Chapitre 5

LES PREDICTIONS ASTRO-CLIMATIQUES

5.1 Introduction

L'Observatoire Européen Austral (ESO) est une organisation scientifique internationale fondée en 1962 regroupant 8 pays membres, dont la France. Le siège de l'organisation est situé à Garching, près de Munich, en Allemagne et regroupe, outre la division scientifique, le centre de coordination européenne du télescope spatial Hubble et le centre de traitement d'images et de contrôle à distance des télescopes à La Silla sur le territoire Chilien. La Silla est une montagne à environ 600km au nord de Santiago, dans la partie méridionale du désert de d'Atacana. Au sommet, est établi l'un des plus grands et des plus modernes observatoire astronomiques du monde. Il compte 14 télescopes optiques, dont le révolutionnaire "New Technology Télescope" de 3.5m de diamètre. Le plus grand des télescopes de La Silla possède un miroir de 3.6m de diamètre. On y trouve aussi un télescope de 15m destiné aux observations des ondes sub-millimétriques, le seul de ce type dans l'hémisphère sud.

Le projet principal en cours à l'ESO est la construction d'un télescope géant d'une ouverture équivalente de 16m nommé "Very Large Télescope" (VLT). Le VLT entrera en opération à la fin des années 1990 au Chili. Il sera constitué, non pas d'un seul mais, de plusieurs télescopes disposés dans une configuration en forme de trapèze. Les quatre télescopes principaux auront des miroirs de 8.2m de diamètre; ils seront accompagnés de plusieurs télescopes auxiliaires de 1.8m. Lorsque la lumière des 4 télescopes est combinée, le pouvoir collecteur global est similaire à celui d'un télescope de 16m de diamètre. Grâce à sa configuration particulière, le pouvoir de résolution du VLT sera équivalent à celui d'un télescope de 200m de diamètre, c'est-à-dire qu'il sera capable de voir un objet d'un mètre situé sur la Lune.

Le VLT est actuellement en construction et sera installé au sommet du mont Paranal de 2660m d'altitude. Paranal est un site primitif situé au Chili dans la région la plus sèche du monde. La très faible humidité et les conditions atmosphériques extrêmement stables à Paranal sont de la plus haute importance pour les observatoires astronomiques. Ces conditions favorables ont été mises en évidence par une étude du site menée sous la direction de Marc Sarazin. Cette étude s'intéresse à la fraction du temps disponible pendant laquelle les conditions d'observation dépassent certains critères d'excellence prédéfinis en fonction du mode d'observation choisi. Pour ce faire, une analyse exhaustive de l'environnement, dite astro-climatologique, est effectuée. Celle-ci regroupe aussi bien les conditions météorologiques au sol (vent, température, humidité), que les caractéristiques de l'atmosphère. Une panoplie d'instruments *ad hoc* permet de décrire l'état instantané de l'atmosphère par le truchement d'instruments météorologiques courants (anémomètres, thermomètres, baromètres, etc) placés sur un mat. A ces instruments conventionnels s'ajoute un "moniteur de mouvement différentiel" dont la fonction est de mesurer la turbulence atmosphérique instantanée : le *seeing*. Le seeing quantifie les modifications subies par une onde lumineuse au cours de sa traversée de l'atmosphère turbulente jusqu'au foyer du télescope.

A l'origine, les résultats de cette analyse présidaient uniquement au choix du site le mieux approprié à l'installation du télescope. Mais désormais, il est question d'adapter la programmation des observations aux caractéristiques atmosphériques par l'entremise d'une station de mesures astro-climatiques du VLT (Astronomical Weather Station-AWS). Cette station de l'an 2000 sera la clé de voûte de la gestion du temps d'observation des futurs télescopes. C'est elle qui délivrera des prédictions de température afin de mettre les instruments et les miroirs en température pour limiter les perturbations locales d'origine thermique, le *local seeing*. C'est elle surtout qui fournira une estimation de la qualité d'observation, en vue de modifier la programmation des observation dans la file d'attente et de modifier l'instrumentation de façon adéquate. Ainsi les astronomes profiteront des conditions les mieux adaptées à leurs observations.

Force nous est d'employer une stratégie opérationnelle qui intègre des modèles statistiques, et/ou cognitifs, capables d'extraire des informations contextuelles dans des conditions sans cesse en évolution (variations diurnales, saisonnières, etc.). Bien sûr, le DRNN est au centre de cette stratégie; avec l'astro-climatologie, il trouvera une source inépuisable d'applications potentielles sur lesquelles évaluer ses performances dans des conditions réelles. En particulier, l'application du DRNN à la température et au seeing à différentes échelles de temps donnera de la chair, une tournure plus concrète, à la station AWS, arrêtée dans son principe, mais dessinée pour l'instant d'une façon toute théorique. Outre le DRNN, des méthodes prédictives plus traditionnelles seront employées parmi lesquelles

- les réseaux non-bouclés du type NAR et NARMA.
- l'auto-régression linéaire multi-dimensionnelle (ARX).
- la reconstruction locale dans l'espace de phase (*k-nearest neighbor*).
- la "copie carbone".

Le passage des suites déterministes uni-dimensionnelles générées par ordinateur aux suites aléatoires multi-dimensionnelles issues du monde réel soulève un certain nombre de difficultés qu'il a fallu aplanir, comme en témoigne la densité de ce chapitre. Parmi les plus embarrassantes figurent la discontuité temporelle due aux fréquentes ruptures des séquences et le nombre relativement restreint de données disponibles.

Les ruptures de séquences

En préambule à cette analyse, discutons brièvement du problème que soulèvent les ruptures de séquences. Nous savons que l'évolution du DRNN est déterminée par son état initial $\mathbf{v}_0$ et par la suite des vecteurs d'entrée, $\mathbf{x}_t$. Aussi, chaque rupture de séquence dans la suite des mesures du processus physique (due à la défaillance d'un maillon de la chaîne d'acquisition de données, à la suspension des mesures, etc.) rompt la période d'échantillonnage et aboutit à une situation où le réseau de neurones est "trompé" sur la nature du passé de la suite. La difficulté que soulève les ruptures de séquence est d'autant plus aiguë lorsque les contingences temporelles s'étendent sur de longues durées. Toutefois, la nature statistique des variations climatiques nous confirme dans l'idée que seules les conditions météorologiques récentes influent sur la conditions à court terme. Ainsi, nous pourrons effacer la mémoire "long terme" du réseau et ne retenir que les événements récents.

La grandeur caractéristique qui fixe l'échelle de temps du réseau est le délai maximal des connexions au sein du réseau, c.a.d. l'ordre des filtre, D. C'est la raison pour laquelle nous optons ici pour une reconstruction de la mémoire sur une période de durée D. En d'autres termes, chaque fois que survient une rupture de séquence, la mémoire du réseau est effacée, l'apprentissage est suspendu, et les D exemples suivants sont employés à reconstituer une mémoire *tronquée*. L'apprentissage reprend à l'issue de la phase de remémorisation. Une conséquence manifeste se dégage de ces observations : si T est la taille de la base de données et S le nombre de ruptures de séquence, la taille effective de la base d'apprentissage se réduit à $T-S*D$ (nous supposons que la base de données est telle que l'intervalle de temps qui sépare deux ruptures de séquence est supérieur à D). Bien entendu, cet argument entre en ligne de compte lorsque se pose le choix de la configuration de réseau au vu du nombre de ruptures de séquences et de la taille de la base de données.

5.2 La précipitation

Dans ce paragraphe, des résultats préliminaires sur l'anticipation de la précipitation sur la montagne de Gamsberg en Namibie sont présentés. Nous avons choisi d'expérimenter sur la séquence des enregistrements de la précipitation effectués à Gamsberg car le site est une alternative à Parnal pour la construction du futur télescope européen VLT. En l'absence de station météorologique autour du site, un modèle anticipatif de la précipitation permettrait de déterminer indirectement la qualité du site.

La base de données limitée, à défaut de permettre une analyse statistique poussée, fournit une opportunité de comparer les approches NAR, NARMA dont nous avons parlé au chapitre précédent. Les objectifs que nous poursuivons se résument à

- valider les équations d'adaptation établies dans le cadre théorique.
- évaluer les performances en comparaison avec les modèles NAR et NARMA à base de réseau non-bouclés.
- estimer la prédictabilité des conditions météorologiques locales au site astronomique.

La base de données est constituée de 180 enregistrements représentant la moyenne saisonnière de la précipitation durant 45 ans. La suite présente des variations périodiques saisonnières aux amplitudes irrégulières. L'ensemble des données sert à la fois de base d'apprentissage et de base de test. Aussi, pour éviter l'*overfitting*, l'élimination des poids est accrue par le biais du paramètre de décroissance des poids fixe à $\epsilon = 5 * 10^{-3}$.

En raison de leurs différences, il est difficile de trouver une base de comparaison entre ces modèles. Pour ne favoriser aucun modèle, l'approche que nous suivons consiste à limiter le nombre de neurones des réseaux à un nombre raisonnable et d'étudier la performance à un pas en avance. La taille de la fenêtre temporelle est arbitrairement fixée à 5 pour le modèle NAR et NARMA et les 5 dernières sorties du réseau NARMA sont réinjectées en entrée comme des entrées supplémentaires. Le nombre de poids ajustable se situe autour de 40, 30 pour le DRNN.

Les 3 modèles que nous avons utilisés sont

- un réseau NAR $5 \times 5 \times 1$.
- un réseau NARMA $10 \times 5 \times 1$.
- un réseau DRNN $1 \times 2 \times 1$ DRNN avec $D = 3$. La configuration du réseau est détaillée Fig. 5.1.

Les données ont été normalisées et centrées. Le choix initial de la matrice de poids a été effectué aléatoirement jusqu'à ce qu'une variance significative de l'activation du neurone de sortie soit observée. Les poids du DRNN ont satisfait aux conditions (3.54).

Table 5.1 résume plusieurs NMSE pour la prédiction à un pas en avant et la sortie de chaque réseau est tracée Fig. 5.2. Bien que nous ayons porté une attention particulière à ne pas désavantager un modèle sur un autre, le DRNN obtient des résultats légèrement meilleurs. Au terme de l'apprentissage, le DRNN est réduit à peau de chagrin (12 connexions sur 31 ont été éliminées) alors que la taille des autres modèles n'a pratiquement pas évolué.

Il est clair, au vu des données, que les amplitudes des oscillations sont les aspects les plus difficiles à prévoir. La prédiction itérée est trop mauvaise pour faire l'objet d'une comparaison sérieuse. Les prédictions itérées avec les réseaux NAR et NARMA convergent rapidement vers l'origine alors que le DRNN fournit une sortie oscillatoire d'amplitude constante.

Modèle	AR	NARMA	DRNN
Nombre initial de poids	36	41	31
Nombre de poids effectifs	35	40	19
NMSE	0.467	0.389	0.295

TAB. 5.1 - : Résultats des prédictions à un pas en avance pour les modèles AR, ARMA et DRNN. Le nombre de poids avant et après élimination est affiché.

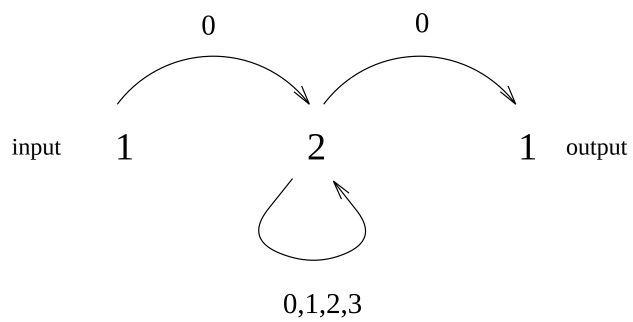

FIG. 5.1 - : Précipitation en Namibie : représentation schématique du DRNN. Le réseau compte 1 neurone d'entrée, 1 neurone de sortie, et deux neurones cachés interconnectés par des liens de délais $d = 0, 1, 2, 3$.

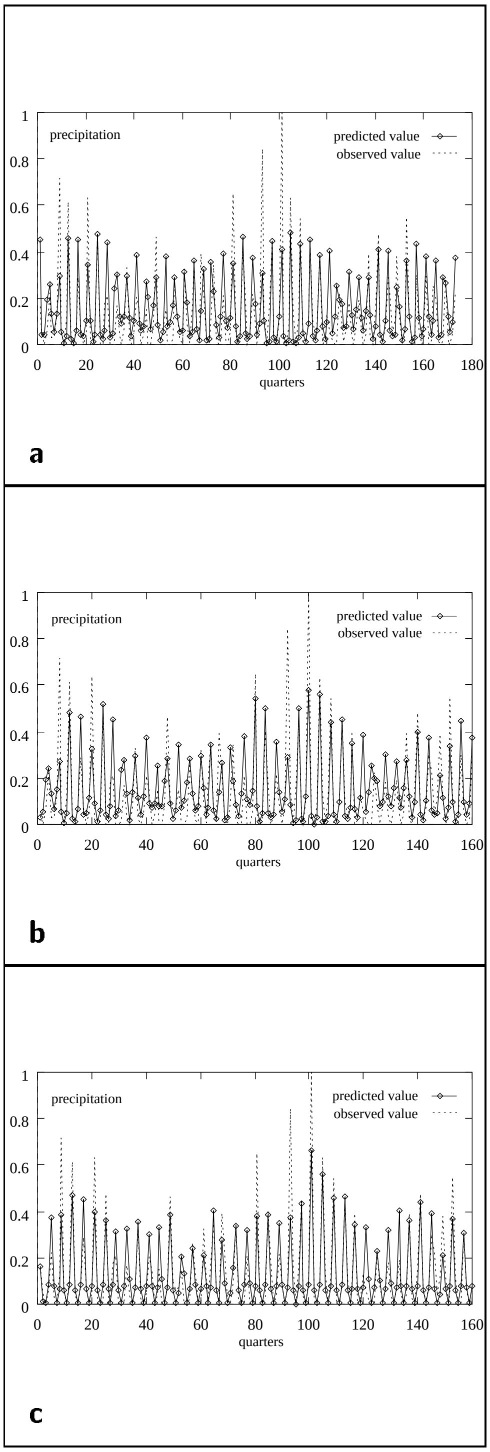

FIG. 5.2 - : Précipitation en Namibie : prédiction avec les réseaux AR, ARMA et DRNN en fonction des trimestres de 1949 à 1989 avec (a) réseau NAR et une fenêtre temporelle de taille 5, (b) réseau NARMA et deux fenêtres temporelles de taille 5 : une pour les observation passées et l'autre pour le bouclage des sorties, (c) le DRNN.

