

II

Le Vent : *Un Problème Réel*

6. Présentation du problème

- 6.1. Le problème Géophysique
- 6.2. Le modèle empirique
- 6.3. Vers une nouvelle approche

7. Approches du problème

- 7.1. Le modèle utilisé
- 7.2. Les approches classiques
- 7.3. L'approche connexionniste

8. Les données

- 8.1. Les Sources de Données
- 8.2. Pré-traitement des données

9. Approche connexionniste

- 9.1. Introduction
- 9.2. Approximation des fonctions de transfert
- 9.3. Approximation des fonctions à l'aide des MLP
- 9.4. Utilisation des MLP pour l'analyse des fonctions de transfert
- 9.5. Interprétation théorique des deux exemples
- 9.6. Conclusion

10. Enchaînement de la machine

- 10.1. La machine neuronale
- 10.2. Levée des ambiguïtés

11. Comparaison avec des méthodes classiques

- 11.1. Les méthodes
- 11.2. Résultats
- 11.3. Conclusion

12. Conclusion

6. Présentation du problème

6.1. Le problème Géophysique

L'importance du vent sur les océans

Les vents sont la conséquence de phénomènes météorologiques responsables de nombreux phénomènes océanographiques. En océanographie, leur importance provient du fait qu'ils forcent les mouvements marins. En météorologie, les vents caractérisent les mouvements de l'atmosphère et résultent de la thermodynamique atmosphérique. En outre, les vents conditionnent les interactions océan-atmosphère, en particulier les flux de chaleur latente et sensible, qui sont une composante essentielle de la détermination du climat régional aussi bien que global. Une meilleure connaissance des vents doit permettre d'améliorer les modèles de circulation Océanique par une prise en compte de forçages plus réalistes et aussi de tester la validité des modèles atmosphériques.

La télédétection océanographique

La télédétection est le nom donné à la mesure à distance de paramètres physiques. Les progrès technologiques ont permis l'apparition d'appareils sophistiqués de mesure à distance. Ces appareils, notamment les radars installés à bord de satellites, permettront l'observation systématique de certains paramètres océaniques; ils impliquent l'enregistrement de grandes quantités de données.

Le premier radar dédié à l'océanographie a été installé à bord du laboratoire Skylab de la NASA en 1972. En 1978 le satellite SEASAT a été lancé embarquant le premier scatteromètre (ou diffusiomètre) destiné à la mesure du vent à la surface de la mer. Malheureusement, ce satellite n'a fonctionné que 3 mois. Ces missions ont malgré tout montré qu'il était

possible de mesurer avec précision les caractéristiques du vent sur les océans avec des scatteromètres.

La décennie des années 90 se présente comme une époque particulièrement propice pour l'étude des océans. En juillet 1991 a été lancé le satellite Européen ERS-1. Ce satellite, placé en orbite polaire à 785 km d'altitude, effectue ses rotations autour de la Terre en 35 jours. Ses missions concernent l'étude des océans et de l'environnement, les prévisions météorologiques et climatologiques, les échanges océan-terre-atmosphère. Parmi ses instruments de mesure, ce satellite possède un scatteromètre, ou diffusiomètre, avec 3 antennes-radar qui doivent lui permettre de mesurer avec précision le vecteur vent à la surface des océans. En 1995, les Etats Unis doivent lancer un appareil analogue : le scatteromètre NSCAT à bord du satellite japonais ADEOS. L'analyse des données fournies par les scatteromètres à bord de ces satellites vont permettre pour la première fois aux océanographes d'avoir une description adéquate des vents qui leur permettra de forcer correctement les modèles océaniques.

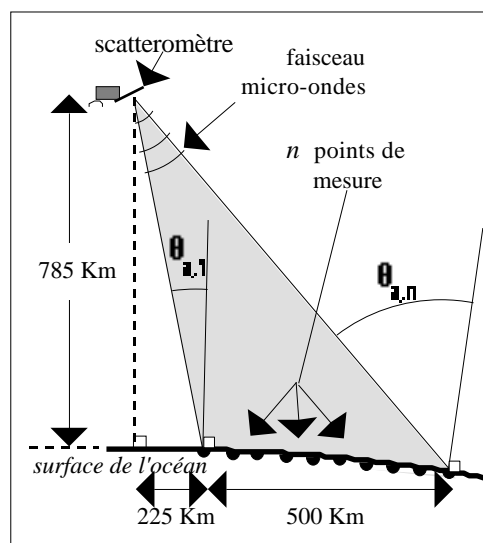


Figure 6.1 : Vue transversale du rayonnement émis par un scatteromètre. Le mouvement du satellite est perpendiculaire au dessin. θ est l'angle d'incidence de la radiation émis par le satellite. Le signal émis par l'antenne a du satellite est reçu avec un angle d'incidence de $\theta_{a,1}$ pour le point le plus proche de la trace du satellite et de $\theta_{a,n}$ pour le point le plus éloigné. Typiquement, pour une résolution spatiale de 50 Km, n est égal à 10.

Classiquement, les scatteromètres émettent vers l'océan un signal polarisé –à une fréquence de 5,3 GHz (Bande C) pour ceux de ERS-1 et près de 14 GHz

(Bande K_V) dans le cas de NSCAT-. Le signal est émis sous la forme d'un faisceau dont l'angle d'ouverture est de l'ordre de $0,5^\circ$ qui balaie un angle d'environ 25° dans la direction pointée par l'antenne émettrice, souvent perpendiculaire à la trajectoire du satellite (Figure 6.1). Il en résulte une fauchée de près de 500 km de large sur la surface des océans. Le signal rétro-diffusé à la surface des océans contient une information qui est fonction de la rugosité de la mer, qui est elle même fonction du vent.

Dans le cas de ERS-1 dont le balayage n'est effectué que d'un seul côté de la trace, le signal rétro-diffusé provient d'une surface parallèle à la trace du satellite, décalée sur le côté d'un angle d'environ 15° (soit sur une bande à la surface de l'océan d'environ 225 km de côté). Le "point" élémentaire de mesure a une résolution de 50 km, autrement-dit il couvre une surface de environ 50 Km de côté. Dans le cas de ERS-1, il y a 19 points de mesure distancés de 25 Km sur la largeur d'une fauchée.

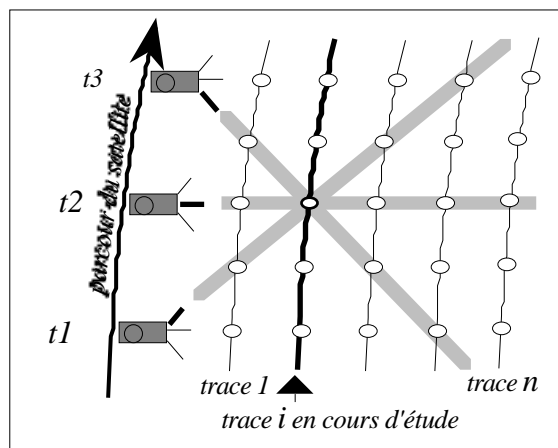


Figure 6.2 : Mesures prises sur un même point avec trois antennes scatteromètres. A chaque instant, chacune des antennes mesure et enregistre n points le long de l'orientation de son axe d'émission. C'est après le temps t_3 que le point signalé sur la trace i aura été complètement mesuré.

Avec plusieurs antennes pointant dans des directions différentes on peut, en profitant du déplacement du satellite réaliser plusieurs mesures en un même point de l'océan –avec des angles d'azimut différents–. Ces mesures seront prises bien entendu en des temps d'observation très légèrement décalés (voir Figure 6.2). Les mesures à partir d'une seule antenne donc avec un seul angle d'azimut, peuvent nous donner des informations concernant la vitesse du vent en un point mais non de sa direction. Avec deux antennes pointant dans des directions différentes, comme il a été confirmé par la

mission SEASAT, il est possible d'atteindre une précision de ± 2 m/s pour la vitesse du vent et d'obtenir la bonne direction parmi les 4 possibles, avec une précision de $\pm 20^\circ$. Le satellite ERS-1 qui compte trois antennes va permettre dans la grande majorité des cas de baisser à 2 le nombre de réponses ambiguës pour la direction du vent [Freilich 85].

6.2. Le modèle empirique

Le signal étudié (σ^0) est le rapport entre l'énergie rétro-diffusée et l'énergie émise. On montre expérimentalement que le σ^0 dépend de l'intensité du vent, de son azimuth (sa direction) par rapport à l'antenne, et de l'angle d'incidence (voir Figure 6.1 et Figure 6.3).

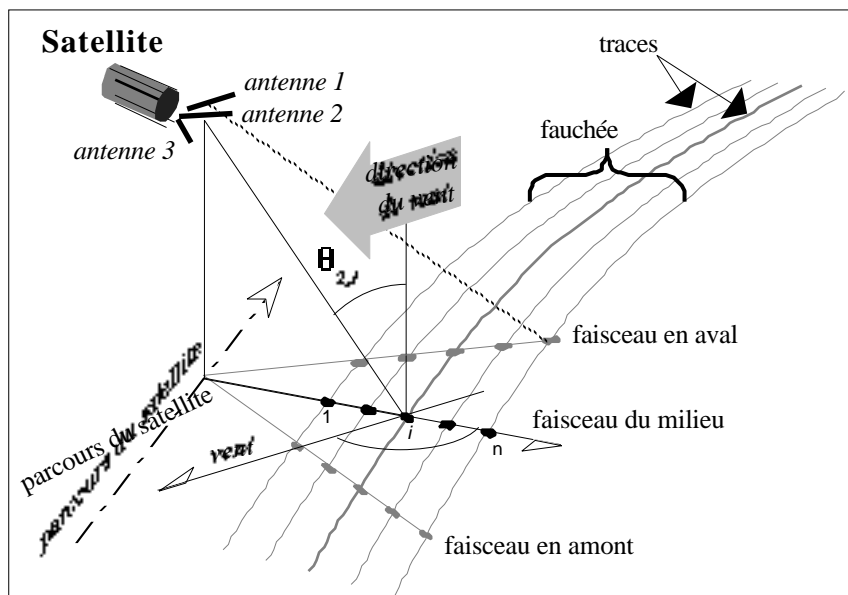


Figure 6.3 : Définition des différentes grandeurs géophysiques : l'angle d'incidence du rayonnement émis par le scatteromètre du satellite; l'angle d'azimut mesuré entre la direction du vent et celle de l'antenne. Pour l'antenne 1, $\theta_{1,1} = 24.7^\circ$ tandis que $\theta_{1,19} = 56.8^\circ$.

Des modèles géophysiques de type empirique ont été mis au point par les géophysiciens : ils proviennent de recherches expérimentales qui ont mis en évidence une forte dépendance entre la mesure du σ^0 en un point de l'océan et le vent en ce point. Ces modèles sont issus d'observations directes sur les grandeurs étudiées. Pour tous les modèles une relation empirique peut être déterminée et résumée par une équation fonctionnelle de la forme :

$$\sigma^0 = G(\theta, \phi, v)$$

où σ^0 représente le rapport entre le signal radar reçu et celui émis. σ^0 est exprimé classiquement en dB.

G est fonction de :

- L'angle d'incidence θ , mesuré dans le plan vertical entre la radiation incidente et la normale à la surface de la mer,
- L'angle d'azimut ϕ , mesuré dans le plan horizontal entre l'orientation de l'antenne émettrice et la direction du vent (Figure 6.3),
- La polarisation (verticale ou horizontale) de la radiation émise et reçue, et
- Le module de la vitesse du vent v .

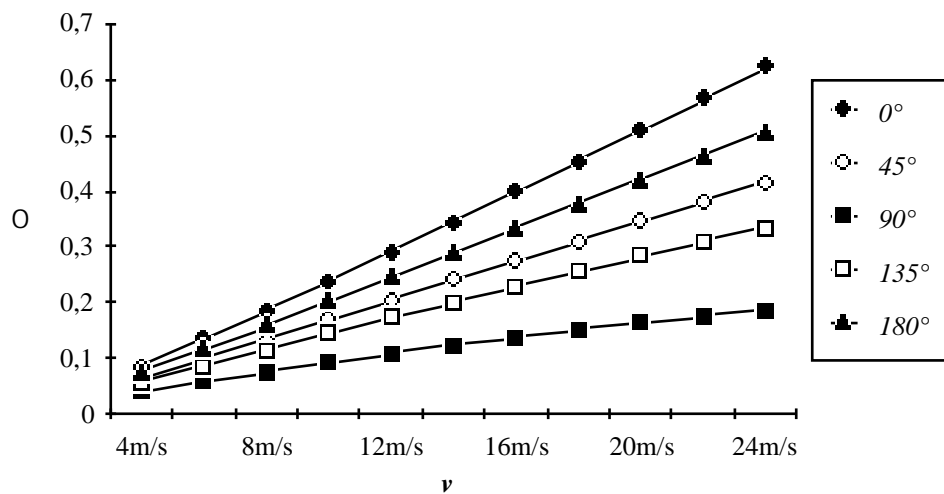


Figure 6.4 : Dépendance du signal rétro-diffusé par rapport à la vitesse du vent. Plusieurs courbes sont montrées pour différentes valeurs de direction du vent. Dans la figure nous avons pris un angle d'incidence constant, correspondant à la trace du milieu ($\theta = 31,8^\circ$).

A partir des mesures prises d'avion lors de vols circulaires sur une région donnée de l'océan, il a été observé que la puissance du signal rétrodiffusé augmentait de manière exponentielle (avec un exposant compris entre 0.7 et 1.5) en fonction de la vitesse du vent et changeait selon l'angle d'azimut. En particulier, cette relation exponentielle ($\sigma^0 \propto v$) est illustrée par la Figure 6.4.

Ceci a incité les géophysiciens à modéliser le signal rétro-diffusé σ^0 par une expression de la forme $A\nu^b$, où b est fonction de θ , et A est fonction de θ (en réalité, sa valeur dépend légèrement de ν lui-même, ce qui rend moins triviale l'expression).

De plus, les observations ont montré qu'un signal rétro-diffusé σ^0 plus puissant était observé lorsque l'on effectuait les mesures dans la même direction du vent, c'est-à-dire lorsque l'angle d'azimut est égal à 0° ou 180° –de telles mesures sont appelés “Upwind” et “Downwind” respectivement–, tandis que pour des observations de vents dans la direction perpendiculaire à l'antenne, le signal rétro-diffusé était minimum (angle d'azimut égal à 90° ou 270° appelé “Crosswind”) (voir Figure 6.5). En mettant en correspondance les observations effectuées lors de vols circulaires avec des observations prises par bateau, des modèles empiriques qui donnent une forme précise à cette fonctionnelle ont pu être développés.

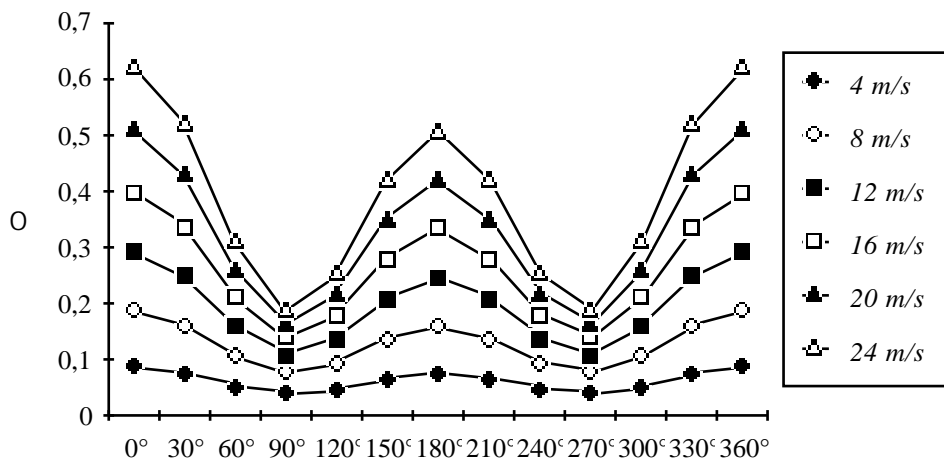


Figure 6.5 : Dépendance du signal rétro-diffusé par rapport à l'angle d'azimut. Plusieurs courbes sont montrées pour différentes valeurs de vitesse du vent. L'angle d'incidence est constant, correspondant à la trace du milieu ($\theta = 31,8^\circ$).

Le modèle empirique choisi pour la présente étude est le modèle CMOD1 développé par A.E. Long [Long 85] qui devrait être utilisé pour calibrer le scatteromètre du satellite ERS-1. Le modèle de Long établit que pour un angle d'incidence l'expression générale de σ^0 peut être approchée de manière satisfaisante par le développement de Fourier suivant :

$$\theta = U \cdot \frac{1 + b_1 \cdot \cos(\alpha) + b_2 \cdot \cos(2\alpha)}{1 + b_1 + b_2} \quad (6.1)$$

où le paramètre U , “Upwind”, correspond à la valeur de θ pris dans la direction du vent (où l’angle d’azimut $\alpha = 0^\circ$). U est égal à :

$$U = A \cdot v$$

Le modèle de Long spécifie que les paramètres A et θ dépendent uniquement de l’angle d’incidence α et que b_1 et b_2 sont fonction à la fois de la vitesse du vent v et de α (voir [Long 85] pour plus de détails concernant le calcul de ces paramètres.)

Actuellement, les différents paramètres du modèle de Long sont déterminés expérimentalement.

Les satellites, qui sont censés effectuer une couverture de la terre en trois jours, fournissent un ensemble de données plus vaste et plus riche en quantité que celles trouvés lors du calibrage des modèles empiriques –recueillis essentiellement par télédétection aéroportée–. En fait, confrontés à ces nouvelles informations, à différentes conditions de vol satellitaire et aéroportée, les modèles empiriques peuvent se montrer inadéquats et nécessiter ainsi un nouveau calibrage de leur paramètres, voire une nouvelle formulation !

En réalité, le problème que nous avons à résoudre est de déterminer le vecteur vent connaissant les trois θ donnés par les trois antennes. On a donc à faire une triple inversion couplée du modèle de Long. Cette inversion est non triviale dans le sens où la fonction inverse n’est pas univoque, ce caractère étant d’autant plus accusé que le bruit est important (voir plus loin). On se propose dans ce qui suit de réaliser cette inversion par réseaux de neurones, la fonction inverse étant considérée comme fonction de transfert.

La détermination des paramètres du modèle connexionniste que l’on a utilisé se fait au cours d’une phase d’apprentissage, à partir d’une base de données scatterométriques. Dans la phase de faisabilité qui constitue l’objet de ce travail, cette base scatterométrique a été simulée à partir de vents fournis par des modèles météorologiques auxquels on a appliqué la formule de Long pour calculer les θ . Le cas du traitement des données réelles, celles mesurées par le satellite, demandera éventuellement un nouveau calibrage des paramètres afin de trouver la vraie fonction qui modélise le phénomène

réel, approximé actuellement par la fonction de Long. Il suffira de refaire un apprentissage avec les nouvelles données; la méthode neuronale trouve automatiquement la nouvelle fonction de transfert et en ajuste les paramètres.

En l'absence du bruit introduit par les appareils de mesure ou par le modèle géophysique, les trois mesures de σ^0 fournies par les trois antennes pour un point donné $(\sigma^1, \sigma^2, \sigma^3)$ devraient permettre de déterminer de façon quasi-unique la vitesse et la direction du vent –à l'exception des points où il existe des ambiguïtés intrinsèques dans la direction (Figure 6.6a)–.

La surface explicative mise en évidence par les physiciens est une surface dans l'espace qui ne permet pas dans tous les cas de calculer la direction du vent de manière unique. Pour une coupe à module de vent constant on obtient une courbe de Lissajous telle que deux directions différentes de $\pm 150^\circ$ en moyenne peuvent être proposées en réponse à certains triplets de mesure (points doubles de la courbe de Lissajous) [Roquet et Ratier 88]. Le nombre de points ambigus est augmenté en présence de bruit (Figure 6.6b).

Le but de notre recherche est donc de dresser une carte des vents exacte par des méthodes neuronales à partir des mesures $(\sigma^1, \sigma^2, \sigma^3)$ observées.

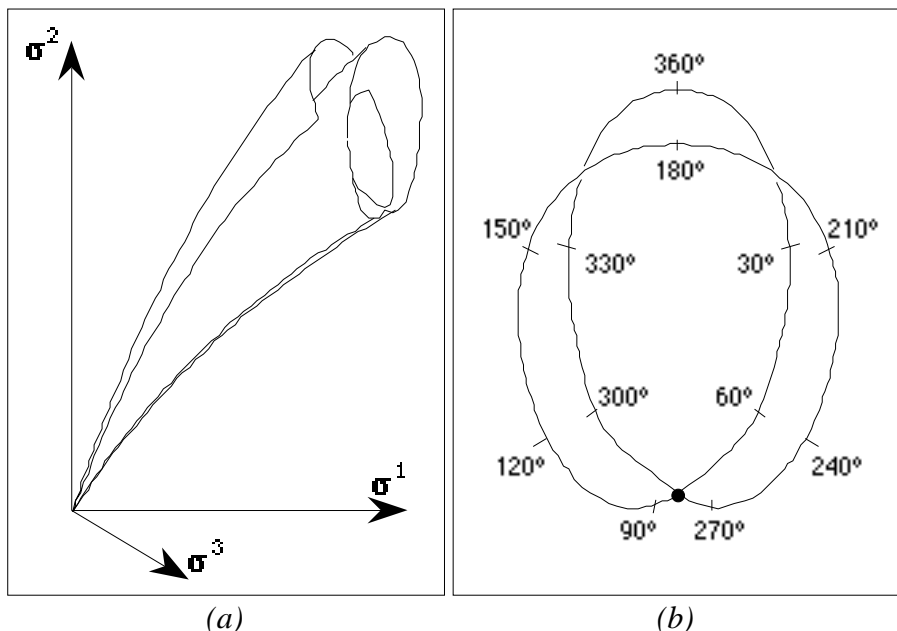


Figure 6.6 : Représentation de la fonction à modéliser : (a) fonctionnelle permettant de calculer la vitesse et la direction du vent en fonction des trois σ^0 mesurés par les trois antennes. (b) courbe obtenue pour une section de la courbe spatiale à vitesse de vent constante.

