

OPTIMISATION D'UN MODÈLE DE RÉSEAU DE NEURONES ROBUSTE À PARTIR DE LA RÉGULARISATION BAYÉSIENNE ET LA MÉTHODE D'AUTOCORRÉLATION POUR LA PRÉDICTION DE LA TEMPÉRATURE D'ÉBULLITION DES ALCANES

A QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIPS MODEL FOR THE BOILING POINT OF ALCANES USING BAYESIAN-REGULARIZED NEURAL NETWORKS

Mohamed NOHAIR* et Mohssine El MORRAKCHI

Laboratoire Catalyse, Synthèse, Chimiométrie et Environnement.
Département de Chimie, Université Hassan II-Mohammedia, Faculté des Sciences et Technique de Mohammedia B.P. 146
20650 Mohammedia-Maroc.
Fax : 212-023-31-53-53

*Correspondance: nohairmohamed@yahoo.fr

Received November 14, 2007

Nous avons utilisé un nouveau modèle de type QSPR pour prédire la température d'ébullition des alcanes, en utilisant les réseaux de neurones et la régularisation bayésienne. Les variables indépendantes utilisées dérivent de la méthode d'autocorrélation. Le descripteur topologique consiste en quatre composantes d'autocorrélation. Nous avons testé le pouvoir prédictif de notre modèle de réseau de neurones par la technique de la validation croisée (leave-20%-out cross-validation method). La technique de la régularisation bayésienne a permis de fournir un modèle d'une grande capacité de généralisation (robustesse) dans la prédiction. La température des 143 molécules d'alcanes utilisées est corrélée de façon très satisfaisante. Le coefficient de corrélation et l'écart type étaient de l'ordre de 99 et 1,71°C respectivement.

We have used a new, robust structure-activity mapping technique, a Bayesian-regularized neural network, to develop a quantitative structure-activity relationships (QSAR) model for boiling point of alkanes. The independent variables used in the modeling were derived solely from the autocorrelation method. The descriptor generated consists of four autocorrelation components. The ability of the prediction was tested by the leave-20%-out cross-validation method. The results show that the method is robust and give satisfied results. The established model allows us the prediction of the boiling point of alkanes with success. Correlation coefficient was close to 0.99 and the standard deviation of 1.71°C.

INTRODUCTION

Récemment, les modèles linéaires largement utilisés pour établir des relations statistiques, sont supplantés par une nouvelle technique basée sur l'intelligence artificielle: ce sont les réseaux de neurones.¹ Ils permettent de proposer un modèle de type cause à effet sans qu'il soit nécessaire de postuler a priori, un modèle de corrélation figé. Néanmoins, ils présentent un inconvénient majeur, il réside dans le surajustement, et donc des capacités prédictives, dans certains cas, affaiblies. Ceci revient à la définition d'une architecture souple de la RNA, c'est-à-dire le nombre de

neurones et le type d'interconnexion entre elles. Il n'existe pas jusqu'à présent dans la littérature de procédure standard pour résoudre ce genre de problème, excepté la technique de l'arrêt immédiat de l'apprentissage¹ et la régularisation bayésienne.²⁻⁴ La première technique pose le problème du choix du meilleur modèle à partir de plusieurs modèles obtenus par le processus de validation. Ce processus consiste à décomposer l'ensemble initial de données en deux sous-ensembles, le premier sert à l'apprentissage et le second permet la validation et la fixation des paramètres de la RNA. La seconde technique est connue par sa robustesse, puisqu'elle ne nécessite

pas un ensemble de données pour la validation. Nous proposons dans ce papier un travail relatif à l'utilisation de la régularisation bayésienne pour la réalisation d'un modèle de type QSPR entre la température d'ébullition des alcanes et un descripteur topologique dérivant de la méthode d'autocorrélation.⁵ Cette technique génère un vecteur formé par un ensemble de composantes d'autocorrélation. La température d'ébullition est un processus régi par la force des interactions de type de Van Der Waals entre les molécules apolaires des alcanes, et le vecteur d'autocorrélation ne reflète qu'une description topologique de cet effet.