5.3 La température

Dans ce paragraphe, nous présentons les expérimentations réalisées sur la prédiction de température à La Silla et à Paranal au Chili. Une des fonctions de la station astronomique AWS est de fournir une estimation fiable de la température à l'ouverture du dôme au couché du soleil. Le dôme ainsi que l'instrumentation doivent être maintenus très précisément à la température extérieure. Il s'agit d'éviter que ne se crée un seeing de coupole, et un seeing des surfaces réflectrices, ou *mirror seeing*, qui résultent d'un gradient de température entre l'air ambiant et l'instrumentation. En d'autres termes, il faut éradiquer les sources de chaleur. Or l'inertie thermique de l'instrumentation est telle que la climatisation ne peut suivre l'évolution de la température en temps réel. C'est la raison pour laquelle une estimation fiable de la température quelques heures en avance, permettrait de commander optimalement le mécanisme de climatisation du dôme.

Depuis 1985, une station d'observation de l'environnement a enregistré un ensemble de variables météorologiques et atmosphériques autour des sites astronomiques. L'expansion continue du processus automatisé d'acquisition de données, réellement effectif depuis 1989, trace la voie à l'évaluation de diverses techniques prédictives. Cette longue séquence de mesures physiques offrent une opportunité de tester la performance des méthodes nouvelles qui s'inscrivent dans le champs d'application des réseaux connexionistes avec les techniques plus traditionnelles issues de l'univers statistique. De surcroît, l'absence d'une prédiction météorologique classique aux alentours des sites chilien rend indispensable le recours à ces techniques anticipatives. De toute les manières, lors même qu'une station météorologique serait implantée sur le site, l'expérience pratique et théorique des météorologistes ne suffit pas à établir une estimation locale précise de la température. Nous supposerons que les enregistrements effectués dans un passé récent portent en eux le potentiel pour prévenir l'évolution à court terme de la température. L'utilisation de modèles non linéaires est fortement motivée par les travaux préliminaires menés par Murtagh et al. [Murtagh 92, Murtagh 93] montrant l'inadéquation des modèles linéaires à la modélisation de la température.

Nous examinons le potentiel du DRNN à inférer les lois implicites qui gouvernent la dynamique de la température locale d'après les régularités présentes dans la séquence des données météorologiques passées. Les résultats obtenus avec le DRNN sont comparés à ceux obtenus avec des méthodes qui sont généralement l'apanage des météorologistes. Outre les modèles linéaires classiques, les techniques de reconstruction de l'espace des phases sont couramment employées. La plus simple d'entre elle, le *k-nearest neighbor*, peut être vue comme une implémentation algorithmique du "sens commun". L'idée clée est de segmenter l'espace de recherche en domaines contigus de telle sorte que la surface varie continuement d'un domaine à l'autre. Une approximation locale est alors obtenue en sélectionnant les k points les plus proches (i.e dont la distance Euclidienne est la plus faible). Enfin, une approximation affine fournit l'estimation recherchée.

En raison du nombre élevé de données disponibles et de la puissance sans cesse croissante de l'informatique moderne, il est désormais possible de s'aventurer sur le terrain des modèles dits "faibles" nécessitant beaucoup de degrés de liberté mais peu de connaissance préalable de la suite étudiée. Néanmoins, nous nous heurtons immanquablement aux problèmes que soulèvent le traitement préliminaire des données issues du monde réel parmi lesquels on compte

- les données manquantes ou incomplètes.
- les erreurs d'enregistrement.
- l'utilité inégale des différentes variables disponibles.
- les fluctuations de la période d'échantillonnage.
- le choix de la normalisation.
- le nombre restreint de données.

A defaut d'heuristiques, le bon sens est notre seul guide pour constituer une base d'apprentissage homogène qui ne soit pas entachée d'erreurs d'enregistrement.

5.3.1 La prédiction au couché du soleil

Nous examinons les problèmes rencontrés lors de la construction des bases d'apprentissage pour la prédiction de la température 24 heures en avance. Une leçon majeure à tirer de l'analyse des suites temporelles est que la *représentation* des données influence de façon décisive la qualité des prédictions obtenues, indépendamment du modèle employé. Une représentation judicieuse peut rendre explicite l'information masquée dans le flot apparemment confus des variables utilisées et mettre en relief les nuages de points opaques. Toutefois, le choix d'une représentation adéquate n'est pas aisé. A défaut de méthode standard, il existe une panoplie d'heuristiques permettant de sélectionner les variables pertinentes, de réduire la dimension de l'espace d'entrée ou de filtrer le bruit que présente les mesures. Bien que ces heuristiques soient le fruit de la pratique, les résultats obtenus sont parfois mitigés.

Les mesures météorologiques accumulées au fil des années forment un ensemble de 165 000 observations à La Silla, et de 140 000 observations à Paranal. Ces mesures sont moyennées sur 20 minutes depuis 1985 dans le cours normal des opérations. Malencontreusement, la séquence des mesures a été interrompue à plusieurs reprises en raison de la construction du site, et des pannes informatiques d'autre part. Dans cette évaluation, une succession de variables atmosphériques mesurées au couché du soleil à 24 heures d'intervalle est employée pour prédire la température au prochain couché du soleil. Nous retenons que les données au couché du soleil (i.e. lorsque le sinus de l'élévation du soleil est nul et de pente négative) lorsque le dôme du télescope est ouvert pour l'observation. Au reste, la prédiction à 24 heures permet d'ignorer les effets diurnaux.

Il est bien connu en théorie du signal que la différentiation contribue à la stationnarité des signaux. Aussi est-il plus astucieux de chercher à prédire la variation de la température entre deux couchés du soleil consécutifs plutôt que de prédire la température elle même. Puisque le DRNN est capable de conserver la mémoire de ses états passés, les variables qu'il convient de présenter à son entrée se limitent aux mesures effectuées au coucher du soleil au jour n pour prédire la variation de température $\Delta t_{n+1} = t_{n+1} - t_n$ au jour $n+1$. Parmi les possibilités qui s'offrent à nous (un vingtaine de mesures) pour définir un vecteur d'entrée de taille raisonnable, le choix suivant rencontre l'adhésion des météorologistes

$$t_n, \Delta p_n, s_n, \sigma_{sn}, d_n \tag{5.1}$$

où

- t_n est la température à 2m au dessus du sol.
- p_n est la pression.
- s_n est la vitesse du vent à 10m au dessus du sol.
- d_n est la direction du vent, exprimée en degrés, à 10m au dessus du sol.
- σ_{sn} représente l'écart type de s_n calculée sur une heure au couché du soleil.

La variation de la pression est révélatrice de l'arrivée d'un front climatique. La vitesse et la direction du vent jouent un rôle déterminant dans l'échange thermique sol-atmosphère.

Les valeurs ont été linéairement normalisées dans l'intervalle [0,1]. Une fonction *tanh* a aussi été employée mais sans succès. Les histogrammes Fig. 5.4 et 5.6 reflètent les distributions non uniformes des variables employées.

Des routines ont été implémentées pour

- constituer une séquence de vecteurs d'entrée à 24 heures d'intervalle.
- radier les vecteurs incomplets.
- éliminer les anomalies ($\Delta t_n < 9^o C$).
- normaliser chaque variable.

Les observations avec au moins une valeur manquante ont été éliminées. Le nombre d'exemples a été réduit dramatiquement par le seul fait des ruptures de séquence. Un filtrage semblait nécessaire d'après les erreurs erratiques de mesure dont l'origine remonte soit aux capteurs physiques, soit plus en aval du processus d'acquisition de données (e.g. erreur de formattage, de manipulation, etc.). Pour cette raison, nous ne retenons que les données telles que $\Delta t_n < 9^o C$ en accord avec une analyse préliminaire de données entreprise sur les sites d'observation.

La quantité de données extraites de la base de données est représentée Fig. 5.3 en fonction de l'époque. Les irrégularités manifestes durant les premières périodes de la campagne d'acquisition de données témoignent des obstacles rencontrés.

Une séquence de 1240 exemples d'apprentissages présentant 124 coupures de séquences a été constituée à La Silla. Les 250 derniers exemples de 1993 on été mis à l'écart pour constituer une base de prédiction sur laquelle seront évaluées les méthodes de prédictives. Seuls 620 exemples ont pu être retenus à Paranal, avec 66 ruptures de séquence. Les derniers 110 exemples ont été mis à l'écart pour constituer une base de prédiction.

5.3.2 Analyse statistique

La stationnarité est une propriété essentielle des suites temporelles. Comme le révèle l'inspection du tracé brut de la variation de température Fig. 5.5.a et Fig. 5.7.a, la non-stationnarité est manifeste même si la différenciation effectuée a très certainement adoucie la dépendance saisonnière du signal. Deux fonctions fondamentales dans le contexte de l'analyse des suites sont tracées : la fonction d'autocorrelation des sorties désirées et la fonction d'intercorrelation des sorties désirées avec les variables d'entrée Fig. 5.8. La vitesse de décroissance de la fonction d'autocorrelation est une mesure approximative de la "mémoire" de la suite. La fonction d'autocorrelation converge rapidement vers l'origine à mesure que le pas augmente, signe du peu d'information présent dans la séquence Δt_n au-delà de 4 jours. Il semble, en outre, qu'aucune périodicité ne soit présente dans la suite[1]. En somme, l'évanouissement de la fonction d'autocorrelation suggère l'utilisation d'une fenêtre temporelle réduite pour une prédiction du type NAR. Mais il n'y a, en principe, pas de moyen rigoureux de choisir la taille optimale de la fenêtre au regard de la fonction d'autocorrelation.

Par ailleurs, la fonction d'intercorrelation entre Δt_n et les 4 variables d'entrée n'indiquent pas de dépendance mutuelle (Fig. 5.8). Δt_n exhibe une distribution de type Gaussienne (Fig. 5.4.a et 5.6.a) et les histogrammes des autres variables sont particulièrement représentatifs. Notons par exemple la claire dichotomie de la direction du vent, signe de la prédominance de l'orientation nord-sud.

Les pics de faible amplitude (Fig. 5.8.e et 5.8.f) qui apparaissent sur les densités spectrales de puissance ne révèlent rien si ce n'est l'effet néfaste des ruptures de séquences mêlé au comportement aléatoire de Δt_n.

5.3.3 Résultats

Dans ce paragraphe, nous exposons les résultats obtenus sur la prédiction au couché du soleil. Les détails concernant les paramètres d'apprentissages ont été vus au § 4.3. Le DRNN employé est un réseau à trois couches $5 \times 5 \times 1$. Les neurones cachés sont complètement interconnectés par des connexions d'ordre 1. A celles-ci s'ajoutent des connexions non bouclées d'ordre 2 entre la couche d'entrée et la couche cachée, et des connexions d'ordre 0 entre la couche cachée et las couche de sortie (l'ordre 0 représente des liens statiques). Le réseau compte 136 liens, inclusion faite des biais. 136 degrés de liberté doivent être adaptés pour approximer une fonction inconnue représentée par un ensemble de 600 exemples à la Silla et de 1000 exemples à Paranal.

La sélection de ces dimensions repose essentiellement sur des essais préliminaires et sur le "bon sens". Remarquons que la sélection optimale des dimensions du réseau est un problème difficile qui fait l'objet de recherches intensives. Dans notre cas, l'ordre des connexions a été choisi en fonction de la prompte décroissance de la courbe d'autocorrelation de Δt_n. Seuls les 2 précédents vecteurs d'entrée sont présentés au réseau à chaque itération; les autres sont codés et mémorisés en interne par les neurones cachés grâce à la

[1]la fonction d'autocorrelation des suites non-périodiques et stationnaires a tendance à décroître exponentiellement.

dynamique induite par les connexions retardées et bouclées.

A défaut d'optimiser l'architecture du réseau, nous nous sommes concentrés sur la réduction de l'erreur quadratique avec cette configuration. Une prédiction "carbon-copy" (i.e. la précédente valeur désirée sert d'estimation courante, $\hat{x}_t = x_{t-1}$), est employée pour fournir un index basique de performance. Par ailleurs, la méthode du *nearest neighbor* n'améliore que très sensiblement les résultats obtenus avec la copie carbone comme le montre Table 5.2. Curieusement, des prédictions de meilleure qualité sont obtenues sur la base de prédiction plutôt que sur la base d'apprentissage à La Silla (Table 5.3). Cela s'explique probablement par le fait que la base d'apprentissage précède la base de prédiction; or les erreurs de mesures furent plus fréquentes dans le passé en raison de problèmes techniques résolus depuis. Il résulte que la base de prédiction à La Silla présente plus de régularités que la base d'apprentissage. A l'opposée, l'erreur normalisée obtenue à Paranal est supérieure à 1.0; lors même que le réseau aurait inféré des règles inhérentes au comportement de la température dans la base d'apprentissage, celles-ci ne s'appliquent manifestement pas à la période de prédiction.

