Chapitre 11

Conclusion

Nous avons mis au point un système de visualisation et de recherche de documents basé sur les cartes auto-organisatrices. Ce système est accessible en ligne aux adresses officielles[1] suivantes : http ://simbad.u-strasbg.fr/A+A/map.pl ; http ://simbad.u-strasbg.fr/ApJ/map.pl ou encore http ://www.journals.uchicago.edu/CDS/ApJ/cgi-bin/bib.pl. Ces cartes regroupent chacune environ 10000 documents parus dans deux revues de référence en astrophysique depuis 1994 : la revue européenne Astronomy and Astrophysics (A&A) et la revue américaine Astrophysical Journal (ApJ).

Nous utilisons les cartes auto-organisatrices pour effectuer une classification des documents dans un tableau bidimensionnel dont chaque case correspond à une classe de documents. Les classifications obtenues sont telles qu'il existe une corrélation spatiale des classes : des documents qui appartiennent à des classes voisines ont des caractéristiques voisines (voir le chapitre 6).

L'organisation spatiale des documents que l'on obtient facilite la découverte du contenu de la base de données textuelles :

- en permettant aux utilisateurs d'évaluer les thèmes abordés dans les différentes zones, grâce à l'affichage des termes les plus fréquents contenus par les documents des zones inspectées ;
- en permettant aux utilisateurs d'explorer un domaine en examinant des zones voisines abordant différents aspects d'un même thème général.

La recherche de documents sur un thème précis peut se faire :

- soit par la visualisation des documents qui contiennent des termes en rapport avec le thème choisi ;
- soit par la localisation d'un document connu traitant du thème en question. On localise ainsi la zone où le thème est abordé.

La validation des résultats

Une comparaison (voir le chapitre 10) entre la carte bibliographique et le système documentaire de la NASA (ADS[2] abstract service) sur leur aptitude à retrouver des documents similaires montre que les deux systèmes donnent des résultats comparables

[1] Les cartes disponibles aux adresses citées ne possèdent pas toutes les fonctionnalités dont il a été question dans cet ouvrage. Une dernière version de notre système est toutefois accessible à l'adresse http ://newb6.u-strasbg.fr/ poincot/cgi-bin/Cartes/bib.pl.

[2] Astrophysics Data System.

en terme de pertinence des documents retrouvés. Il est intéressant de noter que les deux systèmes sont complémentaires puisque chacun d'entre eux retrouve des documents que l'autre ne retrouve pas. Davantage de documents peuvent toutefois être retrouvés grâce à l'interactivité de la carte bibliographique qui facilite l'exploration de la base.

Des champs d'application multiples

Il est bien entendu possible d'utiliser la carte bibliographique avec d'autres données : toute base de données textuelles est en effet susceptible d'être interrogée avec un tel système. C'est ainsi que nous avons pu construire des cartes relatives à des catalogues astronomiques qui sont décrits par les mots-clés d'une liste validée par des spécialistes. Une carte des catalogues a été intégrée dans l'interface d'interrogation de la base de données VizieR http ://vizier.u-strasbg.fr/cgi-bin/VizieR, qui propose avec cette carte trois moyens d'interrogation.

Nous avons également mis en œuvre notre système pour cartographier des sites Internet. Pour cela, nous avons utilisé les réponses de moteurs de recherche sur Internet dans divers domaines, ce qui montre la capacité de notre système à gérer les textes plus généraux. Nous pouvons ainsi visualiser les thèmes traités par les nombreux sites qu'un moteur de recherche comme hotbot[3] peut renvoyer en réponse à une requête générale. Les sites retournés sont classés en fonction du contenu de leur descriptif. De cette manière, l'examen du document moyen de chaque classe obtenue permet d'éliminer ou de conserver les sites par groupes, et réduit la lourde tâche de l'évaluation individuelle des descriptifs (voir l'exemple de l'annexe A).

Il est aussi envisageable d'utiliser ces cartes afin d'obtenir des informations qualitatives par exemple sur l'évolution des thèmes de publications sur une longue période (l'émergence de nouveaux thèmes de recherche), ce que nous n'avons pas pu mettre en évidence, faute du recul nécessaire, puisque la plupart des publications dont nous disposons sous forme électronique sont postérieures à 1994. Nous avons en revanche été en mesure de montrer les orientations de différentes revues : la visualisation des articles parus dans des journaux différents montre clairement les répartitions distinctes des documents, en fonction des thèmes traités. Par exemple, la revue A&A publie proportionnellement plus d'articles sur les sujets stellaires que la revue ApJ qui, elle, est davantage orientée vers les sujets galactiques.

Ces différentes applications sont présentées dans l'annexe A.

Les améliorations et perspectives

La principale amélioration à apporter à notre système est l'automatisation de l'inscription des thèmes abordés dans les articles aux différents endroits des cartes. Une automatisation réduira cette tâche qui est encore effectuée manuellement, à une tâche de vérification, moins coûteuse en temps. De plus, il sera possible de modifier l'affichage des termes en fonction des souhaits de chaque utilisateur (affichage de termes plus ou moins nombreux et précis) ou de ses requêtes (affichage des termes de la requête aux endroits où ils apparaissent).

Pour l'instant, seules les cartes de A&A et ApJ sont disponibles officiellement au CDS (la carte de l'annexe A qui regroupe différentes revues n'est qu'à l'état de développement).

[3] http ://www.hotbot.com

La mise en œuvre de cartes regroupant un grand nombre d'articles (20.000 et plus) nécessitera une refonte du système de recherche par l'utilisation d'un fichier inverse (voir la section 2.2.2.3). Ainsi, les temps d'accès seront réduits et l'utilisation rendue plus confortable. De plus, l'utilisation d'un fichier inverse facilitera la mise en place d'un système d'interrogation par mot-clé plus souple, avec la possibilité de pondérer les termes ou d'effectuer des opérations logiques.

Enfin, diverses améliorations sur le plan ergonomique pourraient être effectuées, et il sera intéressant pour cela de programmer une nouvelle interface en langage Java. Nous serons alors libérés des contraintes et des limitations du langage HTML.

Ces travaux s'inscrivent dans l'optique d'une intégration complète de la carte bibliographique aux services du CDS (cartes construites au vol par exemple), où le passage d'un service à l'autre doit se faire de manière transparente.

Annexe A

Trois applications supplémentaires de la carte bibliographique

A.1 La visualisation des thèmes abordés par différents journaux

Nous présentons ici les résultats de la classification de plus de 22000 articles publiés dans différentes revues (Tableau A.1). Nous avons pratiqué sur ces articles une indexation automatique de laquelle nous avons obtenu un ensemble de 3710 mots-clés.

revue	nb. documents	dates de publications
ApJ	12550	1994 à 1999
A&A	7846	1994 à 1999
PASP	1010	1994 à 1999
MNRAS	459	1999
AJ	859	1998 à 1999
NewA	134	1996 à 1999

Tab. A.1: *Provenance des articles utilisés. Les noms complets des revues sont : Astrophysical Journal, Astrophysics and Astronomy, Publications of the Astronomical Society of the Pacific, Monthly Notices of the Royal Astronomical Society, Astronomical Journal et New Astronomy.*

La figure A.1 montre la fréquence d'apparition des mots-clés dans les documents de la base ainsi que l'histogramme du nombre de mots-clés contenus dans les documents où on retrouve les deux pics (que nous avons déjà rencontrés dans la section 5.3) qui correspondent aux documents pour lesquels on ne dispose pas du résumé pour le premier (le pic à 16 mots-clés par article), et aux autres documents pour le second (49 mots-clés par article).

Les tendances des revues

Chaque image de la figure A.2 représente les documents d'une revue différente. On constate que la répartition de ces documents n'est pas toujours homogène, ce qui montre les différentes orientations de ces revues.

La comparaison des deux premières images (relatives à ApJ et A&A) montre que par rapport à ceux de A&A, les articles publiés dans ApJ traitent plus fréquemment de cosmo-

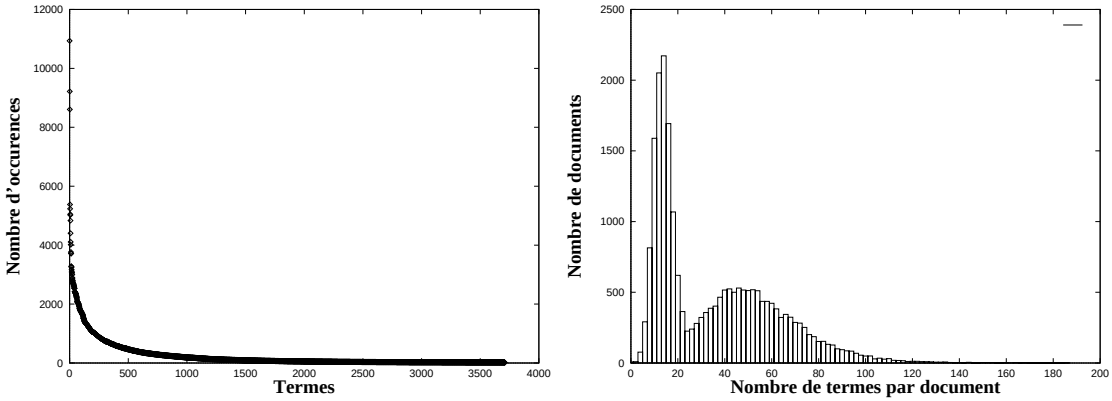

Fig. A.1: *Statistiques sur les 3710 termes d'indexation dans les 22858 documents de la base. Les 5 termes les plus fréquents sont : stars (10934 occurrences), galaxies (8607 occurrences), observations (7096 occurrences), model (5379 occurrences), et lines (5233 occurrences).*

logie et de l'étude des galaxies (les zones proches des deux coins droits de la carte ApJ sont très denses). Il semble que cela se fait au détriment des articles traitant de sujets stellaires dans ApJ dont la carte montre une nette baisse de densité sur une bande qui part du coin inférieur gauche et remonte jusqu'au centre. Cette bande, où est localisée la majorité des articles relatifs aux étoiles, est d'ailleurs très dense pour A&A.

La troisième image montre la répartition des articles des PASP, qui présente principalement une sur-densité dans le coin inférieur gauche, ce qui correspond à des articles très liés à l'instrumentation (descriptions techniques d'instruments). On notera également la faible quantité des documents relatifs au soleil, ainsi qu'une représentation importante des thèmes stellaires et galactiques.

Les articles parus dans les MNRAS (quatrième image) montrent une répartition plus homogène que celle des articles des PASP. On constate néanmoins des pics de densité dans les sujets de la cosmologie, des rayons-X stellaires ainsi que des trous-noirs. Le thème du soleil semble ici encore très peu abordé.

La répartition des articles de l'AJ (cinquième image) montre clairement une absence presque totale de documents relatifs au soleil (la zone vide du quart supérieur droit de l'image). A l'opposé, le quart inférieur droit est très peuplé par des articles relatifs aux galaxies.

Les articles de NewA (sixième image) sont peu nombreux (134 documents), mais la cosmologie semble y être le thème abordé avec une légère prédominance sur les autres.

Remarque : il faut garder à l'esprit que le nombre des documents dont nous disposons pour chaque revue est très variable (par exemple, la carte des documents de ApJ comporte 100 fois plus d'articles que celle relative à NewA). La plus grande couverture des articles de A&A et de ApJ est seulement due à leur nombre plus important.