GENERALITÉS SUR LES DONNÉES ET LA MÉTHODE D'AUTOCCORÉLATION

1. Données

Le choix de l'ensemble des molécules d'alcanes, constituant la base de données, est très important, celle utilisée dans cette étude correspond à 143 valeurs différentes de températures d'ébullition variant de -88,60°C dans l'éthane à 174,12°C dans le décane.⁶ Cette base de données a la particularité de contenir tous les types d'interaction pouvant exister entre deux atomes proches (primaire-secondaire, primaire-tertière, primaire-quaternaire.....) (Tableau I).

Tableau 1

Statistiques relatives à la famille d'alcanes étudiée.

Nombre de molécules	143
T°C maximale	174°C
T°C minimale	-88°C
T°C moyenne	132,80°C
Ecart type	43,33°C

2. Méthode d'Autocorrélation

La méthode d'autocorrélation a trouvé son origine à partir de la fonction mathématique d'autocorrélation $F(\theta)$, qui est définie de la manière suivante :

$$F(\theta) = \sum_{AB} f(x)f(x+\theta)dx \quad (1)$$

θ est une distance qui détermine éventuellement la périodicité de la fonction.

L'introduction de la méthode d'autocorrélation⁷⁻¹¹ dans le domaine des relations structurepropriété et structure-activité¹² avait pour objectif la représentation de la structure d'une molécule par un ensemble de paramètres, qui sont uniques, faciles à calculer, et en plus, ayant une signification structurale. Dans la méthode d'autocorrélation, un composé organique est représenté par un graphe, composé de nœuds et d'arêtes. La relation générale pour calculer une composante d'autocorrélation pour une molécule est définie ci-dessous:

$$P_k = \sum (f(i)f(j))^x \quad (2)$$

x est égal à 1 dans le cas classique de la méthode et égale à 0.5 dans le cas de la méthode d'autocorrélation modifiée. P_k est la composante d'autocorrélation correspondant à la distance topologique k . La contribution atomique $f(i)$ dépend de la propriété choisie dans l'étude. Ces propriétés peuvent être le volume de Van Der Waals (V) ou la surface pour une modélisation de la forme de la molécule, ou encore l'électronégativité pour modéliser l'aspect électronique de la molécule. Elles peuvent être aussi, et tout simplement la connectivité ou la valence atomique.

Nous pouvons alors décrire la molécule par un ensemble de paramètres qui permettent au moins de lever la dégénérescence sur le codage des molécules par exemple. Néanmoins, il reste à donner une interprétation physique de chaque composante. Si par exemple nous étudions une propriété physique additive, tel que le volume de Van Der Waals, la composante V_θ , en référence à P_θ quand $f(i)$ représente le volume atomique de Van Der Waals, ne représente pas la somme des contributions des atomes. C'est pour ces raisons que Chastrette et Tiyal⁷ ont modifié la forme générale de calcul des composantes du vecteur d'autocorrélation. La relation générale devient alors :

$$P_k = \sum_{j \geq i} (f(i)f(j))^{0.5} \quad (3)$$

Cette nouvelle forme a permis d'améliorer nettement les paramètres statistiques lors de l'établissement de relations structure-volume molaire.⁷ Dans ce travail, nous proposons une autre méthode pour établir les composantes du vecteur de la méthode d'autocorrélation, et ce en relation avec le concept du système DARC.¹³ Nous faisons intervenir tout simplement les propriétés des atomes dans le fragment considéré. Seuls les

fragments de distance topologique supérieure ou égale à 2 sont considérés. La relation précédente devient alors:

$$P_K = \sum_{j \geq i} (f(i) (\sum f(k_{ij})) f(j))^{1/n} \quad (4)$$

k_{ij} représente l'atome intermédiaire entre les atomes i et j et n le nombre de ces atomes. $1/n$ est pris pour de telle façon que les composantes aient la même unité de grandeur.

La description d'une molécule par la topologie moléculaire, en l'occurrence la méthode d'autocorrélation ne fait intervenir que son graphe et ses matrices de connexion et de distance. Nous avons élaboré alors un programme qui permet de calculer les composantes d'autocorrélation pour tous les vecteurs cités précédemment à partir de

$$\begin{aligned} T^{\circ}\text{C}_{\text{calc}} = & 0,03e^{-7} + 0,99*V_0 + 0,19*V_1 - 0,05*V_2 + 0,05*V_3 - 0,08*V_4 - 0,16*V_5 \\ & \pm 0,18 \quad \pm 0,11 \quad \pm 0,08 \quad \pm 0,04 \quad \pm 0,02 \quad \pm 0,02 \\ & r=0,99 ; s=5,13^{\circ}\text{C} \end{aligned}$$

De par la valeur des coefficients, la composante V_0 est de loin la plus importante, elle contribue positivement. La contribution des autres variables est non significative, excepté V_1 et V_5 . Les

leurs codes SMILES (Simplified Molecular Input Linear Entry System)¹⁴.