Expérimentation	2 degrés	1 degré	0.5 degré
Réseau neuronal	72.5%	42.5%	20.0%
nearest neighbor	59.8%	37.8%	18.3%
Carbon copy $\hat{t}_n = t_{n-1}$	56.7%	31.3%	18.8%

TAB. 5.2 - : Résultats de la prédiction de la température à 24 heures au couché du soleil à La Silla.

Fig. 5.9 illustre les capacités du DRNN à anticiper les variations de température aux couchés successifs du soleil. Les histogrammes cumulatifs mettent en relief la simple copie carbone avec les prédictions neuronales. Les qualités prédictives du DRNN ne sont pas franchement perceptibles sur ce problème, exception faite de la prédiction de température à La Silla où les variations température sont en règle générale bien anticipées par le réseau.

Les efforts qu'il a fallu fournir pour constituer une base d'apprentissage *relativement*

		NMSE	
à prédire	*site*	apprentissage	prédiction
Δt_n	La Silla	0.75	0.70
Δt_n	Paranal	0.79	1.20
seeing	Paranal	0.81	0.50

TAB. 5.3 - : Erreur quadratique normalisée de la prédiction neuronale sur la base d'apprentissage et de prédiction au couché du soleil.

homogène et cohérente, contrastés aux résultats médiocres auxquels nous aboutissons, ne nous encouragent pas à poursuivre cette voie. Cela confirme du reste le caractère laborieux et relativement infructueux de la prédiction météorologique locale à 24 heures d'intervalle au couché du soleil.

Afin d'augmenter considérablement le nombre de données exploitables et de permettre une comparaison qui soit plus à la faveur des techniques de reconstruction locales, certaines mesures sont adoptées dans le paragraphe suivant.

5.3.4 La prédiction à 24 heures toutes les heures

Selon l'analyse précédente, il semble que les résultats de la prédiction météorologique au couché du soleil soient quelque peu opaques. Il n'est pas clair jusqu'à présent si la base de données contient suffisamment d'information pour modéliser correctement le comportement de la température. Nous n'avons au reste aucune certitude sur la pertinence de la vitesse et de la direction du vent dans la prédiction de la température. L'absence de variables passées dans le vecteur d'entrée pénalise les techniques de reconstruction locale telle que le *nearest neighbor*.

Dans ce paragraphe, nous substituons dans le vecteur d'entrée les variables relatives au vent par des valeurs retardées de la température et de la pression. De plus, la prédiction à 24 heures est déterminée à chaque heure de la journée afin de démultiplier le nombre de données disponibles. Pour minimiser les variations saisonnières, les données utilisées se limitent à l'été austral 1989-1990 durant lequel environ 18000 observations ont été effectuées.

A la sortie désirée Δt_n est associée le vecteur d'entrée suivant

$$t_{n-1}, t_{n-2}, t_{n-3}, p_{n-1}, p_{n-2}, \tag{5.2}$$

Les indices $n-1, n-2$ et $n-3$ désignent respectivement les 24, 48 et 72 heures précédentes. Les procédures d'élagage et de filtrages conduisent à une base de données finale constituée de 5260 tuplets. Les premiers 5000 exemples servent de base d'apprentissage au réseau de neurones. Le reste est utilisé comme base de prédiction.

Parallèlement à l'approche neuronale, la méthode du *k-nearest neighbor* est employée. La méthode consiste à rechercher parmi la séquence des vecteurs de la base, les k termes dont la distance Euclidienne est la plus faible; la moyenne arithmétique des Δt_k associé au vecteur constitue l'estimation de Δt_n. Pour refléter une implémentation réelle, les prédictions ont été menées pour toutes les modèles selon

- E1, la méthode du *k-nearest neighbor* est appliquée aux 1260 données les plus récentes. Le vecteur le plus proche est recherché parmi tous les vecteurs à condition qu'ils soient chronologiquement antérieurs.

- E2, la méthode du *k-nearest neighbor* est appliquée aux 1260 données les plus récentes. Le vecteur le plus proche est recherché parmi les 4000 premiers vecteurs dans la base de données.

- E3, la méthode du *k-nearest neighbor* est appliquée comme en E1 excepté que les termes relatif à la pression sont ignorés dans la distance Euclidienne, seule la température est prise en compte.

- E4, un réseau de neurones statique non bouclé à quatre couches $5 \times 14 \times 14 \times 1$ est utilisé.

- E5, un réseau de neurones statique bouclé à quatre couches $5 \times 10 \times 10 \times 1$ est utilisé. Les connexions rétrogrades relient la troisième couche à la seconde.

- E6, la prédiction se réduit à une "copie carbone" : la précédente valeur désirée Δt_{n-1} sert d'estimation courante Δt_n.

- E7, un modèle auto-régressif multi-dimensionnel (ARX) du type

$$t_n = a_1 t_{n-1} + a_2 t_{n-2} + a_3 t_{n-3} + a_4 p_{n-1} + a_5 p_{n-2} + \epsilon \qquad (5.3)$$

  est employé dans lequel ϵ, est un bruit blanc (i.e. processus Gaussien centré indépendant et identiquement distribué). Les coefficients $a_j, j = 1, \ldots, 5$ sont estimés par la méthode des moindres carrés [Press 88].

L'approche E2, certes moins naturelle, permet de comparer la méthode du *k-nearest neighbor* et la prédiction neuronale sur des bases plus rigoureuses. Les 4000 exemples maintenus à l'écart jouent le rôle de base d'apprentissage dans la terminologie connexioniste. L'expérimentation E3 est conduite pour mesurée l'influence de la pression sur la qualité des prédictions. Les deux réseaux connexionistes statiques, dont l'un est récurrent, comportent chacun environ 300 connexions. Voici un exemple qui met en évidence la versatilité du DRNN et de la procédure TRBP. Dans le cas E4, le DRNN se réduit à un réseau statique non-bouclé tandis que TRBP agit exactement comme la rétro-propagation classique [Rumelhart 86] que nous avons rappelée au § 2.4. Pour E5, il s'agit d'un réseau statique récurrent adapté avec l'algorithme de rétro-propagation récurrente [Almeida 87, Pineda 87, Rohwer 87] vu au § 2.6. Les 4000 premiers exemples constituent la base d'apprentissage, et les 1260 restant forment la base de prédiction.

Table 5.4 résume d'une façon concise la qualité des prédictions obtenues avec les techniques que nous venons d'évoquer. Il apparaît que la présentation spatiale de la mémoire a bouleversé les résultats obtenus au paragraphe précédent. La méthode du *nearest neighbor* génère des prédictions nettement plus précises que les autres techniques. Cela indique que les variables retardées que nous avons substituées ont une influence majeure sur la qualité des résultats. La différence observée entre les réseaux de neurones et la méthode du *nearest neighbor* laisse supposer que la fonction sous-jacente se prête mal à une approximation globale. A ce titre, il est intéressant de noter que le modèle linéaire ARX fournit des prédictions plus erronées que ne le fait une simple copie carbone.

Par ailleurs, les performances des réseaux de neurones statiques sont très similaires à celles des réseaux dynamiques employés au paragraphe précédent. Cela montre plusieurs choses. En premier lieu, l'influence des variables relatives au vent est imperceptible; il semble que seules la température et la pression soient des grandeurs déterminantes. Ensuite, le réseau dynamique ne fait manifestement pas usage d'information à long terme.

On notera aussi que E1 donne de meilleurs résultats que E2 puisque plus d'information historique a été utilisée dans le premier cas. Il y a toutes les raisons à préférer l'arrangement associé à E1 plutôt que E2 dans un scénario pratique. L'utilité de la pression est clairement établie au vu des résultats obtenus avec E3 en contraste à E1.

Il est fort probable que la non-stationnarité évidente due aux variations diurnales soit à l'origine des différences observées. Il s'avère en effet que dans le cas de la reconstruction locale, les vecteurs similaires recherchés parmi les vecteurs passés soient dans la même plage horaire. A l'opposé, le réseau de neurones doit sans cesse s'adapter à des conditions changeantes.

Nous nous garderons de tirer des conclusions définitives. Toutefois, dans le cas de la prédiction à chaque heure de la journée, il semble qu'une fenêtre temporelle de taille réduite appliquée aux données météorologiques soit appropriée aux techniques de reconstruction locale et que les variations diurnales pénalisent les méthodes neuronales.

Expérimentation	2 degrés	1 degré	0.5 degré
E1 - nearest neighbor 1	83.6%	72.7%	62.3%
E2 - nearest neighbor 2	66.7%	48.6%	32.3%
E3 - nearest neighbor 3	68.1%	50.2%	37.1%
E4 - RN $5 \times 14 \times 14 \times 1$ non-bouclé	71.1%	45.1%	22.0%
E5 - RN $5 \times 10 \times 10 \times 1$ bouclé	69.7%	41.8%	21.2%
E6 - copie carbone $\hat{t}_n = t_{n-1}$	63.2%	36.3%	18.8%
E7 - modèle ARX	63.7%	35.8%	16.3%

TAB. 5.4 - : Prédictions de la température à 24 heures à toutes les heures de la journée. Voir le texte pour les détails.

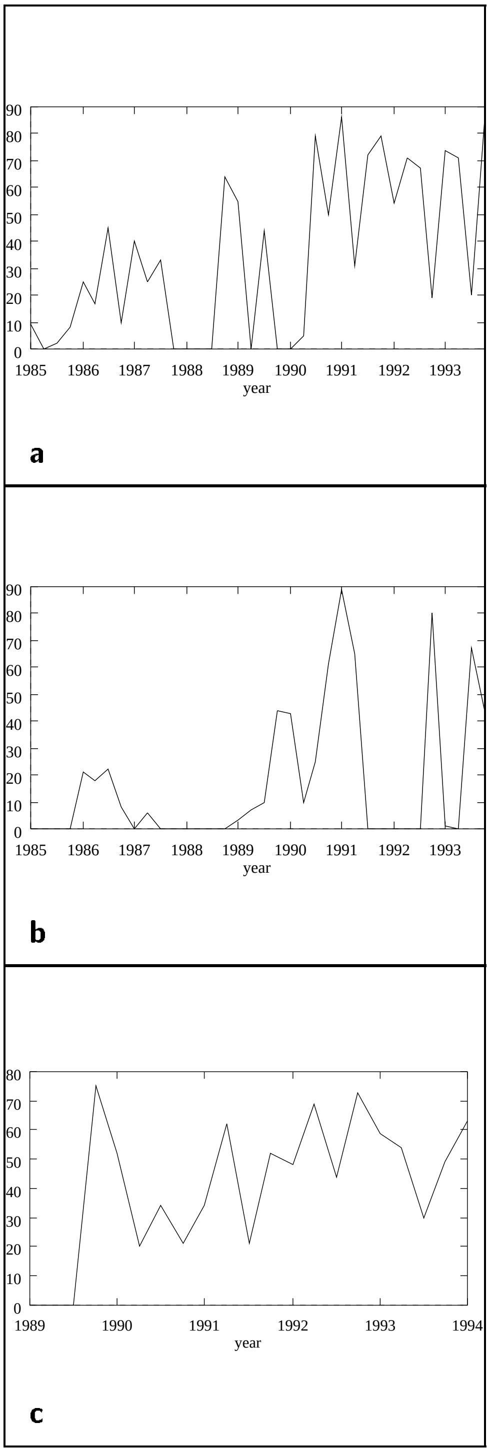

FIG. 5.3 - : Nombre de données exploitées en fonction du temps : (a) température à La Silla, b) température à Paranal, et c) seeing.

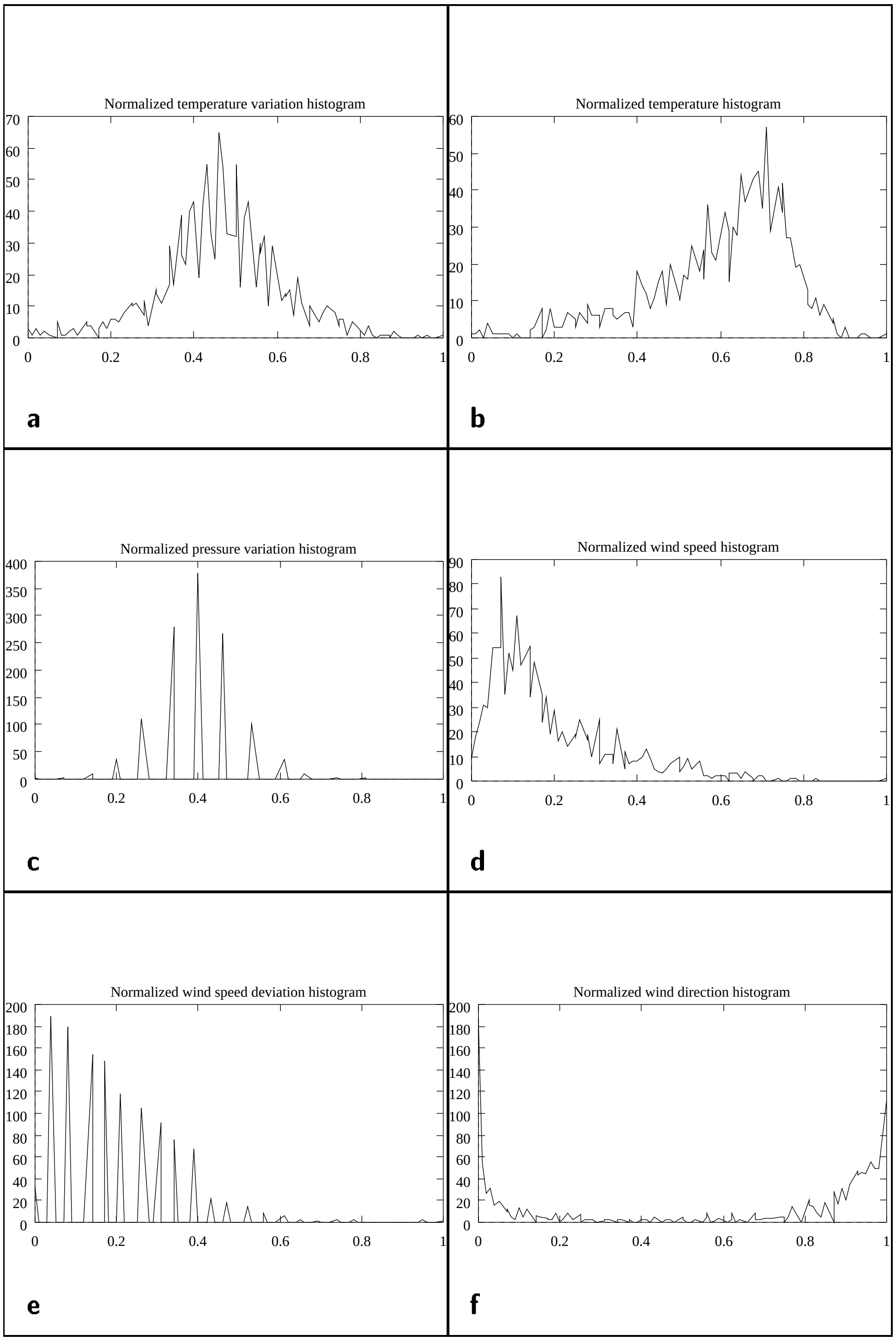

FIG. 5.4 - : Histogrammes cumulatifs des données météo à La Silla au couché du soleil : (a) variation de température, (b) température, (c) variation de pression, (d) vitesse du vent, (e) écart type de la vitesse du vent, (f) direction du vent.