A.2 La classification de catalogues d'objets astronomiques

Nous avons effectué une classification de catalogues d'objets astronomiques à l'aide d'une SOM. La carte obtenue constitue l'une des trois méthodes[1] proposées pour rechercher des catalogues dans la base de données VizieR (http ://vizier.u-strasbg.fr/).

Les vecteurs descriptifs des catalogues sont construits sur la base des mots-clés attribués par les auteurs. La figure A.3 montre les données statistiques sur la répartition des mots-clés dans les catalogues.

On remarque que plus de la moitié des catalogues sont décrits par seulement un à deux mots-clés. De plus, l'examen de ces catalogues montre que c'est souvent par couple que les mots-clés apparaissent (par exemple : "Galaxies" et "Redshifts", "Positional-Data" et "Proper-Motions" ou encore "Galaxies" et "Velocities"). Il en découle de nombreux catalogues décrits par exactement les mêmes mots-clés. Ceci entraîne une classification qui présente des pics de documents (Fig. A.4), dans laquelle les passages graduels d'un thème à l'autre sont moins apparents que ceux que nous pouvons constater dans les classifications de documents (voir la partie 9.1).

Cette carte est mise à jour par ajout simple des nouveaux catalogues (voir le chapitre 8), et, deux ans après sa création, elle contient maintenant près de 2350 catalogues.

A.3 La classification de sites Internet

A.3.1 La provenance des données

Afin de vérifier les qualités de la carte bibliographique appliquée à des données différentes de celles dont nous disposons au CDS, nous nous sommes orientés vers Internet qui présente la plus importante base de données textuelles (notamment) accessible.

Pour pratiquer une première sélection des documents (les sites Internet), nous effectuons une requête sur un moteur de recherche sur Internet (par exemple Hotbot[2] , Altavista[3], Infoseek[4], etc...). Les réponses de ces systèmes de recherche consistent en une liste d'adresses où les sites sélectionnés sont accessibles, accompagnées d'un descriptif.

Deux possibilités s'offrent alors à nous pour constituer les vecteurs descriptifs nécessaires à la classification :

- télécharger les données de chaque site retrouvé pour effectuer une indexation automatique sur leur contenu ;
- effectuer une indexation automatique sur les descriptifs renvoyés par le moteur de recherche.

Les problèmes du temps de transfert et de la place nécessaire pour stocker les données nous on fait opter pour la seconde solution.

A.3.2 Le choix du moteur de recherche

Les classifications que l'on se propose d'effectuer sont basées uniquement sur les descriptifs des sites retournés par les moteurs de recherche. Dans le but d'accéder à un maximum

[1] Les deux autres méthodes de recherche de catalogues sont : a) la recherche par numéro ou abréviation (lorsque ceux-ci sont connus) ; b) la recherche par mot-clé.

[2] http ://www.hotbot.com

[3] http ://www.altavista.com

[4] http ://www.infoseek.com

d'information sur ces sites, nous avons recherché le moteur qui propose les descriptifs les plus long. La figure A.5 montre le nombre moyen de termes contenus dans les descriptifs retournés par les principaux moteurs de recherche. Nous voyons que Hotbot est celui qui renvoie le plus d'information, avec une moyenne de 38 termes par descriptif, c'est donc ce moteur que nous avons utilisé. Hotbot permet en outre d'accéder aux adresse et descriptif de 1000 sites correspondant aux requêtes, ce que ne permet pas Altavista par exemple, qui ne retourne pas plus de 200 sites.

A.3.3 Les résultats

Nous effectuons une indexation automatique sur les descriptifs des 1000 sites que peut retourner Hotbot. Après l'élimination des termes vides et peu fréquents, il reste généralement de 200 à 600 termes d'indexation par lesquels les documents sont décrits. Les figures A.6 et A.7 montrent deux exemples des classifications obtenues.

Commentaires sur ces exemples

Les deux ensembles de descriptifs obtenus avec les requêtes "Jupiter+probe[5]+comet" et "bse[6]" n'ont pas le même degré de diversité :

- bien que les termes des requêtes soient recherchés dans les sites mêmes et non dans leur descriptif, une requête très ciblée comme la première (pour laquelle les sites retournés doivent comporter les trois termes simultanément) conduit à des descriptifs eux aussi très ciblés. C'est pourquoi les termes les plus fréquents, que nous avons reporté sur les images de la figure A.6, sont tous très liés avec le sujet de la requête ;
- la carte correspondant à la requête sur le terme "bse", quant à elle, regroupe des descriptifs de sites très différents. On y retrouve notamment des sites traitant d'économie et le commerce en Inde (Bombay Stock Exchange), d'autres relatifs à une université américaine : la "BS Engineering", et bien sûr d'autres qui traitent du phénomène de la maladie de la vache folle. L'interrêt de la carte bibliographique est que les descriptifs qui traitent de sujets différents appartiennent à des classes différentes, ce qui permet aux utilisateurs de trier rapidement les sites, pour n'examiner que ceux qui correspondent à leurs attentes.

L'utilisation

Les durées d'exécution sont de l'ordre de 3 minutes, dont seulement 40 secondes pour l'indexation des sites et l'apprentissage sur une carte de 6x6 classes. Le temps restant est pris par les 10 interrogations du moteur pour chacune desquelles il renvoie les informations sur 100 sites.

Un tel système permet l'exploration d'un nombre important de descriptifs de sites Internet correspondant à une requête. Dans le cas d'une requête large, et surtout lorsqu'un utilisateur désire avoir connaissance du maximum des informations susceptibles de l'intéresser, une exploration des données à l'aide de la carte bibliographique peut être très efficace.

[5] space probe : sonde spaciale

[6] bse est notamment le sigle de bovine spongiform encephalopathy : l'encéphalie bovine spongiforme.

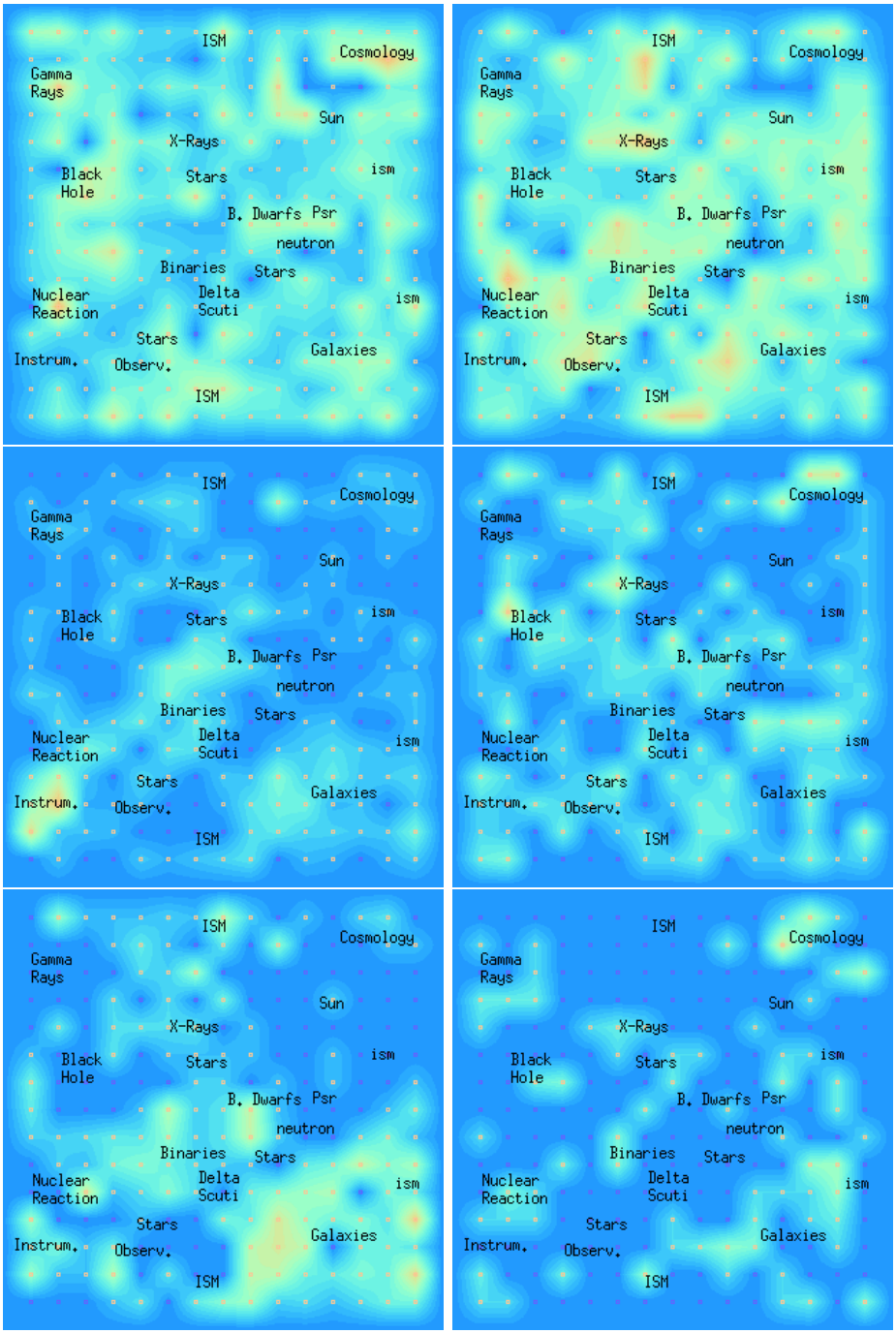

Fig. A.2: *De gauche à droite et de haut en bas, les documents des revues ApJ, A&A, PASP, MNRAS, AJ et NewA.*

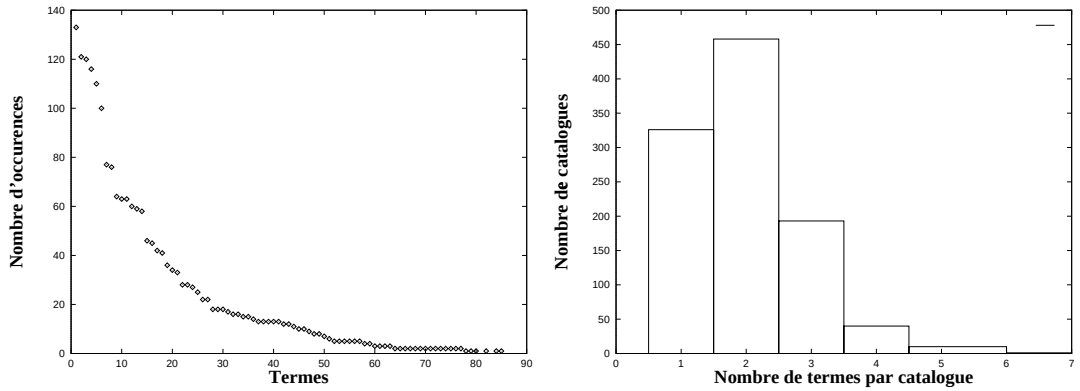

Fig. A.3: *Statistiques des données au moment de l'apprentissage : 1028 catalogues décrits par 85 mots-clés (après élimination des moins fréquents). Les 5 termes les plus fréquents sont : Photometry (133 occurrences), Radio (121 occurrences), Galaxies (120 occurrences), IR (116 occurrences) et Stars (110 occurrences).*