ÉVALUATION PARAMÉTRIQUE DES DESCRIPTEURS PAR RÉGRESSION LINEAIRE

Dans un premier temps, nous avons défini le descripteur topologique, en calculant six composantes de la méthode d'autocorrélation liés au volume de Van Der Waals ($V_0, V_1 \dots V_6$). La régression ne servirait qu'à une évaluation paramétrique du descripteur retenu pour les réseaux de neurones. Avant tout apprentissage, toutes les variables ont été normalisées et centrées pour éviter la surestimation de l'effet des variables ayant des grandeurs importantes. La relation qui en résulte conduit à l'équation suivante :

paramètres statistiques (r et s) sont relativement satisfaisants, mais la distribution de l'erreur (Figure 1) fait apparaître des températures, dont l'erreur est très importante et non significative.

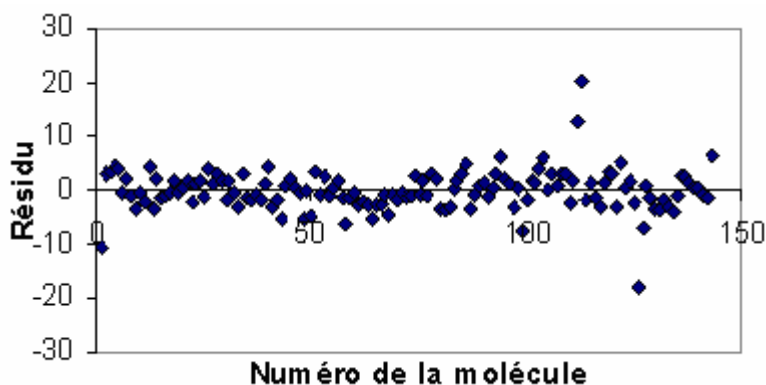


Fig. 1 – Distribution de l'erreur (écart type) entre la température observée et calculée.

Par ailleurs, la matrice de corrélation fait apparaître de fortes corrélations entre les composantes V_i (Tableau 2). Par conséquent, l'interprétation de l'effet des variables explicatives n'a de sens que si la régression ne tient pas compte des éventuelles associations entre elles. Les coefficients de chaque variable doivent être épurés des contributions fournies par les autres variables et représentent, donc, des effets propres.

Pour ce faire, nous avons réalisé une régression pas-à-pas (stepwise regression 'forward selection'). Elle consiste en la construction des 6

modèles possibles à une variable seulement. Nous calculons les critères r et s (coefficient de corrélation et écart type) pour chacun de ces modèles et nous choisissons celui qui minimise le critère. Nous fixons ensuite la variable déjà sélectionnée; (6-1) modèles sont alors construits en ajoutant une seule des variables restantes. Les critères r et s pour chacun de ces modèles sont calculés, et nous choisissons le modèle qui minimise ce critère. Nous continuons le processus ci-dessus jusqu'à ce que les valeurs des critères ne s'améliorent plus. Cette opération de sélection a

permis, en commençant par la variable de forte contribution V_0 , de retenir quatre variables dont les

paramètres statistiques sont tous significatifs (Tableau 3).

Tableau 2

Matrice de corrélation entre les composantes d'autocorrélation et la température d'ébullition.

	V_0	V_1	V_2	V_3	V_4	V_5	T_{ebb}
V_0	1,00						
V_1	0,85	1,00					
V_2	0,79	0,37	1,00				
V_3	0,72	0,44	0,61	1,00			
V_4	0,79	0,64	0,62	0,62	1,00		
V_5	0,62	0,75	0,31	0,13	0,33	1,00	
T_{ebb}	0,98	0,86	0,73	0,75	0,76	0,56	1,00

Tableau 3

Résultat de la régression pas-à-pas

	V_0	V_0, V_2	V_0, V_2, V_4	V_0, V_2, V_4, V_5
(r et s)	(98; 7,6)	(98; 6,98)	(98; 5,52)	(99; 5,15)

Les coefficients des variables retenues sont tous significatifs. La variable V_0 contribue fortement et positivement. Elle peut être attribuée à l'effet de la dimension. En effet quand la dimension de la molécule augmente, les interactions de type de Van Der Waals augmentent et par conséquent la température d'ébullition augmente. Les variables restantes sont faibles et négatives, elles sont probablement liées à l'effet de volume.