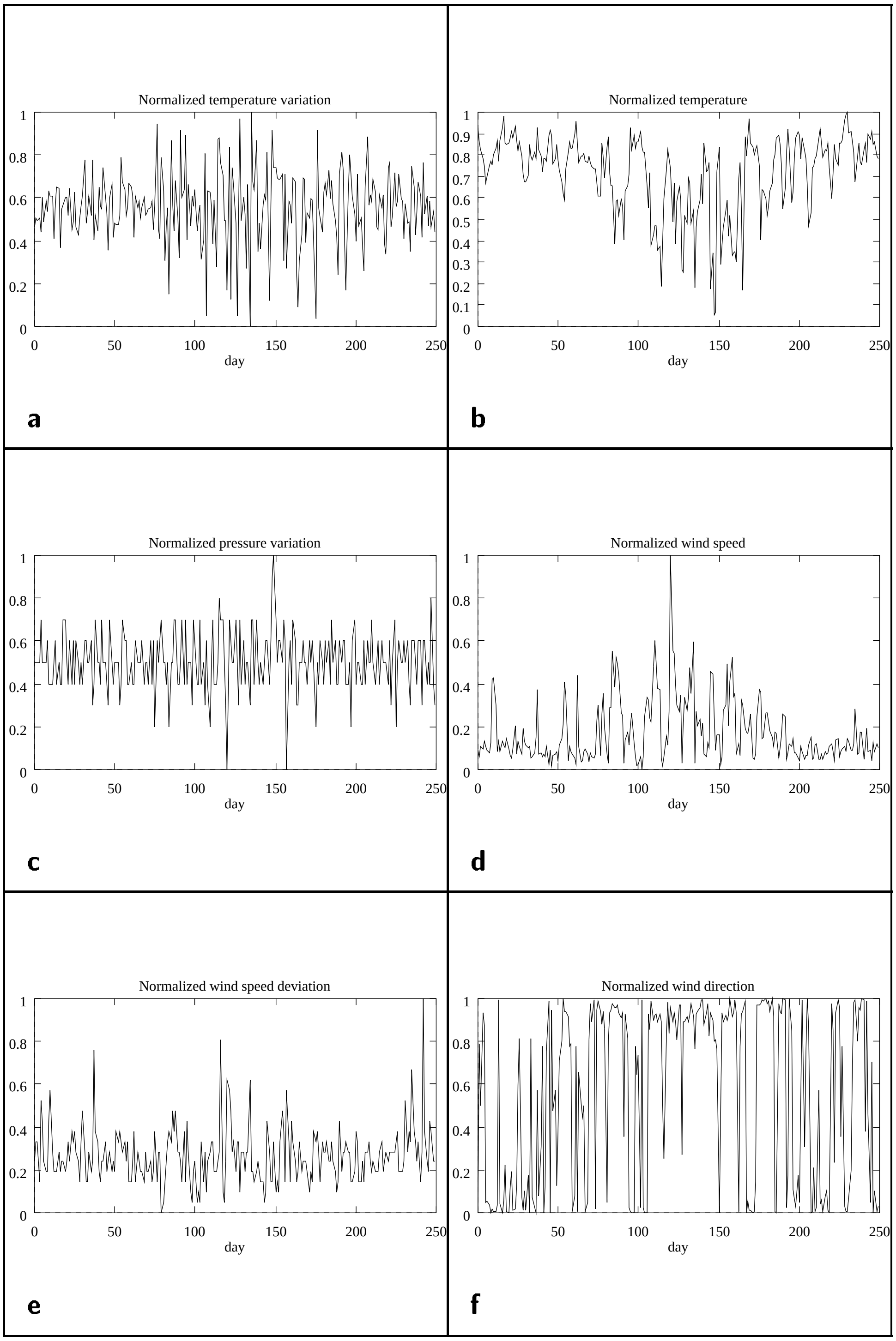

FIG. 5.5 - : Données météo au couché du soleil linéairement normalisées en 1993 à La Silla : (a) variation de température, (b) température, (c) variation de pression, (d) vitesse du vent, (e) écart type de la vitesse du vent, (f) direction du vent.

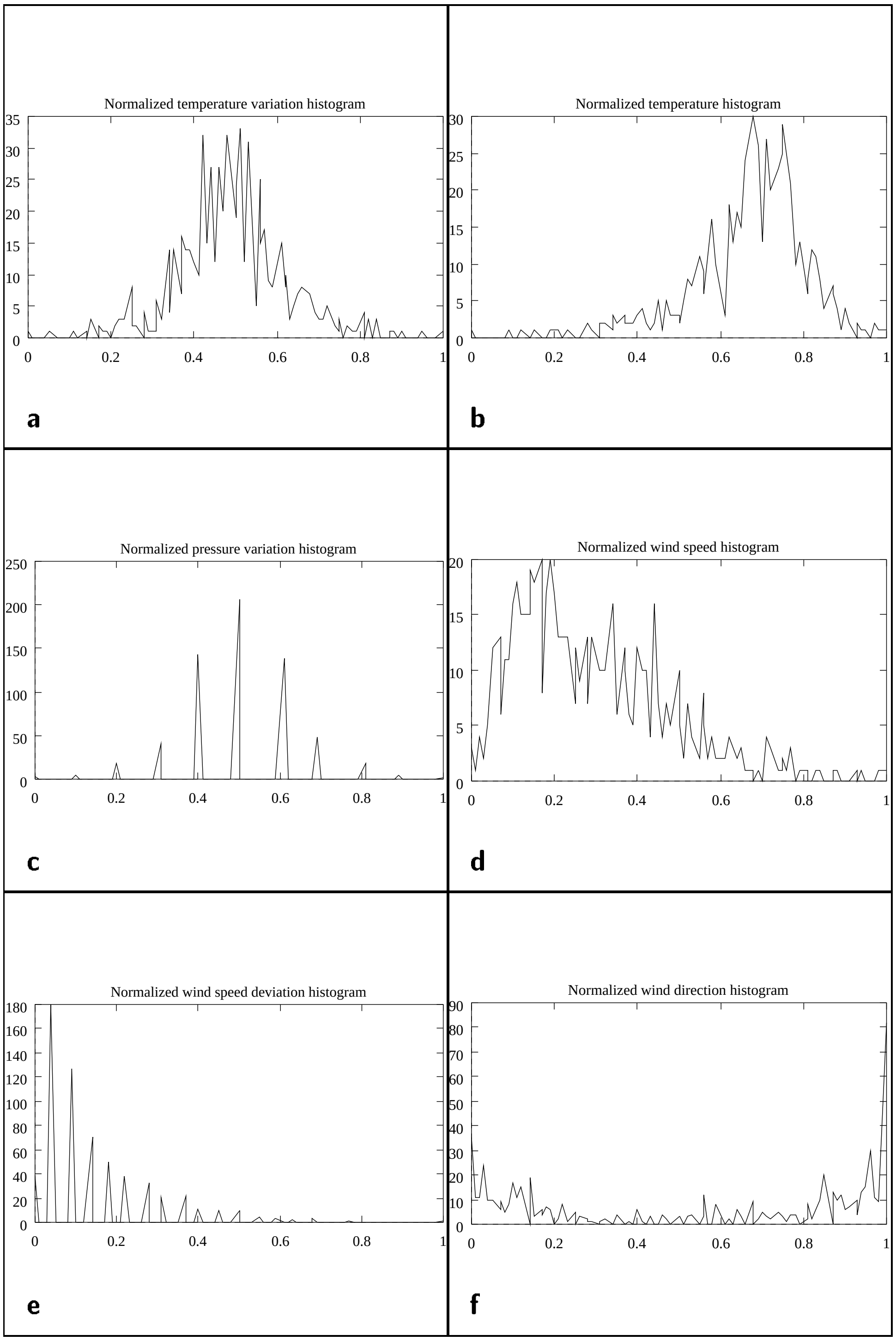

FIG. 5.6 - : Histogrammes cumulatifs des données météo à Paranal au couché du soleil : (a) variation de température, (b) température, (c) variation de pression, (d) vitesse du vent, (e) écart type de la vitesse du vent, (f) direction du vent.

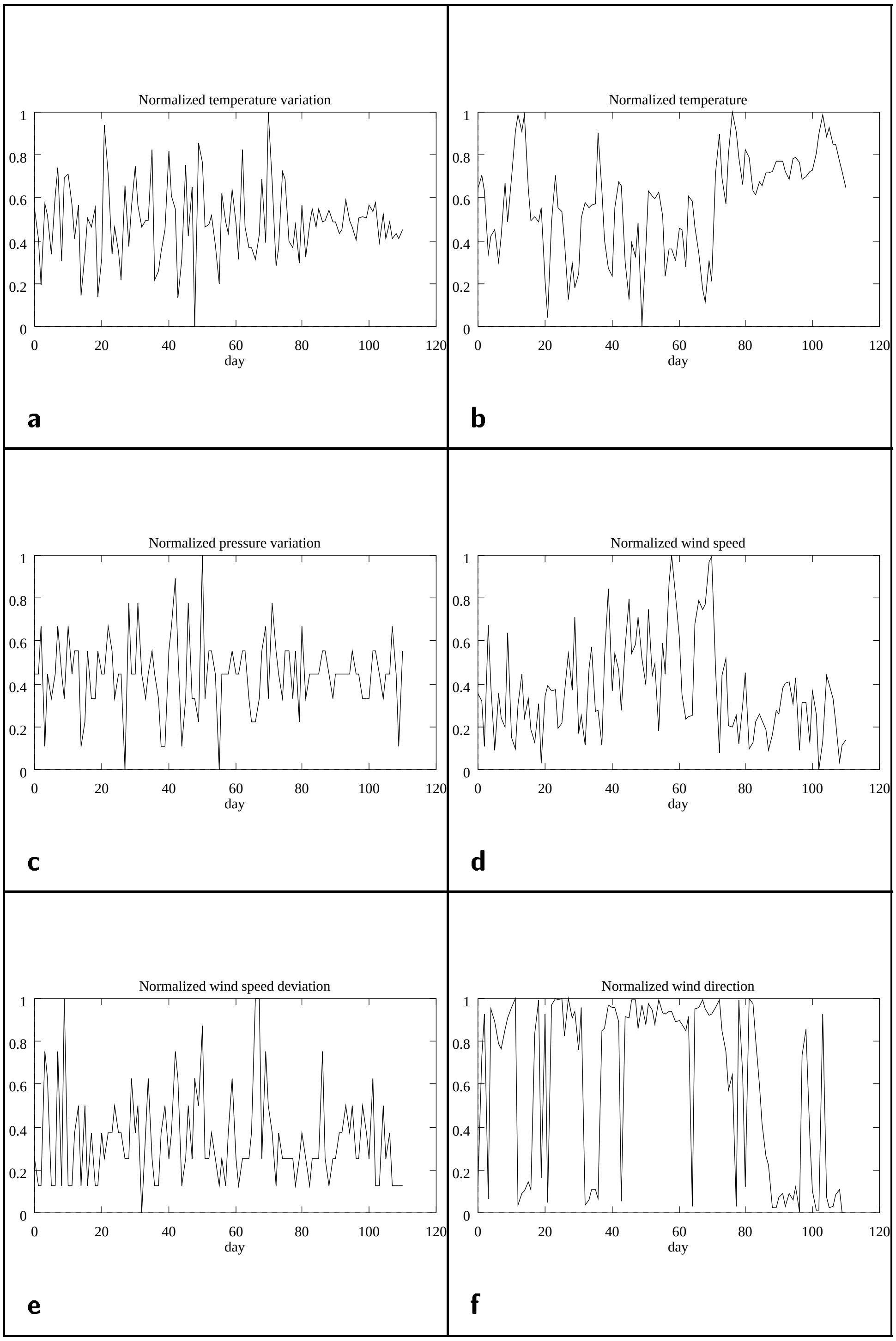

FIG. 5.7 - : Données météo au couché du soleil linéairement normalisées en 1993 à Paranal : (a) variation de température, (b) température, (c) variation de pression, (d) vitesse du vent, (e) écart type de la vitesse du vent, (f) direction du vent.