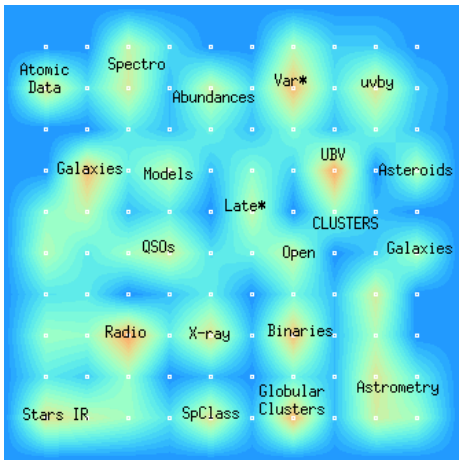

Fig. A.4: *Répartition des catalogues sur une carte de 10x10 classes.*

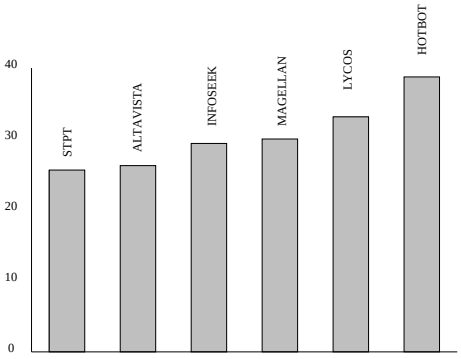

Fig. A.5: *Nombre de termes contenus dans les descriptifs bruts (avant tout traitement) pour différents moteurs de recherche.*

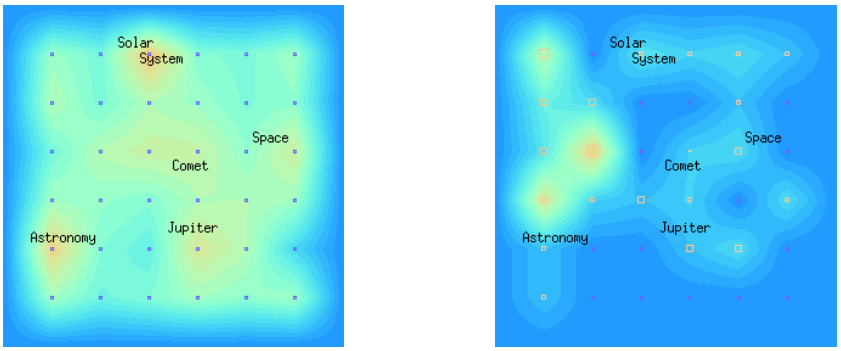

Fig. A.6: *Répartition des sites Internet retournés par Hotbot à la suite d'une requête utilisant les termes "Jupiter", "probe" et "comet". La première image montre l'ensemble des sites, et la seconde montre uniquement les sites dont le descriptif contient le terme "Galileo" (relatif à la sonde spatiale qui porte ce nom).*

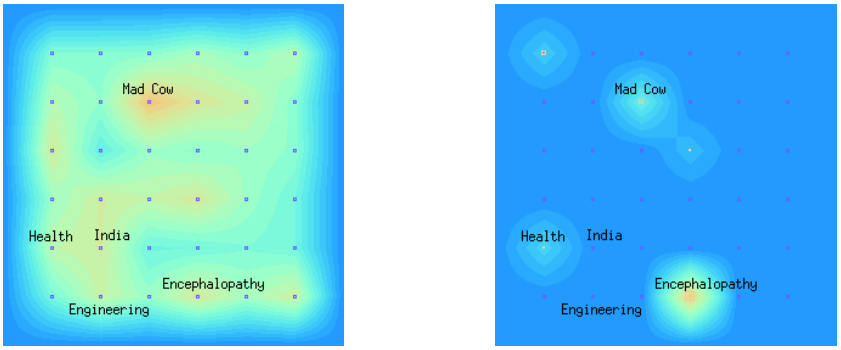

Fig. A.7: *Répartition des sites Internet retournés par Hotbot à la suite d'une requête utilisant le terme "bse". La première image montre l'ensemble des sites, et la seconde montre uniquement les sites dont le descriptif contient le terme "encephalopathy".*

Annexe B

Exemple d'utilisation de la Carte Bibliographique

Nous supposons ici qu'un utilisateur recherche des documents relatifs aux étoiles variables cataclysmiques à l'aide de la Carte Bibliographique, et nous détaillons sa manière de procéder.

Comme nous l'avons vu dans le chapitre 7, il est possible de consulter la Carte Bibliographique de deux différentes manières qui sont :

- l'exploration du contenu des classes, en sélectionnant tour à tour les classes de documents pour en examiner le document moyen (résumé) ;
- l'interrogation, qui consiste à localiser les documents suivant certains critères. Dans l'état actuel de l'interface, deux modes d'interrogations sont possibles : l'interrogation par mot-clé, et l'interrogation par référence (localisation d'un document connu).

Dans le cas de figure que nous décrivons ici, il est conseillé d'utiliser la seconde méthode, l'interrogation, que nous utilisons pour localiser les documents qui contiennent un mot-clé (au moins) relatif au sujet choisi.

La figure B.1 montre comment notre utilisateur a recherché un mot-clé relatif aux variables cataclysmiques. Ne connaissant pas a priori le ou les termes d'indexation qui se rapportent à son interrogation, il choisit d'en rechercher un contenant le terme "cataclysmic". Pour cela, il décide de rechercher les mots-clés qui contiennent la chaîne de caractères "cat". C'est ainsi que le mot-clé "stars :novae,cataclysmic variables" apparait parmi d'autres (Fig. B.1) ; il convient à notre utilisateur qui le sélectionne.

Le résultat de cette requête est représenté sur la figure B.2 qui montre que la plupart des documents concernés sont localisés dans une zone restreinte. La sélection de la classe la plus peuplée de cette zone montre qu'elle comporte un nombre important de documents (173). L'utilisateur choisit alors d'accéder à la carte détaillée qui correspond à cette zone (par l'intermédiaire d'un bouton situé en bas de la partie droite de l'interface, invisible sur la figure B.2). Remarquons que le mot-clé sélectionné est rappelé dans une fenêtre située sous l'image de la carte, fenêtre qui permet en outre l'élimination d'un ou plusieurs des mots-clés de la requête, ainsi que la sélection de la manière de les combiner (AND/OR).

La figure B.3 montre la carte secondaire obtenue. Remarquons que la requête par mot-clé est toujours active (fenêtre située sous l'image de la carte). On remarque une répartition des documents (binaires proches, binaires à éclipse...) qui aide l'utilisateur à apprécier les thèmes secondaires qui sont abordés dans les documents.

L'utilisateur affine sa recherche en s'intéressant maintenant aux étoiles binaires à éclipses. Il a sélectionné le mot-clé correspondant ("stars :binaries :eclipsing") afin de localiser les documents susceptibles de l'interesser (Fig B.4 à gauche).

L'utilisateur accède aux documents via le service bibliographique du CDS (Fig. B.5).

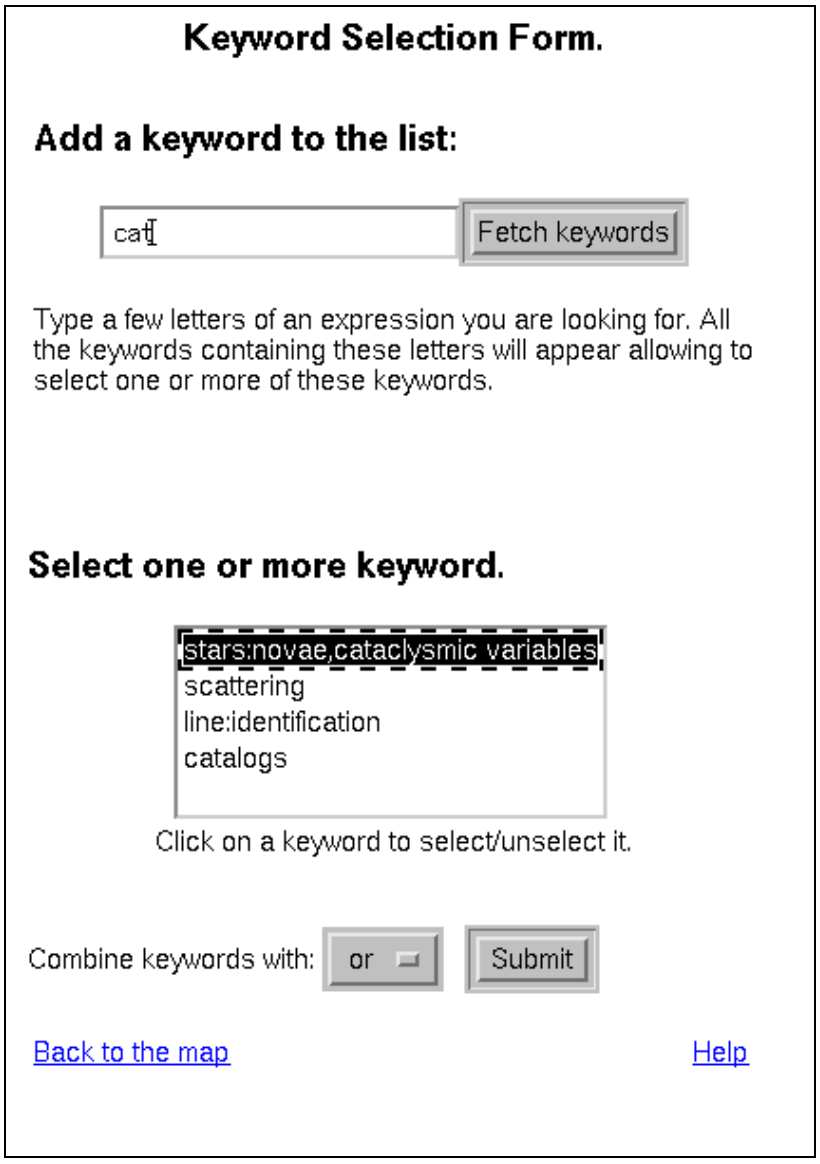

Fig. B.1: *Formulaire de sélection des mots-clés. La fenêtre supérieure permet de saisir un groupe de lettres. La suite de lettres choisie est alors recherchée dans tous les mots-clés. Les mots-clés correspondants (ceux qui contiennent la chaîne de caractères choisie) sont alors retournés dans une seconde fenêtre qui permet leur sélection. Dans l'état actuel de l'interface, plusieurs mots-clés peuvent être combinés entre eux par un OU ou un ET logiques.*

