles qui semblent converger vers un point à l'horizon
Lacunosus	Couche ou amas dans lequel on retrouve des orifices circulaires également distribués; ou encore, des amas ressemblant à un rayon de miel.
Duplicatus	Couche ou amas de dimension horizontale assez étendue et qui laisse voir le soleil ou la lune.
Perlucidus	Couche ou amas d'assez grande dimension avec des trouées distinctes, mais parfois assez petites, et qui laissent voir le soleil, la lune, le ciel ou d'autres nuages.
Opacus	Couche ou amas d'assez grande dimension suffisamment épais pour masquer le soleil ou la lune.

**Table 4** : Classification des nuages selon la variété.

# Annexe B

## La règle de rétropropagation de gradient total

### 1. Règle de Delta

Cette règle a été appliquée à l'adaline (adaptive linear element) qui est un neurone qui possède des valeurs d'activation continues et une fonction d'activation linéaire :

$$v_i = \sum_j w_{ij} x_j + b_i \quad (\text{B.1})$$

$v_i$  : Activation du neurone  $i$ .

$b_i$  : Biais du neurone  $i$ .

$x_j$  : entrée extérieur  $j$ .

Pour n'importe quel échantillon d'entrée, la sortie diffère de la valeur désirée par  $(d_i - v_i)$ , la règle delta utilise une fonction de coût basée sur ces différences pour ajuster les poids :

$$E = \sum_i E_i = \frac{1}{2} \sum_i (d_i - v_i)^2 \quad (\text{B.2})$$

La règle delta trouve toutes les valeurs des poids qui minimisent la fonction d'erreurs par une méthode dite descente du gradient. L'idée est de faire changer les poids proportionnellement au négatif de la dérivée de l'erreur mesurée sur chaque modèle en fonction de chaque poids :

$$\Delta_i w_j = -\eta \frac{\partial E_i}{\partial w_j} \quad (\text{B.3})$$

Où  $\eta$  est le pas d'apprentissage.

La dérivée est :

$$\frac{\partial E_i}{\partial w_j} = \frac{\partial E_i}{\partial v_i} \frac{\partial v_i}{\partial w_j} \quad (\text{B.4})$$

D'après l'équation (B.1) :

$$\frac{\partial v_i}{\partial w_j} = x_j \quad (\text{B.5})$$

Et

$$\frac{\partial E_i}{\partial v_i} = -(d_i - v_i) \quad (\text{B.6})$$

Tel que

$$\Delta_i w_j = \eta \delta_i x_j \quad (\text{B.7})$$

Où  $\delta_i = d_i - v_i$  est la différence entre la sortie désirée et la sortie actuelle pour le neurone  $i$ .

## 2. La rétropropagation

La règle d'apprentissage de Widrow-Hoff (règle de Delta) ne peut s'appliquer que sur des réseaux à une seule couche de poids adaptatifs. C'est donc en étendant la règle de Widrow-Hoff que plusieurs équipes de chercheurs [Le Cun, 1985] et [Werbos, 1974] ont développée un algorithme d'apprentissage appelé rétropropagation du gradient de l'erreur. Elle a été ensuite généralisée par l'équipe de Rumelhart en 1986. Cet algorithme fournit une façon de modifier les poids des connexions de toutes les couches d'un Perceptron Multi Couches (PMC).

Soit le réseau à deux couches décrit par la figure suivante dans lequel les unités de sortie sont notées  $y_k$ , les unités cachées  $v_i$  et les unités d'entrée  $x_j$ , et  $d_k$  sont les sorties désirées.

Les connexions des unités d'entrée aux unités cachées sont notées  $w_{ij}$  et celles des unités cachées aux unités de sortie par  $w_{ki}$

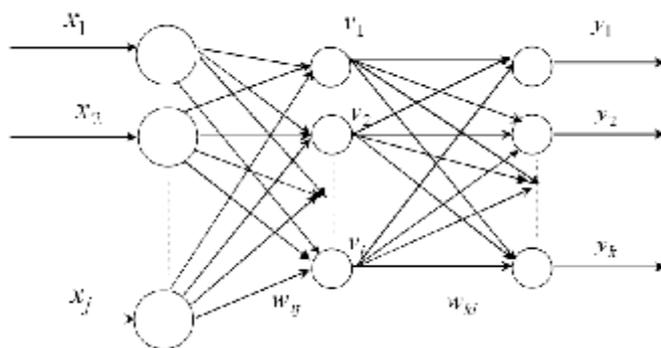


Fig1 : Perceptron Multi Couches.