RÉALISATION DU MODÈLE NEURONAL POUR LA PRÉDICTION DE LA TEMPERATURE D'ÉBULLITION DES ALCANES

1. Introduction

Un neurone est avant tout un opérateur mathématique. Il réalise une somme pondérée suivie d'une fonction non linéaire. La fonction f doit être bornée, continue et dérivable, elle peut aussi avoir la forme d'une fonction seuil si le résultat recherché est de type booléen (soit 0 ou 1). Les fonctions les plus fréquemment utilisées sont les fonctions sigmoïdes. Par ailleurs, il est démontré qu'un réseau de neurones ayant une couche cachée avec une combinaison d'une

fonction sigmoïde et une fonction linéaire est un approximateur universel dans le sens qu'il peut approximer toute fonction multivariable non linéaire et ses dérivées avec une précision fixée, pourvu qu'il soit soumis à un apprentissage efficace.¹⁵

Le principe des réseaux de neurones est le même que n'importe quel phénomène statistique caractérisé par une relation déterministe entre des causes et des effets. Nous présentons au réseau des exemples d'apprentissage, c'est-à-dire des jeux d'activités des neurones d'entrée ainsi que l'activité du neurone de sortie correspondant. L'apprentissage consiste tout simplement à calculer les coefficients de connexion entre les différentes couches de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés, aussi proches que possible des sorties désirées. Pour ce faire, on utilise des algorithmes d'optimisation : ils cherchent à minimiser une fonction de coût qui constitue une mesure de l'écart entre la réponse réelle du réseau et la réponse désirée. Cette optimisation se fait de manière itérative, en modifiant les poids en fonction du gradient de la fonction du coût : le gradient est estimé par une méthode spécifique aux réseaux de neurones, dite méthode de rétropropagation.¹⁶ Les poids sont initialisés

aléatoirement avant l'apprentissage, puis modifiés itérativement jusqu'à obtention d'un compromis satisfaisant entre la précision de l'approximation sur un ensemble de validation disjoint au précédent, et celle obtenue sur l'ensemble des données d'apprentissage.

2. Mise en œuvre des modèles de réseaux de neurones

Pour enseigner une tâche à un réseau, on doit ajuster les coefficients de chaque neurone et

$$J(w) = \sum_{k=1}^N (y^k - g(x^k, w))^2 = \sum_{k=1}^N J^k(w) \quad (5)$$

Où y : température observée, $g(x, w)$ = température calculée.

Pour évaluer ce gradient, il suffit donc d'évaluer le gradient du coût partiel $J^k(w)$ relatif à l'observation k , et de faire ensuite la somme sur tous les exemples. L'algorithme de rétropropagation consiste essentiellement en l'application répétée de la règle des dérivées composées (eq.6).

$$\left(\frac{\partial J^k}{\partial w_{ij}} \right) = \left(\frac{\partial J^k}{\partial v_i} \right) \left(\frac{\partial v_i}{\partial w_{ij}} \right)_k = \delta_i^k x_j^k \quad (6)$$

δ_j^k désigne la valeur du gradient du coût partiel par rapport au potentiel du neurone i pour l'exemple k , il peut être exprimé de la manière suivante (eq.7) :

$$\delta_j^k = \left(\frac{\partial J^k}{\partial v_i} \right) = -2g(x^k, w) \left(\frac{\partial g(x, w)}{\partial v_i} \right)_k \quad (7)$$

La même approche est appliquée aux neurones cachés, les paramètres du réseau sont modifiés par la formule suivante, à l'itération i de l'apprentissage (eq.8) .

$$w(i) = w(i-1) - \mu_i \nabla J(w(i-1)) \quad (8)$$

La quantité μ_i représente le pas du gradient ou pas d'apprentissage.

$$T_{exp} = a * T_{pred} + b.$$

T désigne la température d'ébullition.