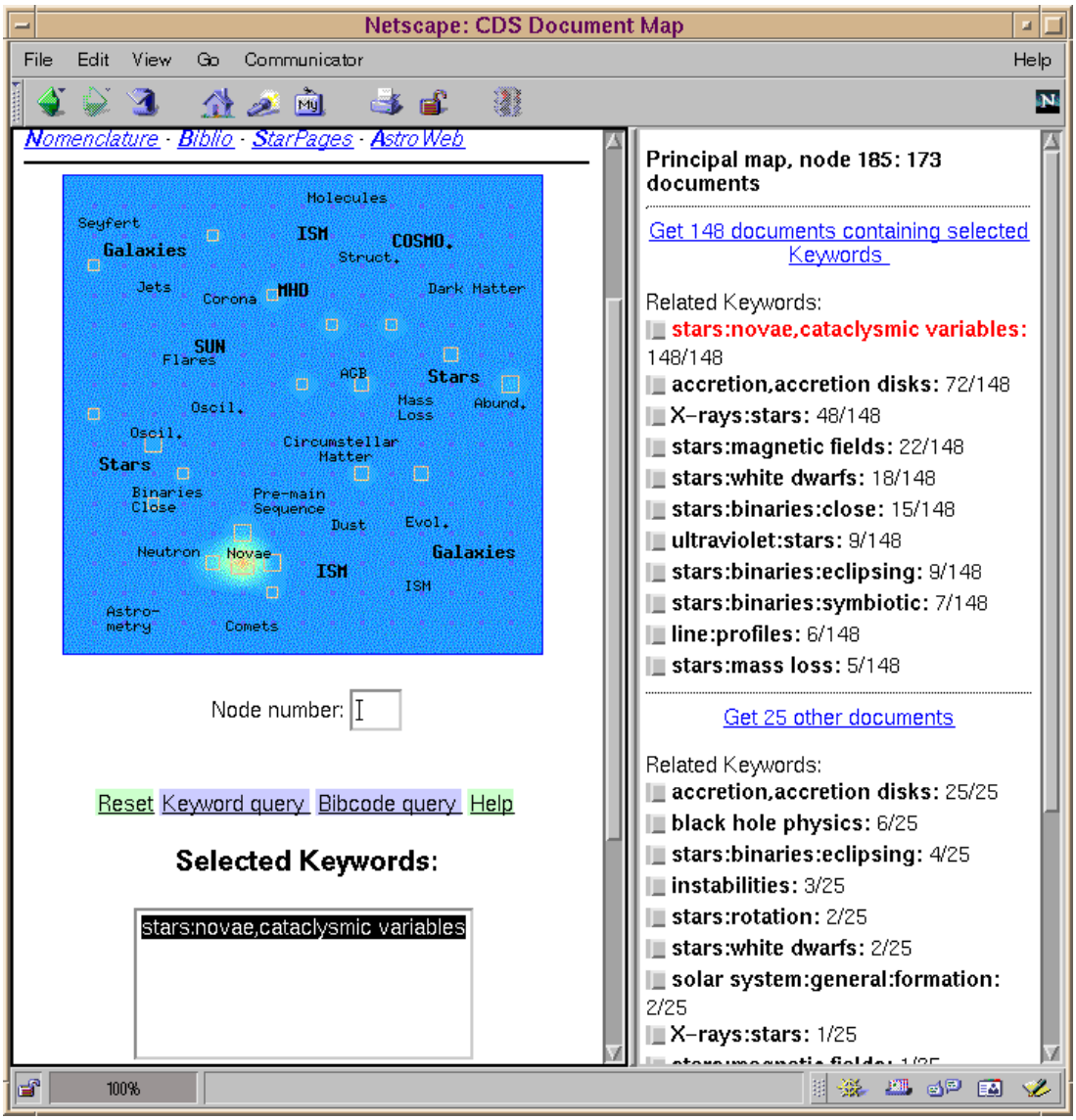

Fig. B.2: *Localisation des documents décrits par le mot-clé : "stars :novae,cataclysmic variables". La taille des carrés représentant les classes est variable, même si un seul mot-clé est demandé : un carré est d'autant plus grand que le document le plus proche de la requête appartenant à la classe correspondante est décrit par peu de mots-clés. Un document décrit uniquement par le mot-clé de la requête engendre donc un carré de taille maximale.*

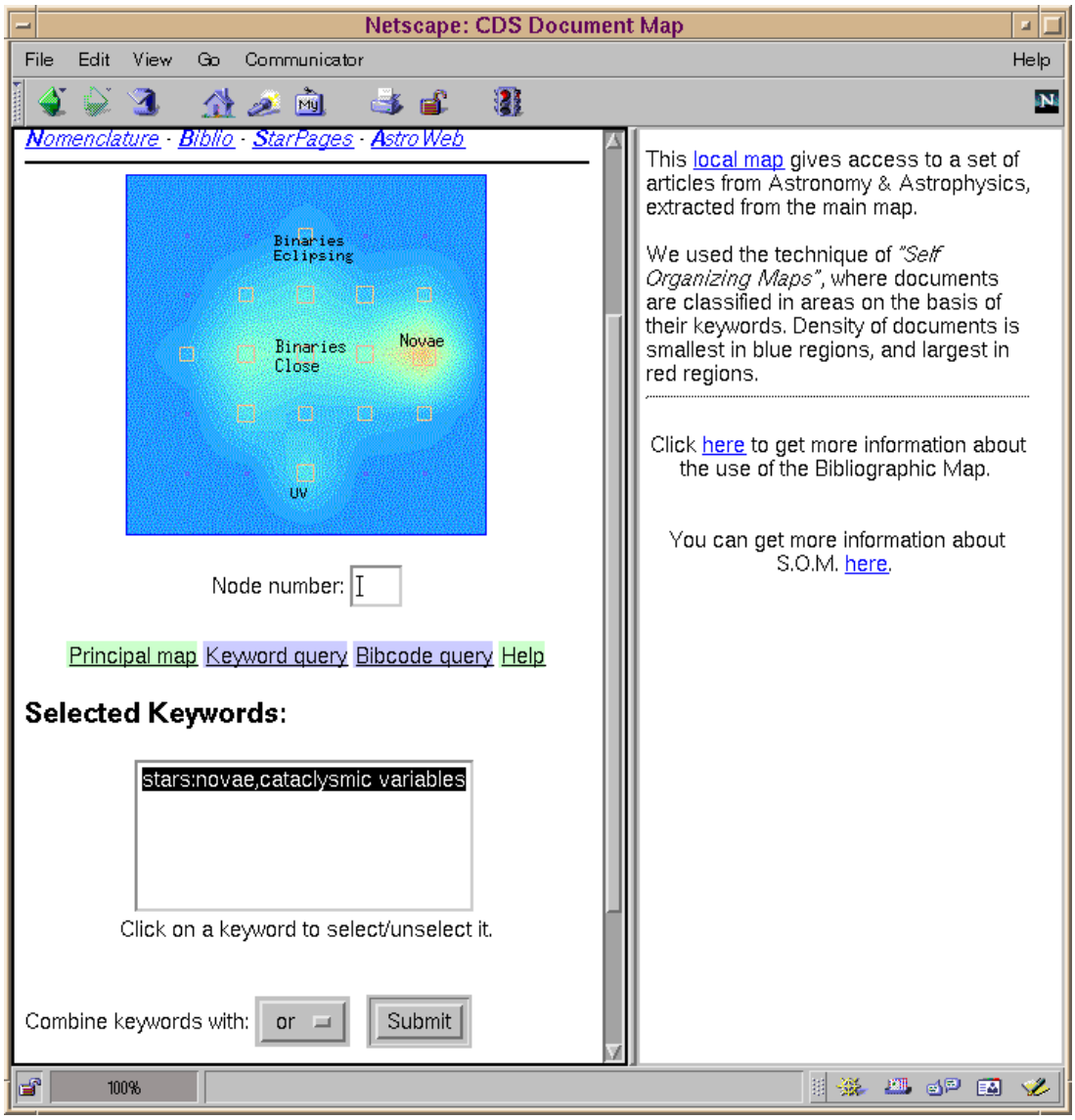

Fig. B.3: *Carte détaillée. La requête est toujours active (fenêtre située sous l'image de la carte).*

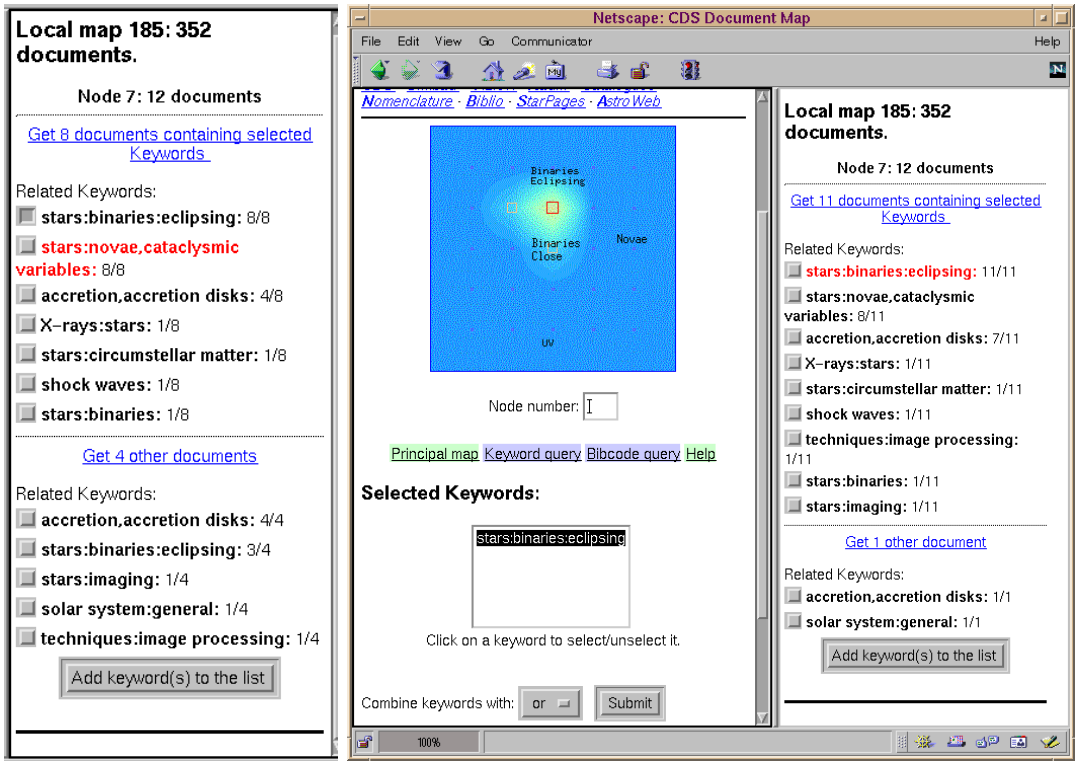

Fig. B.4: *À gauche : sélection directe d'un mot-clé. Il s'agit de la deuxième manière de sélectionner les mots-clés pour effectuer des recherches. Les mots-clés affichés lors de l'examen d'une classe (les plus fréquents dans les documents de cette classe) peuvent être ajoutés à une requête sans passer par le formulaire de choix des mots-clés (Fig. B.1) La requête devient alors "stars :binaries :eclipsing" OU "stars :novae,cataclysmic variables" . À droite : Localisation des documents qui contiennent le mot-clé "stars :binaries :eclipsing". L'autre mot-clé a été simplement éliminé de la requête en utilisant la fenêtre située sous l'image de la carte.*