Cette Figure nous donne une idée de la difficulté de la tâche à résoudre. Calculer la vitesse et la direction du vent en fonction des mesures satellite équivaut à trouver les paramètres v et θ sur la surface de la Figure 6.6 en fonction des coordonnées (x_1, x_2, x_3) .

6.3. Vers une nouvelle approche

Les scatteromètres installés à bord de satellites vont permettre l'observation systématique des océans et donc améliorer la connaissance du vent sur les océans. Cette amélioration devrait déboucher sur une plus grande vraisemblance des modèles océaniques par une meilleure connaissance de la fonction de forçage, et aussi des modèles météorologiques par une assimilation de ces données de vent qui devrait guider ces modèles vers la réalité.

Le problème qui se pose maintenant est donc de définir et de mettre en place des méthodes numériques performantes améliorant la qualité des modèles géophysiques (ou fonctions de transfert) existants. Les procédés choisis pour l'élaboration de ces nouveaux modèles devront permettre de traiter des masses de données importantes et bruitées. En effet, les capteurs que l'on utilise peuvent être fortement bruités, bruits dus à la technologie employée pour la mesure ou aux hétérogénéités rencontrées lors du parcours atmosphérique comme la présence d'humidité dans l'atmosphère. Ce problème peut être abordé par des méthodes numériques relevant de la statistique

Notons dès à présent que certaines de ces méthodes permettent de très bonnes performances sur le plan numérique mais sont difficilement interprétables sur le plan physique. Il est alors possible de concevoir deux phases distinctes complémentaires dans l'élaboration d'un modèle :

- une première étape permet la détermination d'un modèle explicatif du phénomène observé. En particulier elle met en évidence toutes ou la plus grande partie des variables explicatives qui interviennent dans le problème.
- la seconde est essentiellement une étape numérique qui permet la détermination des paramètres du modèle utilisé. Ceux-ci doivent être

extraits à la suite d'une phase d'apprentissage (ou étalonnage) effectuée sur les mesures réelles faites par les appareils de mesure. Le modèle doit utiliser au mieux la richesse des données fournies par les scatteromètres. Cette phase doit enfin produire un modèle opérationnel capable de faire du calcul en temps réel, tout en améliorant la précision de ce calcul.

7. Approches du problème

On présente maintenant d'une façon simplifiée différentes approches pour extraire les caractéristiques du vent sur l'océan en fonction de données scattérométriques. Tout d'abord on décrit les méthodes classiques bâties à l'aide de solutions numériques. Ensuite on présente brièvement les méthodes connexionnistes appliquées au problème. Ces dernières méthodes font l'objet de notre travail et seront expliquées en détail dans les chapitres suivants.

Tant pour l'approche classique que pour l'approche connexionniste, il faut distinguer deux étapes dans le calcul de chaque vecteur du vent : ***l'extraction des solutions ambiguës*** et la ***correction d'ambiguïtés***. En effet, dans la plupart des cas au moins deux vecteurs de vent différents sont trouvés¹. Il faut donc une deuxième étape de calcul afin de lever les ambiguïtés aux résultats obtenus dans la première.

Résoudre le problème des ambiguïtés a donné lieu à un premier travail dans le domaine de télédétection spatiale en utilisant les réseaux neuronaux [Badran et al. 89] [Thiria 89]. Le succès atteint par ces travaux est à l'origine de mes propres travaux dans ce domaine.

D'une façon générale, quelque soit la méthode appliquée à l'extraction des caractéristiques du vent en fonction des mesures radar elle doit avoir un haut rendement en matière de rapidité de calcul. La méthode doit pouvoir traiter un minimum de 4 points par seconde pour permettre un calcul en temps réel du signal reçu. Il est cependant souhaitable d'avoir un rendement 10 fois supérieur [Cavanie et Offiler 86] afin d'envisager des solutions à temps réel.

¹ Dans le cas où les mesures sont prises avec trois antennes-scattéromètres, au moins deux solutions ambiguës, souvent appelées ***alias***, sont trouvés (ce nombre monte à 4 pour les systèmes à deux antennes). Ces solutions sont séparées en direction de 150 à 210 degrés. Il y a toujours une plus probable ou "meilleure" solution et elle représente près de 70 % des cas [Roquet et Ratier 88]. Le 30 % restant de cas, c'est la deuxième solution et même parfois une autre la correcte. Les affirmations précédentes ont été vérifiées par nos résultats.

7.1. Le modèle utilisé

Dans le chapitre précédent nous avons introduit un modèle géophysique qui établissait une relation entre le coefficient de rétro-diffusion σ^0 mesuré par le satellite en un point et la vitesse et la direction du vent dans ce point ainsi que avec l'angle d'incidence du signal dans le point, c'est-à-dire :

$$\sigma^0 = G(\alpha, \theta, \nu) \quad (7.1)$$

Le modèle CMOD1 développé par Long (6.1) [Long 86] correspond à ce type de modèle. Il est le résultat de plusieurs années de recherche et devrait être utilisé pour le décodage du scattéromètre de ERS1 après un éventuel et même probable réétalonnage [Attema 91].

Un autre modèle est celui utilisé dans le programme NSCAT de la NASA. Il s'agit du modèle SASS-1 qui est dans l'essentiel semblable à celui de Long, qui est décrit dans [Chi et Li 88] ².

A présent, tant les méthodes que nous appelons classiques, ou numériques, que les méthodes connexionnistes utilisent, bien que pour des buts différents, un modèle similaire à celui de Long, CMOD1.

7.2. Les approches classiques

Plusieurs techniques ont été testées pour extraire la vitesse et la direction du vent en fonction de données satellitaires. La plupart des techniques utilisent dans leurs calculs une fonction coût construite à partir des écarts entre le signal réel mesuré par le scattéromètre et le signal scattéromérique simulé par une fonction de transfert du type (7.1) correspondant à un vecteur vent mesuré in-situ.

² Dans l'état actuel des recherches nous savons que le modèle utilisé finalement dans le programme ERS-1 est CMOD2. De même dans le programme NSCAT le modèle retenu est SASS-2. On suppose qu'il s'agit de versions plus affinées des versions cités ci-haut (Voir dans [Attema 91] et [Naderi et al. 91] pour plus de détails).

La fonction de coût est typiquement de la forme suivante :

$$J(i, \mathbf{v}) = \sum_{i=1}^3 \frac{\hat{\lambda}_i^0 - \lambda_i^0}{Kp_i}^2 \quad (7.2)$$

où $\hat{\lambda}_i^0$ correspond au coefficient mesuré physiquement par la i -ième antenne du satellite; λ_i^0 est le coefficient calculé à l'aide du modèle (donné par (7.1)) pour l'antenne i ($\lambda_i^0 = G(\mathbf{v}, \mathbf{v}_i)$); et Kp_i est l'écart type normalisé du bruit estimé sur l'antenne i , bruit qui est généré par les instruments de mesure.

La solution consiste à minimiser cette fonction de coût en prenant comme contrôle les paramètres de (7.1). La détermination de différents minimum permet de proposer différentes solutions donc, différents vecteurs.

Plusieurs algorithmes existent pour trouver une (ou plusieurs) solution de (7.1); ils reposent sur le principe de la meilleur approximation ("best-fit"). Leur différence réside dans la façon dont ils calculent les valeurs successives du couple (\mathbf{v}, λ) correspondant à une entrée initiale $(\mathbf{v}_0, \lambda_0)$. Les techniques sont diverses et peuvent aller des méthodes de gradient conjugué ou de Newton, jusqu'à de méthodes très simples, mais rapides, de recherche linéaire et d'interpolations à partir de données tabulées.

7.3. L'approche connexionniste

Comme nous l'avons rappelé dans la partie théorique (chapitre 4) les réseaux MLP fournissent un ensemble de fonctions intéressantes. Le principal résultat utilisé dans cette application est celui de Funahashi [Funahashi 89] qui dit que toute fonction continue d'un compact de \mathbb{R}^n dans \mathbb{R}^p peut être approximée par un réseau multicouche dont les automates des couches cachées utilisent des fonctions sigmoïde et dont les automates de sortie sont linéaires. Le choix de l'architecture, nombre de couches et nombre de cellules cachées, détermine la précision de l'approximation. Les expériences menées ont prouvé que les propriétés pouvaient être expérimentalement étendues à d'autres classes de fonctions, en particulier à la courbe de Lissajous que nous avons étudié (voir Figure 6.6).

Notre but est de démontrer qu'à l'aide des techniques connexionnistes il est possible d'extraire les solutions du vent en fonction des λ_i^0 . D'autre part

nous montrerons que notre méthode est opérationnelle dans le sens où elle respecte les spécifications proposées par les oceanographes. Celles-ci sont de $\pm 20^\circ$ pour la direction du vent et de ± 2 m/s pour la vitesse.

Amélioration de la vitesse de calcul

Un des intérêts de l'approche connexionniste est d'améliorer la vitesse du calcul des vecteurs vent. Nous montrerons comment dans une station de travail de type Sun Sparc 2 notre méthode calcule plus rapidement qu'une méthode variationnel sur un super-calculateur Cray 2. Pour les méthodes classiques [Roquet et Ratier 88], les temps de calcul sont de 0,21~0,24 secondes de CPU sur un ordinateur Cray 2 par vecteur vent. Cependant, une telle estimation n'exploite pas les capacités vectoriels du Cray. Vectoriser le programme utilisé doit pouvoir faire gagner un facteur 10. De notre côté, nous avons calculé le temps CPU de notre application sur une station de travail de type Sun Sparc 2. Nous n'avons pas cherché à optimiser la taille des réseaux, ils sont en fait sur-dimensionnés. En particulier, en ce qui concerne la vitesse nous savons qu'un réseau véritablement minuscule ($3 \times 2 \times 1$ cellules, soit 11 poids de connexions, y compris les seuils) peut répondre avec des résultats très acceptables. Or, le réseau mis au point lors du test de vitesse de la machine complète est plus précis mais beaucoup plus encombrant ($27 \times 25 \times 25 \times 1$ cellules et 1376 poids de connexions). Le logiciel où les réseaux sont implantés est en fait un simulateur de réseaux neuronaux. Le temps total calcul par vecteur vent est de l'ordre de 10 ms de CPU sur une station Sun Sparc 2. Cela est près de 25 fois plus rapide que la méthode de Roquet et Ratier exécuté sur Cray 2.

Ce n'est pas à première vue une comparaison très loyale : différentes méthodes sur différentes machines avec des différentes données (mais du moins a t'on utilisé le même modèle, le modèle de Long, CMOD1 pour la simulation des vraies données). Indépendamment de cette comparaison, examinons de plus près les atouts inhérents à la méthode que nous proposons :

- **L'approche modulaire** : pour le calcul des paramètres en un point de mesure, les réseaux utilisés sont inter-dépendants. L'information calculée par le réseau à un niveau envoie le résultat aux réseaux des niveaux supérieurs. Il n'y a donc pas d'indépendance au niveau de modules séparés et, apparemment, pas de parallélisme possible. Cependant,

l'utilisation d'une machine-réseaux indépendante dédiée à chacun des n points mesurés par le satellite dans le sens de la largeur de la fauchée nous donne la possibilité de calculer indépendamment, et simultanément, n points. Ceci multiplie par ce même facteur le rendement globale de la machine.

- ***Chaque réseau est une machine massivement parallèle.***
- ***L'algorithme donnant la solution peut être implanté en "hardware" :*** en effet, pour nous l'algorithme donnant la solution est en fait l'architecture du réseau, avec ses cellules, ses connexions et les poids associés, les fonctions de transition ainsi que la dynamique de propagation des états. Tout ceci peut très bien être câblé dans des circuit électroniques, ce qui ferait une machine dédié au problème.

En exploitant les multiples avantages de la méthode de réseaux connexionnistes que nous proposons on pourrait avancer (ceci n'est pas simple à estimer de façon précise), un facteur 100 voire 1000 fois par rapport aux vitesses d'exécution actuelles, ce qui le rend approprié à l'exploitation en temps réel, ainsi qu'à son installation dans la source de mesure directement.

8. Les données

8.1. Les Sources de Données

Les données utilisées au cours de nos travaux de recherche proviennent de mesures scattérométriques (σ^0) déterminés par la formule de Long appliquée à des vents simulés par des modèles météorologiques.

La première et plus importante base de données contient des informations qui simulent les fauchées du scatteromètre du satellite ERS-1 puisqu'à l'époque où nous avons entrepris ce travail ERS-1 n'était pas encore lancé. Les champs de vents sont obtenus à partir du modèle de prévision météorologique du Centre Européen de Météorologie de Reading (Grande Bretagne) au dessus duquel on a simulé l'échantillonnage du scatteromètre du satellite ERS-1. L'équipe d'Océanographie et Climatologie LODYC à l'Université de Paris 6 nous a fourni ces champs de vents simulés. Tous les systèmes présentés pour étudier cette application ont été d'ailleurs entraînés avec les données extraites de la base du LODYC.

De manière à rendre comparable les performances de notre étude à celles d'études antérieures, des champs de vent uniformes ont été sélectionnés et des test effectués dans des conditions semblables pour tous les algorithmes à celles décrites dans [Chi et Li 88]. Notamment, des tests à vitesse de vent constante et direction aléatoire.

Les expériences nous permettent de valider l'approche neuronale par rapport à l'approche classique.

8.1.1. Les données fournies par LODYC

La base de données fournie par le laboratoire LODYC est composée par 31 fichiers de données, chaque fichier représentant les mesures prises pendant une journée d'activité du satellite. Il s'agit de données prises sur l'océan

Atlantique sud, mesurées pendant le mois de septembre de l'année 1985. De ces données, plus de 20 000 mesures, une partie a été retenue pour tester les résultats, le reste est disponible pour effectuer l'apprentissage. Nous expliquerons plus bas la constitution de ces ensembles

Chaque fichier contient les informations de six ou neuf fauchées ascendantes et descendantes qui correspondent aux différents passages journaliers du satellite. La Figure 8.1 représente les données contenues dans l'un des 31 fichiers de la base de données LODYC. La fauchée tracée par le satellite est composée par 10 traces de données³ chaque trace étant caractérisée par un triplet de θ d'angles d'incidence $(\theta_1, \theta_2, \theta_3)$ fixe (voir aussi dans le chapitre 6 la Figure 6.3).

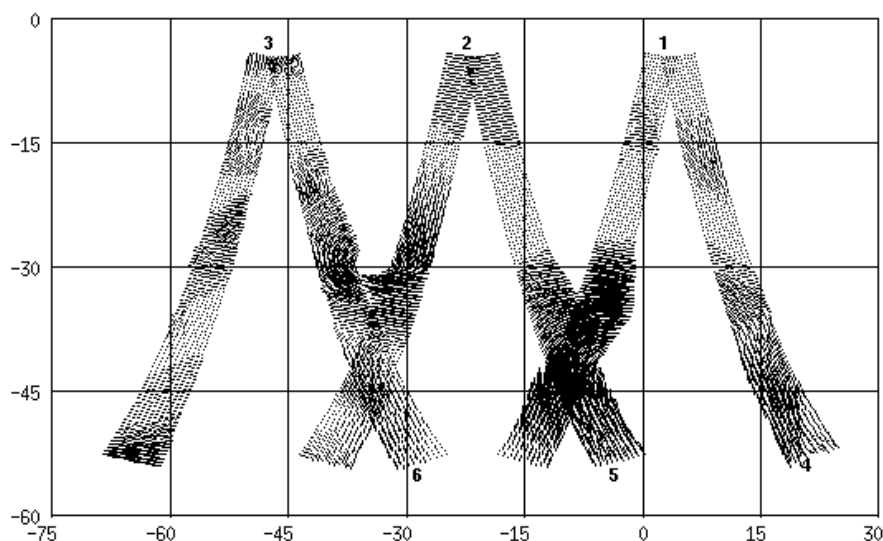


Figure 8.1 : Fauchée du satellite ERS-1 montrant les bandes montantes et descendantes du vent au cours de 3 orbites consécutives. Chaque bande étant composée de dix traces.

Les seuls paramètres qui différencient une trace d'une autre sont les angles d'incidence $(\theta_1, \theta_2, \theta_3)$. Le paramètre θ_0 étant fonction de θ , on trouve en effet, une certaine homogénéité dans les valeurs de θ_0 pour une même trace. Ce qui n'est pas le cas pour des valeurs θ_0 dans de traces différents : *il y*

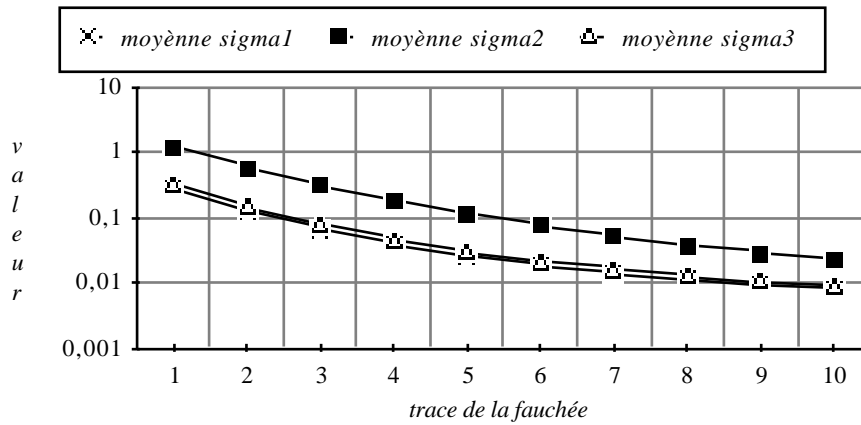
³ Le nombre réel de traces est de 19 ce qui donne une résolution de 25 Km. Les champs de vent fournis par le centre de Reading étaient de 10 traces, ce qui donne une résolution plus faible, de 50 Km, qui est fréquemment utilisé par les océanographes.

a une très forte variation de σ en fonction de θ . La Figure 8.2a montre la moyenne de la valeur des trois paramètres ($\sigma_1, \sigma_2, \sigma_3$) en fonction de la trace observée. Ces moyennes ont été calculées en fonction de la totalité des données. L'échelle des ordonnées est logarithmique. A titre d'exemple, la moyenne du paramètre de la deuxième antenne, σ_2 , est 1,18 et 0,57 respectivement pour la première et la deuxième trace. Elle diminue pour arriver enfin à une valeur de 0,022 pour la dixième trace qui est la plus éloignée de la trajectoire du satellite. On peut très bien comprendre de façon intuitive cette diminution : un grand angle d'incidence implique une vue plus rasante sur l'océan, donc le signal radar rétro-diffusé aura plus de difficulté de revenir en direction du satellite. La caractéristique de la diminution des valeurs est compensée par un codage des données ce qui nous a permis d'améliorer les performances des réseaux entraînés.

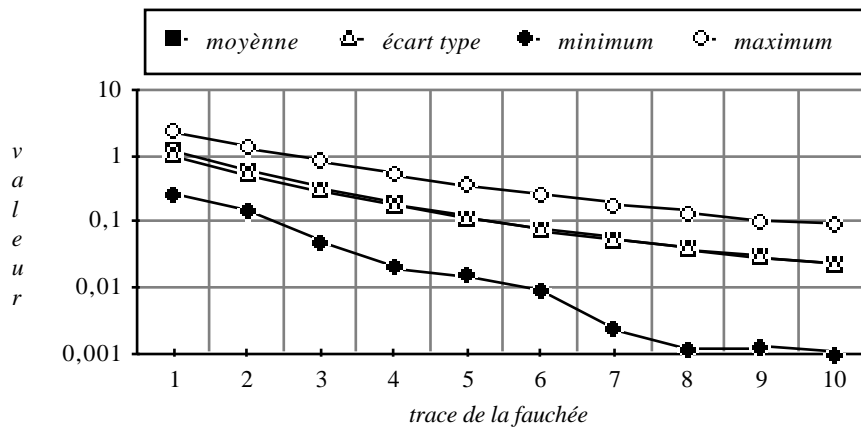
Distribution des données

Outre cette très forte diminution de σ par rapport à l'augmentation de θ , la Figure 8.2 nous montre également une plus grande valeur en moyenne pour les σ , mesurés par la deuxième antenne. Cette antenne pointe perpendiculairement à la trajectoire du satellite et on travaille ici sur une zone de l'océan où, pour la période d'observation, les vents soufflent presque exclusivement dans les directions Est-Ouest et Ouest-Est. Autrement-dit, puisque le satellite se déplace approximativement selon l'axe Nord-Sud, le faisceau émis par l'antenne centrale du satellite mesurera presque toujours des vent "Upwind" –vents soufflant dans la même direction à la mesure effectué par le satellite, i.e. avec un angle d'azimut $\alpha = 0^\circ$ – et "Downwind" –correspondant à un angle $\alpha = 180^\circ$ – angles qui donnent les plus grandes valeurs de σ .

Les histogrammes de la Figure 8.3 montrent la distribution des données par rapport à l'angle d'azimut et au module du vent. Les données correspondent seulement à la trace n° 5; les distributions dans les autres traces sont analogues. La Figure 8.3 montre que le premier ensemble de données –appelé d'apprentissage dans la figure– n'est pas équilibré, autrement-dit, toutes les directions ne sont pas représentées de façon équitables. Nous discuterons plus tard la manière dont il faut choisir "au mieux" dans cet ensemble de quoi définir des ensembles d'apprentissage adaptés.