Les paramètres statistiques sont calculés à partir des relations ci-dessous :

$$s^2 = \frac{\sum (T_{obs} - T_{calc})^2}{n} \quad r^2 = 1 - \frac{\sum (T_{obs} - T_{calc})^2}{\sum (T_{obs} - T_{MEAN})^2}$$

r et s représentent respectivement le coefficient de corrélation et l'écart type de l'erreur.

minimiser ainsi la différence entre la sortie désirée et la sortie effective. Ce procédé impose de calculer la dérivée d'une quantité nommée coût J par rapport aux coefficients de connexion, c'est à dire comment l'erreur varie quand varie chaque coefficient de connexion. La méthode la plus utilisée pour déterminer ces dérivées est la méthode de rétropropagation.

Une fois que les activités de tous les neurones de sortie ont été déterminées par l'ensemble d'apprentissage, le réseau calcule le coût J défini par l'expression suivante (eq.5) :

Cette méthode représente des inconvénients liés par exemple à la possibilité de la faiblesse de la pente de la fonction de coût, donc le minimum peut être dû à ce facteur, plutôt au fait qu'on est au voisinage du vrai minimum de la fonction de coût. Pour remédier ce problème, d'autres techniques sont introduites en se basant sur la méthode des dérivés de second ordre, elle dérive de la méthode de Newton en développant la fonction de coût au voisinage d'un minimum w^* , parmi ces techniques on cite l'algorithme de Levenberg-Marquardt.¹⁷

D'autres techniques sont aussi utilisées, parmi lesquelles celle qui consiste à ajouter au terme de gradient $-\mu \nabla J$ un terme proportionnel à la variation des coefficients à l'itération précédente $\lambda[w(i-1) - w(i-2)]$, le coefficient λ a pour rôle de diriger la variation des valeurs des poids dans la même direction que celle utilisée dans l'itération précédente, et par la suite augmenter la vitesse de convergence et éviter les oscillations de la fonction de coût. Tous ces algorithmes sont implémentés dans le toolbox de Matlab, Neural Network version 6.5.¹⁸

Seuls les variables sélectionnées par la technique de la régression par sélection pas-à-pas sont retenus. La corrélation entre les valeurs calculées (prédites) et observées (expérimentales) est examinée au moyen d'une relation linéaire entre elles :

3. Résultat

L'architecture des réseaux de neurones consiste en une série de trois couches de neurones. La première couche représente les variables retenues précédemment dans le modèle linéaires au moyen de la technique de régression pas-à-pas, la dernière couche représente la température d'ébullition des alcanes. Les paramètres du réseau (les valeurs des poids) sont initialisés à l'aide d'une fonction qui fournit des variables aléatoires. L'apprentissage des données, c'est-à-dire la minimisation de la fonction de coût est réalisée à l'aide de l'algorithme de Levenberg-Marquardt. La couche intermédiaire est composée de sept neurones.

En entraînant le réseau sur l'ensemble des données, nous avons remarqué qu'il n'y a pas de

limite à ses capacités de prédiction, surtout lorsque le nombre de neurones dans la couche cachée augmente, mais nous avons constaté que pour un nombre de neurones dans la couche cachée supérieur à 7, le gain n'est pas significatif. Les paramètres statistiques qui résultent de la relation linéaire entre la température observée et calculée sont performants. Par ailleurs, l'algorithme de Levenberg-Marquardt nécessite à peine une centaine d'itérations pour une convergence totale de la fonction de coût, par rapport à la méthode classique (simple gradient) (plus d'un millier d'itérations). Tous les résultats sont présentés dans le tableau 4, nous avons aussi classé les erreurs dans plusieurs domaines de prédiction.

Tableau 4

Résultat détaillé de l'analyse statistique en utilisant deux algorithmes des réseaux de neurones (gradient simple, et l'algorithme de Levenberg-Marquardt)

Analyse	L'algorithme de Levenberg-Marquardt (100 itérations)	Gradient simple (2000 itérations)
r^2	99	99
s (°C)	0,78	3,5
Histogramme		
$T < 0,1$	32	5
$0,1 \leq T < 0,5$	54	14
$0,5 \leq T < 1$	35	21
$1 \leq T < 2$	18	39
$2 \leq T < 3,5$	4 avec $T < 3,5$	64
$5 \leq T < 10$	0	13
$10 \leq T < 20$	0	4