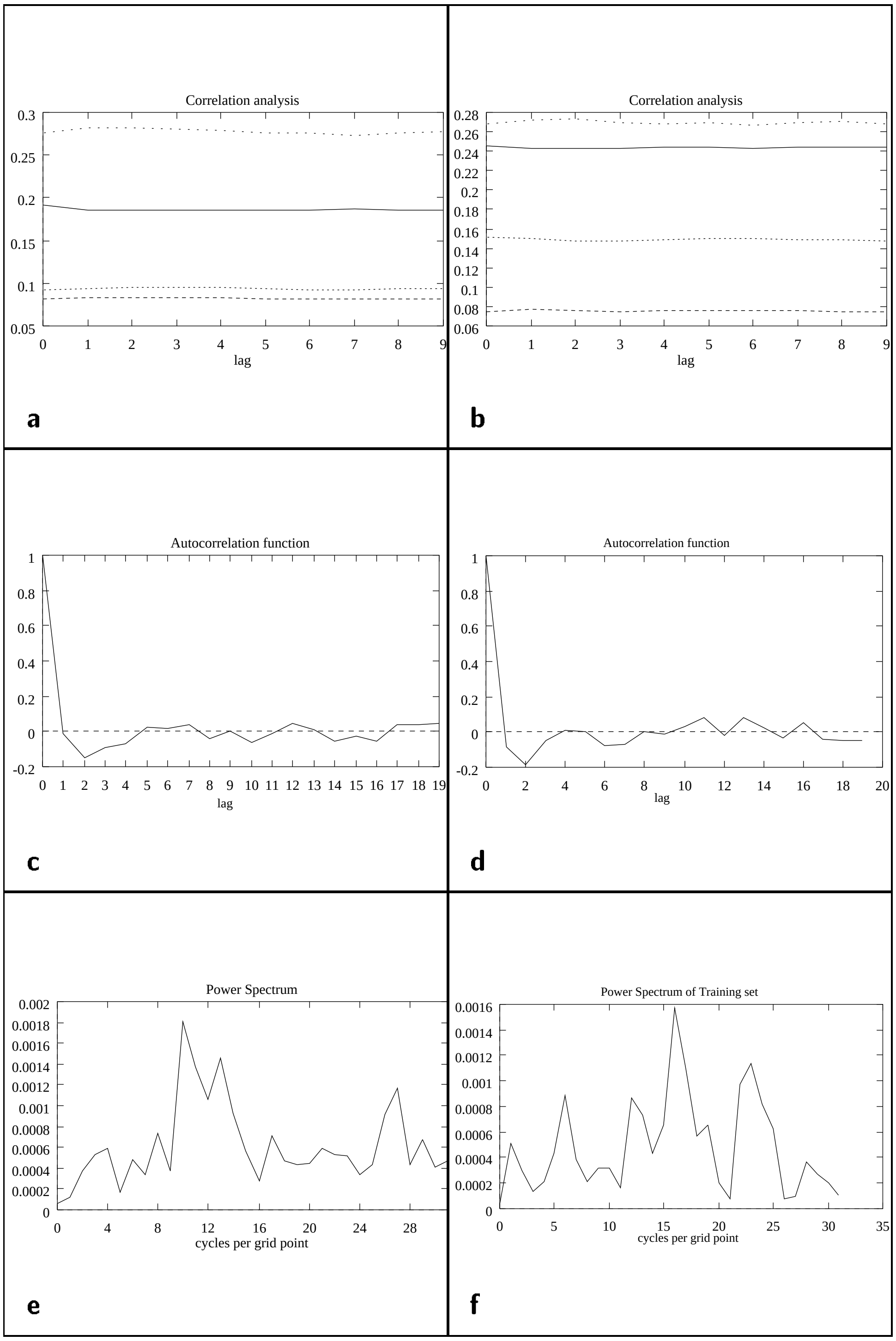

FIG. 5.8 - : Gauche : La Silla - Droite : Paranal.(a),(b) fonction d'intercorrelation entre Δt_n et d_n, Δp_n, s_n et σ_{s_n} en partant du haut. (c),(d) fonction d'autocorrelation de Δt_n. (e),(f) spectre aux fréquences $j * f_c/32$ ($f_c = 1/48\ h^{-1}$ est la fréquence de Nyquist)

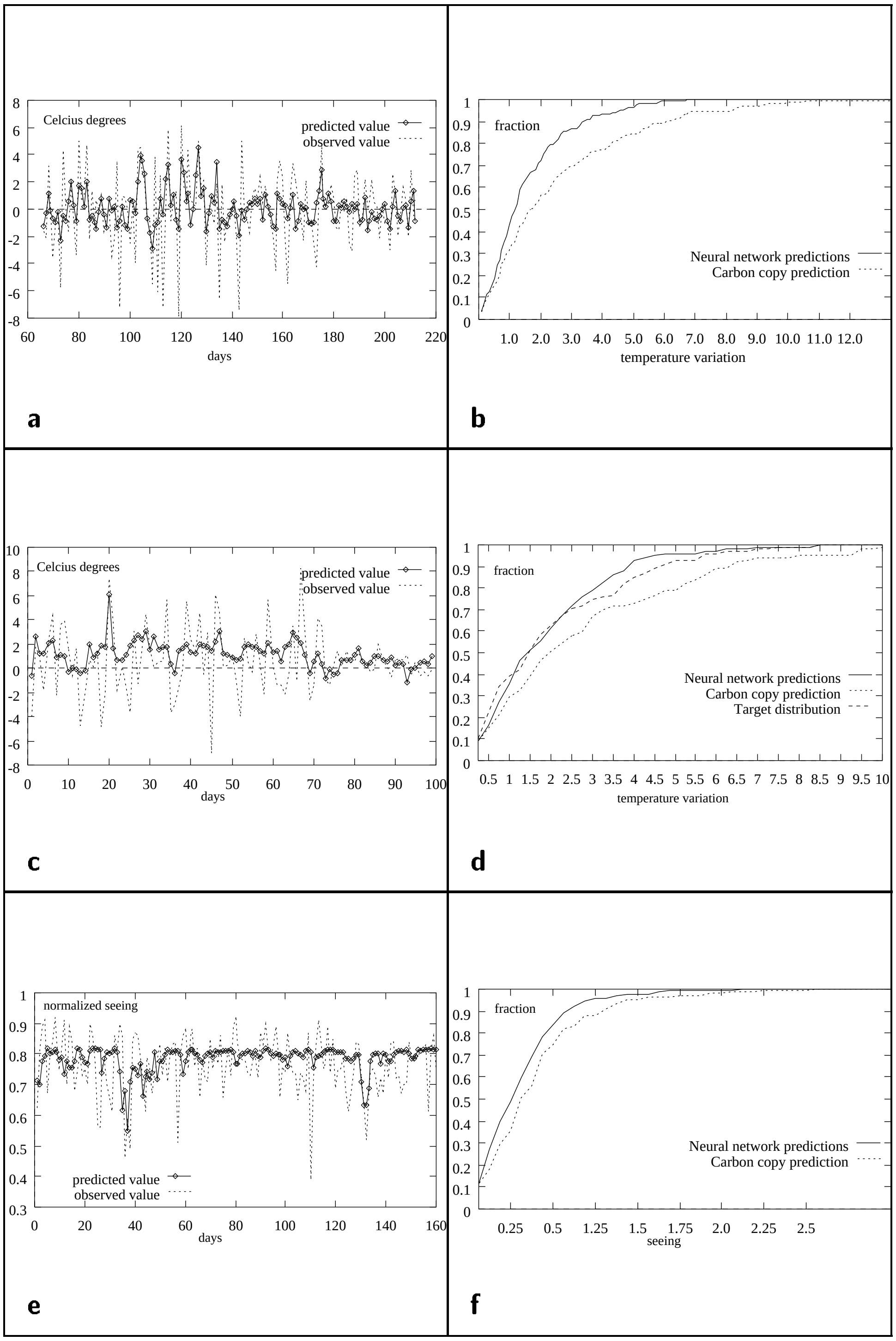

FIG. 5.9 - : Prédiction avec le DRNN et histogrammes cumulatifs correspondants : (a) et (b) Δt à La Silla, (c) et (d) Δt à Paranal, (e) et(f) seeing moyen à Paranal.

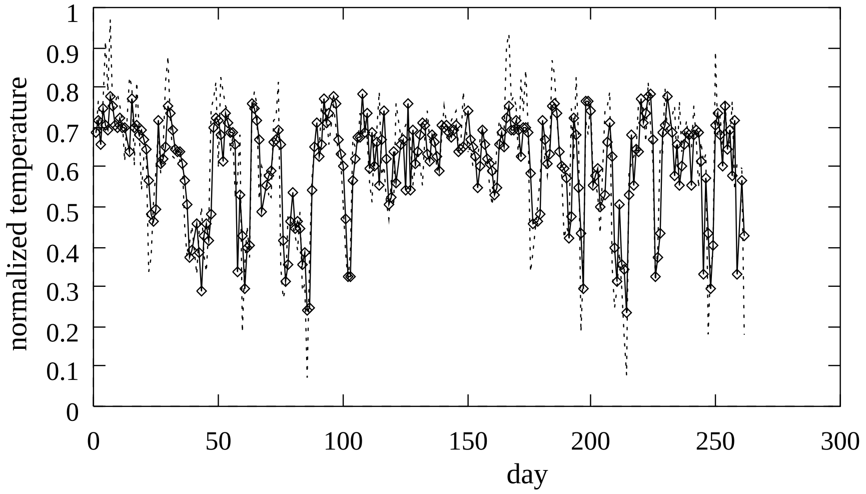

FIG. 5.10 - : Prédiction de Δt 24 heures en avance à chaque heure de la journée à Paranal. En pointillé, la variation observée; en trait plein, la prediction.

5.4 Le seeing

5.4.1 La prédiction à 24 heures du seeing moyen

Dans la continuité des expérimentations menées sur la température, nous cherchons à estimer le seeing astronomique 24 heures en avance sur le site de Paranal de 1989 à 1994. L'analyse menée par Murtagh et al. [Murtagh 93] montre qu'il est illusoire de chercher à prédire la valeur quantitative du seeing 24 heures en avance compte tenu de l'absence d'une relation claire entre les conditions météorologiques au sol avec le seeing. Par ailleurs, Jorgenson et al. [Jorgenson 91], dans leur analyse de la turbulence atmosphérique, laissent entendre que les variations du front d'onde incident présentent certains aspects d'un processus chaotique.

Lorsqu'on regarde le comportement erratique et confus du seeing et, en sus, le nombre relativement faible de mesures dont nous disposons, il parait sage de se restreindre à l'estimation du seeing moyen au cours de la nuit suivante dans cette étude préliminaire. L'objet de cette étude est de fournir une indication vague, un jour en avance, de la qualité d'observation. Etant donnée la difficulté de la tâche dévolue au réseau de neurones, une large éventail de variables météorologiques est injecté en entrée du réseau. L'histogramme du seeing moyen, tracé Fig. 5.11, ressemble à une distribution log-normale rencontrée fréquemment dans les phénomènes naturels.

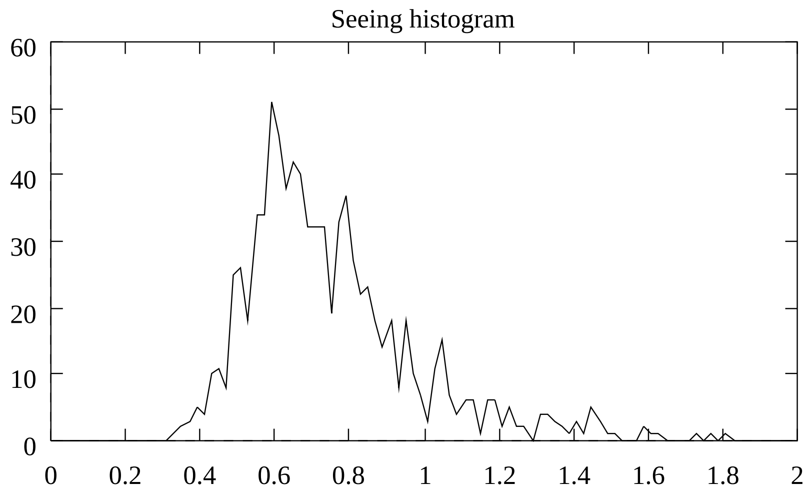

FIG. 5.11 - : Histogramme du seeing moyen à Paranal entre 1989 et 1994.

Pour la prédiction du seeing moyen au cours de la nuit qui sépare le jour n du jour $n+1$, le tuplet suivant est constitué de

$$\max(t_n),\ \max(p_n),\ \max(s_n),\ \max(\sigma_{sn}),\ \max(rh_n),$$
$$\min(t_n),\ \min(p_n),\ \min(s_n),\ \min(\sigma_{sn}),\ \min(rh_n),$$
$$\mathrm{aver}(d_n)$$

max, *min* et *aver* désignent respectivement le maximum, le minimum et la valeur moyenne entre le lever et le coucher du soleil au cours de la journée n. Le seeing moyen n'est calculé que lorsque plus de 4 mesures ont été faites durant la nuit. 860 vecteurs ont été collectés à partir de la base de données originale. Parmi les 860, 160 ont été mis de côté pour évaluer les performances du modèle.