(a)

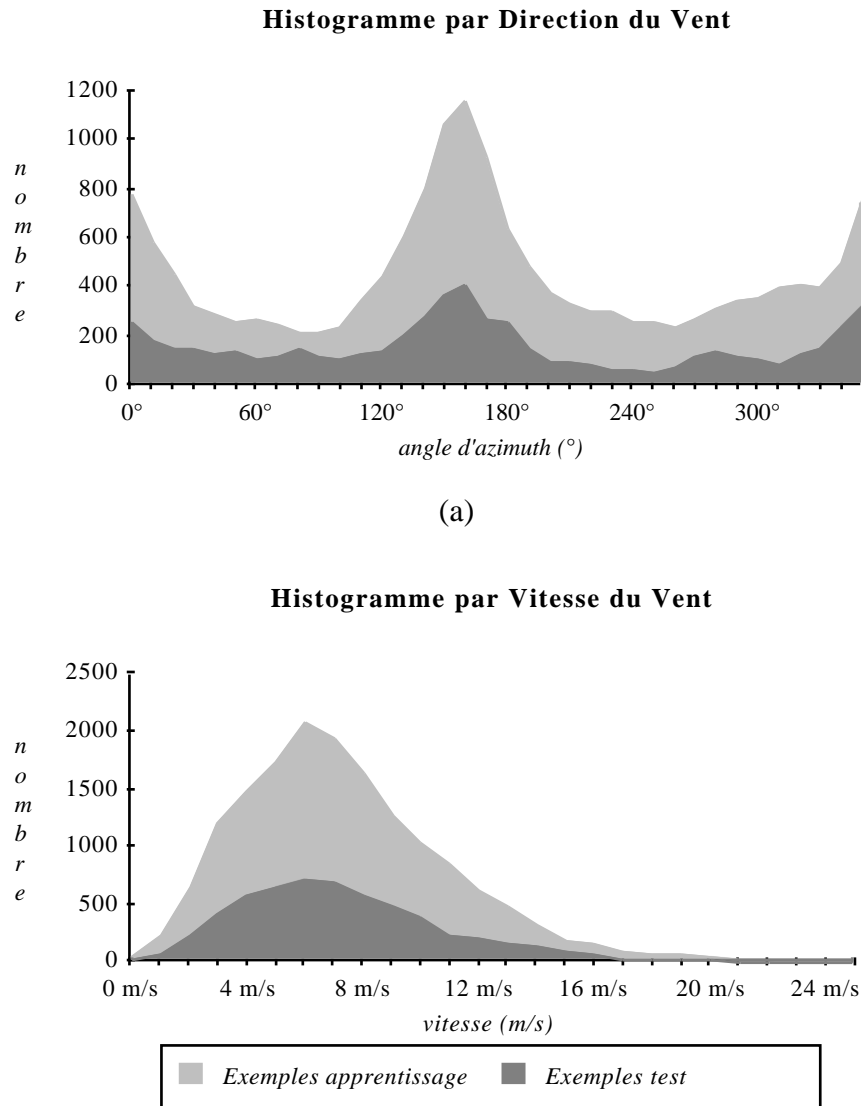


(b)

Figure 8.2 : Statistiques sur la Base de Données : (a) Valeurs moyennes des triplets (σ_1 , σ_2 , σ_3) par trace de la fauchée. La trace 1, plus proche à la trajectoire du satellite, a un angle d'incidence plus petit. La puissance du signal écho-radar rétro-diffusé par ce point est la plus importante. Pour le signal écho-radar mesuré par l'antenne du milieu σ_2 , sa moyenne va de 1,18 à 0,022, tandis que pour l'antenne en aval la moyenne de σ_1 varie entre 0,28 et 0,008. Les moyennes pour les mesures des antennes des extrêmes, σ_1 et σ_3 sont presque identiques, les courbes se superposent. (b) Calculs statistiques sur le paramètre σ_2 - antenne du milieu qui pointe perpendiculairement la trace du satellite-. Dans la figure la moyenne est superposée à l'écart type.

Due à la faible quantité de données pour les très basses et très hautes vitesses nous avons limité la base de données à des vents compris entre 4 et 20 m/s. Un vent au dessous de 4 m/s est considéré comme faible et peu énergétique vis à vis des transferts océan-atmosphère et par conséquent de très faible intérêt pour les océanographes. En revanche, les vents forts, de

plus grand intérêt, ne sont pas en mesure d'être appris, leur occurrence statistique n'étant pas significative



(b)

Figure 8.3 : Distribution des données. (a) Histogramme des directions : Les données avec un angle d'azimut proche de 0° et 180° sont largement plus représentées. Les abscisses indiquent l'angle de direction en dizaines de degrés. (b) Histogramme des vitesses : les basses et moyennes vitesses sont plus fréquents. Les histogrammes correspondent aux données sur la trace du milieu seulement (trace n° 5).

Différentes traces : proche, du milieu et extrême

La dépendance de θ par rapport à l'angle d'incidence fait que le calcul des caractéristiques du vent en des points appartenant à des traces

différentes (différent) (Figure 8.2) doit être traité comme des problèmes séparés, d'où la nécessité de dédier une "machine réseaux" par angle d'incidence. La trace choisie pour la plupart de nos essais est celle du milieu, la trace n° 5. Cependant, des tests ont été réalisés pour une trace plus proche de la trajectoire du satellite, la n° 2, et, par symétrie, pour la trace n° 8, correspondant à la plus éloignée.

Angle d'Incidence par trace :	Trace n° 2 "trace proche"	Trace n°5 "trace du milieu"	Trace n° 8 "trace extrême"
1, antenne en aval	29,5°	41,9°	51,5°
2, antenne du milieu	21,7°	31,8°	39,2°
3, antenne en amont	29,5°	41,9°	51,5°

Table 8.1 : Angle d'incidence par trace du satellite.

Base d'apprentissage

Sur les 31 cartes correspondant aux jours de Septembre, 22 cartes ont été sélectionnées pour créer l'ensemble d'apprentissage. Les 9 autres ont été utilisées pour les tests.

Les premiers essais ont mis en évidence un déséquilibre de la base d'apprentissage : le système apprenait très bien certaines directions et très mal d'autres. Ceci était une conséquence directe du fait qu'il y avait trop de données appartenant à une classe et trop peu à d'autres. Cette propriété est mise en évidence dans la partie théorique (§4.2.2). La précision de la classification obtenue étant liée à la probabilité d'apparition d'une valeur. Les calculs définitifs ont été faites avec des bases de données équilibrées. La base d'apprentissage été constitué approximativement de 6 000 exemples qui diffèrent selon les expériences effectuées. Nous préciserons à chaque fois de quelle manière ils ont été constitués.

Base de test

Les 9 cartes utilisées pour former les différents ensembles de test forment un ensemble représentatif de la totalité. On y retrouve notamment le même déséquilibre rencontré dans les bases d'apprentissage.

La présence du déséquilibre dans la base de données dédiée aux tests n'est pas un grave problème (mais il ne faut pas le négliger non plus). Un déséquilibre à l'apprentissage génère une meilleure reconnaissance des directions redondantes. Un pareil déséquilibre dans le test donne des bases de test plus "faciles", les performances globales sont souvent plus élevées que pour la base d'apprentissage. Une vérification des performances par classe s'impose. La base de test est constituée approximativement de 5 000 exemples.

8.2. Pré-traitement des données

Données brutes

Une des grandes capacités des réseaux de neurones est celle de pouvoir traiter les données brutes, telle qu'elles sont fournies par les appareils de mesure, sans aucun pré-traitement !.

Cependant, des meilleurs résultats peuvent être obtenus à l'aide d'un prétraitement simple sur les données. Un prétraitement souvent utilisé consiste à centrer et réduire les variables afin de ramener sa moyenne à zéro et son écart type inférieur à un. Ainsi, la plupart des valeurs se trouveront dans l'intervalle $[-1, 1]$ ce qui s'est montré expérimentalement convenable. Tout ceci est lié bien évidemment à la valeur initiale des poids des connexions et aux caractéristiques de la fonction de transition utilisée.

La fonction de transition utilisée dans nos travaux est la fonction tangente hyperbolique (ou sigmoïde) calibrée comme le montre la Figure 2.7 ($a = 1,7159$ et $k = 1,333$) afin d'avoir un comportement presque linéaire dans l'intervalle $[-1,1]$.

Normalisation des données

Un approche classique en statistique multidimensionnelle est de centrer et de réduire les données par variable de manière à ce que chacune d'entre elles amène la même information. Les variables que nous avons à traiter sont les (x_1, x_2, x_3) correspondant à des paramètres pris par trois antennes différentes mesurés sur un même point (voir Figure 6.2). De plus, pour traiter cette application nous avons été amené à faire intervenir des informations

contextuelles : les données relatives à plusieurs points, donc à plusieurs traces, ont été utilisés simultanément. Plus particulièrement un voisinage de 3×3 points a été retenu autour du point d'intérêt (voir Figure 8.4).

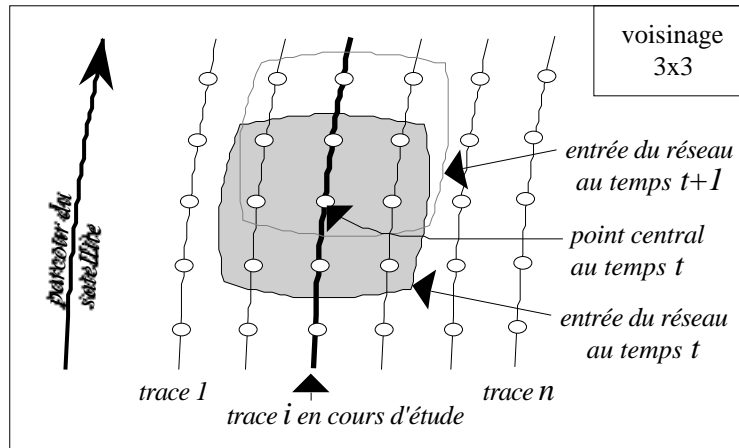


Figure 8.4 : Prise en compte du contexte. Pour chacun des paramètres (x_1, x_2, x_3) les valeurs dans un voisinage 3×3 autour du point à l'étude sont prises à chaque instant.

Ces paramètres ont une dépendance spatiale, chaque triplet de mesures (x_1, x_2, x_3) représente un point dans l'espace sur la courbe représenté dans la Figure 6.6a. Une normalisation par variable x_i affecterait cette dépendance spatiale et ferait apprendre au réseau une version déformée de la Figure 6.6. Nous avons donc pris les trois (x_1, x_2, x_3) ainsi que l'information contextuelle comme une seule variable en ce qui concerne la normalisation. Comme une conséquence, la moyenne et l'écart type dont nous parlons dans cette section correspond à la moyenne et l'écart pris sur tout l'ensemble de données et toutes les "variables" de chaque donnée.

La zone (presque-)linéaire des fonctions sigmoïde utilisées dans notre démarche se trouve dans l'intervalle $[-1, 1]$. Au delà de cette intervalle on entre dans une zone fortement non linéaire pour arriver finalement dans une zone de saturation. Si l'on tombe dans la zone de saturation, dûe à des valeurs des entrées trop importantes ou à une mauvaise initialisation des poids (trop importantes aussi), ou même à cause d'une trop grande correction des poids dûe à une erreur en sortie trop importante, l'algorithme d'apprentissage risque de donner des solutions biaisées. Il est donc recommandé pour éviter cela, de coder l'information utilisée par le simulateur –les entrées et les sorties– afin de les placer dans cet intervalle.

Il est aussi important de centrer les données autour de zéro. On peut ne pas le faire mais ce qui peut produire un apprentissage moins efficace et plus lent.

Une méthode utilisée couramment en statistique pour coder des données est de les centrer et les réduire. Le même procédé est souvent utilisé pour les données présentés aux réseaux. Rappelons que les réseaux de neurones font partie de méthodes statistiques.

La procédure de centrer et réduire se fait en deux pas :

- **centrer les données** : la moyenne des données est ramenée vers zéro. Pour ceci on remplace chaque élément par lui même moins la moyenne.
- **réduire les données** : s'assurer que la plupart des données sont codées dans l'intervalle $[-1, 1]$. Chaque nouvelle donnée centrée est divisée par l'écart type et multipliée par 0,66. Ceci produit un écart type de 0,66 sur le nouvel ensemble de données centré et réduit.

Ceci nous assure, du moins dans le cas d'une distribution gaussienne, que la plupart de données se trouveront dans l'intervalle $[-1, +1]$.

Une telle normalisation en tenant compte de la globalité des entrées permet de conserver la relation entre les différentes variables à l'entrée du système. On applique aux bases d'apprentissage et de test exactement la même fonction de normalisation : la moyenne et l'écart type de la base d'apprentissage sont utilisés aussi dans la réduction des différentes bases de test. Toutefois, pour les données relatives à chaque trace du satellite, la moyenne de la base d'apprentissage correspondante a été appliquée.

Cette technique de normalisation est celle qui nous a donné les meilleurs résultats des performances du système. La rapidité d'apprentissage est supérieure au cas où la normalisation n'est pas effectuée.

9. Approche connexionniste

9.1. Introduction

Nous avons présenté dans les chapitres précédents le problème géophysique tel qu'il se pose aux océanographes ainsi que les données qu'ils possèdent pour mettre au point des solutions opérationnelles. Ce point est d'une particulière importance ici puisqu'il s'agit d'un problème réel dont les solutions sont loin d'avoir été trouvées jusqu'à aujourd'hui faute de mesures réelles. Les premières modélisations proposées par les océanographes ont donc été élaborées à partir de mesures simulées par les modèles météorologiques et des mesures exactes, en nombre limité, obtenues à partir de bateaux. Ces mesures ne sont pas, de par la manière dont elles sont acquises, distribuées de manière hétérogène aussi bien spatialement que temporairement. Les modélisations proposées peuvent donc être très imprécises, et c'est ce que semble prouver l'arrivée des premières données scattérométriques. Les premières vérifications, qui datent de Mars 1992, tendent à montrer que la forme analytique de la fonction est trop imprécise pour traiter de ces données et du bruit dont elles sont entachées.

L'étude que nous présentons a commencé en 1989, nous avons donc utilisé, comme les océanographes, les données disponibles à cette époque en particulier la Fonction de Long et les données simulées. Cependant, la portée du travail présenté est plus générale puisque nous tenterons de montrer que la méthode employée ne dépend pas de manière exclusive des fonctions et des données utilisées ici pour résoudre le problème. Nous présentons une méthodologie, fondée sur l'utilisation des réseaux, qui pourra être utilisée intégralement à l'arrivée des données réelles pour tracer les cartes de vent, mais qui pourra également traiter des problèmes similaires qui apparaissent en géophysique. En effet, la généralité de cette méthode réside dans ses capacités d'apprentissage supervisé; il devient possible de déterminer automatiquement, à partir d'exemples de mesure, les nouveaux paramètres nécessaires. On met ainsi à la disposition des océanographes un outil

puissant : non seulement la méthode numérique est précise et simple, mais l'apprentissage peut être effectué de manière évolutive. La précision du modèle peut, dans la réalité, être affinée petit à petit avec l'arrivée de nouvelles valeurs expérimentales.

Les problèmes qui apparaissent dans les sciences expérimentales lorsque l'on doit analyser des mesures représentatives d'un phénomène à étudier peuvent souvent se ramener à l'élaboration de fonctions de transfert. Le but principal de l'étude est de montrer que les réseaux de neurones constituent un outil utile pour la recherche de ces fonctions. Nous présentons ensuite une méthodologie générale qui peut être utilisée de façon systématique. Son utilisation permet de construire des "machines neuronales" dont les éléments de bases sont des réseaux de neurones. Ces techniques confèrent aux systèmes que l'on construira les qualités suivantes qui apparaissent comme essentielles dans le développement de machines dédiées pour des grandes applications. Il s'agit de :

- **la modularité** : Les systèmes que nous proposons de développer sont constitués d'un ensemble modulaire de réseaux de neurones indépendants qui collaborent à la réalisation globale d'une tâche complexe. Pour mettre l'accent sur cette notion d'architecture modulaire nous parlerons désormais de "machine neuronale" pour désigner ces systèmes, bien qu'au stade actuel il s'agisse d'un projet purement logiciel. Les aspects matériels pourraient être envisagés ultérieurement.
- **le parallélisme** : les différents modules réseaux fonctionnent chacun naturellement de façon parallèle, leur coopération se fait en alternant les modes parallèles et séquentiels. L'ensemble constitue une architecture spécialisée massivement parallèle.
- **l'efficacité** : la conception modulaire et la coopération de machines simples permet à moindre coût d'atteindre des performances élevées.

Cette étude propose un projet original dans le domaine des réseaux de neurones. Il s'agit de construire pour une application donnée un **système de réseaux modulaire** permettant le traitement complet de l'application. La puissance de la modélisation neuronale est qu'elle favorise, par son formalisme unifié, l'utilisation de la modularité. En effet, bien que de nombreux réseaux aient été utilisés pour résoudre des problèmes ponctuels, faire collaborer différents modules à une tâche globale constitue une étape

supplémentaire dans l'utilisation de ces réseaux. Cette collaboration pose de nouveaux problèmes pour la conception et la mise en oeuvre des réseaux, cependant la réussite d'un tel projet doit permettre d'envisager des améliorations de performances importantes par rapport aux méthodes actuelles [Chi et Li 88] et [Roquet et Ratier 88].

Le modèle développé illustre une méthodologie générale, elle sera développée sur le problème de télédétection spatiale présenté plus haut et qui appartient à un domaine en expansion. L'utilisation d'une telle machine par rapport aux techniques actuelles nous permet d'avancer un chiffre variant entre 100 et 1000 pour le gain en temps de calcul. Il s'agit d'une application qui est à la fois de taille importante et réalisable en l'état actuel du domaine. Ce dernier point est extrêmement important, car s'il est simple de proposer des systèmes intégrés correspondant à des chaînes de traitement, il est beaucoup plus compliqué de définir des applications raisonnables et réalisables en temps limité. Ces dernières années ont ainsi vu de nombreux projets qui se sont rapidement avérés irréalistes comme par exemple la réalisation de chaînes complètes de traitement neuronal dans des domaines complexes comme le traitement de l'image ou de la parole.

Nous avons dans la première partie développé tous les résultats théoriques qui vont nous permettre maintenant de montrer que notre méthodologie est générale et qu'un même raisonnement peut être appliqué avec succès pour résoudre de nombreux problèmes. Comme nous l'avons déjà dit une importante part des problèmes qui se présentent en géophysique se ramènent à la détermination d'une fonction de transfert, la première partie sera donc dédiée à ce sujet. La seconde partie présentera la "machine neuronale" qui permet de traiter dans son entier l'application de télédétection spatiale qui a été présentée.

De manière générale une telle approche peut être reproduite pour des problèmes similaires. Plusieurs études qui reprennent la méthodologie et des parties de la "machine neuronale" présentée ici ont déjà été utilisées sur des sujets proches en utilisant des données réelles et non simulées. Les comparaisons effectuées avec d'autres techniques classiques ont montré la validité de l'approche. Ces résultats ne seront pas présentés ici, nous avons choisi d'exposer la méthodologie et les justifications théoriques qui prouvent la généralité de la méthode.

9.2. Approximation des fonctions de transfert

9.2.1. Les modèles explicatifs

Dans une première approche, modéliser un problème revient à déterminer les différentes variables capables d'expliquer le phénomène étudié. Il importe à ce stade de comprendre grossièrement les différentes variations que présentent les variables explicatives au cours des différentes situations observées. Cette phase de modélisation est essentielle pour la compréhension du phénomène physique, elle débouche sur la détermination d'un premier modèle. L'expression mathématique se fait la plupart du temps sous la forme d'une fonctionnelle:

$$T(\mathbf{x}) = \mathbf{y}$$

où \mathbf{x} représente le vecteur des données expérimentales mesurées, \mathbf{y} le vecteur des valeurs à expliquer et T est une fonction vectorielle.

T peut être vue comme la fonction de transfert qui permet d'expliquer le phénomène à partir des observations et l'on fait le plus souvent, de manière sous-jacente, une hypothèse de continuité sur elle. Les points obtenus par expériences permettent la détermination des paramètres et assurent la validité de la relation en tout point.

La donnée d'une expression analytique est encore la manière la plus concise de comprendre les dépendances multiples qui interviennent dans les phénomènes physiques. Le plus souvent, les fonctions choisies à ce stade sont simples et incluent des hypothèses de linéarité. Les résultats numériques obtenus de cette manière ne peuvent représenter de manière très précise la réalité puisque le modèle est trop simplifié par rapport à la réalité.

Notons que ceci est exactement ce qui arrive pour la fonction de Long qui fait l'objet de cette étude. Les réseaux de neurones, qui permettent d'utiliser des fonctions non linéaires, peuvent jouer un rôle important dans une seconde phase de modélisation pour atténuer ces défauts.

9.2.2. Les modèles opérationnels

On recherche maintenant des modèles numériques capables d'améliorer les modèles existants. Si l'on suppose que le nombre de mesures disponibles

est très grand, il doit être possible de trouver une fonction de transfert précise qui permette la prédiction. Plus exactement, l'existence de la fonctionnelle T étant jugée acceptable pour expliquer le phénomène physique, nous pouvons maintenant accepter que la fonction de transfert proposée soit une fonction complexe des variables sélectionnées précédemment. On abandonne donc ici des hypothèses restrictives comme la linéarité; les réseaux multicouches quasi-linéaires fournissent alors, comme nous l'avons vu dans le chapitre 4, des bases de fonctions puissantes très adaptés pour traiter ce problème. En effet, outre les propriétés d'approximateur universel, les propriétés statistiques qu'ils présentent les rendent très résistants au bruit et donc très adaptés au traitement des données réelles entachées par les erreurs de mesures et le bruit relatif aux grandeurs observés. De nombreuses expérimentations menées dans des domaines comme ceux de la reconnaissance de la parole ou de l'image ont démontré les capacités de ces réseaux à traiter des masses importantes de données numériques [Bennani 92]. Dans un grand nombre de problèmes de classification, ils se sont montrés capables de performances de généralisation supérieures à celles obtenues par des procédés plus classiques existants. Ces performances confirment les résultats théoriques et justifient l'utilisation des réseaux pour l'approximation de fonctions de transfert.

De plus, la détermination du modèle peut nécessiter une phase de mise au point ou d'apprentissage longue, mais doit nécessairement être rapide au moment de son utilisation en vue de la prédiction (par exemple calcul de la direction du vent à partir des θ). Le principal objectif n'est plus comme dans la première phase d'élaborer un modèle physique pour rendre compte des relations existantes entre les différentes grandeurs, mais de fournir un système permettant de calculer, en temps réel, les sorties des modèles précédents avec un maximum de précision possible. Des logiciels adaptés rendent l'utilisation des réseaux de neurones facile et rapide, ils sont donc une alternative intéressante en modélisation. De plus, comme nous l'avons déjà mentionné, l'introduction de machines parallèles, transputers, etc. permet d'envisager la construction de machines permettant d'effectuer un grand nombre de calculs en temps réel.

La "machine neuronale", représente la machine dédiée au traitement du problème complet : construction des cartes de vent à partir des signaux scattérométriques. Avant de la présenter, nous montrerons les résultats obtenus sur des problèmes intermédiaires dont le traitement est nécessaire à

l'élaboration de l'ensemble. Chaque problème peut se formuler comme la recherche d'une fonction de transfert complexe. Les résultats théoriques du chapitre 4 seront donc utilisés pleinement. Les résultats partiels présentés dans ce paragraphe seront donc réunis et utilisés par la "machine neuronale".