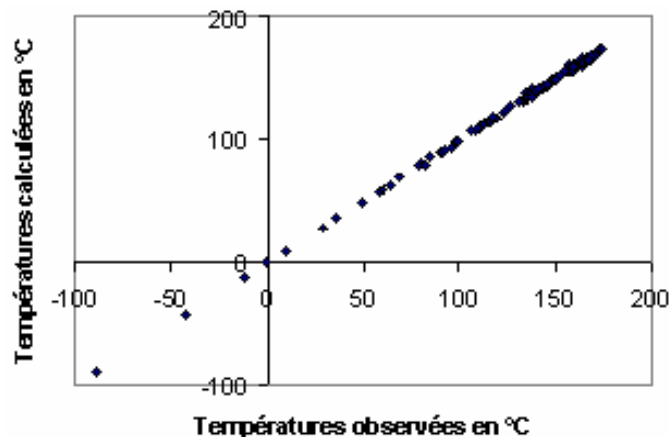


Fig. 2 – Températures observées et calculées pour l'ensemble des données en utilisant l'algorithme de Levenberg-Marquardt.

Notons que ces paramètres résultent d'un apprentissage sur la totalité de la base de données. Malheureusement, nous ne disposons d'aucun moyen de vérifier la nature de la fonction à approximer par les réseaux de neurones. Ce résultat peut résulter d'un apprentissage aléatoire sans aucune régularisation, d'autant plus que les valeurs des températures sont parfois très distantes entre deux molécules consécutives. Pour pallier à cet inconvénient, nous avons réalisé une partition de l'ensemble des données en deux ensembles. Le premier sert à l'apprentissage du modèle, et le deuxième pour en fixer les paramètres. Nous effectuons un arrêt de l'apprentissage lorsque l'erreur sur l'ensemble de la validation (le second ensemble) augmente. Les paramètres statistiques régressent, mais néanmoins restent très satisfaisants ($r=0,99$ et $s=1,9^{\circ}\text{C}$). Cette méthode peut conduire à plusieurs modèles, et ce autant que le nombre de partitions possibles de la base de données initiale.

4. Généralisation par Cross –Validation ou Validation Croisée

Notre objectif est de choisir un modèle capable d'apprendre les données qu'on lui propose, mais aussi qui offre une bonne généralisation à des fins de prédiction, en d'autre terme éviter le surajustement lors d'un test. Pour ce faire, une des techniques utilisées consiste à décomposer l'ensemble de données en trois sous ensemble, le

Nombre de neurones dans la couche cachée
Nombre d'itération:

: 7 neurones
: 100 moyenne

La relation entre la température observée et calculée conduit à une parfaite corrélation, avec des paramètres statistiques très satisfaisants ($r=0,99$ $s=2^{\circ}\text{C}$). Rappelons que ce test porte sur la totalité des données. D'autre part, la nature de la décomposition de l'ensemble des données permet d'évaluer la performance du modèle. Lorsque les ensembles du test sont composés à partir de données prélevées un par un à intervalles égales (saut de quatre molécules), les paramètres

premier sert à l'apprentissage du modèle, le deuxième pour ajuster le modèle et éviter le surajustement en effectuant un arrêt prématuré quand son erreur augmente lors du calibrage. Le troisième sous ensemble sert à tester la performance du modèle. Cette technique est lourde à réaliser et pose la problématique du choix du deuxième sous ensemble pour le calibrage du modèle. Dans notre cas, nous utilisons la méthode basée sur la technique de régularisation bayésienne par modération des poids, elle consiste à pénaliser les valeurs élevées des poids en modifiant la fonction de coût. Cette technique force les paramètres (les poids) à ne pas prendre des valeurs élevées, et par conséquent à éviter le surajustement.

Les résultats de la décomposition effectuée auparavant restent arbitraires puisqu'ils sont limités à un ensemble de test fixe. Donc, nous avons testé le pouvoir prédictif de notre modèle connexioniste par la technique de la validation croisée (leave-20%-out cross-validation method). Dans cette procédure, nous retirons 20% de l'échantillon de base qui servira de test et le reste à l'apprentissage. Cette opération est répétée cinq fois jusqu'au test de la totalité des données disponibles. Pour toutes les décompositions, le choix des valeurs de poids initiales n'a presque pas d'effet, et on observe une convergence maximale de la fonction de coût après une centaine d'itérations en moyenne.

statistiques sont nettement améliorés ($r= 0,99$, $s=1,7^{\circ}\text{C}$) (Tableau 5), ceci est dû essentiellement au fait que la température de l'ensemble d'apprentissage représente une parfaite empreinte des données dans l'ensemble du test. Tous les types d'interaction pouvant exister entre les atomes (primaire-secondaire, primaire-tertière, primaire-quaternaire.....) se retrouvent dans les deux ensembles.