L'entrée  $j$  a pour valeur  $x_j^p$  lorsque le modèle de données  $p$  est présenté au réseau. Ces valeurs peuvent être binaires (0/1 ou +1/-1) ou continues. La valeur de sortie de l'unité cachée  $i$  est donnée par :

$$v_i^p = f_i(a_i^p) = f_i\left(\sum_j w_{ij}x_j^p\right) \quad (\text{B.8})$$

Les unités de sortie ont comme valeur :

$$y_k^p = f_k(a_k^p) = f_k\left(\sum_i w_{ki}v_i^p\right) \quad (\text{B.9})$$

Les fonctions d'erreurs partielles et globales sont alors définies par :

$$E^p = \frac{1}{2} \sum_i (y_k^p - d_k^p)^2 \quad \text{et} \quad E = \sum_p E^p \quad (\text{B.10})$$

La minimisation de la fonction d'erreur globale va se faire par une descente de gradient, par conséquent, après la présentation de tous les vecteurs d'entrée de la base d'apprentissage, nous modifierons la valeur de chaque connexion par la règle suivante:

$$\Delta w = -\eta \frac{\partial E}{\partial w} \quad (\text{B.11})$$

Cette règle d'apprentissage est généralement appelée la règle de delta généralisée.

Dans l'expression (B.10), seule la sortie  $y_k$  dépend du paramètre  $w$ . Selon la position des poids des connexions, deux cas se présentent :

**a. Cas des connexions entre la couche cachée et celle de sortie  $w_{ki}$**

Pour le cas des neurones de sortie, l'expression (B.11) devient fonction du paramètre  $w_{ki}$  qui influe uniquement sur la sortie du neurone d'indice  $k$ . Nous pouvons donc décomposer la dérivée de l'expression (B.11) par :

$$\frac{\partial E^p}{\partial w_{ki}} = \frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial a_k^p} \frac{\partial a_k^p}{\partial w_{ki}} = (y_k^p - d_k^p) f'_k(a_k^p) v_i^p \quad (\text{B.12})$$

L'expression (B.12) devient alors :

$$\Delta w_{ki} = \eta \sum_p f'_k(a_k^p) (d_k^p - y_k^p) v_i^p \quad (\text{B.13})$$

### b. Cas des connexions entre la couche d'entrée et la couche cachée $w_{ij}$

Pour le cas des neurones cachés, l'expression (B.11) est fonction du paramètre  $w_{ij}$  qui influe non seulement sur la sortie du neurone  $i$  de la deuxième couche, mais aussi sur tous les neurones  $k$  de la couche de sortie (en aval) qui lui sont connectés. On obtient alors l'équation suivante :

$$\frac{\partial E^p}{\partial w_{ij}} = \frac{\partial E^p}{\partial v_i^p} \frac{\partial v_i^p}{\partial a_i^p} \frac{\partial a_i^p}{\partial w_{ij}} = \frac{\partial E^p}{\partial v_i^p} f_i'(a_i^p) x_j^p \quad (\text{B.14})$$

Le premier terme de cette expression devient alors :

$$\frac{\partial E^p}{\partial v_i^p} = \sum_k \frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial v_i^p} = \sum_k \frac{\partial E^p}{\partial y_k^p} \frac{\partial y_k^p}{\partial a_k^p} \frac{\partial a_k^p}{\partial v_i^p} = \sum_k (y_k^p - d_k^p) f_k'(a_k^p) w_{ki} \quad (\text{B.15})$$

On obtient alors la modification des poids :

$$\Delta w_{ij} = \eta \sum_p (f_k'(a_i^p) x_j^p (\sum_k (d_k^p - y_{ki}^p) f_k'(a_k^p) w_{ki})) \quad (\text{B.16})$$

Après avoir calculé la variation des poids des connexions pour tous les neurones de sortie (B.13), on calcule alors la variation des poids des connexions de la couche cachée (B.16), on met ainsi à jour les poids des connexions de la couche de sortie jusqu'à la couche d'entrée, on rétropropage ainsi le signal d'erreur. C'est de là que vient le nom de cet algorithme :

« rétropropagation du gradient de l'erreur ».

Du fait de sommer les  $\Delta w_{ij}$  pour tous les vecteurs  $p$  de la base d'apprentissage puis de remettre à jour les poids avec la variation totale ainsi calculée, l'algorithme est appelé gradient total.

## BIBLIOGRAPHIE

Allan H. Murphy: "The Finley Affair: A Signal Event in the History of Forecast Verification". Prediction and Evaluation Systems, Corvallis, Oregon, 1996.

Antoine Mahul : « Apprentissage de la qualité de service dans les réseaux multiservices : applications au routage optimal sous contraintes ».Thèse de doctorat, Université Blaise Pascal - Clermont-Ferrand II, 2005.

Arnaud. Yand Laloë.F., 1991: « Détermination de la phase dynamique des nuages convectifs (croissance, décroissance) à partir de leur indice volumique et de leur surface sur une image instantanée ».

Battiti. R : "First and second-order methods for learning: between steepest descent and newton's method". Neural Computation, 4(2):141–166, 1992.

Bengio.Y : Cours IFT6265 : « Poids Partagés, TDNNs et réseaux à convolution ». Université de Montréal - Département d'informatique et de recherche opérationnelle, 1997.

Box. G, Jenkins. G. M, Reinsel. G: "Time Series Analysis: Forecasting & Control". Prentice Hall, 1994.

Bontron. G : « Prévision quantitative des précipitations ». Thèse de doctorat, LTHE, 2004.

Claude Touzet : « Contribution à l'étude et au développement de modèles connexionnistes séquentiels de l'apprentissage ». Thèse de doctorat, 1990.