9.3. Approximation des fonctions à l'aide des MLP

Dans ce paragraphe nous montrons que la modélisation neuronale permet d'obtenir des résultats numériques précis. Ceci tient au fait que les fonctions qu'elle manipule sont suffisamment complexes. Nous verrons aussi que la diversité des architectures possibles conduit à une analyse originale des phénomènes étudiés. Les résultats recherchés ne sont plus la précision mais la connaissance supplémentaire apportée.

Dans le paragraphe suivant les deux sous-problèmes que sont la détermination de la vitesse et de la direction du vent à partir des données scattérométriques seront étudiés séparément. Pour chacun d'entre eux, nous déterminerons la fonction de transfert la plus précise possible et nous mettrons en valeur la méthodologie utilisée pour permettre l'acquisition du maximum d'informations possibles sur le phénomène.

9.3.1. Rappel sur l'approximation réelle par MLP

Un réseau MLP est caractérisé par son architecture et la valeur de ses poids. Ceci définit une famille F de fonctions fortement non-linéaires (voir §4.1.1).

L'approximation d'une fonction de transfert multidimensionnelle $T(x)$ ($\mathbb{R}^n \rightarrow \mathbb{R}$) avec un MLP peut être obtenue en utilisant des architectures spécifiques (donc des familles F particulières) dont le premier niveau comporte n automates et le dernier 1. Le nombre de couches et le nombre d'automates dans chaque couche est fixé en fonction de la difficulté de la fonction T à approximer. Une fois l'architecture A choisie, la meilleure approximation de T au sens introduit dans le chapitre 4 doit être déterminée dans la famille F_A ainsi déterminée. Une fois la phase de minimisation achevée, calculer $T(x)$ signifie forcer les états de la couche d'entrée aux valeurs des vecteurs x et calculer les états des automates des couches cachées

et de la couche de sortie. Si la famille F_A est assez riche, la valeur de sortie calculée doit s'approcher "au mieux" à la fonction $T(\mathbf{x})$.

Les résultats théoriques que nous avons rappelé au chapitre 4 montrent que sous certaines hypothèses une approximation à epsilon près de T peut être trouvée à l'intérieur d'une famille F_A qu'il convient de déterminer. En effet, si la théorie prouve l'existence du MLP qui peut approximer, elle n'explique pas comment le trouver. Toutes les expériences menées pendant cette thèse permettent de voir que pour des phénomènes physiques réels dont les variations ne sont pas trop brusques l'architecture du MLP est de petite taille à partir du moment où le problème à traiter a été découpé de façon modulaire et ne comporte pas trop de variables à prendre en compte en entrée.

La phase d'apprentissage permet de déterminer la matrice de poids \mathbf{W} , donc de choisir dans F_A la fonction $F(\cdot, \mathbf{W})$ qui approxime le mieux T . Ainsi $T(\mathbf{x}) \approx F(\mathbf{x}, \mathbf{W})$ pour tout \mathbf{x} où T est défini. Pour ceci nous déterminons la fonction $F(\cdot, \mathbf{W})$ qui minimise la fonction de coût :

$$C(m, \mathbf{W}) = \sum_{k=1}^m \left\| F(\mathbf{x}_k, \mathbf{W}) - T(\mathbf{x}_k) \right\|^2 \quad (9.1)$$

où la somme est prise sur l'ensemble des m formes d'apprentissage. Il s'agit ici de l'ensemble de valeurs possibles pour la fonction $T(\mathbf{x})$, donc de couples $(\mathbf{x}, T(\mathbf{x}))$.

Nous donnons maintenant les différents résultats numériques obtenus pour déterminer les deux fonctions de coût qui nous intéressent. L'objectif de cette thèse étant de présenter une méthodologie d'utilisation des réseaux pour la construction de machine neuronales, nous mettrons l'accent sur les différentes facilités qu'ils fournissent quand on traite de phénomènes physiques.

9.3.2. Résultats numériques

Détermination de la vitesse du Vent

La base de données satellitaires utilisée lors des expériences suivantes est composée de 10 traces de données, chaque trace ayant des paramètres qui la différencient des autres. L'angle d'incidence étant différent pour chaque trace, on entraîne un réseau par trace. La trace choisie pour la plupart de nos essais est celle du milieu, la trace n° 5 ($\alpha = 31,8^\circ$). Cependant, des tests ont été réalisés dans une trace proche à la trajectoire du satellite, la n° 2 ($\alpha = 21,7^\circ$), et par symétrie par rapport à la 5-ème trace, dans la n° 8 ($\alpha = 39,2^\circ$), plus éloignée que celle du milieu. Les résultats présentés indiqueront à chaque fois la trace pour lesquels les expériences ont été effectuées. Les performances obtenues dépendent de la position de la trace, les traces extrêmes présentant des plus importantes difficultés. D'autre part, tous les réseaux utilisés sont des MLP totalement connectés, chaque MLP sera donc repéré par son nombre d'automates sur chacune des couches : A-3×2×1 représente un réseau avec une couche cachée de deux neurones, trois neurones en entrées et un en sortie.

Une approche simple

Une première expérience va consister à rechercher une architecture minimale permettant de traiter les signaux reçus le long d'une trajectoire. Il s'agit donc maintenant d'appliquer les résultats théoriques et d'approximer directement la fonction de transfert $T_1 : \mathbb{R}^3 \rightarrow \mathbb{R}$ telle que $(x_1, x_2, x_3) \mapsto v$. Nous chercherons donc à déterminer la plus petite architecture qui en fournisse des résultats satisfaisants. Les performances obtenues à l'aide de cette architecture fourniront une borne inférieure permettant de juger de la qualité des améliorations qui seront apportées par la suite. Le réseau retenu est très simple, il s'agit de A-3×2×1. Les trois entrées représentent les trois signaux (x_1, x_2, x_3) obtenus pour chaque point de la trace considérée, la sortie est un codage de la vitesse recherchée codée entre -1 et +1. De manière à prouver la dépendance extrême de l'apprentissage en fonction du traitement des données, plusieurs codages ont été testés. Nous donnons les performances obtenues pour trois d'entre eux. Bien entendu, de manière à rendre les expériences comparables, les ensembles d'apprentissage et de test restent les mêmes quelles que soient les expériences présentées. Par ailleurs nous signifions que l'ensemble des associations $(x, T_1(x) = v)$ permettant

d'effectuer les calculs représentent les mêmes points de la carte. Le vecteur x est bien entendu différent selon le codage utilisé ou bien la représentation choisie.

Données centrées et réduites par variable

Les entrées sont centrées indépendamment par variable autour de zéro. Autrement-dit la moyenne de chacune des variables d'entrée (x_1, x_2, x_3) est ramenée à zéro.

L'apprentissage est rapide. Le système arrive très rapidement, en 2 ou 3 passages de la base d'apprentissage, à des performances optimales. Poursuivre l'apprentissage ne permet pas d'augmenter les performances; en 200 passages de la base d'apprentissage les performances du système n'augmentent que de 2 %.

Nous présentons dans la Table 9.1 les résultats obtenus en fonction de la précision recherchée. Une mesure prédite est considérée comme exacte pour une précision r si elle diffère de la réponse recherchée de $\pm r$ m/s. Les spécifications demandées par les organismes qui étudient ce problème sont de 2 m/s. Pour montrer la précision des MLP nous avons également étudié le cas $r=1$. Les performances représentent le pourcentage de l'ensemble pour lequel les réponses sont exactes.

Les performances sont meilleures sur l'ensemble de test car celui-ci est d'une certaine manière plus facile que l'ensemble d'apprentissage. En effet, dans ce dernier, toutes les directions et toutes les vitesses, même celles qui sont très rares ou plus complexes, ont été représentées de manière égale. Dans l'ensemble de test, 9 cartes (5041 vecteurs) ont été découpées indépendamment de ces valeurs.

Trace n° 5 :	précision de 1 m/s	précision de 2 m/s
performances app.	67,6 %	92,5 %
test	77,9 %	97,3 %

Table 9.1 : Performances de la vitesse pour des données réduites par variable.

Données brutes

Une deuxième série d'expériences a consisté à utiliser le même réseau avec maintenant les données brutes telles qu'elles sont originellement, nous ne faisons pas ni centrage ni réduction. L'apprentissage est plus lent. Cependant après 20 passages de la base d'apprentissage les performances sont nettement meilleures.

Trace n° 5 :	précision de 1 m/s	précision de 2 m/s
performances app.	72,9 %	100 %
test	86,6 %	100 %

Table 9.2 : Performances de la vitesse pour des données brutes.

Cette amélioration prouve que lors de l'utilisation de variables corrélées, il importe de conserver "au mieux" les rapports existants entre ces variables. L'utilisation des données brutes simplement recentrées pour permettre une bonne utilisation des sigmoïdes est préférable à une manipulation comme le centrage et la réduction qui changent de façon trop importante les corrélations existantes. Dans le problème traité, les variables sont liées physiquement : il s'agit de mesures prises à des angles différentes sur un même point. Modifier ce rapport équivaut à altérer la géométrie du problème.

Nous avons trouvé expérimentalement pour des réseaux à deux couches cachées, à un nombre égal de paramètres, une plus grande capacité à résoudre des problèmes d'une certaine complexité que pour des réseaux à une seule couche cachée. L'utilisation d'une architecture légèrement plus grosse A-3×2×2×1 permet une nouvelle et dernière amélioration de ces performances :

Trace n° 5 :	précision de 1 m/s	précision de 2 m/s
performances app.	78,9 %	100 %
test	88,1 %	100 %

Table 9.3 : Performances de la vitesse pour des données brutes et architecture à 2 couches cachées.

Données centrées et réduites globalement

L'entrée est centrée autour de zéro selon la moyenne de la globalité des données. La corrélation entre les variables d'entrée (x_1, x_2, x_3) est mieux conservée que dans le cas de réduction par variable, tout en assurant des valeurs d'entrée plus aisément utilisables avec la fonction sigmoïde utilisée.

L'apprentissage n'est pas si rapide que dans le premier cas mais il est plus direct que dans le cas de données brutes. De plus, on constate une amélioration dans la précision des résultats à architectures égales (A-3×2×1).

Trace n° 5 :		Codage par variable	Données Brutes	Codage Global
performances à une précision de 1 m/s	app.	67,6 %	72,9 %	80,1 %
	test	77,9 %	86,6 %	88,5 %
performances à une précision de 2 m/s	app.	92,5 %	100 %	99,8 %
	test	97,3 %	100 %	100 %

Table 9.4 : Comparaisons des performances de la vitesse pour des données codées différemment L'architecture des trois réseaux est identique : A-3×2×1.

La performance de 88,5 % avec une précision de 1 m/s obtenue sur la base de test sera donc retenue pour la suite comme la performance minimale possible pour la détermination de la vitesse du vent.

L'apport du contexte

L'approche naïve que nous avons présentée cherchait à approximer directement la fonction T_1 ; les techniques neuronales permettent de prendre aisément en compte d'autres facteurs. La première amélioration proposée consiste à utiliser les corrélations spatiales qui apparaissent de manière évidente au niveau des signaux. Pour un point donné, nous avons introduit la connaissance des mesures radar des points voisins et découpé des fenêtres sur la fauchée. Les fenêtres utilisées sont ici de taille 3×3. Cette étude doit être considérée comme une étude de faisabilité, nous n'avons donc pas cherché à optimiser la taille de cette fenêtre. La fonction de transfert est donc maintenant $T_1 : R^{27} \rightarrow R$, le vecteur d'entrée étant composé de 9×3 éléments correspondant aux valeurs des triplets (x_1, x_2, x_3) du point considéré et de ses voisins pour chaque antenne (voir chapitre précédent Figure 8.4).

Le réseau solution est maintenant beaucoup plus important, l'augmentation du temps de calcul qui en résulte étant cependant largement compensée par une amélioration des performances de 4%. La précision requise pour l'approximation de T_1 est d'une importance capitale dans le traitement du problème complet de détermination des cartes de vent. En effet, les valeurs calculées à ce stade seront utilisées comme données par d'autres modules réseaux. L'enchaînement des modules de la "machine neuronale" pose le problème de la composition des erreurs de calcul. Une première étude indique que la sensibilité des modules enchaînés tels que nous les avons utilisés est grande, il convient donc d'obtenir une précision maximale pour les modules inférieurs constitués par ceux qui déterminent la vitesse du vent. Nous discuterons des améliorations possibles dans les perspectives d'avenir présentées à la fin de la thèse, notons dès à présent qu'il s'agit de rechercher une approche modulaire comme celle qui est présentée ici, mais optimisée de façon globale.

Nous donnons maintenant les performances obtenues pour quelques unes des expériences menées. Ces performances ne sont pas optimales et nous présenterons au §9.4 une nouvelle décomposition permettant une amélioration sensible des performances. La Table 9.5 permet de suivre l'amélioration des performances obtenue lors de la prise en compte du contexte dans le problème du calcul de la vitesse du vent.

Trace n° 5 :	sans contexte		avec contexte	
	réseau 3x2x1	réseau 3x2x2x1	réseau 27x15x1	réseau 27x25x25x1
performances app.	72,9 %	78,9 %	81,3 %	87,2 %
test	86,6 %	88,1 %	90,4 %	92,0 %

Table 9.5 : Apport du contexte au calcul de la vitesse Résultats sur données brutes à une précision de la vitesse de ± 1 m/s.

Cette première étude nous permet de montrer l'efficacité et la précision des méthodes neuronales dans l'approximation de fonctions de transfert complexes. Notre but est maintenant de démontrer qu'il est possible d'utiliser une telle technique à d'autres fins, et ainsi mettre à la disposition des physiciens une méthodologie qui permet l'étude de la complexité des phénomènes étudiés.

Une étude des performances obtenues en fonction de la vitesse du vent par le MLP utilisé permet de comprendre que la fonction de transfert recherchée présente un changement important autour de 12 m/s. Pour cette étude, nous avons généré initialement une base de données composée des échantillons représentant toutes les vitesses (de 4 à 20 m/s). Par la suite, nous avons généré deux autres bases de données, une avec les données de basse vitesse (de 4 à 12 m/s) et l'autre avec les fortes vitesses (de 12 à 20 m/s). Les tests de performances sont affichés dans la Figure 9.1.

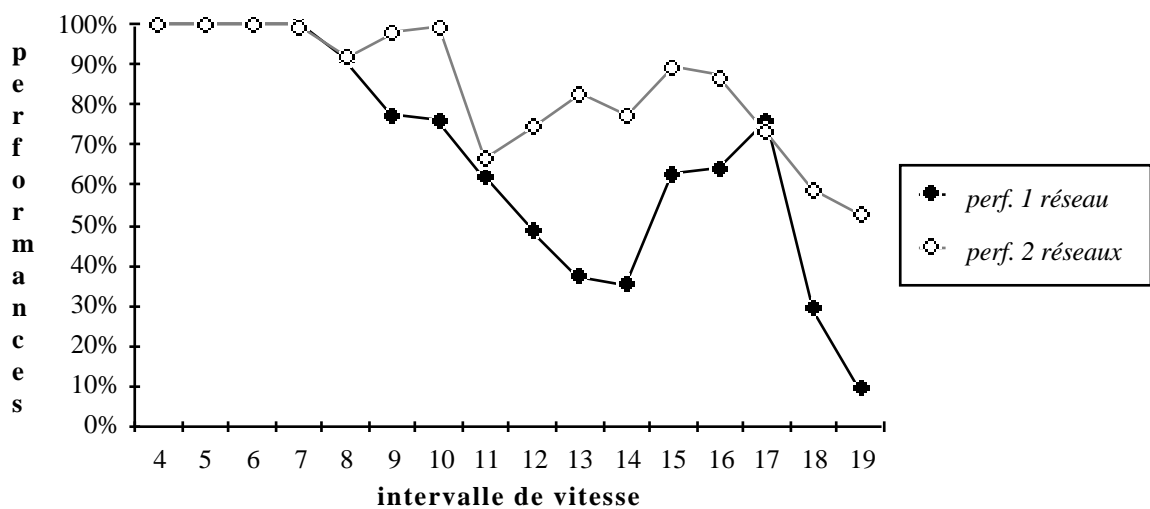


Figure 9.1 : Performances de la vitesse du vent par intervalle de vitesse. Comparaison des performances entre un réseau entraîné sur tout l'intervalle de vitesses et deux réseaux spécialisés en basses et hautes vitesses respectivement.

Cette valeur frontière indique que la nature de la fonction est différente selon que la vitesse considérée est petite ou grande. Nous chercherons par la suite à utiliser au mieux une telle information et à traiter séparément ces deux cas. Nous regarderons en particulier quelle amélioration peut être obtenue si l'on recherche une approximation par morceaux de T_1 .

9.4. Utilisation des MLP pour l'analyse des fonctions de transfert

Il arrive qu'un phénomène physique ne soit connu qu'à travers un ensemble de mesures et qu'il soit difficile de proposer une forme simple à la

fonction de transfert sous-jacente. Pour de tels phénomènes, toute indication comme l'estimation de l'adéquation de certaines variables ou de la "complexité" de la fonction de transfert est fondamentale dans la connaissance de ceux-ci.

Dans ce paragraphe, nous exposons les expériences réalisées en utilisant la méthodologie développée au chapitre 4 qui permet de juger de la "complexité" des fonctions de transfert : il devient possible de savoir si le phénomène réel doit être représenté par une fonction simple ou multivoque.

Nous traiterons les deux problèmes qui nous concernent : la détermination de la vitesse et de la direction du vent. Nous ne présentons ici que les expériences donnant des performances optimales. Celles-ci ont été obtenues à l'aide des architectures représentées dans la Figure 9.2.

Notons que tous les réseaux expérimentés en classification ont été entraînés en utilisant la métrique pondérée et en interprétant les courbes d'activation obtenues en sortie du réseau. Cette interprétation des sorties du réseau permet de proposer plusieurs valeurs possibles pour la fonction $T(x)$ à approximer et de leur associer des coefficients de vraisemblance où, si nécessaire, des probabilités. Lorsque la fonction sous-jacente est multi-valuée, la détermination d'une valeur unique pour $T(x)$ est en effet inadéquate. L'ajout de bruit dû à l'utilisation d'instruments de mesures rend les ambiguïtés impossibles à lever, les différentes distributions se recouvrant. Il est donc le plus souvent impossible de déterminer avec certitude la valeur exacte de la fonction. L'utilisation de MLP pris comme classifieur, exposée au chapitre 4, fournit une méthode probabiliste pour traiter ce problème. D'autre part, chaque signal est associé à plusieurs solutions et l'étude du nombre moyen de solutions proposées par le MLP permet une estimation des points doubles, voire triples, de la fonction de transfert recherchée. Ce mode d'étude est intéressant pour la connaissance du phénomène lui-même et pour la mise au point des différents traitements qui doivent être effectués par la suite pour automatiser le décodage des signaux.

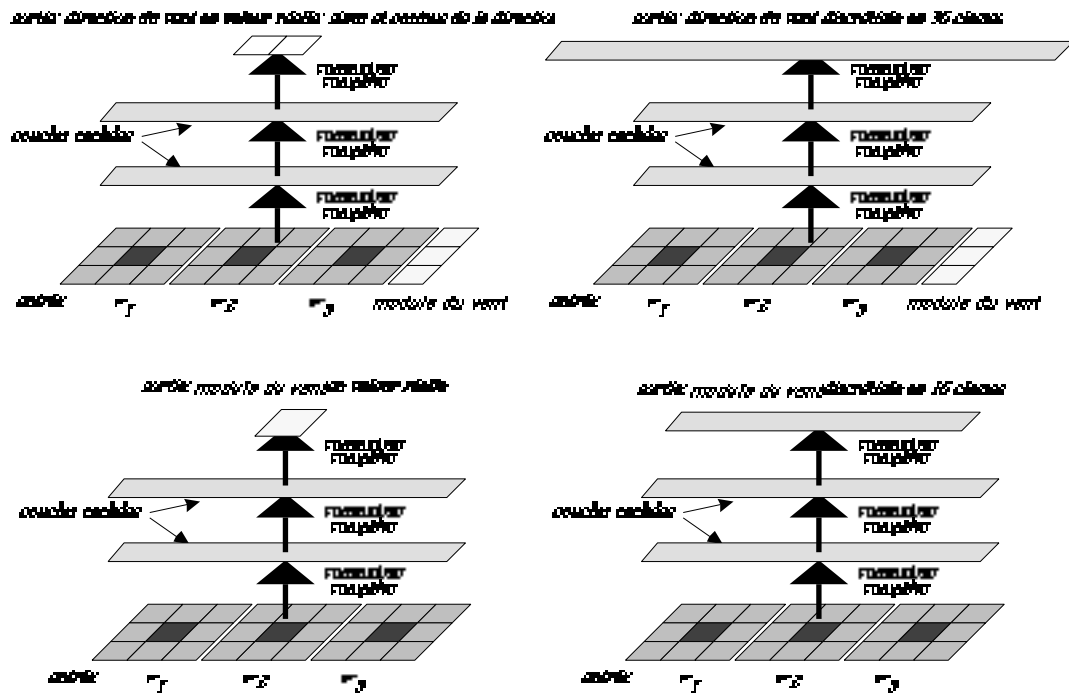


Figure 9.2 : architectures qui montrent la différence entre les réseaux réels et les réseaux classifieurs. Pour le calcul de la vitesse réelle nous avons retenu un réseau à sortie réelle (architecture A-27×25×25×1) tandis que pour la direction nous utilisons une sortie discrète (architecture A-30×25×25×36), celle-ci contenant plus d'information que la sortie réelle. Nous présentons aussi dans la figure le réseau à architecture discrète testé pour le calcul de la vitesse (A-27×25×25×16) ainsi que le réseau réel testé pour la détermination de la direction (A-30×25×25×2).

Détermination de la vitesse du vent

Dans les expériences que nous présentons, nous voulons montrer également que l'approche discrète permet d'obtenir une bonne précision numérique. Pour cette raison, nous comparerons les différentes performances obtenues par le classifieur à celles obtenues dans l'approche réelle présentée précédemment.