Un réseau statique non-bouclé à 3 couches $11 \times 11 \times 1$ entièrement connecté est employé; le réseau compte 144 poids au total, inclusion faite des biais. Un réseau à délais ne s'avère pas nécessaire dans ce cas de figure. Fig. 5.9 illustre les performances du réseau de neurones. Hormis une erreur quadratique relativement faible (Table 5.3), le tracé des prédictions révèle que dans certains cas isolés, le bon seeing est anticipé correctement par le réseau. Ces relations, même très ponctuelles, entre la météo au sol et le bon seeing augurent d'ores et déjà des résultats encourageants lorsque les variables d'entrée seront filtrées et codées selon les principes de la logique floue.

5.4.2 La prédiction floue du seeing

Compte tenu des résultats mitigés de la prédiction à 24 heures du seeing moyen, une orientation différente est choisie. Dans l'optique d'améliorer la qualité de l'information présentée au réseau, le problème est scindé en deux partie : i) la prédiction des conditions météorologiques locales un jour en avance, et ii) le "nowcasting" du seeing, c'est-à-dire l'anticipation du seeing en fonction de données météo *au même instant*. Notre étude se limite à la seconde partie. En termes statistiques, le nowcasting est une régression multi-dimensionnelle, i.e. la régression du seeing est effectuée sur l'ensemble des variables météorologiques. Nous laissons à une station météorologique ou à un modèle prédictif *ad hoc* le soin de nous fournir des estimations fiables.

Les variables météorologiques

Les variables météorologiques employées sont :

- la direction du vent.
- la vélocité du vent à 10m et à 2.5m au-dessus du sol.
- l'écart type de la vélocité du vent sur 20min avec une période d'échantillonnage de 2s.
- le seeing astronomique à 5m.
- la température, la pression et l'humidité à 2m.
- la température au sol à -0.1m.

Les corrélation importantes entre ces variables nous amènent à ne retenir que

- la vélocité du vent (ms^{-1}) et l'écart type à 10m au-dessus du sol.

– le seeing astronomique à 5m.

– la température, la pression (mB) et l'humidité (%) à 2m.

Le seeing est défini par la largeur à mi-hauteur de l'image stellaire avec un télescope parfait de 0.5 μm de longueur d'onde au zénith. Une valeur médiane durant la nuit est calculée pour chacune de ces variables. La médiane s'interprète comme une moyenne non pondérée par les valeurs que prennent ces variables.

Les données utilisées s'étendent du printemps austral 1989 à l'automne 1991 à Cerro Paranal. En raison de la construction du site d'observation à Paranal, les mesures au cours des quatre saisons suivantes ont été effectuées à Cerro NTT, une autre site Chilien. Les données rassemblées forment une base de 605 tuplets.

Un codage flou référencé aux saisons

Le problème de la représentation de l'information joue un rôle déterminant. C'est la raison pour laquelle nous avons adopté une forme codage intuitive et facilement interprétable. Les variables ont été segmentées en trois catégories : une valeur faible (= 0), une valeur modérée dans l'intervalle $]0,1[$, et une valeur élevée (= 1). Les seuils de 33% et 67% délimitent ces catégories de façon à repartir équitablement les données. En d'autres termes, on compte 33% de 0, 33% de de 1, et 34% de valeurs intermédiaires. Il est fort probable que les estimations météorologiques catégorisées, et donc relativement vagues, puissent être générées systématiquement par des modèles météorologiques relativement simples.

Le codage du seeing est référencé à la saison. Une saison est un ensemble d'environ 90 nuits consécutives. A l'issue du filtrage et de la sélection des tuplets valides, seules 50 nuits environ demeurent exploitables. Aussi, nous choisissons de déterminer les seuils de 33% et 67% en fonction des 50 dernières nuits. Cette approche est robuste : les valeurs extrêmes ne jouent pas un rôle disproportionné. En codant le seeing de cette manière, on peut associer le bon seeing aux valeurs nulles et inversement, sans perdre de vue que ce bon et mauvais seeing se réfèrent à la saison. Concernant les techniques de reconstruction locale dans l'espace des phases, la répartition uniforme des vecteurs dans chaque catégorie donne plus de sens à la distance Euclidienne pour le calcul des vecteurs les plus proches. Ce calcul était, jusqu'alors, influencé par la distribution des variables considérées. Mais allons plus loin et considérons le codage linéaire par morceaux employé comme un codage flou; la notion traditionnelle de probabilité qui sous-tend l'utilisation la distance Euclidienne doit être redéfinie. L'objet du paragraphe suivant est de rappeler ces notions dont nous ferons usage lorsque le DRNN sera évalué grâce aux résultats obtenus avec la méthode du *k-nearest fuzzy neighbors*.

Définition de la Possibilité

Les qualificatifs "élevée", "intermédiaire" et "faible" condensent au stricte minimum l'information que recellent les variables météorologiques. L'emploi de ces quantités floues requiert malgré tout la redéfinition de la mesure, au sens mathématique du terme, sur laquelle est basée la notion de probabilité. Pour être compatible avec la forme de codage que nous avons adoptée, la mesure de Jacard [Miyamoto 90] est employée. Nous ne rentrons pas dans le détail de cette mesure (couramment employée du reste en analyse de données), seule une nouvelle définition de la probabilité est introduite. Afin d'éviter toute confusion, la transposition de la notion de probabilité dans l'univers flou est désignée par le terme de *possibilité* noté par $\mathcal{P}oss$. Aussi, la probabilité conditionnelle devient possibilité conditionnelle.

Considérons les variables, A_i, B_i, et C_i, où l'index $i = 1, \ldots, N$ correspond à un enregistrement dans notre cas. A_i peut être vue comme la température mesurée à l'instant i; B_i, le seeing; etc. Nous avons

$$\mathcal{P}oss(A \cap B) = \frac{1}{N} \bigcap_{i=1}^{N} (A_i, B_i) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{Min}(A_i, B_i) \tag{5.4}$$

$$\begin{aligned} \mathcal{P}oss(A|B \cap C) &= \bigcap_{i=1}^{N} (A_i, B_i, C_i) / \bigcap_{i=1}^{N} (B_i, C_i) \\ &= \sum_{i=1}^{N} (\mathrm{Min}(A_i, B_i, C_i)) / \sum_{i=1}^{N} (\mathrm{Min}(B_i, C_i)) \end{aligned} \tag{5.5}$$

Ces notions sont discutées plus en détail dans [Aussem 95b, Murtagh 94].

L'analyse des correspondances

L'analyse des correspondances (CA) est une méthode de réduction de la dimension de l'espace d'entrée pour faciliter l'analyse et la représentation données. Il s'agit d'une approximation dans la mesure où on considère généralement la projection d'un espace multi-dimensionnel sur un plan. La proximité des points projetés sur ce plan est révélateur de la corrélation entre les variables (ou parfois de relations fortuites); l'éloignement reflète en revanche l'opposition (par exemple entre le bon et le mauvais seeing).

Les variables considérées par l'analyse des correspondances sont les variables floues que nous avons citées. Bien que nous omettions les détails de cette analyse, il est intéressant de remarquer que chaque variable "floue" a été dupliquée, et son double complémenté à 1 (1− la variable) (Table 5.5). Cette complémentarité entre les paires de valeurs, certes considérablement redondante, facilite l'interprétation. Les couples de variables complémentaires sont positionnés symétriquement vis-à-vis de l'origine dans l'espace initial. Puisque l'origine représente un profile moyen, plus le couple de variables est positionné à proximité de l'origine, et moins le pouvoir discriminant de ces variables sera grand.

Fig. 5.12 et 5.13 visualisent les projections planaires des variables effectuées au cours de 11 saisons étudiées de 1989 à 1992. Cette représentation, à valeur descriptive, permet de rechercher les relations intéressantes présentant un intérêt potentiel. Les occurrences significatrices sont énumérées Table 5.6. En référençant le codage aux saisons, une liste de co-occurrences révélatrices a pu être exhibée [Aussem 95b], indiquant que un nombre limité de variables peut être associé à un bon seeing. Les deux critères flous employés pour sélectionner les relations présentées reflètent le dilemme suivant : combien de fois un bon seeing est-il associé à cette occurrence de variables; et quelle est la *possibilité* de cette occurrence? Certaines relations en Table 5.6 sont très encourageantes. En automne 1992, les variables s1L, s1dL et pH (respectivement faible vitesse de vent, faible écart type de la vitesse du vent, et pression élevée) sont associées dans 90% des cas à un bon seeing (i.e. sL); une telle “constellation” de variable survient avec une possibilité de 31%. Il y a cependant de nettes différences durant la même saison à des années différentes, ce qui limite la pertinence de ces résultats. Ces co-occurrences éphémères pourraient être capturées par un modèle en mesure de replacer ces variables dans un *contexte* temporel; notre espoir dans le DRNN repose sur ces apparitions spontanées et de courtes durées de relations conjoncturelles entre le bon seeing et des conditions météorologiques fugaces.

s1H	vitesse du vent – élevée
s1L	vitesse du vent – faible
s1dH	écart type de la vitesse du vent – élevé
s1dL	écart type de la vitesse du vent – faible
rhH	humidité relative – élevée
rhL	humidité relative – faible
t1H	température – élevée
t1L	température – faible
pH	pression – élevée
pL	pression – faible
seeH	seeing – élevé
seeL	seeing – faible

TAB. 5.5 - : Variables floues utilisées.

Saison	No. de cas	Bon seeing étant donné ...	Probabilité conditionnelle du bon seeing	Possibilité d'apparition des variables
Spr89	81	s1L	0.68	0.35
		s1dL	0.61	0.27
		s1L et s1dL	0.62	0.20
Sum89	77	pH	0.64	0.32
		s1dL	0.69	0.38
		s1L	0.68	0.35
		pH, s1dL et s1L	0.75	0.24
Aut90	48	rhL	0.76	0.35
Win90	69	rhL	0.63	0.32
		t1H	0.63	0.31
		pH	0.63	0.33
Spr90	47	s1L	0.58	0.29
		s1dL	0.67	0.36
		s1L et s1dL	0.67	0.27
Sum90	38	rhL	0.52	0.27
		t1H	0.62	0.31
		rhL et t1H	0.60	0.21
Aut91	61	s1L	0.61	0.29
		s1dL	0.70	0.41
		s1L et s1dL	0.60	0.25
Sum92	55	s1L	0.76	0.40
		s1dL	0.77	0.38
		pH	0.68	0.35
		s1L, s1dL et pH	0.88	0.24
Aut92	66	s1L	0.78	0.40
		s1dL	0.80	0.43
		pH	0.67	0.37
		s1L, s1dL et pH	0.90	0.31
Win92	44	s1L	0.89	0.46
		s1dL	0.87	0.42
		rhL	0.72	0.38
		s1L, s1dL et rhL	0.96	0.27
Spr92	69	s1L	0.74	0.38
		s1dL	0.68	0.33
		s1L et s1dL	0.72	0.29

TAB. 5.6 - : Résultats de l'analyse des corrélations.

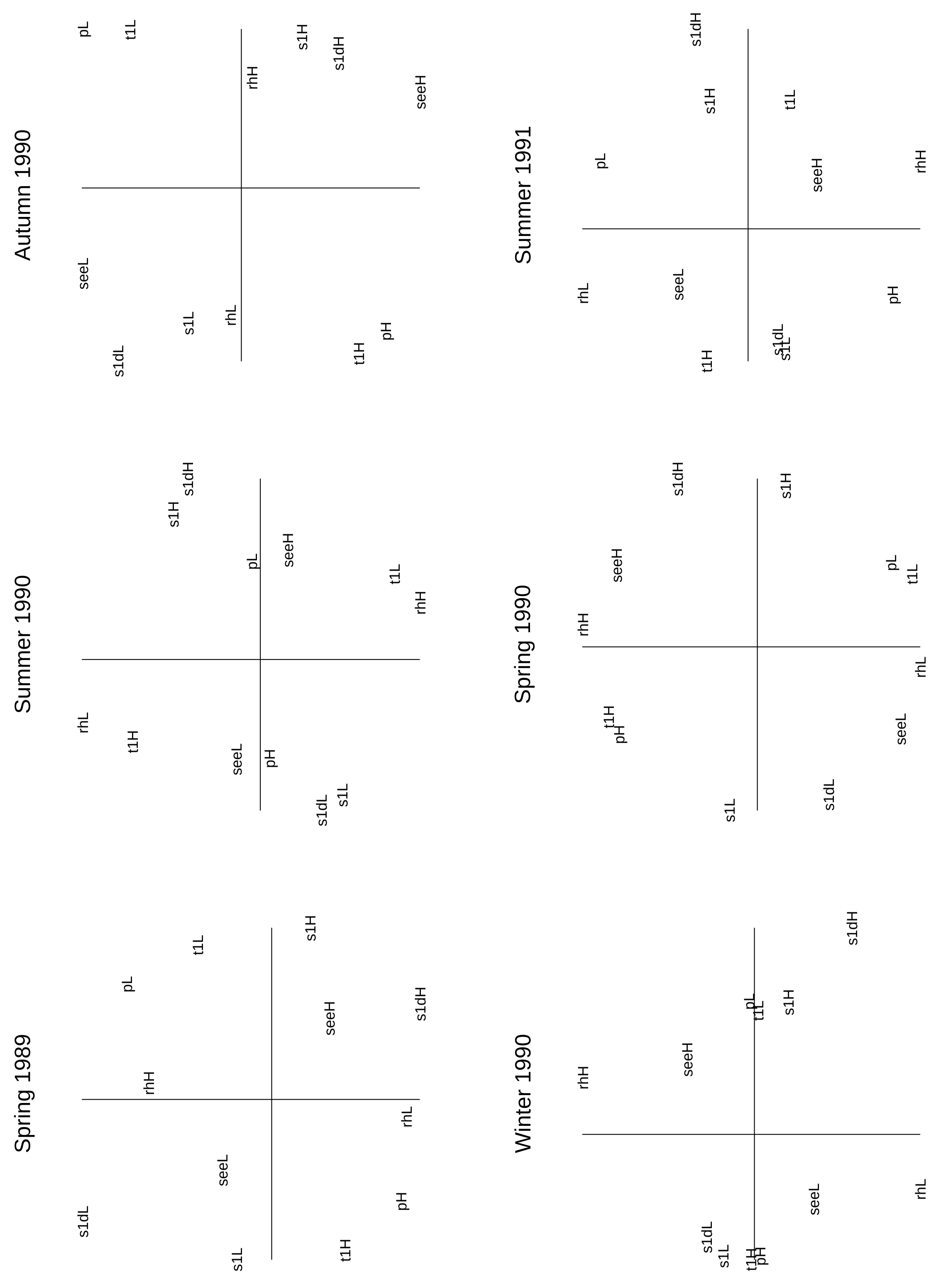

FIG. 5.12 - : Analyse des correspondances sur la pèriode 1989-1991.