Coiffier. J : « Un demi-siècle de prévision numérique du temps ». La Météorologie 8<sup>em</sup> série - n° 30, juin 2000.

Dreyfus. G : « Les réseaux de neurones », ESPCI paris, 1998.

Dreyfus. G, Martinez. J.M, Samuelides . M, Gordon. M. B, Badran. F,Thiria. S, Hérault. L : « Réseaux de neurones : Méthodologie et applications ». Eyrolles, Paris, 2002.

French. M, Krajewski. W, and Cuykendal. R.R : “Rainfall forecasting in space and time using a neural network”. J. Hydrol., 137: 1-31, 1992.

Gong Zhenbin : “Application of Artificial Neural Networks for Rainfall Forecasting in Fuzhou City “. MEE09203.

Guillot. B, Lahuec. J. P et Dagorne.D : « Satellite et surveillance de climat ». Atlas de veille climatique : 1986-1994, ORSTOM-Météo-France, Paris, 91 pages, 1994.

Guillaume Bontron: “Prévision quantitative des précipitations: Adaptation probabiliste par recherché d’analogues”. Thèse de doctorat, Institut national polytechnique de Grenoble, 2004.

Hassoum, M.H: “Fundamentals of artificial neural networks”. MIT Press, Cambridge, 1995.

Haykin. S: « Neural Networks: A comprehensive Foundation ». Macmillan, 1994.

Hervé. N : « Introduction à la télédétection ». Agrocampus, Rennes, 2007.

Hertz, Krogh & Palmer: “Introduction to the theory of neural computation “. Addison Wesley, 1991.

Hung. N. Q, Babel. M. S, Weesakul. S and Tripathi. N. K: “An artificial neural network model for rainfall forecasting in Bangkok, Thailand ». Hydrol. Earth Syst. Sci., 13, 1413–1425, 2009.

J.L. Bertrand-Krajewski : « Cours d’Hydrologie Urbaine : la pluie ». URGC, INSA de Lyon, 2007.

Kin. C. Luk, James E. Ball and Ashish Sharma:”An application of artificial neural network for rainfall forecasting”. Water Research Laboratory, School of Civil Engineering, University of New South Wales.

LeCun. Y, Bottou. L, Bengio. Y and Haffner. P, “Gradient-Based Learning Applied to Document Recognition”. Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

Luk. K.C:“An application of artificial neural networks for rainfall forecasting“. Mathematical and Computer Modeling Journal, 33, 883-699, 2001.

Maier. R. H and Dandy. G. C: “Comparison of various methods for training feed-forward neural network for salinity forecasting”. *Water Resources Research* , 35(8), 2591–2596, 1999.

Marc Parizeau : « Réseaux de neurones ». Université Laval, 2004.

Minsky. M. L and Paper. S. A: “Perceptrons”. Expanded Edition, MIT Press, 1988.

Personnaz. L, Rivals. I : « Réseaux de neurones formels pour la modélisation, la commande et la classification ». Edition CNRS, Paris, 2003.

Poisson. E, Viard-Gaudin. C : « Réseaux de neurones à convolution : Reconnaissance de l’écriture manuscrite non contrainte ». *Valgo 2001* (ISSN 1625-9661), n° 01-02, Oct. 2001.

Ramirez. M.C.V, Ferreira. N.J and Velho. H. F: “Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region“. *Journal of Hydrology*, 301, 146–162, 2005.

Rumelhart. D, Hinton. G and William. R : “ Learning representations by back-propagating errors”. *Nature*, 323:533–536, 1986.

Sauvageot. H: « Radar météorologique ». Aetech house, Boston, 361, pp.

Schenkel, Guyon. I, Henderson. D : “On-Line Cursive Script Recognition using Time Delay Neural Networks and Hidden Markov Models”. *Machine Vision and Applications*, pp. 215–223, 1995.

Selami. N et al.: « Etude des structures nuageuses ». JIG’2007-3èmes Journées Internationales sur l’informatique Graphique 195-201, 2007.

Sevruk, B., and Klemm, S.: “Types of standard precipitation gauges”. *Precipitation Measurement, WMO/IASH/ETH Workshop on Precipitation Measurement*, B. Sevruk, ed., Swiss Federal Institute of Technology, ETH-Zentrum, CH-8092 Zurich, 227–232, 1989.

Thiria, Gascuel, Lechevallier & Canu : « Statistiques et méthodes neuronales ». Dunod, 1997.

Weibel. A, Hanazawa.T, Hinton.G, Shinkano. K: “Phoneme recognition using time-delay neural networks”. *IEEE Transactions on Acoustics, Speech and Signal Processing*, volume 37 pp. 328-339, 1989.

<http://www.iro.umontreal.ca/~bengioy/ift6265/convo/convo.html>.

[http://www.meteo.fr/temps/domtom/antilles/pack-public/cyclone/tout\\_cyclone/sommaire.html](http://www.meteo.fr/temps/domtom/antilles/pack-public/cyclone/tout_cyclone/sommaire.html)

<http://www.meteo.org/phenomen.html>

<http://www.lameteo.org/meteo-9.html>

<http://education.meteofrance.com>