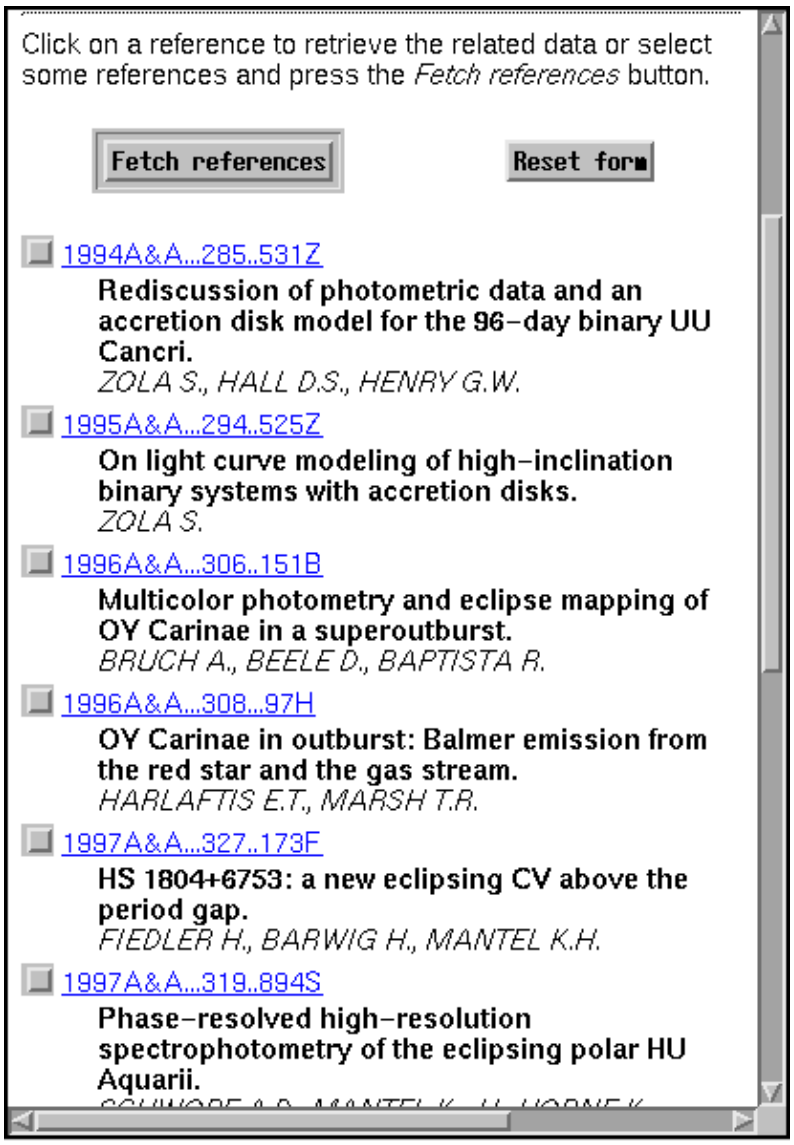

Fig. B.5: *Liste des documents contenus dans la classe choisie. Ils sont classés par ordre de pertinence à la requête (pertinence système) décroissante.*

Annexe C

Les algorithmes de descente de gradient

La plupart des algorithmes d'apprentissage des réseaux de neurones sont basés sur une descente de gradient d'une fonction d'erreur (aussi appelée fonction de coût) : $Err(w_i)$. Le but de ces apprentissages est d'adapter les paramètres w_i des réseaux de manière à minimiser cette fonction.

Nous allons ici décrire la descente de gradient dans le cas général d'une fonction scalaire $f(\overrightarrow{w})$ de plusieurs variables w_i que l'on cherche à minimiser.

Développement limité de f

Écrivons le développement limité de f en un point $\overrightarrow{w(t)}$:

$$f(\overrightarrow{w(t)} + \Delta\overrightarrow{w}) = f(\overrightarrow{w(t)}) + \overrightarrow{grad}\left(f|_{\overrightarrow{w}=\overrightarrow{w(t)}}\right).\Delta\overrightarrow{w} + o(\| \Delta\overrightarrow{w} \|)$$

Pour des variations suffisamment petites de $\overrightarrow{w}$, on peut écrire :

$$\begin{aligned} \Delta f|_{\overrightarrow{w}=\overrightarrow{w(t)}} &= f(\overrightarrow{w(t)} + \Delta\overrightarrow{w}) - f(\overrightarrow{w(t)}) \\ &\simeq \overrightarrow{grad}\left(f|_{\overrightarrow{w}=\overrightarrow{w(t)}}\right).\Delta\overrightarrow{w} \end{aligned}$$

L'algorithme de descente du gradient

Comme nous voulons minimiser f, nous sommes menés à choisir $\Delta\overrightarrow{w}$ colinéaire à $\overrightarrow{grad}\left(f|_{\overrightarrow{w}=\overrightarrow{w(t)}}\right)$, et de sens opposé, soit :

$$\Delta\overrightarrow{w} = -k.\overrightarrow{grad}(f) \text{ avec } k > 0.$$

D'où l'algorithme de descente du gradient :

1. Choisir un point de départ $\overrightarrow{w(0)}$ proche du minimum supposé, ou aléatoirement.
2. Répéter :
$$\overrightarrow{w(t+1)} = \overrightarrow{w(t)} - k.\overrightarrow{grad}(f|_{\overrightarrow{w}=\overrightarrow{w(t)}})$$
tant que le minimum n'a pas été suffisamment approché.

Descente de gradient et optimisation

Pour certains problèmes, comme l'apprentissage des réseaux de neurones, la fonction à minimiser est de la forme :

$$g(\overrightarrow{w}, \mathcal{E}) = \sum_{i=1}^{N} f(\overrightarrow{w}, \overrightarrow{e_i})$$

où $\mathcal{E}$ est un ensemble de N vecteurs $\overrightarrow{e_i}$. Nous sommes ici face à un problème d'optimisation qui consiste à trouver le vecteur $\overrightarrow{w}$ qui minimise g étant donné l'ensemble $\mathcal{E}$.

Si on applique la formule de la descente du gradient sur g, on obtient :

$$\begin{aligned} \Delta\overrightarrow{w} &= -k.\overrightarrow{grad}(g) \\ &= -k.\sum_{i=0}^{N} \overrightarrow{grad|_w}\,(f(\overrightarrow{w}, \overrightarrow{e_i})) \end{aligned} \qquad \text{(C.1)}$$

Deux algorithmes de descente du gradient de g sont envisageables :

La descente de gradient déterministe

1. Choisir un point de départ $\overrightarrow{w(0)}$ proche du minimum supposé, ou aléatoirement.
2. Répéter :
$$\overrightarrow{w(t+1)} = \overrightarrow{w(t)} - k.\sum_{i=0}^{N} \overrightarrow{grad|_w}\,(f(\overrightarrow{w}, \overrightarrow{e_i}))$$
tant que le minimum n'a pas été suffisamment approché.

Il s'agit ici simplement de l'algorithme précédent, adapté à la forme particulière de g.

La descente de gradient stochastique

1. Choisir un point de départ $\overrightarrow{w(0)}$ proche du minimum supposé, ou aléatoirement.
2. Répéter :
   - choisir au hasard un vecteur $\overrightarrow{e_i}$ de $\mathcal{E}$
   - calculer :

$$\overrightarrow{w(t+1)} = \overrightarrow{w(t)} - k.\overrightarrow{grad|_w}\,(f(\overrightarrow{w}, \overrightarrow{e_i}))$$

   tant que le minimum n'a pas été suffisamment approché.

Ce second algorithme sera utilisé dans tous les cas où on ne souhaite pas traiter les données de $\mathcal{E}$ dans leur ensemble, mais individuellement, contrairement à l'algorithme de descente de gradient déterministe. La méthode stochastique est indiquée dans les cas suivants :

- lorsque l'ensemble $\mathcal{E}$ est très grand, ce qui rend la méthode déterministe difficile à utiliser en raison de l'importance des ressources informatiques qu'elle demande ;
- dans les cas où toutes les données ne sont pas connues en même temps, par exemple lorsqu'elles sont générées suivant une distribution (lorsque seule cette distribution est connue), ou encore lorsque ce sont des données issues d'une expérience et qu'elles doivent être utilisées dès qu'elles sont générées.

Nous voyons dans le chapitre 3 que dans la majorité des cas, les algorithmes d'apprentissage des réseaux de neurones sont équivalents à des méthodes de descente de gradient stochastiques.

Annexe D

Généralisation de la règle du delta

Nous avons vu dans le chapitre 3 que l'algorithme d'apprentissage de l'ADALINE correspondait à l'algorithme de descente stochastique du gradient de l'erreur quadratique pour un neurone doté d'une fonction d'activation linéaire.

Nous allons voir maintenant ce que devient cet algorithme pour les cas où la fonction d'activation F des neurones n'est pas linéaire.

La descente du gradient de l'erreur quadratique Err s'écrit :

$$w_i(t+1) = w_i(t) - k.\frac{\partial Err}{\partial w_i} \tag{D.1}$$

L'expression de l'erreur quadratique à la sortie d'un neurone d'activation A en approximation de la valeur désirée R est la suivante :

$$\begin{aligned} Err &= \frac{1}{2}.(\Delta)^2 \\ &= \frac{1}{2}.\left(R - F(\sum_i w_i.e_i + \theta)\right)^2 \end{aligned}$$

où F est la fonction d'activation (généralement, la fonction sigmoïde), soit :

$$\begin{aligned} \frac{\partial Err}{\partial w_i} &= -e_i.F'(\sum_i w_i.e_i + \theta).(R - F(\sum_i w_i.e_i + \theta)) \\ &= -e_i.\Delta.F'(\sum_i w_i.e_i + \theta) \end{aligned} \tag{D.2}$$

Si on remplace le terme $\frac{\partial Err}{\partial w_i}$ dans l'équation (D.1) par sa valeur en (D.2), on obtient la règle du delta généralisé :

$$w_i(t+1) = w_i(t) + k.e_i.\Delta.\underbrace{F'(\sum_i w_i.e_i + \theta)}_{\text{terme de généralisation}}$$

Annexe E

Les paramètres de l'apprentissage : quelques résultats

E.1 Expressions des fonctions $\alpha(t)$ et $h(r,t)$

Voici l'expression des fonctions que nous avons utilisées pour $\alpha(t)$ et $h(r,t)$ afin de déterminer un couple de paramètres permettant une classification où $T1$ et $T2$ sont minimisés simultanément de manière optimale.

$\alpha(t)$			$h(r,t)$
			en cloche
$\alpha(t) = \alpha_{\text{init}}.\left(\frac{\alpha_{\text{fin}}}{\alpha_{\text{init}}}\right)^{\frac{t}{t_{\text{fin}}}}$	A	a	$h(r,t) = \exp(\frac{-r^2}{\sigma(t)})$ avec $\sigma(t) = \sigma_{\text{init}}.\left(\frac{\sigma_{\text{fin}}}{\sigma_{\text{init}}}\right)^{\frac{t}{t_{\text{fin}}}}$
			en cône
$\alpha(t) = \alpha_{\text{init}}.(1-\frac{t}{t_c})$ si $t < tc$ sinon, $\alpha(t) = \alpha_{\text{fin}}$	B	b	$h(r,t) = 1 - r/r_c(t)$ si $r < r_c(t)$, 0 sinon. et $h(0,t) = 0$ avec $r_c(t) = r_{\text{init}}.(1-\frac{t}{t_c})$ si $t < t_c(t)$, 0 sinon.
			en chapeau mexicain
$\alpha(t) = \frac{U}{V+t}$	C	c	$h(r,t) = \frac{\sin(R(t))}{R(t)}$ avec $R(t) = \frac{2\pi.r}{r_c(t)-1}$ et $r_c(t) = (\frac{1}{r_{\text{init}}})^{\frac{t}{t_{\text{fin}}}}$