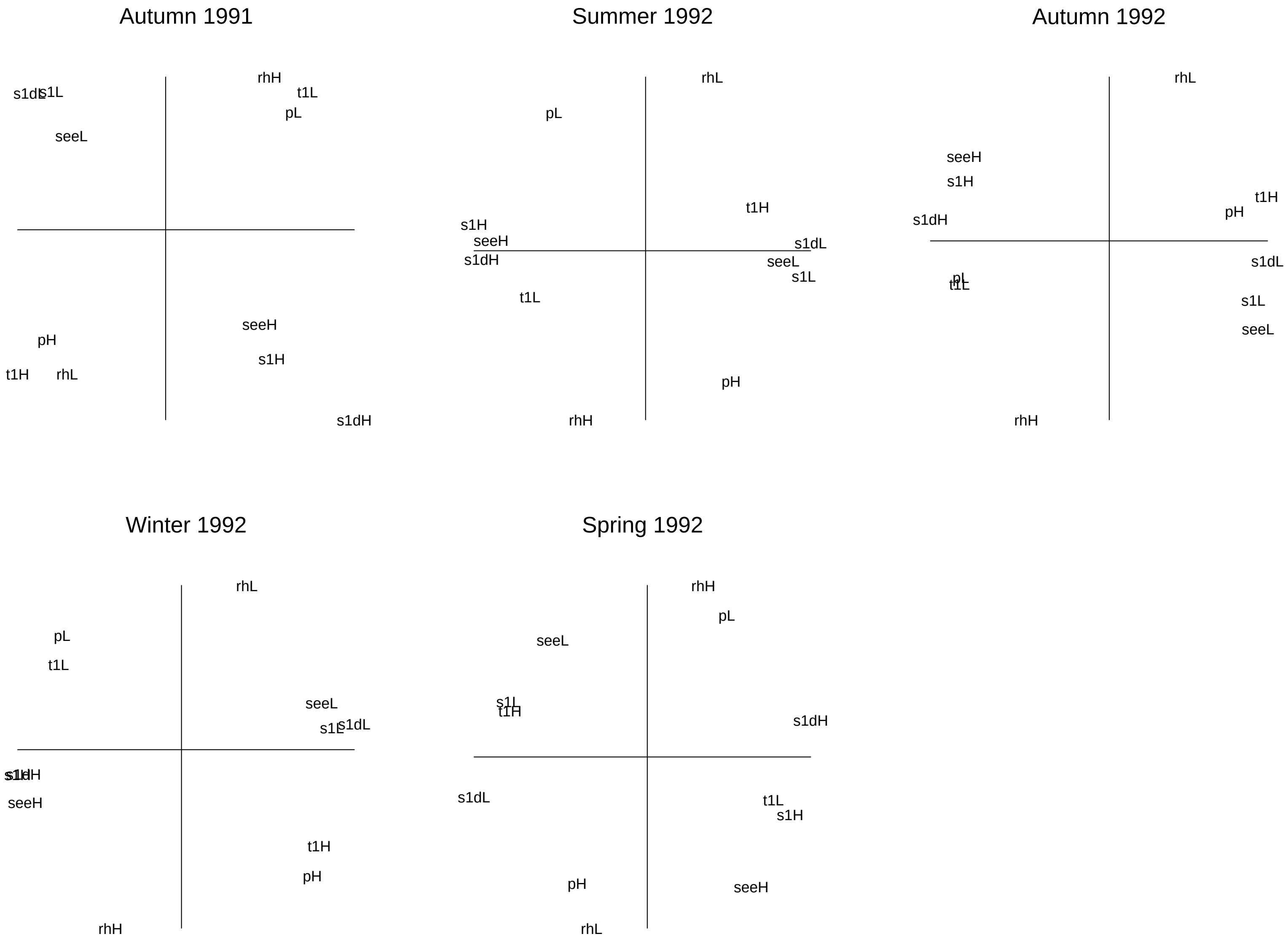

FIG. 5.13 - : Analyse des correspondances sur la période 1991-1992.

Prédictions par reconstruction locale

La méthode du nearest neighbor est étendue pour s'accommoder du codage flou. Etant donné un ensemble de variables météorologiques, le seeing flou associé au tuplet le plus similaire stocké dans la base de données fournit l'estimation du seeing désiré. Pour être compatible avec le codage utilisé, une mesure floue, dite de Jacard, [Miyamoto 90, Murtagh 93, Murtagh 94] est employée. La mesure de Jacard a donné de meilleurs résultats que la distance Euclidienne floue au cours de tests préliminaires. Sa définition, pour les séquences $\{A_i\}$ et $\{B_i\}$ s'écrit

$$\bigcap_i (A_i, B_i) / \bigcup_i (A_i, B_i) = \sum_i \mathrm{Min}(A_i, B_i) / \sum_i \mathrm{Max}(A_i, B_i) \tag{5.6}$$

Lorsque des similarités équivaluées apparaissent, les valeurs du seeing correspondantes sont moyennées selon la définition habituelle de la moyenne. La recherche des vecteurs similaires s'effectue parmi les vecteurs chronologiquement antérieurs. Les résultats sont donc indexés par le nombre de données collectées car la probabilité de trouver un vecteur similaire est proportionnelle à la densité de l'espace d'entrée.

Les résultats obtenus sont présentés exhaustivement dans [Aussem 95a, Murtagh 94] et se résument de manière concise dans ce qui suit

- La probabilité conditionnelle d'observer un bon (resp. un mauvais seeing), sachant qu'un bon seeing (resp. un mauvais seeing) a été estimé, est de 48.2% (resp. 22.0%).
- Dans 40.2% des cas, le seeing estimé et le seeing observé sont dans la même catégorie.
- Dans 9.6% des cas, le seeing estimé et le seeing observé sont dans des catégories opposées (bon-mauvais ou mauvais-bon).

Il faut préciser, au demeurant, qu'une stratégie simpliste consistant à prédire un seeing constant conduit à un taux de réussite de 33%.

Une approche connexioniste adaptée au coût

Dans ce paragraphe, nous nous tournons vers la méthode prédictive fondée sur le DRNN. Au-delà de la quête légitime pour la performance pure, notre intérêt se porte sur l'optimisation du temps d'observation alloué. Jusqu'à présent, la procédure ne teint guère compte des exigences des astronomes exprimées en terme de qualité d'observation. Une analyse préliminaire des coûts a été menée visant à réduire les pertes de temps occasionnées par un usage inopportun du télescope. L'idée consiste à sélectionner l'observation dont les exigences sont compatibles avec l'estimation faite du seeing. Table 5.7 associe un coût à chaque situation rencontrée.

Alors que la méthode du *fuzzy nearest neighbor* génère des estimations non-équivoques du seeing, la sortie du réseau de neurones peut être interprétée en fonction de l'objectif de performance retenu. Lorsque le réseau de neurones est entraîné sur le problème du seeing flou, sa sortie $o(t)$, à chaque instant t, est à valeur dans $]0,1[$, correspondant à une valeur intermédiaire du seeing. C'est pourquoi deux seuils (s_1, s_2) sont requis pour interpréter la valeur de sortie. La prédiction est considérée comme nulle lorsque $o(t) < s_1$, intermédiaire lorsque $s_1 \leq o(t) \leq s_2$, et enfin élevée dès que $o(t) > s_2$. La valeur de ces seuils est déterminée de façon à minimiser la fonction coût sur la base d'apprentissage. Les seuils sont ensuite appliqués à la base de prédiction pour évaluer les performances réelles du modèle. Les valeurs de ces seuils pourraient être ajustées *en ligne* dans une implémentation pratique à mesure que de nouvelles données sont collectées, parmi les tuplets les plus récents. En conséquence, les résultats présentés pourraient être améliorés dans une application pratique. Nous insistons sur le fait que la fonction de coût peut être continuellement adaptée pour satisfaire les exigences évolutives des astronomes.

Dans cette optique, plusieurs couples de seuils sont déterminés en relation avec les stratégies suivantes

- P1 - maximiser la taux de réussite, lorsque l'estimation et le seeing observé sont dans la même catégorie.
- P2 - maximiser la probabilité conditionnelle d'avoir un bon seeing lorsqu'un bon seeing est prédit.
- P3 - minimiser la fonction de coût définie Table 5.7.

estimation \ vrai seeing	bon	modéré	mauvais
bon	-1	+1	+1
modéré	+1/2	0	+1/4
mauvais	+1/2	+1/4	-1/4

TAB. 5.7 - : Table des coûts associés à chaque possibilité. Une valeur positive reflète une situation indésirable.

De nombreuses mesures intéressantes peuvent être extraites des tables de contingences 5.8 à 5.10. Nous les décrivons au paragraphe suivant d'une manière élégante et concise en fonction des stratégies que nous avons citées.

La prédiction neuronale

D'après la fonction d'autocorrelation Fig. 5.18, des corrélations à court terme sont présentes dans le seeing flou en dépit des ruptures de séquences. Un DRNN est donc employé pour inférer les dépendances conjoncturelles entre les variables météorologiques locales et le seeing flou. Etant donné la chute rapide de la fonction d'autocorrelation, l'ordre des connexions est restreint à de faibles valeurs. Le modèle possède 126 liens, biais compris. Des expérimentations ont été conduites auparavant avec des réseaux statiques non-bouclés mais les résultats sont meilleurs avec le modèle bouclé [Aussem 95a, Aussem 95b, Aussem 95c].

estimé \ seeing observé	bon	modéré	mauvais
bon	68	42	31
modéré	109	117	104
mauvais	27	49	58

TAB. 5.8 - : Table des contingences sur l'ensemble de la base de données du seeing entre les prédictions obtenues avec le *k-nearest neighbor* et les valeurs observées - Catégorie identique : 40.2%, erreur partielle : 50.2% incorrecte : 9.6%.

estimé \ seeing observé	bon	modéré	mauvais
bon	22	4	1
modéré	12	13	12
mauvais	1	15	25

TAB. 5.9 - : Table des contingences sur la base de prédiction du seeing entre les prédictions obtenues avec le DRNN et les valeurs observées en appliquant la stratégie P1 - Catégorie identique : 57.1%, erreur partielle : 41.0% incorrecte : 1.9%.

estimé \ seeing observé	bon	estimé	mauvais
bon	3	0	0
modéré	31	24	20
mauvais	0	9	18

TAB. 5.10 - : Table des contingences sur la base de prédiction du seeing entre les prédictions obtenues avec le DRNN et les valeurs observées en appliquant la stratégie P2 - Catégorie identique : 42.9%, erreur partielle : 57.1% incorrecte : 0%.

estimé \ seeing observé	bon	modéré	mauvais
bon	27	9	2
modéré	11	8	11
mauvais	1	16	25

TAB. 5.11 - : Table des contingences sur la base de prédiction du seeing entre les prédictions obtenues avec le DRNN et les valeurs observées en appliquant la stratégie P3 - Catégorie identique : 57.1%, erreur partielle : 40.0% incorrecte : 2.8%.

Les NMSE sur la base d'apprentissage et la base de prédiction sont respectivement de 0.73 et de 0.56 respectivement Fig. 5.14. Curieusement, de meilleurs résultats sont obtenus sur la base de prédiction probablement en raison des irrégularités dans le processus de mesure en début de campagne du seeing. Les histogrammes cumulatifs des prédictions avec la technique neuronale avant et après le seuillage, ainsi que la copie carbone sont tracés Fig. 5.15. La sortie seuillée du réseau de neurones est tracée Fig. 5.16.

- Pour la stratégie P1, la probabilité conditionnelle d'avoir un bon seeing (resp. du mauvais seeing) lorsqu'un bon seeing est prédit est de 81.5% (resp. 3.7%). La fréquence de l'estimation d'un bon seeing est de 25.7%.
- Pour la stratégie P2, la probabilité conditionnelle d'avoir un bon seeing (resp. du mauvais seeing) lorsqu'un bon seeing est prédit est de 100% (resp. 0%). La fréquence de l'estimation d'un bon seeing est de 2.9%.
- Pour la stratégie P3, la probabilité conditionnelle d'avoir un bon seeing (resp. du mauvais seeing) lorsqu'un bon seeing est prédit est de 71.0% (resp. 7.9%). La fréquence de l'estimation d'un bon seeing, soit le rapport nombre d'estimations du bon seeing/nombre total d'estimations, est de 36.2%.