Nous avons vu que la fonction T_1 que nous étudions présentait une discontinuité, aussi deux séries d'expériences ont été réalisées :

- la première approxime la fonction dans son ensemble pour des vents variant de 4 m/s à 20 m/s.
- la seconde approxime la fonction par morceaux en réalisant des réseaux séparés, l'un pour les petites vitesses (4 à 12 m/s) et l'autre pour les

grandes vitesses (12 à 20 m/s). Les performances totales des deux méthodes ont alors été testées en pondérant les performances de chaque réseau par les effectifs des deux classes ayant servi à l'apprentissage et au test (petits vents et grands vents).

Trois ensembles d'apprentissage ont permis l'entraînement des 3 réseaux : vitesse entre 4 et 20 m/s, 4 et 12 m/s, 12 et 20 m/s. Ces ensembles sont constitués respectivement de 6616, 6176, 440 exemples; dans les deux derniers cas la répartition est homogène à raison de 772 et 55 vecteurs par vitesse de vent, nombres imposés par le nombre d'exemples existant sur les 22 cartes d'apprentissage dont nous disposons. Pour le premier réseau classifieur les valeurs de $T_1(x)$ ont été discrétisées en 16 valeurs ($p=16$), chacune représentant un intervalle de 1 m/s, et pour chacun des autres réseaux 8 intervalles de 1 m/s ont été utilisés ($p=8$). L'ensemble d'apprentissage pour la tâche globale est donc de 6616 vecteurs.

La Table 9.6 donne les performances en pourcentage pour l'apprentissage et pour la généralisation obtenues pour l'approximation en réel de la fonction pour les trois réseaux pris séparément et pour la réunion des deux réseaux spécialisés (basse vitesse, haute vitesse). La Table 9.7 donne les performances correspondantes en classification.

Les résultats montrent que l'approximation par classification entraîne une légère perte dans la précision des prédictions. Cependant, pour une précision de 2 m/s, qui est celle prévue par les spécifications officielles, les performances sont semblables. L'approche par classification constitue néanmoins une aide intéressante dans l'étude d'un phénomène inconnu. Dans le cas de la détermination de la vitesse du vent, il est possible de voir qu'il n'existe pas d'ambiguïtés intrinsèques (Figure 9.3), les courbes d'activation obtenues à la sortie du MLP présentent en moyenne un pic significatif. Ces courbes seront différentes dans le cas de la détermination de la direction et elles permettront de conclure à l'existence de points doubles sur la courbe (Figure 9.4).

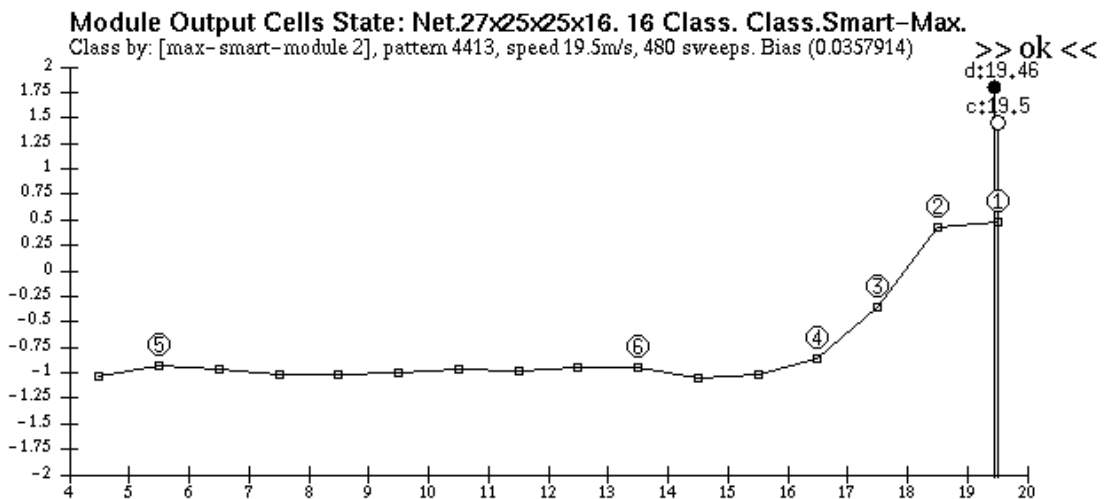
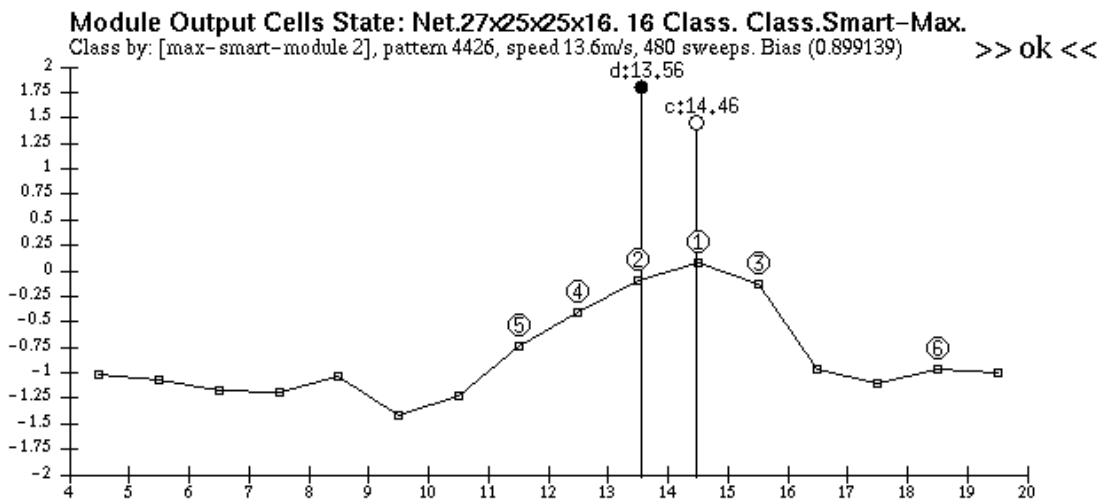
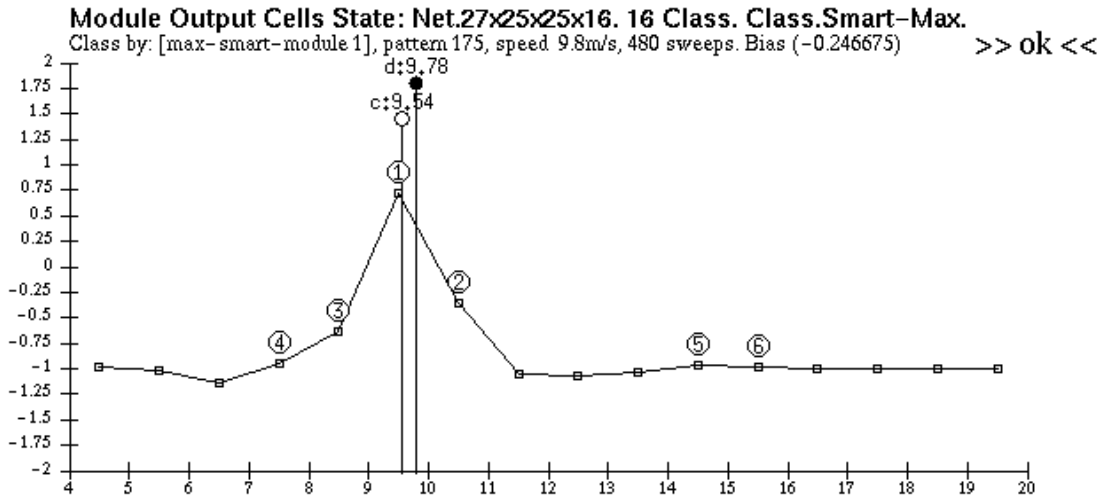


Figure 9.3 : Approximation par classification de la vitesse du vent. Dans la plupart des cas un seul pic bien distinct est représenté par les états des neurones de sortie.

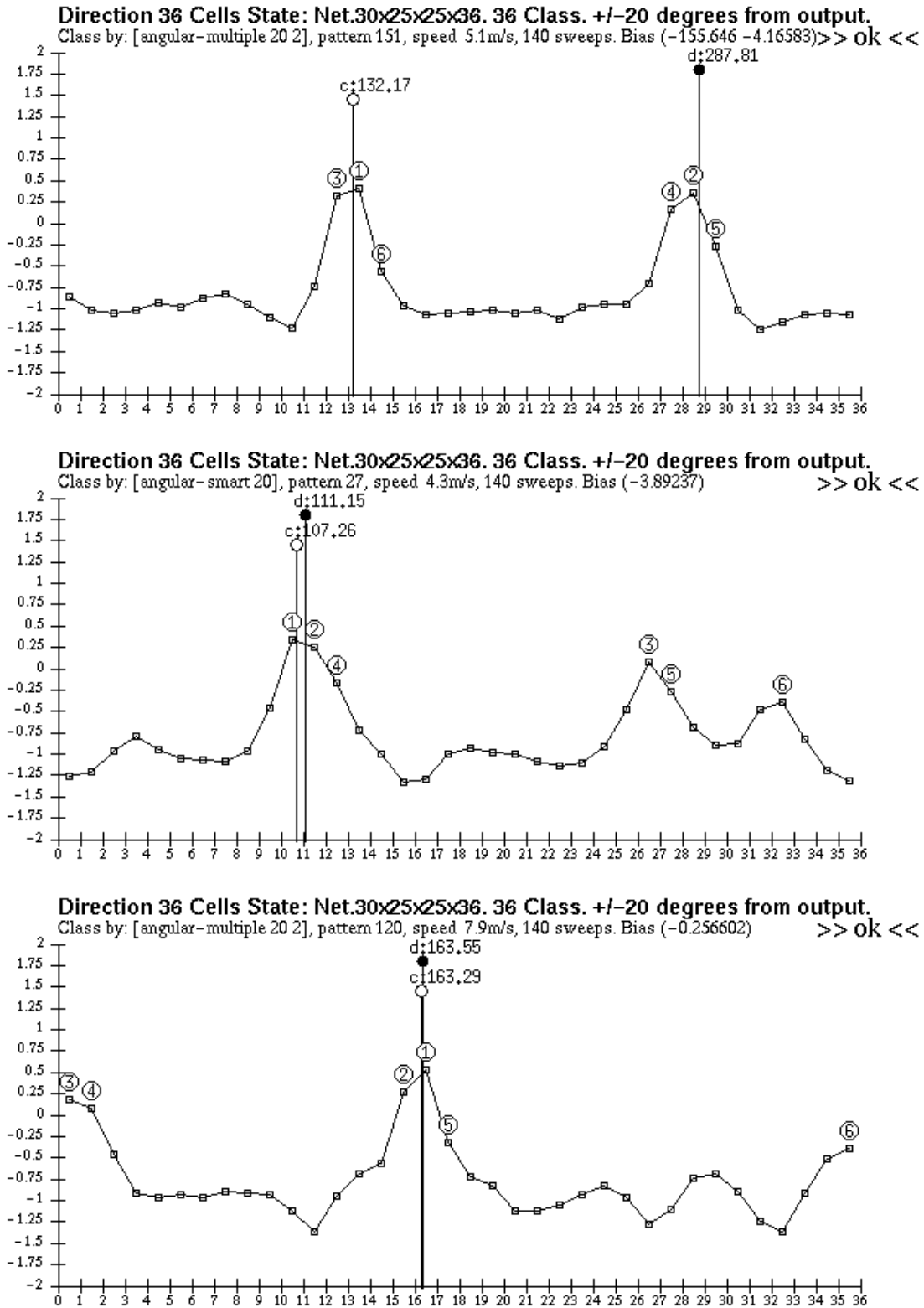


Figure 9.4 : Approximation par classification de la direction du vent. Deux pics et même plus sont représentés par les états des neurones de sortie.

Nous présentons dans les Tables 9.6 et 9.7 les résultats obtenus en fonction de la précision recherchée.

Bande 5 — Vitesse du Vent — sortie réelle						
3 réseaux, apprentissage sur 4 -12 m/s, 12 -20 m/s et 4 -20 m/s						
précision	précision	ens.	réseau 4 -12 m/s	réseau 12 -20 m/s	ens. réseaux 4-12+12-20 m/s	réseau 4 -20 m/s
±1 m/s	performances	app.	99,3%	98,9%	99,3%	98,0 %
		test	99,5%	91,1%	98,4%	97,4 %
±2 m/s	performances	app.	100,0%	100,0%	100,0%	100 %
		test	100,0%	99,7%	100,0%	99,8 %
biais moyen		app.	0,05 m/s	-0,01 m/s	0,05 m/s	0,04 m/s
		test	0,06 m/s	-0,11 m/s	0,04 m/s	0,04 m/s
Erreur RMS moyen		app.	0,18 m/s	0,19 m/s	0,18 m/s	0,27 m/s
		test	0,19 m/s	0,42 m/s	0,22 m/s	0,29 m/s
architecture			27x25x25x1	27x25x25x1	—//—	27x25x25x1
age			560 sweeps 3 458 560 it.	1400 sweeps 616 000 it.	—//—	450 sweeps 2 977 200 it.
cas			3-05Lb	3-05Hb	—//—	3-05_cra2
nb.patterns		app.	6176 772 exo/class	440 55 exo/class	6176+440=66 16	6616
		test	4348	675	4348+675=50 23	5023

Table 9.6 : Performances de la vitesse du vent, sortie réelle.

Bande 5 — Vitesse du Vent — classification						
3 réseaux, apprentissage sur 4-12 m/s, 12-20 m/s et 4-20 m/s						
précision	précision	ens.	réseau 4-12 m/s	réseau 12-20 m/s	ens. réseaux 4-12+12-20 m/s	réseau 4-20 m/s
±1 m/s	performances	app.	92,7 %	98,6 %	93,1%	94,5 %
	± 1 m/s	test	95,5 %	90,7 %	94,9%	91,3 %
±2 m/s	performances	app.	100,0 %	98,6 %	99,9%	98,6 %
	± 2 m/s	test	100,0 %	95,9 %	99,4%	97,3 %
biais moyen		app.	-0,09 m/s	-0,01 m/s	-0,08 m/s	0,04 m/s
		test	-0,05 m/s	0,13 m/s	-0,03 m/s	0,02 m/s
Erreur RMS moyen		app.	0,44 m/s	-0,01 m/s	0,41 m/s	0,34 m/s
		test	0,41 m/s	0,13 m/s	0,37 m/s	0,42 m/s
architecture			27x25x25x8	27x25x25x8	—//—	27x25x25x16
age			45 sweeps 277 920 it.	490 sweeps 215 600 it.	—//—	480 sweeps 3 175 680 it.
cas			3-05Lclp_clas	3-05Hclp_clas	—//—	3-05clp_class
nb.patterns		app.	6176 772 exo./class	440 55 exo./class	6176+440=6616	6616
		test	4348	675	4348+675=5023	5023

Table 9.7 : Performances de la vitesse du vent, sortie discrète.

Le biais moyen s'exprime en m/s pour la vitesse et en *degrés* pour la direction. Il est la moyenne de la différence entre la réponse calculée par le réseau et la réponse désirée. Pour la vitesse il est calculé par : $bias_k = ||v_c|| - ||v_d||$, tandis que pour la direction il est calculé par : $bias_k = (v_c) - (v_d)$, où v_d est le vecteur vent exact – désiré – et v_c le vecteur prédit – calculé – par le réseau. De même, $||u||$ et (u) représentent respectivement le module et l'angle du vecteur u ; notons que le biais peut être négatif, il nous donne la tendance générale du système.

$$bias = \frac{1}{m} \sum_{k=1}^m bias_k \quad (9.2)$$

L'erreur e_{rms} , en m/s , est la moyenne du module du vecteur de l'erreur effectué par le réseau. Définissons le vecteur erreur par : $\bar{e} = v_c - v_d$, et par $||\bar{e}||$ sa norme. L'erreur e_{rms} est la moyenne de cette norme sur les m exemples et est donné par l'expression suivante :

$$e_{rms} = \frac{1}{m} \sum_{k=1}^m ||\bar{e}|| \quad (9.3)$$

Détermination de la direction du vent

Les mêmes expériences ont été effectuées pour étudier le second problème : la détermination de la direction du vent. Il s'agit maintenant d'approximer la fonction de transfert suivante :

$$T_2: \mathbb{R}^3 \rightarrow \mathbb{R} \quad \text{telle que } (x_1, x_2, x_3)$$

Nous avons vu dans la présentation du problème qu'à module de vent constant, il s'agit d'une courbe de Lissajous qui comporte un grand nombre de points doubles. Les ensembles d'apprentissage ont maintenant été équilibrés par direction. Chaque intervalle de 10° dans la direction est représenté par 164 exemples. Nous donnons à chaque expérience les meilleures architectures ayant permis d'obtenir les meilleures performances. Comme dans la détermination de la vitesse, une première série d'expériences permet d'évaluer l'apport du contexte. Nous donnons ces premières performances qui déterminent une borne inférieure. L'approximation de T_2 étant recherchée en valeurs réelles, les sorties du réseau codent le sinus et le cosinus de l'angle recherché. Les architectures utilisées ont donc 3 cellules en entrée "les signaux issus de chaque antenne", et deux en sortie. Les meilleurs résultats sont obtenus à partir de l'architecture à deux couches cachées A-3×15×15×2. Pour estimer les performances nous utilisons les précisions suivantes :

- $\pm 20^\circ$ qui est la précision désirée par les océanographes elle nous permettra de comparer nos résultats à ceux de différentes méthodes utilisées par ceux-ci pour calculer la direction du vent.
- $\pm 10^\circ$ pour juger de la précision que peuvent fournir en comparaison les réseaux.
- $\pm 30^\circ$ pour comprendre comment est approximée T_2 .

L'apprentissage a été effectué sur un ensemble comprenant des vitesses situées sur tout l'intervalle [4, 20 m/s]. Cependant, pour pouvoir étudier comme dans le premier problème le comportement des vents forts et des vents faibles, les tests sont effectués en divisant en deux les ensembles d'apprentissage et de test selon cette classification.

Bande 5 — Direction du Vent — sortie réelle							
SANS contexte et SANS module du vent à l'entrée — apprentissage sur 4-20 m/s							
		test sur 4 - 12 m/s		test sur 12- 20 m/s		test sur 4 - 20 m/s	
précision	ensemble	perf.	Err.RMS	perf.	Err.RMS	perf.	Err.RMS
±10°	app.	17,6 %	7,68 m/s	21,0 %	10,60 m/s	18,0 %	8,06 m/s
	test	20,2 %	7,27 m/s	20,3 %	9,55 m/s	20,2 %	7,58 m/s
±20°	app.	33,8 %	7,68 m/s	51,0 %	10,60 m/s	36,0 %	8,06 m/s
	test	38,4 %	7,27 m/s	58,8 %	9,55 m/s	41,2 %	7,58 m/s
±30°	app.	42,9 %	7,68 m/s	60,5 %	10,60 m/s	45,2 %	8,06 m/s
	test	47,4 %	7,27 m/s	68,1 %	9,55 m/s	50,2 %	7,58 m/s
nb.patterns	app.	5129		775		5904	
	test	4348		675		5023	
architecture		3x15x15x2					
age		310 sweeps 1 830 240 it.					
cas		3-05_NoENoM_15x15					

Table 9.8 : Performances de la direction du vent, sortie réelle sans contexte et sans module du vent à l'entrée .

La même démarche que pour la détermination de la vitesse amène à ajouter un contexte spatial. Nous considérerons une même fenêtre de 3×3 autour du signal. Les architectures utilisées ont donc maintenant 9×3 = 27 entrées et deux sorties. Les architectures sont maintenant A-27-25-25-2, les performances obtenues présentant la même amélioration des performances de 3% observées précédemment.

Bande 5 — Direction du Vent — sortie réelle							
AVEC contexte et SANS module du vent à l'entrée — apprentissage sur 4-20 m/s							
		test sur 4-12 m/s		test sur 12-20 m/s		test sur 4-20 m/s	
précision	ensemble	perf.	Err.RMS	perf.	Err.RMS	perf.	Err.RMS
±10°	app.	27,0 %	5,41 m/s	37,9 %	6,74 m/s	28,5 %	5,59 m/s
	test	24,1 %	5,98 m/s	35,7 %	6,98 m/s	25,7 %	6,11 m/s
±20°	app.	44,7 %	5,41 m/s	63,2 %	6,74 m/s	47,2 %	5,59 m/s
	test	41,2 %	5,98 m/s	62,7 %	6,98 m/s	44,1 %	6,11 m/s
±30°	app.	55,4 %	5,41 m/s	75,7 %	6,74 m/s	58,1 %	5,59 m/s
	test	51,2 %	5,98 m/s	72,9 %	6,98 m/s	54,1 %	6,11 m/s
nb.patterns	app.	5129		775		5904	
	test	4348		675		5023	
architecture		27x25x25x2					
age		360 sweeps 2 125 440 it.					
cas		3-05_sigma					

Table 9.9 : Performances de la direction du vent, sortie réelle avec contexte et sans module du vent à l'entrée.

Une seconde amélioration consiste maintenant à donner comme information supplémentaire la vitesse du vent. Une telle approche, qui est à considérer dans l'optique de la "machine neuronale" dont le premier module calcule la vitesse du vent, permet de simplifier l'espace dans lequel doit s'effectuer la recherche de la direction cherchée (voir Figure 6.6). De manière à rendre cette information assez importante, comparée à celles amenées par les 9×3 signaux (v_1, v_2, v_3) , la vitesse du vent est répétée 3 fois, une fois par antenne. Les entrées des réseaux considérés sont donc maintenant constituées par $9 \times 3 + 3 = 30$ neurones. Notons que dans ce chapitre nous cherchons à tester séparément chacun des sous problèmes mis en évidence et que la vitesse utilisée, qui correspond au signal mis en entrée, est une valeur exacte. Le problème de l'enchaînement des différents modules est traité dans le chapitre suivant.

Les mêmes expériences qui mettent en lumière l'importance de cette donnée supplémentaire figurent dans les tableaux de performances qui suivent. Nous donnons cette fois-ci pour illustrer la généralité de la démarche les résultats optimaux obtenus pour les bandes 2, 5 et 8 (Tables 9.10, 9.11 et 9.12). On observe dans ces tableaux une progression des performances qui tient au fait que la difficulté de la courbe étudiée croit avec l'angle d'incidence. L'architecture utilisée pour l'approximation de T_2 a été choisie constante de manière à illustrer ce phénomène, il s'agit de $A-30 \times 25 \times 25 \times 2$.