Tab. E.1: *Expressions des fonctions $\alpha(t)$ et $h(r,t)$.*

E.2 Variations de $T1$ et $T2$ en fonction de $\alpha(t)$ et $h(r,t)$

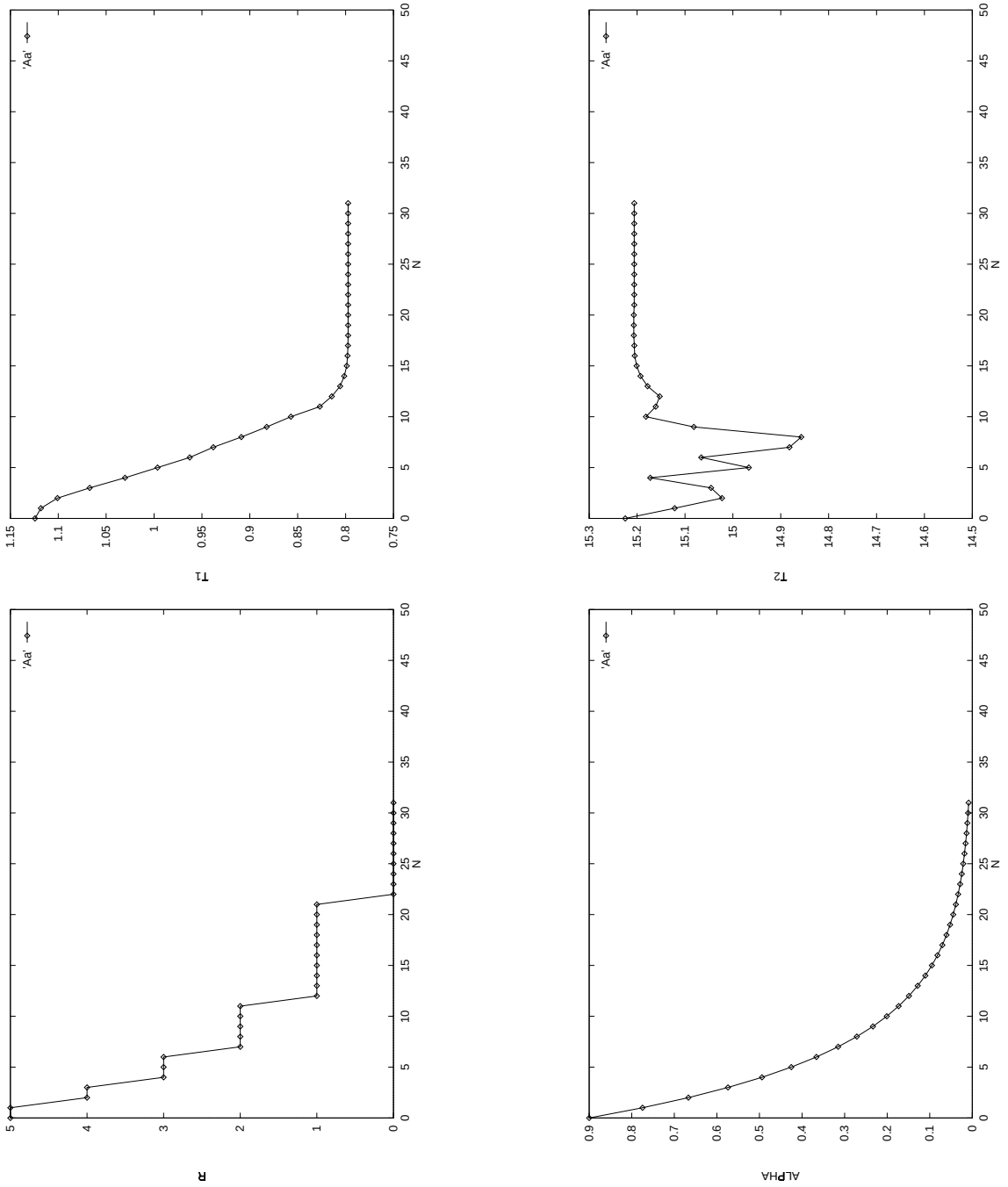

Fig. E.1: *Variations de $T1$, $T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Aa.*

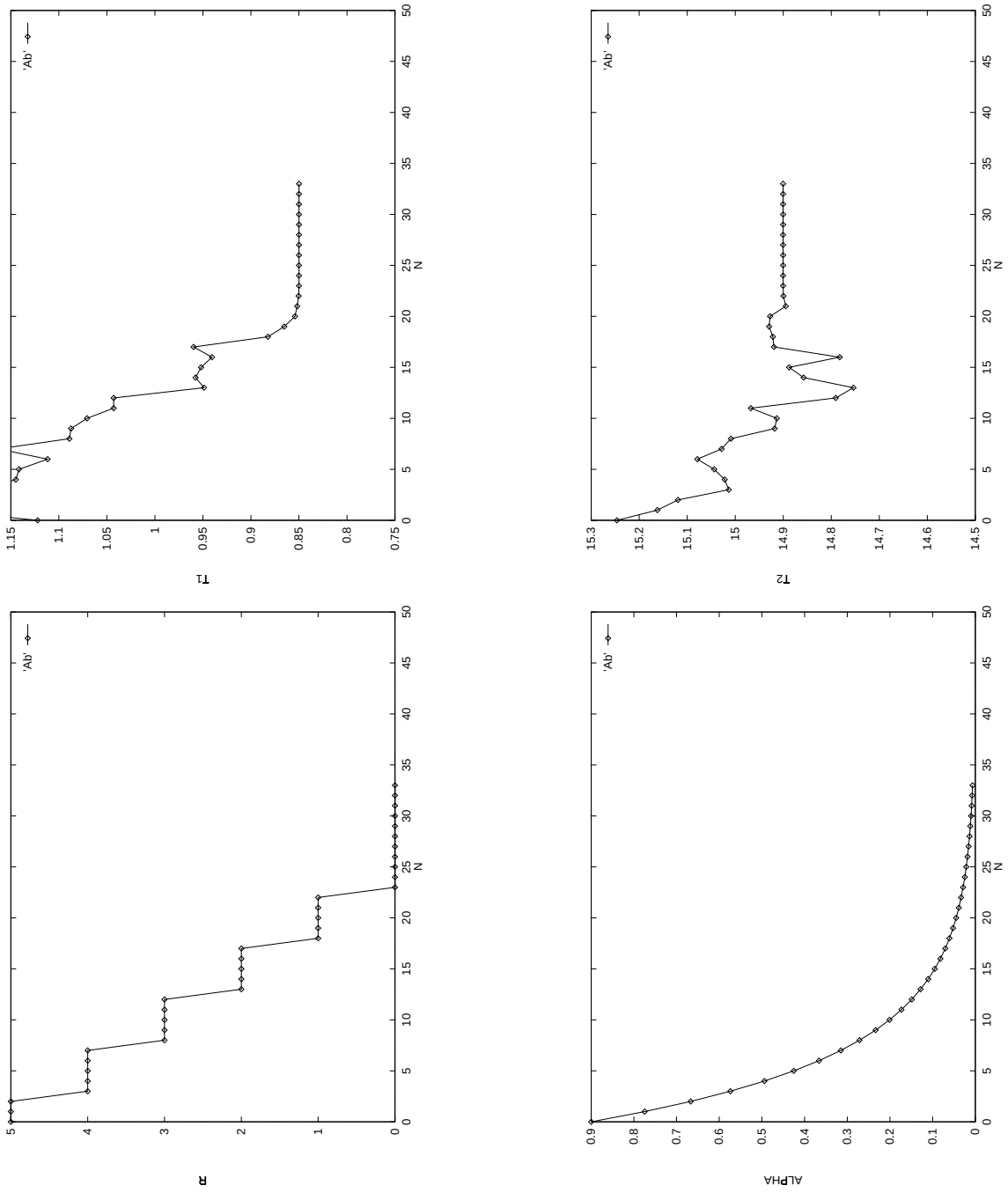

Fig. E.2: *Variations de $T1, T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Ab.*

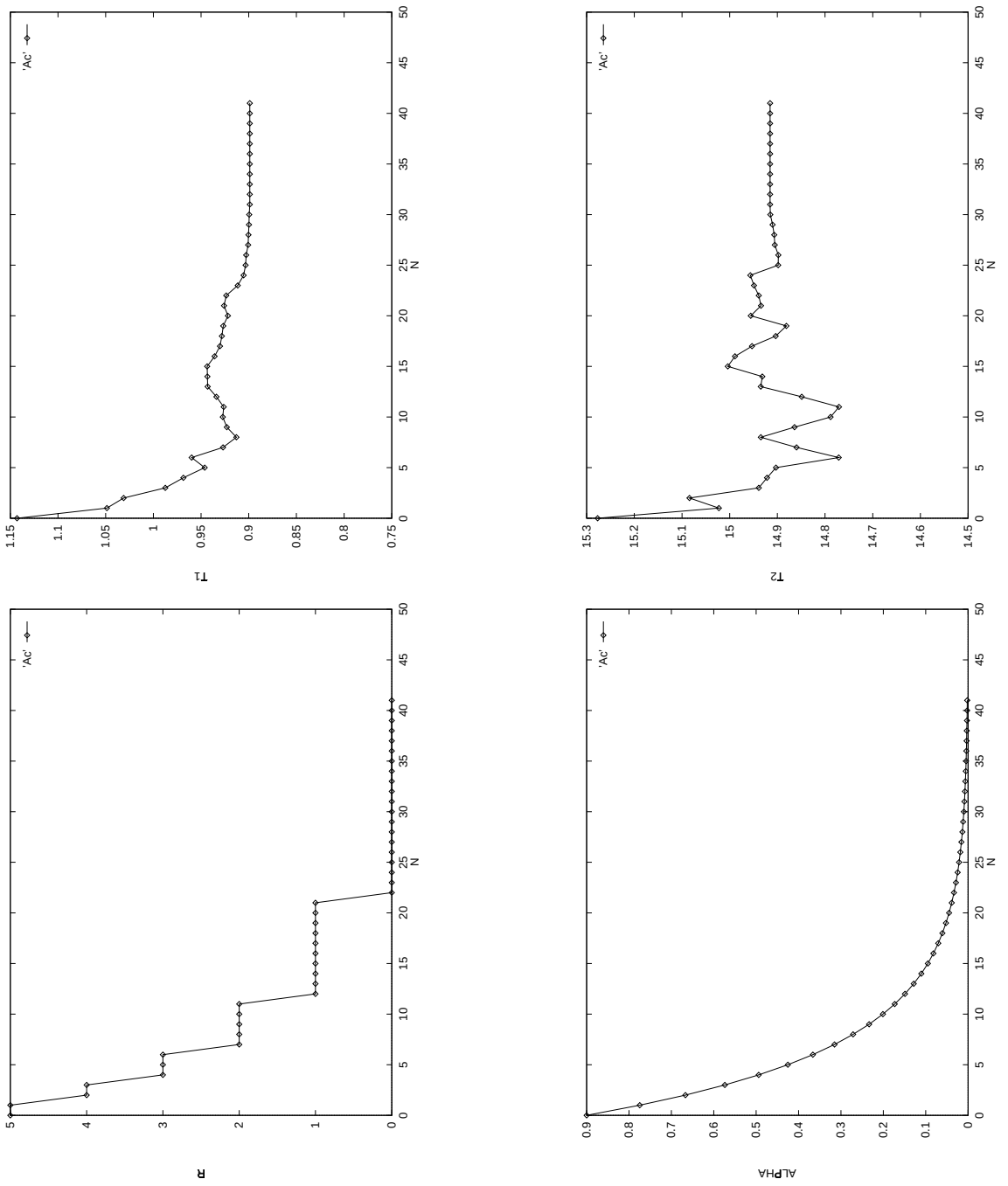

Fig. E.3: *Variations de $T1$, $T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Ac.*