Ces résultats, manifestement positifs, améliorent très nettement l'approche locale décrite au paragraphe précédent. La décroissance des termes rétro-propagés dans le réseau adjoint, y^n_{n-j}, est illustrée Fig. 5.17. A valeur indicative, les coût respectifs de ces trois stratégies sont de -15, -11.75 et -14.75. Le fait que P1 ait un coût moindre que P3 n'est pas paradoxal dans la mesure où le coût de P3 a été minimisé sur la base d'apprentissage; or les mesures de performance ont été faites sur la base de prédiction. Un développement plus exhaustif est fourni dans les références citées.

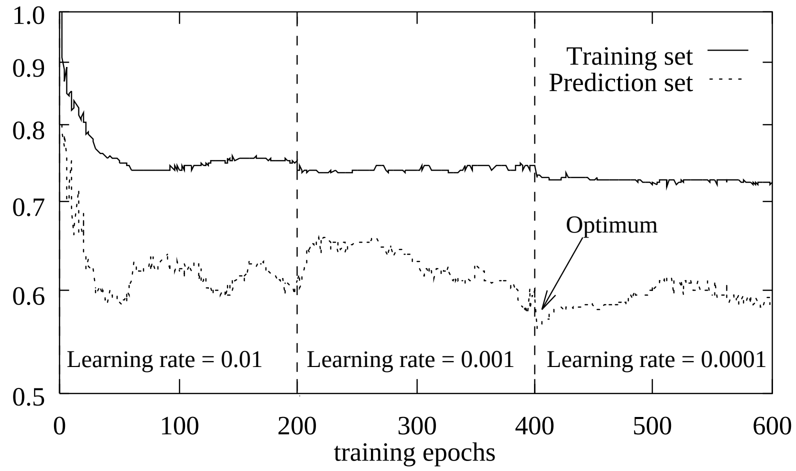

FIG. 5.14 - : NMSE sur les bases d'apprentissage et de prédiction du seeing flou en fonction des itérations. Curieusement de meilleurs résultats sont obtenus sur la base de prédiction.

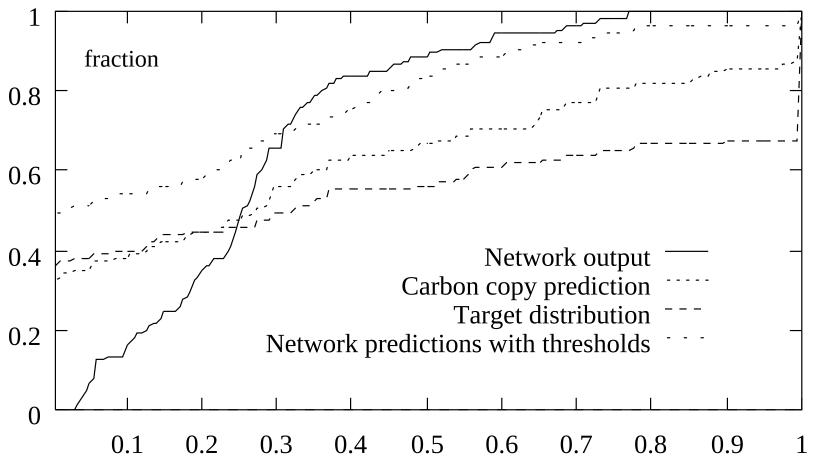

FIG. 5.15 - : Histogrammes cumulatifs des erreurs de prédiction du seeing flou en 1992. En trait plein les sorties brutes du DRNN, en pointillé les sorties seuillées.

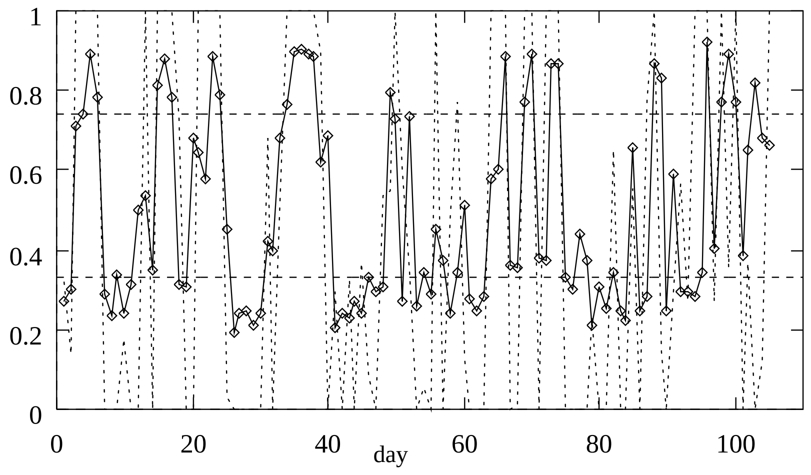

FIG. 5.16 - : Prédictions du seeing flou : en pointillé les valeurs, observées, en trait plein les prédictions.

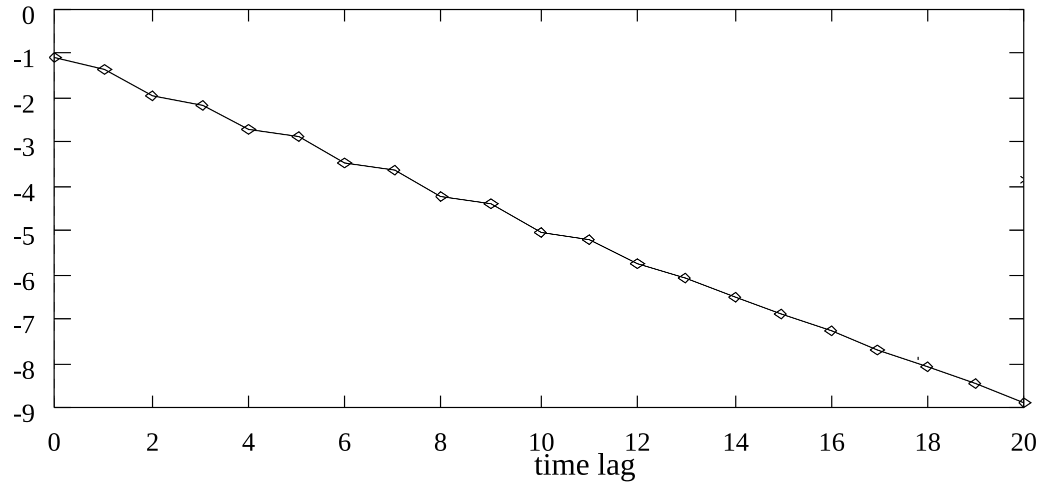

FIG. 5.17 - : Seeing : Illustration du gradient rétro-propagé : $< \mathrm{Log}_{10}(\|y^n_{n-j}\|_\infty) >$ est tracé en fonction de l'itération j.

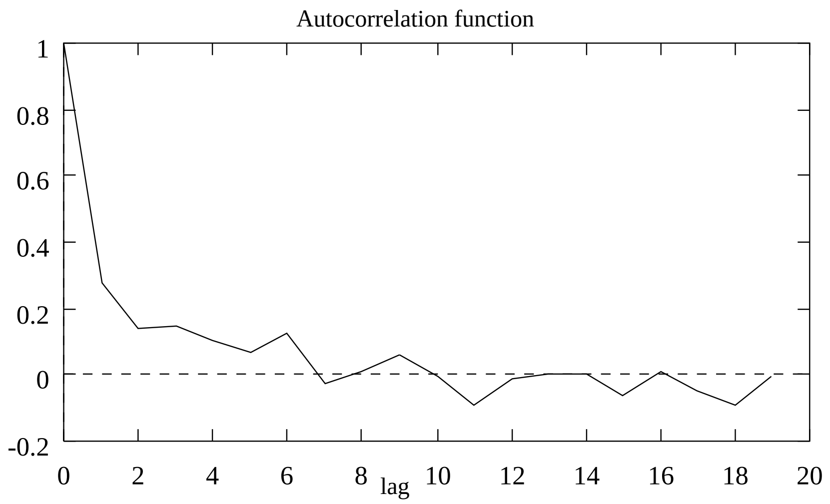

FIG. 5.18 - : Fonction d'autocorrelation du seeing flou.

5.5 Discussion

Il s'avère que les deux méthodologies, cf. le DRNN et la méthode du *nearest neighbor*, diamétralement opposées sur le plan conceptuel, le sont aussi par la teneur des résultats obtenus. Il -serait intéressant de s'interroger sur les raisons de cette différence afin de sélectionner le modèle le mieux adapté aux circonstances.

La qualité des prédictions obtenues avec la méthode du k-nearest neighbor repose sur un nombre suffisamment grand de prototypes quasi identiques dans la suite. C'est empiriquement une précondition au succès de cette approche. En effet, Farmer et al. [Farmer 88] ont analysé en détail les avantages et inconvénients de la reconstruction locale. La considération qui motive le choix de cette technique a trait à la densité des données dans l'espace multi-dimensionnel. A quel point cette méthode est-elle adaptée aux données?

Cela dépend non seulement du nombre de points N mais aussi de la dimension D de l'espace, du temps de prédiction et du rapport signal/bruit S. Supposons que l'on cherche à approximer une fonction analytique f avec la méthode du nearest neighbor, c.a.d. une approximation locale du premier ordre. L'erreur, dépendant de la dérivée première $f'(x)$, est proportionnelle à ϵ, l'espace élémentaire entre les données. Maintenant si $\epsilon \simeq N^{1/D} < S^{-1}$ alors la qualité des prédictions est limitée par le bruit. En revanche, si $\epsilon > S^{-1}$, alors la précision des prédictions est limitée par le nombre de données. Dans ce dernier cas, si f est inversible, l'erreur quadratique moyenne normalisée E vérifie

$$E \sim N^{-1/D} \tag{5.7}$$

Il vient d'après cette formule que l'erreur augmente de 30% lorsque 5000 points sont utilisés au lieu de 1000 dans la prédiction de la température à 24 heures.

Ainsi cette approche est inadaptée lorsque la dimension de l'espace d'entrée est trop grande comme c'est le cas dans la prédiction du seeing moyen. La non-stationnarité nuit par ailleurs à la validité des estimations parce que la surface à approximer dépend du temps. Par ailleurs, lorsque plusieurs sorties sont associées à une entrée précise se pose alors le problème du choix de l'estimateur car la moyenne arithmétique n'est pas nécessairement un choix judicieux. Une manière de contourner le problème est d'ajouter des observations météorologiques retardées au détriment de la dimension du problème. Des améliorations significatives ont été accomplies concernant la prédiction de la température grâce à l'ajout de valeurs retardées.

La faible densité des points et le niveau de bruit dans les mesures plaident en faveur de la technique neuronale. D'autres considérations concernent la stationnarité et surtout le nombre de données disponibles. L'approximation globale lisse les variations irrégulières de façon à éliminer le bruit. En outre, l'apprentissage en temps-réel permet de s'adapter aux propriétés statistiques du moment, ce qui contourne le piège de la non-stationnarité. Néanmoins, la plupart des représentations globales telles que les réseaux connexionistes stagnent lorsque sont ajoutés des degrés de liberté supplémentaires (i.e. ordre du système) et demeurent quasiment insensibles au nombre de données passé un certain stade. A l'origine du problème est l'apprentissage du bruit, *overfitting* en anglais.

5.6 Conclusion

Dans ce chapitre nous nous sommes intéressés à la prédiction des conditions environnementales des télescopes terrestres avec le modèle DRNN. L'objet de cette étude a été de fournir une évaluation préliminaire des performances prédictives de la future Astronomical Weather Station dont l'installation est envisagée sur le site de Paranal à la fin de ce siècle. En outre, nous avons posé les premiers jalons d'une stratégie opérationnelle vouée à la gestion des programmes d'observations astronomiques.

Les qualités prédictives du DRNN ont été évaluées à la lumière des résultats obtenus avec d'autres méthodes neuronales, statistiques et, en particulier, la technique de reconstruction locale du *k-nearest neighbor*.

Les tentatives de prédiction au couché du soleil un jour à l'avance se sont avérées difficiles avec le DRNN en raison du faible nombre de données exploitables : des résultats mitigés ont été obtenus avec la température. Quant au seeing, il apparaît qu'un excellent seeing est dans certaines circonstances prévisible un jour à l'avance. Afin de remédier au manque de données disponibles, la prédiction a été effectuée à chaque heure de la journée pour déterminer si la cause était le manque de données ou la nature trop aléatoire de la température à cette échelle de temps. De meilleurs résultats ont été obtenus, en particulier avec la méthode du nearest neighbor.

Dans un second temps, nous avons cherché à anticiper la qualité d'observation moyenne au cours de la nuit suivante en supposant que nous avions à notre disposition une estimation fiable des conditions météorologiques 24 heures en avance, fournie par une station météorologique *ad hoc*. Cette technique appelée *nowcasting* permet de scinder la tâche prédictive du seeing en deux sous problèmes indépendants : 1) la prédiction météorologique et 2) la régression non-linéaire du seeing sur les variables météorologiques au même instant. Des résultats fort encourageants ont été obtenus avec le DRNN en adoptant une codage flou : dans 80% des cas, la valeur qualitative du seeing moyen est anticipée correctement et il est possible de réduire le taux de prédictions erronées à zero. La méthodologie neuronale permet en outre à l'astronome de définir la fonction coût la mieux adaptée aux exigences du moment.

Le clivage conceptuel entre la méthode du nearest neighbor et le DRNN est à l'origine des résultats observés. Alors que la première privilégie une reconstruction locale de l'estimateur, l'approche neuronale effectue une approximation globale. Une analyse est menée dans l'intention d'établir laquelle de ces deux techniques complémentaires est la mieux adaptée aux circonstances. Il ressort que la reconstruction locale se prête bien au cas où un grand nombre de données stationnaires sont disponibles, alors que le DRNN sied aux situations plus alambiquées, où il est nécessaire d'extraire d'un nombre restreint de données, des informations contextuelles et ce, en dépit des variations diurnales et saisonnières.