Bande 2 — Direction du Vent — sortie réelle							
apprentissage sur 4-20 m/s							
		test sur 4-12 m/s		test sur 12-20 m/s		test sur 4-20 m/s	
précision	ensemble	perf.	Err.RMS	perf.	Err.RMS	perf.	Err.RMS
±10°	app.	48,8 %	2,60 m/s	52,9 %	3,66 m/s	49,3 %	2,75 m/s
	test	48,0 %	2,98 m/s	41,7 %	4,63 m/s	47,1 %	3,20 m/s
±20°	app.	72,4 %	2,60 m/s	80,2 %	3,66 m/s	73,5 %	2,75 m/s
	test	70,2 %	2,98 m/s	73,1 %	4,63 m/s	70,6 %	3,20 m/s
±30°	app.	81,1 %	2,60 m/s	89,4 %	3,66 m/s	82,3 %	2,75 m/s
	test	78,1 %	2,98 m/s	85,6 %	4,63 m/s	79,1 %	3,20 m/s
nb.patterns	app.	5201		847		6048	
	test	4297		662		4959	
architecture	30x25x25x2						
age	440 sweeps 2 661 120 it.						
cas	3-02_eq						

Table 9.10 : Performances de la direction du vent, sortie réelle-Bande proche.

Bande 5 — Direction du Vent — sortie réelle							
apprentissage sur 4 - 20 m/s							
		test sur 4 - 12 m/s		test sur 12- 20 m/s		test sur 4 - 20 m/s	
précision	ensemble	perf.	Err.RMS	perf.	Err.RMS	perf.	Err.RMS
±10°	app.	53,0 %	2,29 m/s	52,3 %	3,63 m/s	52,9 %	2,46 m/s
	test	51,8 %	2,60 m/s	48,7 %	4,53 m/s	51,4 %	2,86 m/s
±20°	app.	75,2 %	2,29 m/s	78,6 %	3,63 m/s	75,6 %	2,46 m/s
	test	73,6 %	2,60 m/s	77,0 %	4,53 m/s	74,0 %	2,86 m/s
±30°	app.	84,3 %	2,29 m/s	89,3 %	3,63 m/s	84,9 %	2,46 m/s
	test	82,2 %	2,60 m/s	84,1 %	4,53 m/s	82,4 %	2,86 m/s
nb.patterns	app.	5129		775		5904	
	test	4348		675		5023	
architecture	30x25x25x2						
age	365 sweeps 2 154 960 it.						
cas	3-05_eq						

Table 9.11 : Performances de la direction du vent, sortie réelle-Bande du milieu.

Bande 8 — Direction du Vent — sortie réelle							
apprentissage sur 4-20 m/s							
		test sur 4-12 m/s		test sur 12-20 m/s		test sur 4-20 m/s	
précision	ensemble	perf.	Err.RMS	perf.	Err.RMS	perf.	Err.RMS
±10°	app.	60,0 %	1,66 m/s	69,7 %	2,37 m/s	61,3 %	1,76 m/s
	test	53,9 %	2,11 m/s	63,0 %	3,80 m/s	55,1 %	2,33 m/s
±20°	app.	82,5 %	1,66 m/s	90,5 %	2,37 m/s	83,6 %	1,76 m/s
	test	78,4 %	2,11 m/s	81,3 %	3,80 m/s	78,8 %	2,33 m/s
±30°	app.	89,5 %	1,66 m/s	95,8 %	2,37 m/s	90,3 %	1,76 m/s
	test	85,7 %	2,11 m/s	89,7 %	3,80 m/s	86,3 %	2,33 m/s
nb.patterns	app.	5029		803		5832	
	test	4365		678		5043	
architecture	30x25x25x2						
age	470 sweeps 2 741 040 it.						
cas	3-08_eq						

Table 9.12 : Performances de la direction du vent, sortie réelle-Bande extrême.

Les performances de 74% de bonnes directions sur la bande 5 constituent une performance optimale pour la courbe apprise; en effet, 72% des points de la courbe initiale utilisés dans ces simulations peuvent être déterminés sans ambiguïtés. On peut attribuer le nombre légèrement supérieur de points retrouvés à l'utilisation du contexte qui a permis de lever certaines ambiguïtés. L'optimisation de la taille du voisinage n'ayant pas été effectuée, il est possible que l'utilisation d'un voisinage plus large permette d'augmenter encore ces performances.

La progression des performances obtenues dans ces dernières expériences prouve la pertinence des informations prises en compte. Pour la suite des expériences, nous utiliserons toujours l'information spatiale sur un voisinage 3×3 et la connaissance de la vitesse du vent comme information supplémentaire.

La présence des points doubles de la courbe T_2 va nous permettre d'illustrer la puissance de la méthode proposée. Nous allons montrer qu'une telle méthodologie permet non seulement de faire apparaître la présence de points doubles, mais aussi de calculer les différentes valeurs possibles de manière optimale.

Recherche des points multiples

Dans ce paragraphe, nous allons, comme pour la détermination de la vitesse, mettre en oeuvre la technique de discrétisation pour approximer la fonction T_2 . Nous rappelons que cette fonction est prise sous une forme permettant d'inclure les corrélations spatiales et la connaissance de la vitesse : il s'agit donc d'une fonction différente de T_2 qui s'applique de $\mathbb{R}^{30} \rightarrow \mathbb{R}$; pour simplifier la compréhension du texte nous continuerons à la désigner par T_2 . De la même manière, nous continuerons à appeler T_2 la version discrète de cette fonction de $\mathbb{R}^{30} \rightarrow \mathbb{R}^p$ où p représente le nombre d'intervalles de discrétisation considérés. Dans les expériences que nous présentons, les intervalles considérés ont une largeur de 10° et $p = 36$. L'architecture testée est A-30×25×25×36 pour pouvoir comparer les performances obtenues à celles de 74% de bonnes directions avec $\pm 20^\circ$ de précision trouvées dans le paragraphe précédent. Pour permettre cette comparaison, les ensembles d'apprentissage et de test sont donc identiques à ceux des premières expériences et les performances sont données à la fois sur l'ensemble tout entier et séparément pour les hautes et les basses vitesses.

Bande 5 — Direction du Vent — classification							
apprentissage sur 4-20 m/s							
		test sur basse vitesse 4-12 m/s		test sur vitesse forte 12-20 m/s		test sur l'ensemble 4-20 m/s	
précision	ensemble	perf.	Err.RMS	perf.	Err.RMS	perf.	Err.RMS
±10° 1 pic	app.	62,8 %	5,17 m/s	76,0 %	7,01 m/s	64,5 %	5,41 m/s
	test	66,1 %	4,58 m/s	78,7 %	5,87 m/s	67,8 %	4,75 m/s
±10° 2 pics	app.	92,9 %	0,89 m/s	97,7 %	1,16 m/s	93,5 %	0,92 m/s
	test	91,5 %	0,93 m/s	94,8 %	1,68 m/s	91,9 %	1,03 m/s
±20° 1 pic	app.	66,1 %	5,17 m/s	76,6 %	7,01 m/s	67,5 %	5,41 m/s
	test	70,6 %	4,58 m/s	80,9 %	5,87 m/s	72,0 %	4,75 m/s
±20° 2 pics	app.	98,2 %	0,68 m/s	99,0 %	1,02 m/s	98,3 %	0,72 m/s
	test	98,4 %	0,66 m/s	98,2 %	1,35 m/s	98,4 %	0,75 m/s
±20° 3 pics	app.	99,6 %	0,52 m/s	100 %	0,78 m/s	99,7 %	0,55 m/s
	test	99,5 %	0,55 m/s	99,4 %	1,01 m/s	99,5 %	0,61 m/s
±20° 4 pics	app.	99,9 %	0,49 m/s	100 %	0,78 m/s	99,9 %	0,53 m/s
	test	99,9 %	0,52 m/s	99,7 %	0,92 m/s	99,9 %	0,57 m/s
nb.patterns	app.	5129		775		5904 164 exo./class	
	test	4348		675		5023	
architecture	30x25x25x36						
age	140 sweeps 826 560it.						
cas	3-05clp_eq						

Table 9.13 : Performances de la direction du vent, sortie discrète, réseau unique.

Pour un vecteur d'entrée donné, l'ensemble des pics significatifs de la courbe d'activation des sorties sont classés par ordre décroissant. Nous appellerons n_j le nombre de solutions probables que l'on considère. Les Tables 9.13 et 9.14 donnent les performances obtenues en fonction de n_j , c'est à dire qu'un point est considéré comme bien classé si la bonne solution se trouve avec une précision de $\pm\rho$ dans les n_j solutions proposées par l'algorithme. Nous rappelons que la précision requise est de $\pm 20^\circ$, la Table 9.13 montre donc que $n_j = 4$ permet de retrouver toutes les valeurs désirées. Les algorithmes classiquement utilisés pour résoudre ce problème nécessitent de calculer jusqu'à 6 solutions. Ce nombre n_j est d'une grande importance dans la phase de désambiguation au cours de laquelle il faut choisir parmi ces solutions celle qui est la bonne.

Bande 5 — Direction du Vent — classification — 2 réseaux							
apprentissage séparé sur 4-12 m/s et sur 12-20 m/s							
		Basse Vitesse app&test 4-12 m/s		Vitesse Forte app&test 12-20 m/s		tests calculés sur l'ensemble	
précision	ensemble	perf.	Err.RMS	perf.	Err.RMS	perf.	Err.RMS
±10° 1 pic	app.	62,7 %	5,50 m/s	86,9 %	3,86 m/s	66,9%	5,21 m/s
	test	69,3 %	4,63 m/s	83,9 %	4,33 m/s	71,3%	4,59 m/s
±10° 2 pics	app.	97,0 %	0,61 m/s	99,7 %	0,56 m/s	97,5%	0,60 m/s
	test	97,5 %	0,51 m/s	96,1 %	1,20 m/s	97,3%	0,60 m/s
±20° 1 pic	app.	63,9 %	5,50 m/s	87,1 %	3,86 m/s	68,0%	5,21 m/s
	test	70,6 %	4,63 m/s	85,8 %	4,33 m/s	72,6%	4,59 m/s
±20° 2 pics	app.	98,9 %	0,53 m/s	99,9 %	0,56 m/s	99,1%	0,54 m/s
	test	99,4 %	0,45 m/s	98,2 %	1,14 m/s	99,2%	0,54 m/s
±20° 3 pics	app.	100 %	0,40 m/s	100 %	0,53 m/s	100,0%	0,42 m/s
	test	99,9 %	0,39 m/s	99,3 %	0,90 m/s	99,8%	0,46 m/s
±20° 4 pics	app.	100 %	0,40 m/s	100 %	0,53 m/s	100,0%	0,42 m/s
	test	100 %	0,38 m/s	100 %	0,84 m/s	100,0%	0,46 m/s
nb.patterns	app.	5184 144 exo./class		1098 30 exo./class		6282	
	test	4348		675		5023	
architecture		30x25x25x36		30x25x25x36		-	
age		120 sw. 622 080 it.		1900 sw. 2 086 200 it.		-	
cas		3-05Lclp		3-05Hclp		-	

Table 9.14 : Performances de la direction du vent, sortie discrète, deux réseaux.

Une expérience supplémentaire parallèle à celle effectuée pour l'approximation de T_1 montre que les performances peuvent encore être affinées en considérant 2 réseaux séparés, le premier dédié au calcul de la direction des cas correspondant aux basses vitesses et le second aux hautes vitesses. Les performances moins bonnes obtenues par le réseau hautes vitesses pourraient être facilement améliorées si la base d'apprentissage était augmentée, ceci n'était pas possible dans le cadre de cette thèse puisque nous ne disposions pas de ces données.

La Figure 9.5 analyse les erreurs effectuées par le réseau, elle montre la distribution sur l'ensemble d'apprentissage de la valeur absolue de l'erreur qui existe entre la vraie valeur de $T_2(x)$ et la réponse la plus probable calculée par le réseau. Cette figure montre qu'avec une précision de $\pm 20^\circ$, 72% des réponses sont bien retrouvées par le premier pic de la courbe de sortie. D'autre part, les erreurs effectuées, quand la première réponse est erronée, se situent autour de $\pm 150^\circ$. Cette valeur correspond exactement aux

ambiguïtés de la courbe théorique. Comme la Table 9.14 montre clairement que $n_j = 2$ suffit pour obtenir 99,2% de bonnes réponses, on peut en général n'utiliser que les deux valeurs plus probables, qui diffèrent de $\pm 150^\circ$, ce qui correspond également aux caractéristiques du modèle.

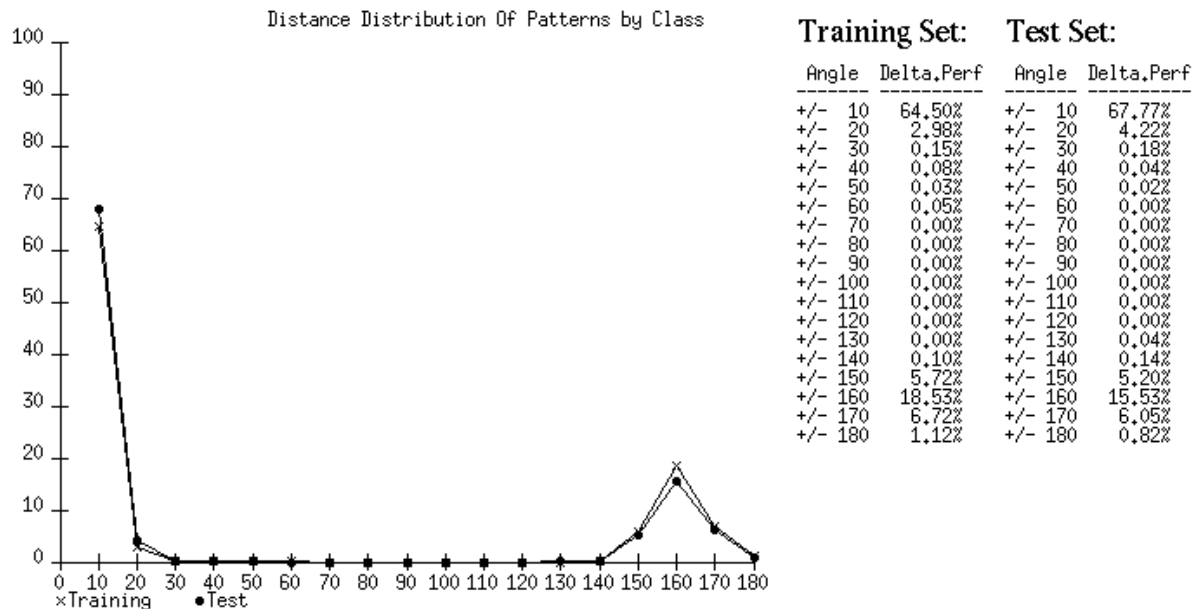


Figure 9.5 : Distribution des erreurs effectuées dans la détermination de la direction du vent. Le point dans la colonne a indique le pourcentage de points dont la réponse calculée par le réseau est entre a et $a-10^\circ$ degrés au maximum de la réponse exacte. Clairement la courbe démontre que le réseau effectue des erreurs d'angle en moyenne de $\pm 150^\circ$ à $\pm 160^\circ$.

Ce second exemple montre avec force les qualités qui peuvent être attendues par une approche par classification dans l'approximation de fonctions multivoques. En l'absence d'information *a priori* sur le comportement de la fonction de transfert que l'on étudie, l'approche par classification donne des renseignements sur la nature physique et mathématique des phénomènes étudiés, renseignements que ne procurent pas l'approche réelle. Ces informations peuvent être précieuses et servir à augmenter la précision de la détermination du phénomène en orientant soit la stratégie de modélisation, soit même la stratégie expérimentale. Dans l'exemple que nous venons de résoudre, on pourra utiliser les deux directions ambiguës avec leur coefficient de vraisemblance dans une autre phase permettant de lever l'ambiguïté [Badran et al. 89].

9.5. Interprétation théorique des deux exemples

Nous avons illustré sur deux problèmes réels la précision qu'il était possible d'obtenir dans l'approximation d'une fonction de transfert lorsqu'on l'exprimait sous forme discrétisée. La procédure d'interprétation des valeurs de sortie nous a permis d'obtenir les valeurs réelles de la fonction. De plus, cette approche permet l'investigation de la fonction elle-même et met en évidence sa nature simple ou multivoque, par exemple elle donne les différentes ambiguïtés intrinsèques liées au problème. D'autre part les propriétés statistiques des MLP leur permettent une prise en compte du bruit réel.

Nous donnons les justifications théoriques qui expliquent les différents résultats mis en évidence dans le courant de ce chapitre. Les expériences relatées ont été effectuées sur des données simulées à partir de modèles numériques, nous montrerons également que les mêmes justifications théoriques s'appliquent à l'utilisation de cette technique pour traiter de données réelles.

9.5.1. Précision de l'Approximation

Nous traiterons surtout dans ce paragraphe le cas de la détermination de la direction du vent, les mêmes démonstrations peuvent être appliquées pour la détermination de la vitesse. Nous employerons les notations suivantes :

(x_1, x_2, x_3) représente un point quelconque de l'espace à trois dimensions
 $\mathbb{R}^3 = \left\{ (x_1, x_2, x_3) \mid x_i \in \mathbb{R} \right\}$,

v : vitesse du vent,

θ : direction du vent –angle que fait le vent avec le nord géographique–,

$\hat{p}_i(\theta, v)$: $(\hat{p}_1(\theta, v), \hat{p}_2(\theta, v), \hat{p}_3(\theta, v))$ un point de la surface correspondant au vent à une direction et vitesse données (θ, v) ⁴.

⁴ La valeur de \hat{p}_i dépende aussi de θ , l'angle d'incidence du signal, qui n'a pas été prise en compte dans ces estimations car nous travaillons à θ constant.

C_j : $j = 1, \dots, 36$ les 36 intervalles de 10° correspondant aux vents dont l'angle est tel que $10(j-1) < 10j$.

La base d'apprentissage a été générée de manière telle que toutes les vitesses et toutes les directions soient représentées uniformément. Nous avons donc les hypothèses suivantes :

- Les couples (θ, v) ont été générés d'une façon aléatoire et selon la loi uniforme sur le rectangle $\theta \in [0, 360]$ $v \in [4, 20]$. La fonction densité du couple (θ, v) est donc :

$$g(\theta, v) = \frac{1}{360} \frac{1}{16} \mathbb{1} \quad (9.4)$$

où $\mathbb{1}$ est la fonction caractéristique du rectangle .

- Si $\hat{\theta}(\theta, v) : (\hat{\theta}_1(\theta, v), \hat{\theta}_2(\theta, v), \hat{\theta}_3(\theta, v))$ est le triplet "réel" -calculé par la formule de Long- et $\theta : (\theta_1, \theta_2, \theta_3)$ est le triplet généré, alors $(\theta_i - \hat{\theta}_i(\theta, v))$ suit une loi normale $N(0, s)$. Ainsi :

$$p(\theta_i | \theta, v) = \frac{1}{(s\sqrt{2})^3} e^{-\frac{\|\theta - \hat{\theta}(\theta, v)\|^2}{2s^2}} \quad (9.5)$$

On a alors : $p(\theta, v) = p(\theta_i | \theta, v) g(\theta, v)$ et

$$p(\theta, v) = \int_0^{360} p(\theta_i, v) d\theta_i \quad (9.6)$$

D'autre part

$$p(C_i, v) = \int_{10(i-1)}^{10i} p(\theta, v) d\theta \quad (9.7)$$

De (9.6) et (9.7) nous pouvons déduire

$$p(C_i, v) = \frac{p(C_i, v)}{p(\theta, v)}$$

et par conséquent les fonctions de décision de Bayes :

$$g_i(x, v) = \frac{p(C_i, x, v)}{p(x, v)} - \sum_{j \neq i} \frac{p(C_j, x, v)}{p(x, v)} \quad (9.8)$$

pour $i = 1, \dots, 36$.

La surface $S = \{ (x, v) \in \mathbb{R}^2 \}$ est un compact et chaque forme est générée en ajoutant une perturbation gaussienne à un point de S , on peut donc supposer qu'elles appartiennent toutes à un voisinage compact K de S . Les fonctions de décision de Bayes g_i qui sont définies par (9.6), (9.7), (9.8) sont donc continues et définies sur le compact K [4, 20].

Le résultat sur l'approximation des fonctions par MLP (§4.1) montre qu'il est possible de trouver une architecture de réseau pouvant générer une fonction $F(\cdot, W) = (F_1(\cdot, W), \dots, F_p(\cdot, W))$ qui approche à ϵ près les fonctions g_i de décision de Bayes. Le résultat sur l'approximation des fonctions de décision de Bayes par les MLP (§4.2.2) montre que l'algorithme de la rétro-propagation réalise une approximation des g_i .

Les résultats précédents ne donnent aucune indication sur l'architecture à employer et donc sur la possibilité de résoudre dans la pratique une telle approximation. Toute la difficulté réside dans la détermination de l'architecture adéquate pour les fonctions g_i données. La complexité du réseau dépend de la nature des fonctions à approximer; au §4.1 nous avons vu que l'**amplitude de continuité** (ϵ)⁵ des fonctions f_i est un facteur important pour mesurer la taille maximale du réseau pouvant approximer avec une précision de ϵ . Ainsi dans le cas d'un réseau de type BL (ayant des unités cachées avec des fonctions de transition à seuil [Blum et Li 91]) la taille maximale du réseau est inversement proportionnelle à ϵ .

Les résultats précédents (9.4), (9.5), (9.6), (9.7) et (9.8) montrent que les fonctions de décision de Bayes pour la détermination de la direction du vent sur les données simulées sont des fonctions continues et simples à exprimer. Ceci justifie les bonnes approximations obtenues avec un réseau simple de deux couches cachées de 25 automates chacune (voir Table 9.13 et 9.14).

⁵ L'amplitude de continuité d'une fonction $f(x)$ définie dans un compact K est définie par $\epsilon(K) = \text{SUP} \{ \|x - x'\|; x \text{ et } x' \in K \text{ et } \|f(x) - f(x')\| < \epsilon \}$.