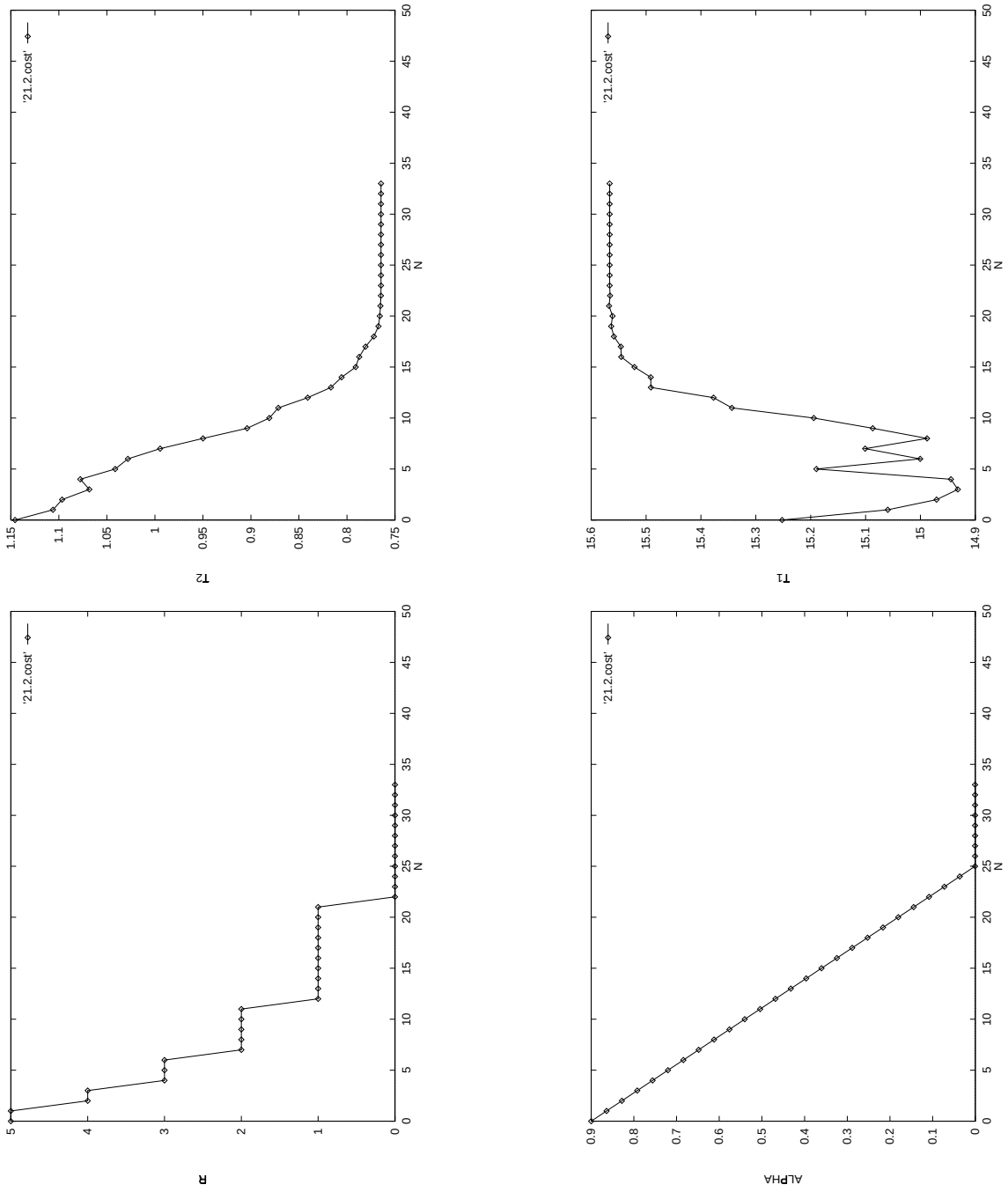

Fig. E.4: *Variations de $T1, T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Ba.*

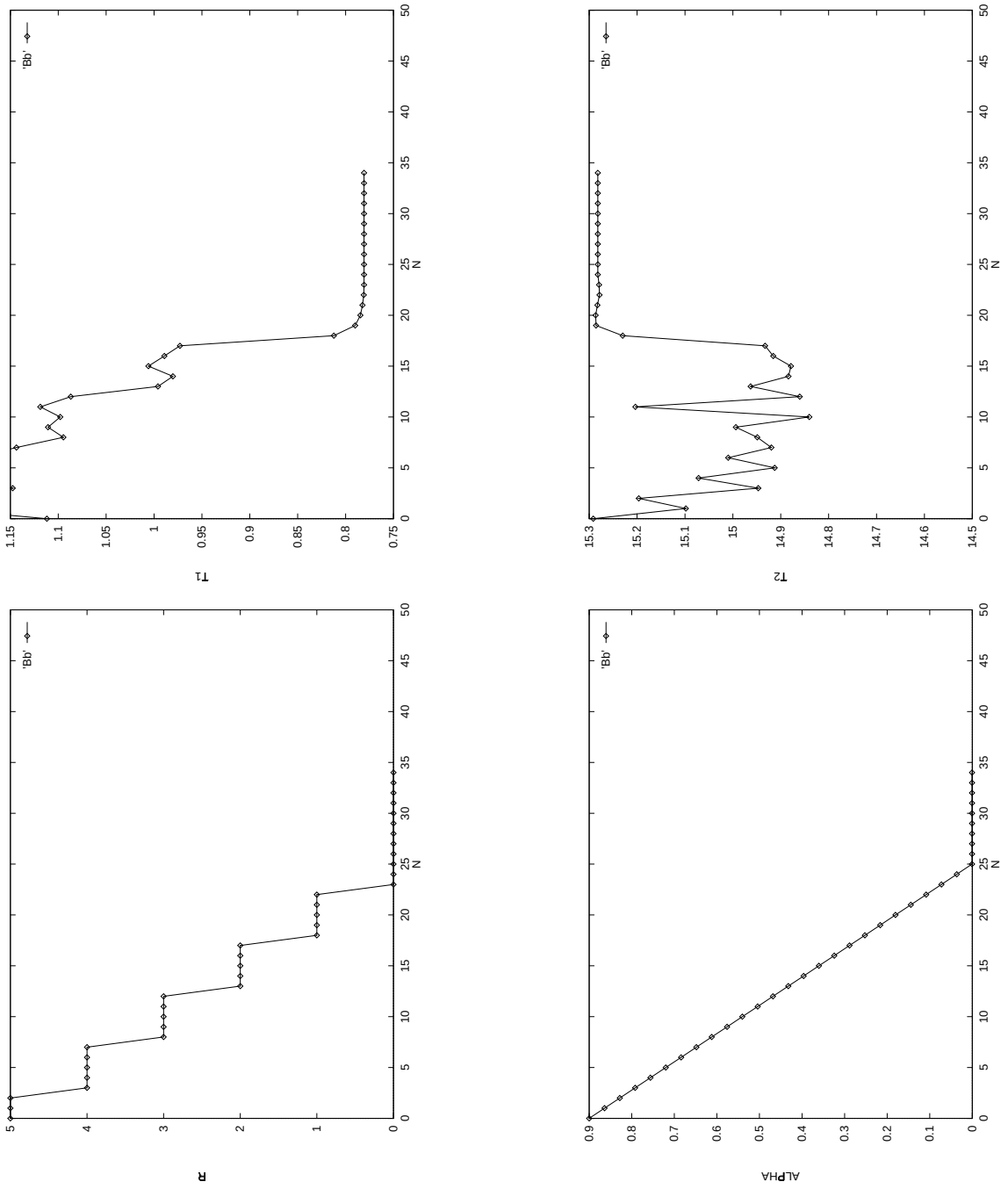

Fig. E.5: *Variations de $T1$, $T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Bb.*

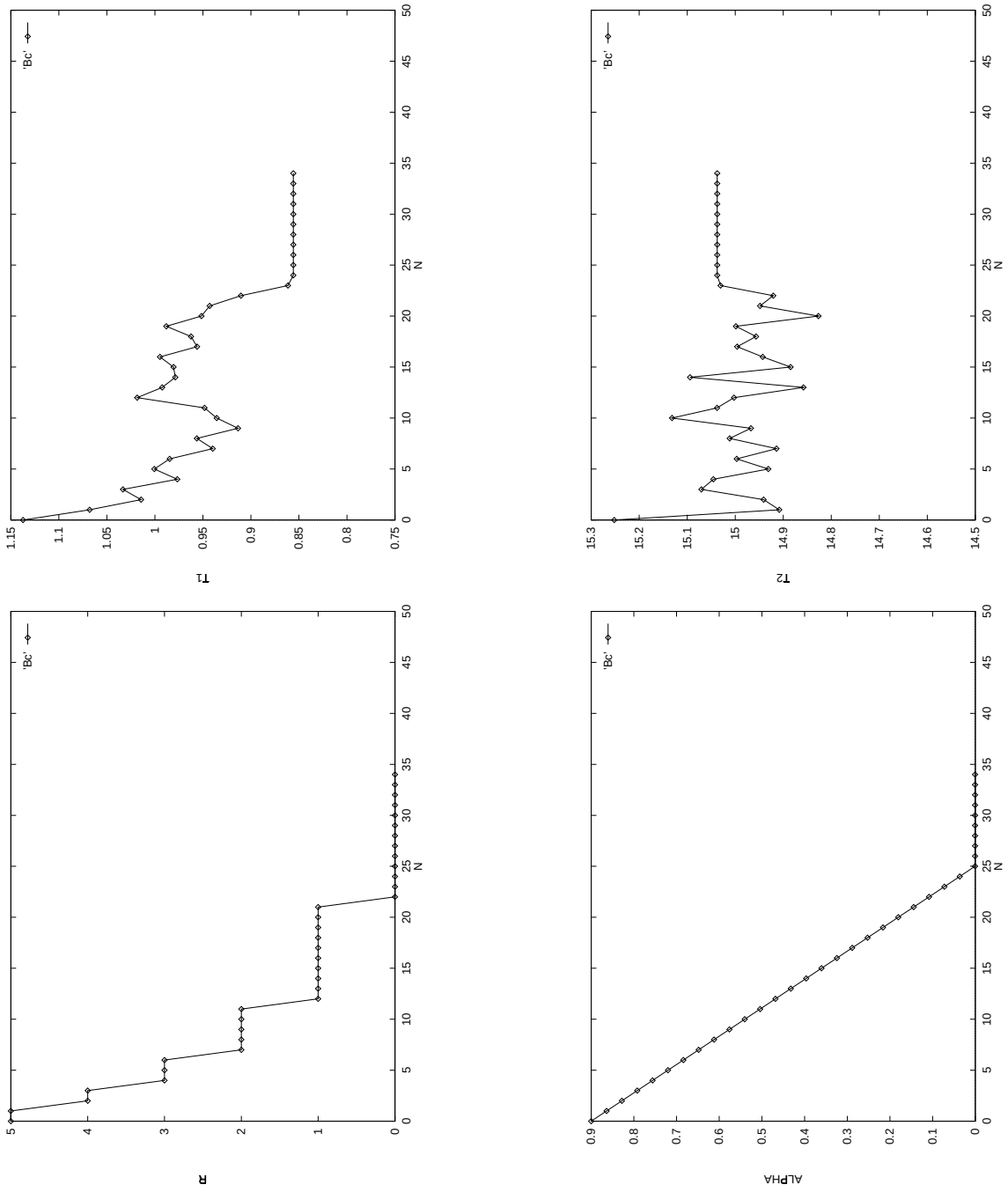

Fig. E.6: *Variations de $T1, T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Bc.*