Tableau 5

Analyse statistique de la Cross-validation 20% par la technique de la régularisation bayésienne

Paramètres statistiques		Différents domaine de prédiction					
r	s	T<0,1	0,1≤T<0,5	0,5≤T<1	1≤T<2	2≤T<5	5≤T<6
99	1,7°C	10 cas	42	38	29	22	2

CONCLUSION

Nous avons montré dans ce travail qu'il est possible de relier la température à la topologie de la molécule par la méthode d'autocorrélation. Les composantes ne reflètent qu'une variation topologique de l'effet de Van Der Waals régissant la température d'ébullition. Nous ne disposons d'aucun outil mathématique capable d'épurer l'effet d'un composant de la contribution des autres composantes. Néanmoins la régression linéaire avec la technique de sélection pas-à-pas a conduit à un vecteur avec deux types de composantes. Ils sont simples d'une part mais surtout capables de tenir compte de toutes les interactions entre les molécules, ajouter à cela, leur forte sélectivité. Par ailleurs les réseaux de neurones avec la régularisation bayésienne fournissent un modèle robuste capable de prédire la totalité des molécules avec une grande performance, un coefficient de corrélation élevé avec une erreur très faible pour la validation croisée et le test de l'ensemble des données. Il est essentiellement indépendant de l'architecture des réseaux de neurones. Il n'est nul besoin d'effectuer une partition de l'ensemble des données pour la validation de l'apprentissage et par la suite la fixation des paramètres des réseaux de neurones.

REFERENCES

1. M.H. Hassoum, (Eds.), "Fundamentals of artificial neural networks", MIT Press, Cambridge, 1995.
2. D. J. C. Mackay, *Comput. Neural Syst.*, **1995**, 6, 469-505.
3. D. J. C. Mackay, *Neural Comput.*, **1992**, 4, 415-447.
4. D. J. C. Mackay, *Comput. Neural Syst.*, **1995**, 6, 469-505.
5. G. Moreau, et P. Broto, *Nouv. J. Chim.*, **1980**, 4, 757-764.
6. R. C. Weast et S. M. Shelby, (Eds), "Handbook of Chemistry and Physics", Chemical Rubber Company, Cleveland, 1990, 70th edn.
7. M. Chastrette, F. Tiyal et J. F. Peyraud, *C.R. Acad. Sci. Paris, Ser. II*, **1990**, 310, 514-515.
8. D. Zakarya, M. Nohair et H. Nyassi, *C. R. Acad. Sci. Paris, Ser.IIc*, **2000**, Vol.3, 839-842.
9. M. Nohair et D. Zakarya, *J. Mol. Model.*, **2003**, Vol.9, 365-371.
10. J. Devillers, D. Domine et C. Guillon, *Eur. J. Med. Chem.*, **1998**, 33, 659-664.
11. R. Todeschini et V. Consonni (Eds), "The handbook of molecular descriptors", Wiley-VCH, New York, 2000.
12. D. Zakarya, et M. Nohair, *Rev. Roum. Chim.*, **1999**, 47, 713-722
13. J. E. Dubois, C. Mercier et A. Panaye, *Acta. Pharm. Jugosl.*, **1986**, 36, 135-169.
14. D. Weininger, *J. Chem. Inf. Comput. Sci.*, **1988**, 28, 31-36.
15. G. J. Dreyfus, M. Martinez, M. Samuelides, F. Gordon, S. Badran, S.Thiria et L. Héroult (Eds), "Réseaux de neurones - Méthodologie et applications", Eyrolles, 2002.
16. S. Rudolph, *Neurocomputing*, **1997**, 16, 1-22.
17. M. T. Hagen et M. Menhaj, *IEEE Trans. Neural Networks*, **1994**, 989-993.
18. MATLAB 7.0, *The MathWorks, Inc.*, Natick, MA. **2002**.