Le même raisonnement dans son ensemble peut être repris afin de justifier les résultats obtenus pour déterminer la vitesse du vent.

9.5.2. Généralisation au cas réel

La fonction de Long représente une modélisation du phénomène réel; l'étude de l'ordre de complexité des g_i associées au problème réel va permettre d'étendre la validité des résultats précédents à des données réelles.

En effet, le même calcul que précédemment montre qu'il suffit de connaître $g(\theta, v)$ et $p(\theta | \theta_0, v)$ pour déterminer les g_i .

La fonction densité $g(\theta, v)$ mesure la densité de répartition de vent réel dans l'hémisphère étudié. La nature même du phénomène météorologique permet de supposer que cette fonction est continue et à variation "lente".

La fonction de Long, parce qu'elle modélise le phénomène, est une indication sur la complexité de (θ, v) qui est la fonction de transfert de l'antenne et donc sur celle de $p(\theta | \theta_0, v) = h(\theta - \hat{\theta}(\theta, v))$ où h est la fonction densité de l'erreur ajoutée par les antennes.

On peut donc en conclure que dans la réalité les fonctions de décision de Bayes g_i sont du même ordre que celles attachées aux données simulées et que les données réelles pourront être traitées par des architectures du même ordre.

D'autre part les formules précédentes montrent clairement que les fonctions de décision de Bayes prennent en compte à la fois la répartition statistique du vent, l'erreur de mesure et la fonction de transfert : ceci montre que cette méthode permet d'allier pour un même problème l'aspect de l'approximation numérique à l'aspect du traitement statistique.

9.6. Conclusion

Comme nous venons de le voir l'approche neuronale pour la recherche des fonctions de transfert et en particulier l'approche par discrétisation est un outil efficace pour l'étude d'un phénomène physique. Nous avons montré que la discrétisation, bien que légèrement moins précise au point de vue numérique donne des performances tout à fait comparables. L'étude

statistique des points multiples permet de comprendre le phénomène réel sous-jacent et de proposer des modélisations plus adéquates sous forme de fonctions analytiques. Toute recherche de fonction de transfert peut donc commencer par cette approche discrète qui donne une bonne estimation de la complexité de la fonction cherchée et des performances qui peuvent être attendues. Selon les résultats, et si la précision numérique semble fondamentale pour la suite de l'étude que l'on effectue, il est alors possible de passer à l'approximation à l'aide de réseaux donnant des valeurs réelles. Les architectures et les ensembles d'apprentissage mis au point durant la première phase sont une bonne estimation de ceux nécessaires pour cette seconde étape. La phase d'apprentissage qui constitue toujours la phase la plus longue quand on utilise les méthodes neuronales est ici beaucoup plus simple.

L'approche neuronale fournit également un bon outil permettant de juger de l'adéquation des variables utilisées pour expliquer le phénomène. Si les ensembles d'apprentissage utilisés sont statistiquement représentatifs, des mauvaises performances en généralisation prouvent la non adéquation ou le manque d'informations apportées par les variables prises en compte.

10. Enchaînement de la machine

Nous venons de montrer de quelle manière il était possible de calculer séparément à partir des signaux scattérométriques les deux grandeurs qui nous intéressent : la vitesse et la direction du vent. Nous allons montrer dans le chapitre suivant comment la modélisation neuronale permet d'enchaîner les méthodes mises au point de manière à avoir une méthode automatique de décodage.

10.1. La machine neuronale

Le problème est de dresser une carte des vents exacte à partir des mesures (v_1, v_2, v_3) observées.

Nous utiliserons les MLP mis au point dans la section précédente pour réaliser cette phase d'inversion.

Comme nous l'avons vu dans le chapitre précédent, la détermination de la vitesse et de la direction du vent mènent à deux problèmes distincts d'une difficulté différente. La "machine neuronale" est composée de manière à tenir compte de cette information et de celles exposées précédemment. Elle comprend trois modules distincts qui communiquent. Le premier module détermine la vitesse du vent en chaque point de la fauchée. Les résultats sont communiqués au second module comme données supplémentaires, ils permettent avec les signaux scattérométriques de calculer la direction du vent. Comme nous l'avons vu, la détermination de la direction du vent ne permet pas de calculer de manière déterministe la direction du vent, le troisième module permet à l'aide des résultats obtenus par les deux autres de dresser les cartes complètes de vent. L'ensemble des trois modules constituent donc le procédé automatique recherché permettant de dresser automatiquement les cartes de vent (Figure 10.1).

Une fauchée est composée d'un nombre n de trajectoires différentes, dans notre exemple $n = 10$. De façon à tenir compte de la dépendance très forte des

mesures par rapport à une trajectoire donnée (i.e. l'angle d'incidence) chaque module est constitué de n systèmes distincts mais similaires. Un système d'un module donné est attaché de manière spécifique à une trajectoire.

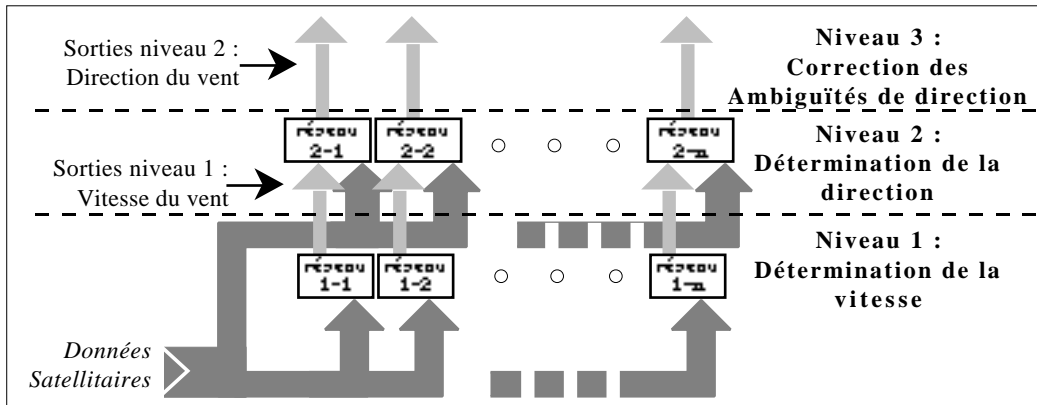


Figure 10.1 : "machine neuronale" permettant l'établissement des cartes de vent. Les réseaux du premier niveau vont permettre la détermination de la vitesse du vent en tout point de la fauchée. Les résultats obtenus seront fournis comme données supplémentaires aux réseaux du deuxième niveau qui calculeront alors la direction du vent. Un ensemble additionnel de réseaux sera dédié dans le troisième niveau à lever les ambiguïtés dans la direction du vent. Chaque colonne de réseaux est attaché de manière spécifique à une trajectoire.

Dans un module chaque système permet d'approximer une fonction de transfert du même type puisque le phénomène physique sous-jacent est identique. La différence essentielle provient des paramètres de cette fonction de transfert qui sont différents pour chaque trajectoire. Chaque système est composé de réseaux de neurones dont les architectures seront donc identiques. Les valeurs des connexions seront fixées à partir d'ensembles d'apprentissage élaborés à partir des mesures prises sur les trajectoire correspondantes. Dans la suite nous présentons les résultats obtenus sur la trace n°5, cependant les performances que nous avons mises en évidence dans le chapitre précédent prouvent qu'une telle approche est valable. Tous les test effectués sur des traces différentes ont montré que les performances obtenues étaient bien similaires.

L'architecture d'un système générique de chaque système est maintenant présentée; chaque système constitue un procédé automatique pour approximer une fonction de transfert.

Dans le premier module un système que nous noterons $S1$ permet de calculer la vitesse du vent en m/s le long d'une trajectoire. Le premier module

est donc constitué de 10 de ces systèmes. Puisque le but de cette méthode est d'obtenir un procédé automatique pour calculer cette vitesse en fonction des signaux, chaque système S_1 sera composé de 3 réseaux différents R_1 , R_2 , R_3 (Figure 10.2).

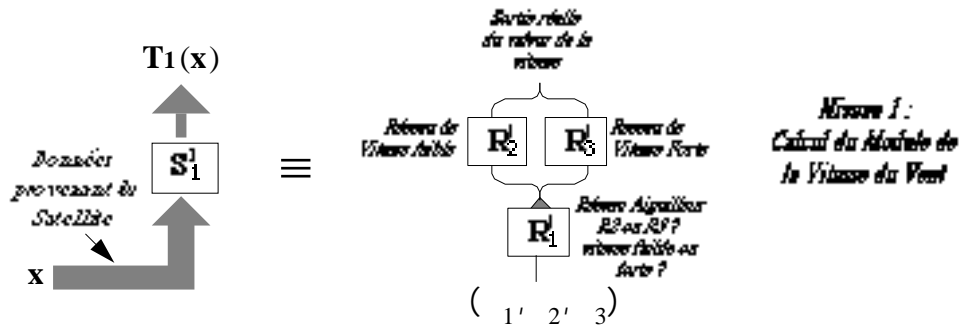


Figure 10.2 : système s_1 de calcul de la vitesse du vent Le système s_1 est composé par 3 réseaux de neurones : le premier, R_1 , est un réseau aiguilleur entre basses et hautes vitesses. Les autres, R_2 et R_3 , correspondent aux réseaux de calcul de basses et hautes vitesses respectivement.

Pour prendre en compte les différents résultats du chapitre 10, nous tiendrons compte de la non-linéarité de la fonction T_1 , celle-ci sera donc approximé par deux réseaux différents R_2 et R_3 . Nous utiliserons donc $A-27 \times 25 \times 25 \times 2$ pour les deux intervalles de vitesse [4 m/s, 12 m/s] et [12 m/s, 20 m/s] de la même manière qu'ils ont été définis précédemment. La précision du calcul étant fondamentale au moment du passage des valeurs entre le premier et le second module, nous avons choisi d'utiliser des réseaux qui fournissent en sortie des valeurs réelles.

Le problème pour calculer v est de mettre au point un procédé qui sélectionne automatiquement le réseau R_2 ou R_3 qui doit effectuer le calcul. Ceci a été réalisé en entraînant un nouveau réseau aiguilleur R_1 dont la fonction est de décider à partir des signaux si le vent considéré est faible ([4 m/s, 12 m/s]) ou fort ([12 m/s, 20 m/s]). Le réseau R_1 est un classifieur qui calcule la probabilité connaissant le signal sur une fenêtre 3×3 autour du point considéré pour la vitesse du vent en ce point d'être faible ou fort. C'est un MLP entièrement connecté dont les inputs sont les mêmes que ceux de R_2 et R_3 (9×3 automates correspondant aux $(1', 2', 3')$); il possède deux couches cachées complètement connectés de 25 cellules chacune et deux automates de sortie chacun correspondant à l'un des intervalles considéré. Pour un signal

donné, la cellule de sortie la plus active indique lequel des deux MLP $R2$ ou $R3$ doit être utilisé pour effectuer le calcul exact de la vitesse.

Un même nombre d'exemples par palier de vitesse (à une résolution de 1 m/s) à été utilisé lors de la composition des ensembles d'apprentissage de ces réseaux. Pour le réseau $R2$, basses vitesses, l'ensemble d'apprentissage est composé de 772 exemples par classe, soit un total de 6176 exemples. Pour le réseau $R3$, vitesses fortes, l'ensemble d'apprentissage est composé de 55 exemples par classe, soit un total d'à peine 440 exemples. Il s'agit d'ensembles équilibrés par vitesse. Pour le réseau aiguilleur $R1$ l'ensemble d'apprentissage est composé par l'union des ensembles utilisés dans $R2$ et $R3$.

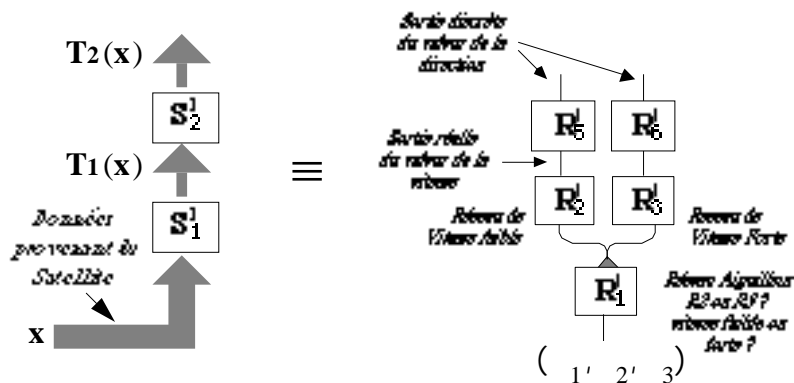


Figure 10.3 : système $s1 + s2$ de calcul du vecteur vent. Le système $s2$ est superposé au système $s1$ afin de calculer la direction. Le réseau $R5$ et $R6$, reçoivent les données concernant la vitesse de $R2$ et $R3$, respectivement. La sortie des réseaux $R5$ et $R6$ est la direction sous forme discretisée.

Dans le second module deux systèmes $S2$ différents ont été testés

- $S2$ est constitué par un seul réseau que nous appellerons $R4$ qui permet une fois reçu la vitesse calculée par le premier module et le signal de calculer quel que soit la vitesse du vent plusieurs directions possibles. $R4$ est ici le réseau classifieur présenté au chapitre précédent dont l'architecture est $A-30 \times 25 \times 25 \times 36$ et dont l'ensemble d'apprentissage contient des vent fort et faibles.
- $S2$ est constitué de deux réseaux $R5$ et $R6$ (Figure 10.3), il s'agit cette fois ci des deux réseaux classifieurs entraînés séparément sur les hautes et les basses vitesses. L'architecture pour chacun des réseaux est comme précédemment $A-30 \times 25 \times 25 \times 36$. Notons qu'il n'est plus besoin ici

d'aiguillage, la sortie du réseau vitesse active sélectionne le réseau direction qui lui est associé. Il faut également signaler que le temps de calcul nécessaire pour effectuer le calcul pour un point donné est exactement le même que pour le premier système *S2* présenté. En effet, seul l'un des deux réseaux *R5* ou *R6* est actionné et effectue son calcul.

Dés que les deux modules ont effectués leurs calculs, les valeurs sont présentées au troisième module dont le rôle est de choisir parmi les différentes directions celles qui sont le plus probables.

10.1.1. Résultats

Nous devons pour apprécier les performances de notre machine neuronale simuler totalement son fonctionnement. Les réseaux utilisés sont ceux dont l'apprentissage a été décrit dans la partie §9.3 et qui ont été sélectionnés ici pour faire partie de la machine. Les performances que nous présentons ici sont donc celles obtenus en simulant le comportement de chacun des réseaux qui constituent les deux modules et en effectuant exactement le choix qui est fait au niveau du réseau *R1*. Les performances obtenues par chaque réseau sont présentées séparément de façon à montrer le fonctionnement de cette machine.

vitesse en m/s	nb.exemples	aiguillés sur basse vitesse	aiguillés sur haute vitesse
4	614	614	
5	689	689	
6	710	710	
7	735	735	
8	579	579	
9	470	470	
10	324	317	7
11	227	207	20
12	178	54	124
13	171	17	154
14	130	7	123
15	74		74
16	50		50
17	37		37
18	17		17
19	18		18
total	5023	4399	624

Table 10.1 : nombre de données par palier de vitesse. La table montre également les performances du réseau aiguilleur : le nombre d'exemples appartenant à une classe et étant bien classés par le réseau aiguilleur. La précision est de ± 1 m/s.

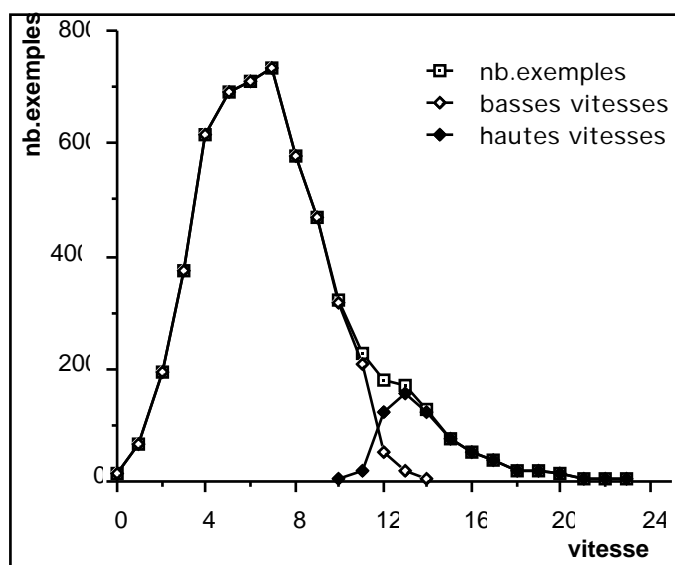


Figure 10.4 : histogramme de l'ensemble de test et résultats du réseau aiguilleur. Nombre d'exemples + aiguillage basses et hautes vitesses Intervalles de 1 m/s, entre 0 et 24 m/s.

Performances du réseau R1

intervalle de vitesse	nb.réel d'exemples (réseau4-20)	nb.exemples aiguillés basse vitesse ⁶ (réseau 4-12)	nb.exemples aiguillés haute vitesse (réseau 12-20)	performances du réseau aiguilleur
4-20 m/s	5023	-	-	97,9 %
4-12 m/s	4348	4399⁷	-	99,4 %
12-20 m/s	675	-	624⁸	88,4 %

Table 10.2 : performances du réseau aiguilleur. Résultats par groupes Basse et Haute vitesse.

Performances des réseaux R2 et R3

Bande 5 — Vitesse du Vent — sortie réelle				
— test sur données issues du réseau aiguilleur —				
2 réseaux, apprentissage sur 4 - 12 m/s et 12 - 20 m/s				
	Réseau Basses Vitesses		Réseau Hautes Vitesses	
précision	perform.	err.RMS	perform.	err.RMS
±1 m/s	98,9 %	0.202 m/s	89,9 %	0.432 m/s
±2 m/s	99,7 %	0.202 m/s	99,0 %	0.432 m/s

Table 10.3 : performances du réseaux R2 et R3. Les performances sont calculés sur l'ensemble de données correspondant sélectionné par le réseau aiguilleur.

- ⁶ Nombre de données sélectionnées par le réseau aiguilleur. L'aiguillage est faite selon le critère suivant : les exemples reconnues comme ayant une vitesse de vent inférieur à 12 m/s forment la base de test des réseaux dites de basse vitesse. Les autres, exemples reconnues comme ayant une vitesse supérieure ou égal à 12 m/s forment donc, la base de test des réseaux de haute vitesse.
- ⁷ Il s'agit en réalité du nombre d'exemples avec une vitesse de vent supérieur à 4 m/s et reconnues par le réseau aiguilleur comme étant dans la classe des basses vitesses (<12 m/s).
- ⁸ Le nombre d'exemples avec une vitesse de vent inférieure à 20 m/s et associés à la classe de hautes vitesses (> 12 m/s).

Performances du réseau R4

Bande 5 — Direction du Vent— classification					
— test sur toutes les données : Basses + Hautes vitesses —					
1 seul réseau — apprentissage sur 4 - 20 m/s					
précision	vitesse	biais moyen $S_{\text{calculé}} - S_{\text{desiré}}$	ecart type du biais	erreur RMS (en m/s)	performance s tt.vitesses 76 m/s)
±10° 1 pic	4-20 m/s	0,88°	85,6°	5,24	61,2 %
	4-12 m/s	1,03°	87,4°	5,00	59,2 %
	12-20 m/s	-0,09°	72,4°	6,80	74,4 %
±10° 2 pics	4-20 m/s	-0,80°	16,9°	1,73	84,2 %
	4-12 m/s	-0,73°	17,2°	1,52	83,6 %
	12-20 m/s	-1,23°	15,8°	3,15	88,4 %
±20° 1 pic	4-20 m/s	0,88°	85,6°	5,24	69,5 %
	4-12 m/s	1,03°	87,4°	5,00	68,4 %
	12-20 m/s	-0,09°	72,4°	6,80	76,9 %
±20° 2 pics	4-20 m/s	-0,82°	17,0°	1,274 m/s	95,8 %
	4-12 m/s	-0,76°	17,2°	1,12	96,0 %
	12-20 m/s	-1,23°	15,8°	2,27	94,5 %
±20° 3 pics	4-20 m/s	-0,64°	8,3°	1,07	97,6 %
	4-12 m/s	-0,51°	8,1°	0,94	97,9 %
	12-20 m/s	-1,47°	9,8°	1,95	95,9 %
±20° 4 pics	4-20 m/s	-0,65°	7,1°	0,90	99,0 %
	4-12 m/s	-0,58°	7,0°	0,77	99,3 %
	12-20 m/s	-1,11°	7,3°	1,75	97,3 %

Table 10.4 : performances du réseau R4. Les performances sont calculés sur l'ensemble de données issues des réseaux R2 et R3. La valeur réel calculé de la vitesse est donnée en entrée de R4.

Performances des réseaux R5 et R6

Bande 5 — Direction du Vent— classification				
— test sur toutes les données : Basses + Hautes vitesses —				
Deux Réseaux Direction: Basses + Hautes vitesses				
— apprentissage sur 4 - 12 m/s et 12 - 20 m/s —				
précision	biais moyen $S_{\text{calculé}} - S_{\text{désiré}}$	ecart type du biais	erreur RMS (en m/s)	performance s basses vitesses 153 m/s)
±10° 1 pic	-4,67°	90,2°	5,35	58,4 %
±10° 2 pics	-1,00°	8,7°	1,30	86,2 %
±20° 1 pic	-4,67°	90,2°	6,35	66,4 %
±20° 2 pics	-1,00°	8,7°	0,990 m/s	96,7 %
±20° 3 pics	-0,73°	6,8°	0,74	98,9 %
±20° 4 pics	-0,77°	6,64°	0,71	99,4 %

Table 10.5 : performances des réseaux R5 et R6 Les performances sont calculés sur l'ensemble de données issues des réseaux R2 et R3. La valeur réel de la vitesse calculé par R2 est ajoutée en entrée de R5. La sortie de R3 est ajoutée en entrée de R6.

10.1.2. Interprétation des Résultats

Le réseau R1 effectue quelques erreurs (97,9% de bonne classification), il faut remarquer que ces erreurs se situent autour du seuil de 12 m/s. Une telle imprécision peut difficilement être évitée, la précision des réseaux R2 et R3 étant de 2 m/s.