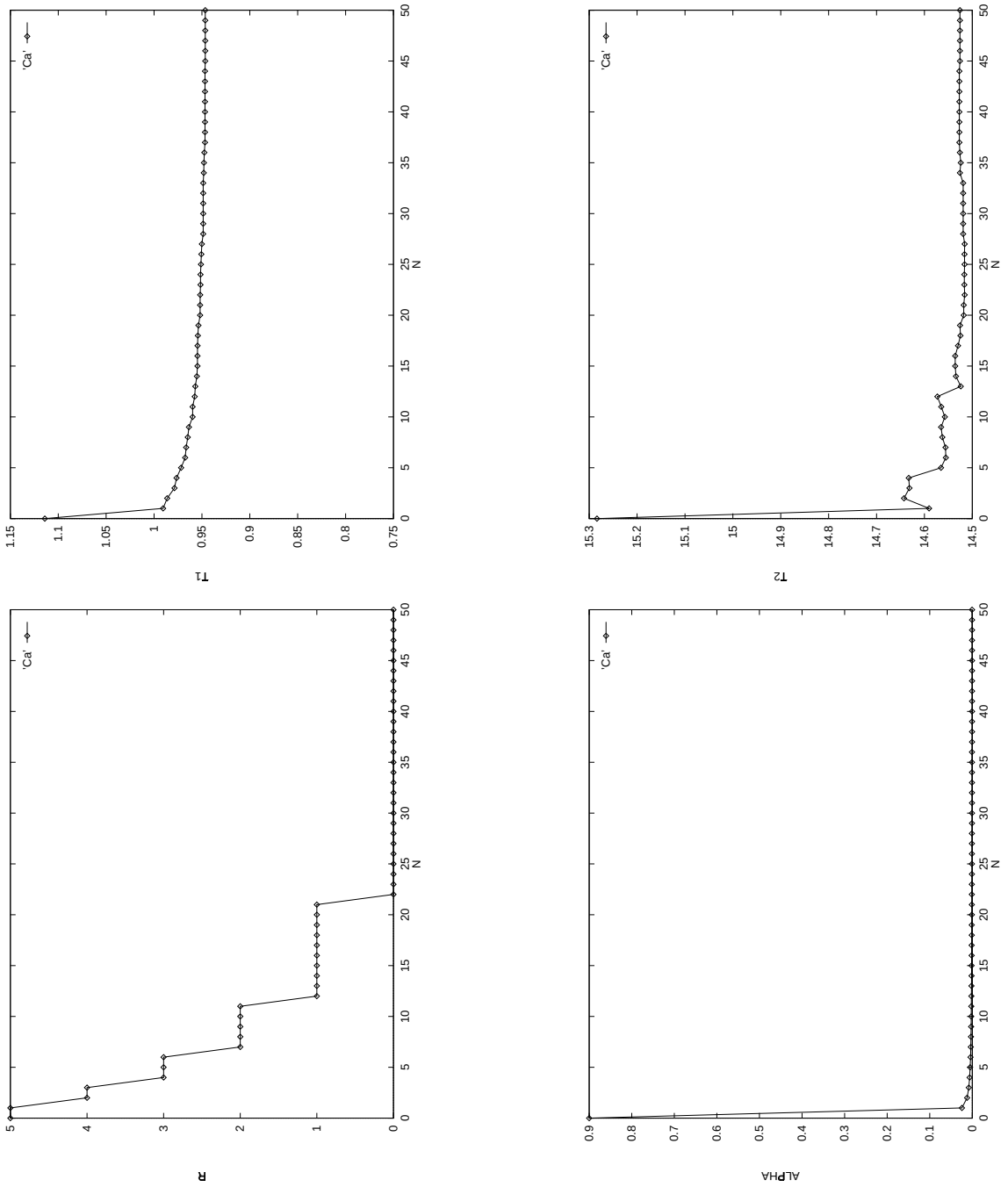

Fig. E.7: *Variations de $T1$, $T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Ca.*

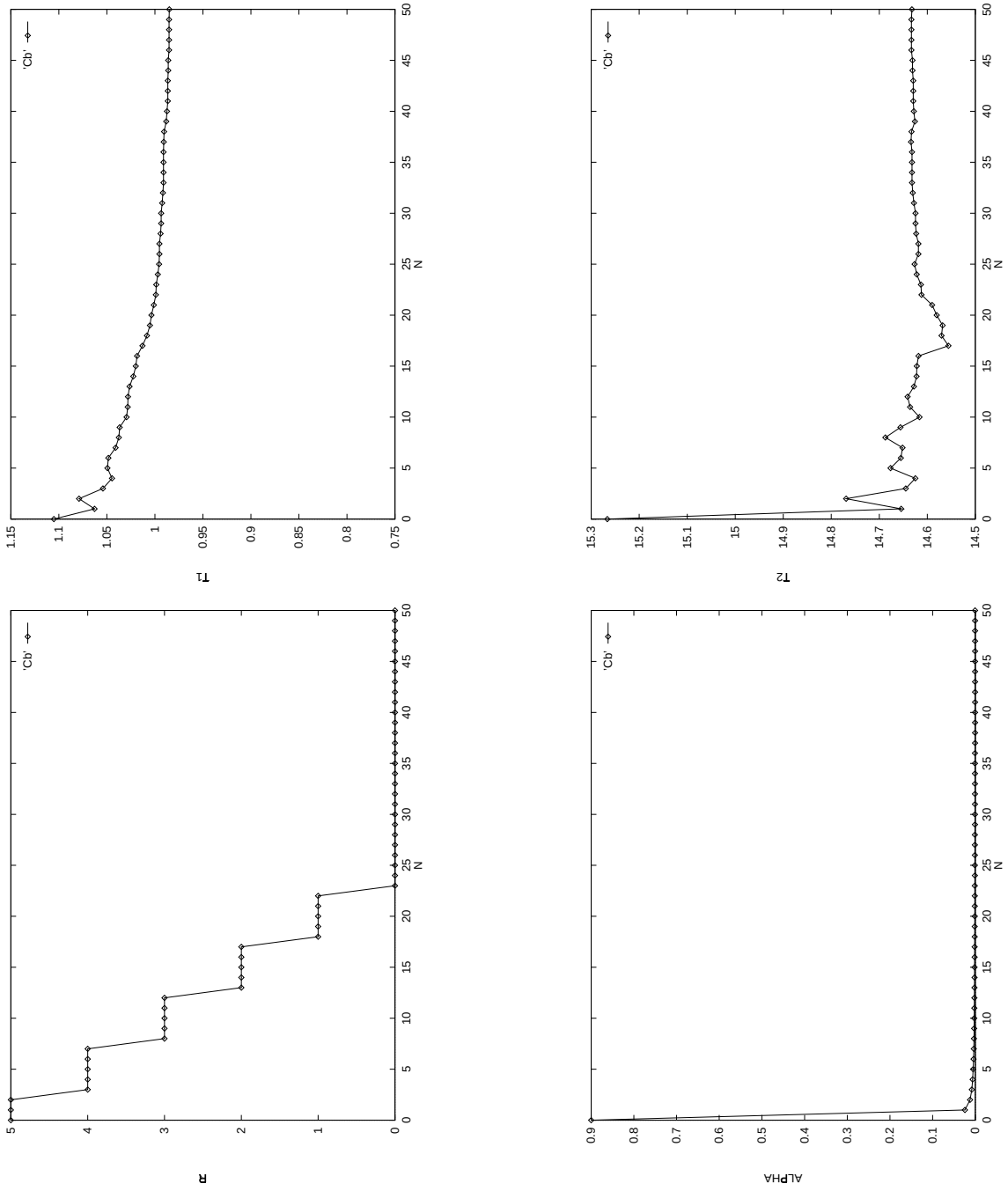

Fig. E.8: *Variations de $T1, T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Cb.*

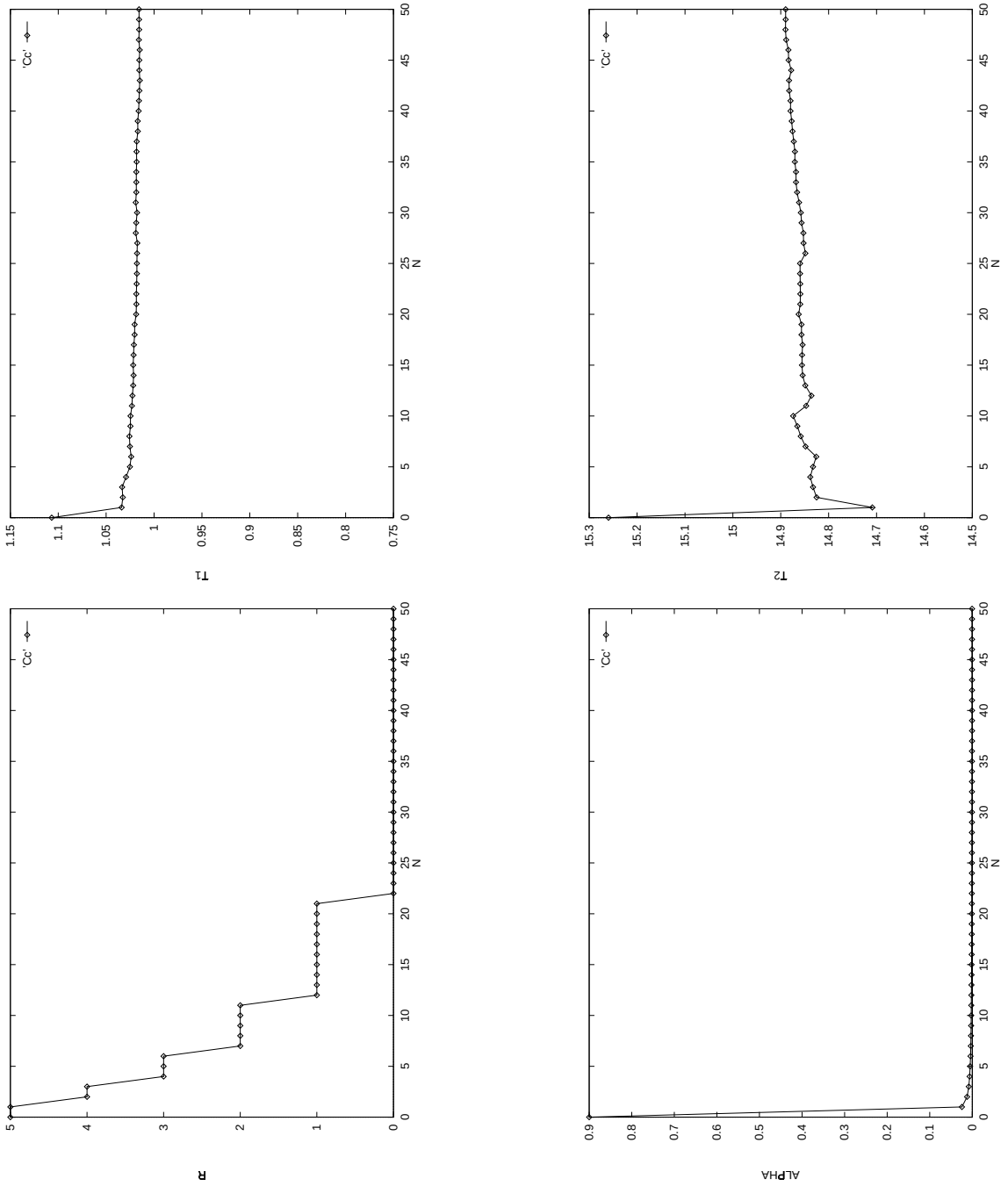

Fig. E.9: *Variations de $T1$, $T2$, ainsi que du rayon de corrélation R et du paramètre d'apprentissage α au cours de l'apprentissage, pour le couple de fonctions Cc.*