Dues aux erreurs de l'aiguillage, les performances de R2+R3 sont donc légèrement inférieures à celles présentés sans celui-ci. Notons que pour la précision demandée (2 m/s) les résultats restent très bons. La dégradation des performances se voit cependant dans la précision de la valeur prédite : l'erreur e_{rms} qui était de 0,22 m/s pour l'expérience comparable présentée dans le chapitre précédent, est maintenant deux fois plus importante : 0,432 m/s. Comme nous l'avons déjà mentionné cette précision va se révéler fondamentale au moment du passage au module suivant.

Les performances les meilleures qui représentent les résultats de l'enchaînement des deux modules sont données par la deuxième méthode qui approxime la fonction T_2 en deux parties distinctes selon la vitesse qui est calculée. Le pourcentage de bonne prédiction est maintenant de 96,3%, si l'on autorise deux solutions possibles, avec une erreur e_{rms} de 1,15 m/s sur

l'ensemble de test, il est à comparer aux 99,2 % ($e_{rms} = 0,54$ m/s) fourni par la meilleure des méthodes précédentes.

L'enchaînement des différents réseaux conserve donc des performances tout à fait satisfaisantes et justifient pleinement l'utilisation de cette méthode. Bien que nous ayons ajouté aux données simulées le bruit prévu pour les appareils de mesures, il est probable qu'au moment de l'utilisation des données réelles ce bruit soit plus important ou très différent. La phase de mise au point devra donc être recommencée et certainement ajustée. Cependant la méthode elle même devrait permettre une interprétation du phénomène.

La constitution de cartes correctes des vents nécessite maintenant une dernière étape. Les données calculées par les deux modules que nous venons de présenter ne permettent dans un premier temps que l'établissement d'un "first guess", c'est à dire de la carte la plus probable. Ce dernier module n'a pas été mis au point dans cette thèse. En fait un premier travail, qui constitue une approche académique du problème, a été réalisé préalablement aux débuts de cette thèse. Les résultats qui en ont résulté et qui prouvaient l'adéquation de la méthode au problème ont donné naissance au travail complet qui vient d'être présenté. Nous voulons dans ce dernier paragraphe donner de la "machine neuronale" une idée complète, pour cette raison nous décrirons rapidement la méthodologie utilisée dans ce premier travail. Il illustrera la grande souplesse que procurent la modélisation neuronale et le nombre important de traitements différents qui peuvent être réalisés à partir d'un même modèle. Bien entendu pour être totalement efficace cette méthode devra être adaptée par la suite aux résultats que nous venons de présenter.

10.2. Levée des ambiguïtés

Le choix entre les différentes solutions possibles qui seul permet l'élaboration de cartes de vent exactes constituent un premier travail préalable à cette thèse qui est décrit dans l'article [Badran et al. 89]. Nous retraçons rapidement la méthode utilisée qui a été élaborée à partir de données académiques avant ce travail et qui doit être adapté dans le futur pour permettre de traiter les résultats que nous venons de présenter.

Deux types d'ambiguïtés qui correspondent aux ambiguïtés théoriques du modèle que l'on cherche à apprendre ont été considérées. Les ambiguïtés qui apparaissent au moment de l'interprétation des mesures prises par les trois antennes dont dispose ERS-1 et NSCAT étaient spécifiés comme faisant respectivement un angle de $180^\circ (\pm 30^\circ)$ et de $\pm 90^\circ (\pm 30^\circ)$ avec la direction vraie du vent. Pour lever ces ambiguïtés deux machines différentes qui permettent de traiter séparément la levée de chacune de ces ambiguïtés ont été construites. Les ensembles d'apprentissage nécessaires à chaque machine sont différents, mais les mécanismes nécessaires à l'élaboration de chacune d'entre elles sont identiques. Nous rappelons ici la méthodologie commune suivie pour générer les données et l'architecture employée.

10.2.1. Les données

La mesure du vent en un point est liée étroitement aux mesures effectuées aux points géographiquement proches. Pour une ambiguïté donnée, une machine permet de détecter en chaque point si le vecteur considéré a une direction exacte ou bien erronée en fonction d'une fenêtre 5×5 découpée autour de ce point. Une telle fenêtre revient à prendre en compte les 25 mesures des vents uniformément réparties dans un carré de 200 Km de côté, étant donné que les vents sont fournis sur des pixels de 50×50 Km. L'expérience a montré que l'on obtenait ainsi une information suffisante pour détecter les erreurs de mesure. Afin de prendre en considération les effets de bord dûs à la portée des antennes, trois types de fenêtres ont été considérées.

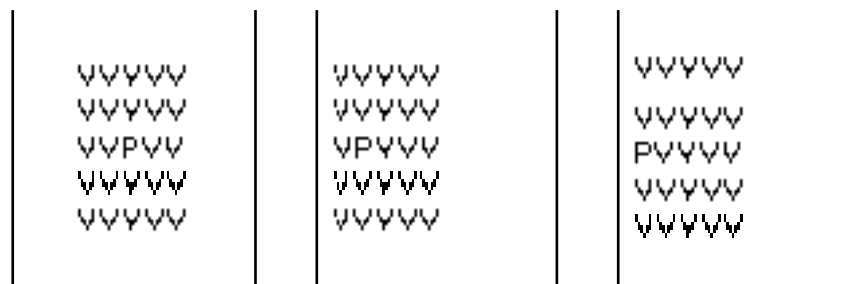


Figure 10.5 : types de voisinage utilisés pour le levé des ambiguïtés. P indique le point qui est examiné et V ceux qui sont pris en compte en fonction de la position du point P par rapport aux bords de la fauchée.

Une fenêtre représente le voisinage d'un point, selon la position de ce point par rapport au bord de la fauchée elle sera centrée (type 2) ou non (type 0 ou 1)

autour du point. La fenêtre de type 0 regarde un point situé exactement sur le bord de la fauchée, celle de type 1 un point situé à 50 Km de ce bord (voir Figure 10.5). Les effets de bord sont nombreux puisque nous ne connaissons pas la carte des vents sur un hémisphère complet, mais le long de fauchées de 450 Km de large, ce qui représente une bande relativement étroite.

Une machine sera donc composé de trois réseaux différents. A chaque point de la carte correspond l'un de ces trois réseaux qui lui permet de modifier ou non sa direction, il est choisi en fonction de la position du point dans la fauchée.

Le temps nécessaire au redressement d'une ambiguïté sur une carte entière est donc faible, il est dans nos expériences proportionnel au nombre de points pour lesquels on effectue les mesures.

Le développement d'architectures spécialisées qui autorisent un calcul en parallèle permettra de réduire ce temps dans deux directions différentes : à l'intérieur du réseau en mettant en parallèle les calculs intermédiaires, sur la carte en traitant plusieurs points à la fois. L'utilisation de ces nouveaux calculateurs permettra de réduire de manière très importante des temps de calculs déjà courts. Il est montré plus tard que le redressement complet d'une carte nécessite trois passages de machine. La levé des ambiguïtés des cartes de vents peut donc être effectué en temps réel sur des stations de travail.

10.2.2. L'architecture

Chacun des réseaux construits possède la même architecture : il possède 3 couches (voir Figure 10.6).

- La couche d'entrée contient 50 cellules : les 25 premières sont forcées aux valeurs des *sinus* des 25 vecteurs représentatifs des vents de la fenêtre considérée. Les 25 suivantes correspondent aux *cosinus* de ces mêmes vecteurs. Le codage en ces composantes a été retenu car il permet d'éviter la brusque discontinuité présente au voisinage de 2π dans le cas où l'angle est pris comme entrée.
- La couche cachée contient 25 cellules : ce nombre a permis d'obtenir de très bons résultats à l'apprentissage et à la généralisation, mais on n'a pas cherché à trouver une architecture minimale. Une étude ultérieure doit être menée pour déterminer une telle architecture car elle

permettrait de diminuer la somme des calculs effectués aussi bien à l'apprentissage qu'à la généralisation.

- La couche de sortie contient 2 cellules, chacune d'entre elles correspond à l'une des réponses attendue : la direction proposée est vraie, la direction proposée est fausse.
- Les connexions sont complètes entre les deux niveaux.

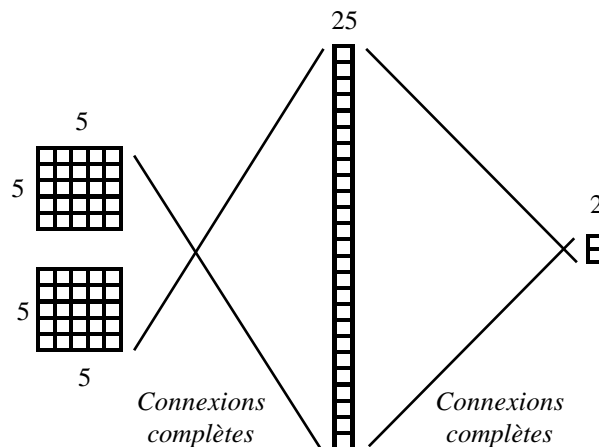


Figure 10.6 : architecture utilisée pour le levé des ambiguïtés.

10.2.3. L'ensemble d'apprentissage

Les réseaux d'automates faisant partie des méthodes d'apprentissage supervisée, nous avons entraîné des réseaux à lever automatiquement les ambiguïtés en lui présentant un certain nombre d'exemples de la tâche à accomplir : association (carte ambiguë, carte non ambiguë). Pour cela, un ensemble d'apprentissage a été déterminé à partir d'une carte représentant les huit fauchées produites par le satellite au cours d'une journée dans laquelle des régions représentatives des différentes configurations qui peuvent se présenter ont été découpées. La représentation de manière égale des principales structures météorologiques de moyenne latitude (perturbations cycloniques, anti-cycloniques, fronts, etc) à été notamment prise en compte.

Pour chaque ambiguïté on a constitué des cartes bruitées sur lesquelles seule cette ambiguïté apparaît. Les régions sélectionnées pour

l'apprentissage ont donné naissance à trois ensembles d'apprentissage d'environ 2000 exemples chacun : un pour chaque type de fenêtre. La prise en compte des deux ambiguïtés amène donc à considérer six ensembles, pour chacun d'entre eux il a été indiqué si la mesure était erronée ou pas. Chacun de ces ensembles représente les entrées à partir desquelles ont été entraînés les six réseaux distincts.

10.2.4. Simulation des ambiguïtés

Les mesures produites par NSCAT et ERS-1 n'étant pas encore accessibles, celles qui nous ont été fournies proviennent de modèles météorologiques et sont donc extrêmement régulières. Pour modéliser avec plus de vraisemblance les valeurs qui seront fournies au système quand elles proviendront de capteurs réels, nous avons ajouté à la direction du vent et en chaque point un bruit gaussien de moyenne nulle et d'écart type 10° . Pour simuler les ambiguïtés à 180° nous avons nous-mêmes inversé de manière aléatoire une certaine proportion de vecteurs de la carte initiale. Dans les expériences présentées ce pourcentage s'est élevé à 20%. Les mêmes expériences ont été répétées avec un taux de 30% d'inversions : les résultats numériques se sont montrés comparables. La seconde ambiguïté étant beaucoup moins fréquente que la première, nous avons mené l'apprentissage en introduisant 7% d'erreurs dans la direction du vent, nous avons effectué une rotation de $\pm 90^\circ$ de manière aléatoire à partie de la carte initiale. Le nombre d'exemples à rectifier étant trop faible, vu le taux d'erreurs envisagé, il a été doublé leur nombre en dupliquant chacun de ces exemples dans l'ensemble d'apprentissage de la seconde machine.

10.2.5. Généralisation

Chacun de ces six réseaux a été testé sur des données nouvelles de manière à illustrer ses facultés de généralisation. Ces tests ont été réalisés sur des cartes de vents complètes choisies à des dates d'abord proches puis éloignées de celle qui a servi à l'apprentissage; des cartes présentant de nombreuses perturbations ont été prises. Les réseaux construits se sont montrés robustes pour tous les tests effectués.

10.2.6. Les résultats

L'apprentissage a été mené sur l'ensemble des 2000 exemples et sur les 3 types de voisinage pour la même journée de et pour chacune des ambiguïtés séparément.

type de voisinage	1 Septembre	18 Août	16 Août	6 Août
type 2	98,00 %	97,50 %	98,00 %	98,00 %
type 1	99,47 %	98,50 %	99,12 %	98,30 %
type 0	98,93 %	98,20 %	99,00 %	98,14 %

Table 10.6 : taux de généralisation pour la levée des ambiguïtés à 180°.

type de voisinage	1 Septembre	18 Août	16 Août	6 Août
type 2	99,95 %	99,95 %	99,95 %	99,95 %
type 1	97,86 %	97,70 %	97,60 %	97,23 %
type 0	99,50 %	99,37 %	99,60 %	98,96 %

Table 10.7 : taux de généralisation pour la levée des ambiguïtés à $\pm 90^\circ$.

La validité des résultats concernant les voisinages de type 2 a été vérifiée sur l'ensemble de la carte complète du même jour (18 août) et sur trois cartes complètes simulant les données de trois journées différentes (6 août, 16 août, 1 septembre) les résultats sont données dans les tableaux 1 et 2. Dans tous les cas, les cartes considérées ont été bruitées de la même manière que pour l'ensemble d'apprentissage : bruit gaussien de moyenne nulle et d'écart type 10° , auquel on a ajouté une perturbation de 20% pour la première ambiguïté (20% de vecteurs du vent ont été inversés de manière aléatoire) et de 7% pour la seconde.

Ce pourcentage représente le nombre de points pour lesquels la direction du vent est exacte lorsque la reconnaissance est achevée sur la carte entière. On peut interpréter ces résultats comme représentant la correction locale qui peut être effectuée à l'aide des réseaux, ils représentent donc une borne inférieure de l'efficacité de la machine que nous allons construire. En effet, une prise en compte des directions des vents sur un carré de 250 Km de côté ne représente pas une information suffisante pour obtenir une carte totalement exacte.

10.2.7. Conclusion

Nous voyons que le travail effectué dans la thèse change de manière sensible le problème. Nous possédons maintenant une plus grande connaissance des solutions possibles. En particulier nous connaissons pour un point donné les 4 solutions possibles ainsi que leur vraisemblance. La méthode que nous mettrons au point cherchera encore à utiliser des filtres neuronaux et un processus de relaxation. Il cherchera à utiliser pendant la relaxation les valeurs précises proposées par la machine neuronale.

11. Comparaison avec des méthodes classiques

11.1. Les méthodes

Plusieurs algorithmes ont été proposés afin d'estimer le vent proche de la surface de l'océan. Tous ces algorithmes se fondent sur le rapport existant entre le paramètre θ^0 mesuré en un point sur l'océan et le vecteur –vitesse et direction– vent en ce point. Afin de prouver la validité de notre approche connexionniste, nous montrons dans la Table 11.1 une comparaison des performances de la méthode connexionniste contre sept autres algorithmes proposés pour estimer le vent. Tous ces méthodes dites classiques sont fondées sur la minimisation des moindres carrés pondérés (weighted least squares minimization) : **WLSL** est un algorithme de minimisation des moindres carrés pondérés défini dans le domaine des logarithmes, **LS** est l'algorithme des moindres carrés, **WLS** est l'algorithme des moindres carrés pondérés, **AWLS** est une méthode ajustable des moindres carrés pondérés, **L1** est une méthode fondée sur la norme L1, **LWSS** est l'algorithme des moindres carrés sur la vitesse du vent et **ML** est l'algorithme de vraisemblance maximal. Une description complète de ces algorithmes est donnée dans l'article de Chong-Yung Chi and Fuk K. Li [Chi et Li 88]. Les simulations ont été réalisées suivant le processus suivant : les entrées sont générées suivant le même processus que pour la méthode connexionniste⁹. L'erreur dans le vecteur vent e est décrit dans la section §9.4..

Couramment, entre deux et quatre réponses simultanées, ou solutions ambiguës, sont obtenues en utilisant l'algorithme d'estimation du vent. La

⁹ Un vecteur de vent (v, θ) est choisi et, en fonction d'une relation empirique, le paramètre θ^0 est calculé à l'aide d'une fonction empirique. Le paramètre θ^0 obtenu est bruité par un bruit gaussien estimé. La donnée résultante sera prise pour faire le processus inverse : trouver (v, θ) en fonction du θ^0 bruité.

solution, on dit souvent l'*ambiguïté*, la plus proche du vrai vecteur vent est choisie comme la bonne estimation. L'erreur RMS statistique de $\|e\|$, noté par e_{rms} , a été calculé sur 10 000 estimations indépendantes.

11.2. Résultats

La Table 11.1 montre que les résultats obtenus par la méthode connexionniste sont très bons par rapport aux autres techniques. Bien que le processus de génération des données entre les méthodes numériques et la méthode connexionniste est le même, la fonction empirique utilisé n'est pas la même, mais elle est équivalente puisque toutes les deux simulent le même phénomène. Une deuxième différence dans la génération des données vient de la nature du bruit ajouté aux "mesures". En effet, les algorithmes pour la génération du bruit différent légèrement mais ils sont toutefois équivalents. Vus ces différences nous interprétons les résultats dans la Table 11.1 comme indicatifs de l'ordre de magnitude de la précision qui est possible d'atteindre avec les méthodes connexionnistes.

Vitesse	WLSL	ML	LS	WLS	AWLS	L1	LWSS	A- 30×25×25×36
Basse	0.92	0.66	0.67	0.74	0.69	0.63	1.02	0.32
Moyenne	0.89	0.85	1.10	1.31	0.89	0.98	0.87	0.40
Haute	3.71	3.44	4.11	5.52	3.52	4.06	3.49	1.18

Table 11.1 : erreur RMS (en m/s) pour sept algorithmes d'estimation du vent et méthode neuronale. Ces algorithmes sont décrits dans [Chi et Li 88]. Les résultats sont donnés pour différentes vitesses du vent.

Toutes les méthodes classiques sont fondés sur l'hypothèse qu'il existe une fonction analytique qui établit le rapport : $((v, \theta))$. La méthode connexionniste est plus générale et ne dépend pas de cette hypothèse. Comme nous l'avons fait remarquer avant, la vitesse de calcul, qui est un aspect important dans la désambiguation du vent, est assez élevée. Le calcul d'un vecteur par la méthode connexionniste est fait en près de 12 ms avec un ordinateur Sun SparcStation 2.

Nous concluons que les réseaux de neurones peuvent faire aussi bien que les meilleures des méthodes classiques. A ceci nous ajoutons la capacité des réseaux de neurones de modéliser des problèmes non-linéaires complexes,

sans aucune information *a priori*, et à leur capacité de prendre en compte des dépendances spatiales.

11.3. Conclusion

Plusieurs méthodes d'extraction du vent en fonction des mesures radar ont été comparées dans la section précédente. Les meilleurs résultats sont obtenus avec la méthode fondée sur les réseaux neuronaux. L'application traitée est complexe. Nous pouvons dire que les réseaux neuronaux sont un puissant outil pour modéliser une vaste classe de fonctions de transfert. D'ailleurs, les réseaux neuronaux sont capables de résoudre beaucoup de problèmes trouvés en Physique. Un avantage de taille des réseaux de neurones est leur capacité à extraire de l'information dans un environnement bruité quelque soit le bruit.

Les réseaux de neurones présentent d'autres avantages. Il leur est très facile de prendre en compte d'autres paramètres même si leur dépendance ne peut pas être établie sous la forme d'une équation. Il suffit d'ajouter des nouvelles cellules à la couche d'entrée et de les connecter aux couches cachées. Dans le cas du scatteromètre il pourrait être facile d'examiner la dépendance de la solution à de nouvelles variables comme par exemple les vagues, qui sont supposées influencer la fonction de transfert à basse vitesse du vent.

La méthodologie développée ici sera appliquée aux données scatterométriques fournies par le satellite ERS-1 aussitôt que nous aurons la possibilité d'examiner ces données. La détermination et l'étalonnage du modèle géophysique peut être contourné. En fait, le modèle géophysique est inclus dans l'architecture de la "machine neuronale" décrite précédemment. La "machine neuronale" décrite peut être mise au point pendant une phase d'apprentissage sur les sorties fournies par des modèles météorologiques comme le modèle PERIDOT, utilisé par la Météorologie Nationale de France. C'est lors de cette confrontation avec le monde réel que l'utilité des réseaux neuronaux sera définitivement évaluée.

12. Conclusion

Au début de ce travail de thèse, le retour du Perceptron et l'apparition d'autres techniques neuronales, ont suscité un enthousiasme effréné et sans doute excessif. On commençait juste à envisager l'utilisation des techniques connexionnistes sur des problèmes liés au monde réel.

Dans cette thèse, nous nous sommes intéressés plus particulièrement à l'étude des architectures de type Perceptron Multicouches entraînées avec l'algorithme de la rétro-propagation du gradient.

Nous avons exploré les possibilités d'utilisation des réseaux connexionnistes pour un problème de télédétection ou en général d'approximation des fonctions de transfert. Il s'agit d'une tâche ardue. Cette complexité rend en particulier difficile, sinon impossible, l'utilisation d'un réseau unique. Cela nous a amené à élaborer des architectures modulaires qui reflètent une division logique de la tâche globale en sous-tâches plus simples à traiter. Nous avons alors étudié la possibilité d'incorporer des connaissances *a priori* sur le problème, dans les architectures connexionnistes.

Dans cette architecture modulaire de réseaux connexionnistes, chaque module est dédié au traitement d'une sous-tâche. La conception d'une telle architecture nous a permis, d'une part, de montrer la validité opérationnelle de l'outil "réseaux neuronal" sur un problème réel, et d'autre part, de faciliter l'analyse du comportement des différents modules. Elle nous a permis aussi de nous rendre compte des avantages et inconvénients ainsi que des limitations des architectures du type Perceptron Multicouches.

Cette approche modulaire nous a permis de mettre en œuvre un système performant.

Au-delà des résultats concernant une application particulière, le travail que nous avons présenté a contribué à éclaircir quelques problèmes liés à l'apprentissage dans les architectures de type perceptron multicouches, à

analyser leur comportement sur un problème complexe et à proposer des solutions pour lever ces limitations.

Mais bien des problèmes restent à résoudre, comme la décomposition automatique de la tâche globale et la coopération pendant l'apprentissage des différents modules composant le système afin d'atteindre des solutions optimales.

Annexes

A. Incorporating knowledge in multi-layer networks:

