

Des neurones dans le champagne

Vincent Steinmetz, Bruno Cauquil et Félix Bocquet

La Champagne est une région viticole particulière dans le vignoble français, caractérisée par sa situation septentrionale et sa production de vins effervescents. Des 250 millions de bouteilles produites environ en 1994, un tiers est exporté. Dans l'élaboration des vins mousseux, et du champagne en particulier, l'équilibre acide joue un rôle primordial pour la qualité du produit final. Il intervient sur la stabilité du vin, sur ses qualités gustatives et ses aptitudes au vieillissement. Le pH des vins est aussi un facteur clé de la réussite de la prise de mousse (fermentation en bouteille), et doit être impérativement pris en considération.

A partir d'un moût de raisin, le niveau d'acidité du futur vin peut être modulé par le choix du type de vinification, notamment en réalisant ou non une fermentation malolactique (FML), comme décrit en figure 1, ci-dessous.

Actuellement, la plupart des vinificateurs réalisent la fermentation malolactique qui est parfaitement maîtrisée depuis quelques années (Valade et Laurent, 1993). Au cours de cette fermentation, les bactéries lactiques dégradent l'acide malique en acide lactique, réduisent l'acidité du vin et lui confèrent une plus grande stabilité micro-biologique.

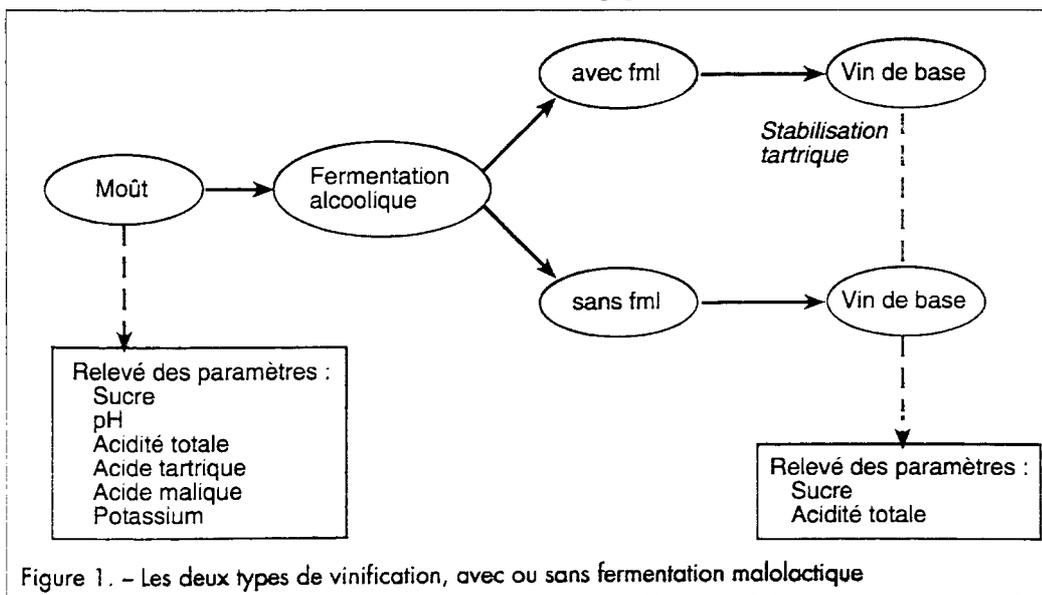


Figure 1. - Les deux types de vinification, avec ou sans fermentation malolactique

**V. Steinmetz
et B. Cauquil**

Cemagref
Génie des équipements
agricoles et alimentaires
BP 5095
34033 Montpellier Cx

F. Bocquet

Comité interprofessionnel
du vin de Champagne
5, rue Henri-Martin
51200 Epernay

Ce schéma classique de vinification ne s'adapte pas forcément à tous les millésimes. Lorsque cette désacidification biologique est trop importante, le vinificateur a alors recours à des techniques correctives (ajout d'acide tartrique, essentiellement). Une alternative consiste à ajuster le niveau d'acidité par l'assemblage (traditionnel en Champagne) de vins ayant subi ou non la fermentation malolactique. Cette solution exige de prédire dès la réception de la vendange l'évolution de l'acidité, afin de choisir très tôt la stratégie de vinification à appliquer à chaque cuve. La prédiction de l'acidité potentielle du vin apparaît donc comme un point clé de la maîtrise de la qualité du vin de Champagne. C'est cette prédiction qui fait l'objet d'une étude dont nous rapportons ici les principaux résultats.

Contrôler l'acidité pour la qualité des vins de Champagne

Dès 1987, le Comité interprofessionnel du vin de Champagne (CIVC) a mis en place un échantillonnage annuel de moûts représentatifs des principaux secteurs du vignoble champenois et de leur encépagement. Cet échantillonnage comprend l'analyse physico-chimique de chaque moût, et celle des vins correspondants ayant suivi ou non la fermentation malolactique. Ces mesures sont réalisées au laboratoire à partir de mini-vinifications et, à ce jour, un total de 510 vins ont été ainsi analysés (tableau 1).

Six paramètres analytiques ont été retenus pour caractériser l'acidité des moûts et prédire le pH et l'acidité totale des vins. Le pH et l'acidité totale des moûts sont l'expression chimique du potentiel acide des vins. L'acidité des moûts est due principalement à deux acides organiques (acide malique et acide tartrique), et au potassium qui est le principal cation du vin. Les autres acides et les sels

minéraux, très peu abondants, peuvent être négligés. Enfin, le degré alcoolique potentiel constitue une indication du niveau de maturité de la vendange et caractérise en quelque sorte le millésime. On connaît donc en partie les paramètres intervenant lors de la vinification, mais leurs interactions restent complexes et partiellement connues. La prédiction du pH et de l'acidité totale sur les vins stabilisés relève encore d'une approche empirique.

Vers une modélisation de l'évolution du potentiel acide

Le CIVC a développé dans un premier temps des modèles statistiques permettant de prédire ces paramètres. Ces modèles ne fournissent pas une précision suffisante, précision imposée par les appareils de mesure analytique ($\pm 0,04$ unités pour le pH, et de $\pm 0,2$ g H_2SO_4/l pour l'acidité totale). Par exemple, un modèle de régression quadratique permet de prédire, dans le meilleur des cas, le pH à $\pm 0,1$ unités près, et l'acidité totale à ± 1 g H_2SO_4/l . Cette limitation dans la prédiction peut être due aux phénomènes non linéaires présents dans les réactions intervenant au cours de la vinification (fermentation, stabilisation tartrique...), que les modèles statistiques utilisés, essentiellement linéaires, ne peuvent pas prendre en compte.

Afin d'améliorer cette précision, le CIVC s'est associé avec la division Génie des équipements agricoles et alimentaires du Cemagref de Montpellier afin de développer des modèles basés sur les réseaux de neurones. L'application des techniques neuronales (voir encadré) aux processus biologiques, qu'il s'agisse de classification (Steinmetz et Delwiche, 1993) ou de prédiction (Bochereau et Sablayrolles, 1993), permet en effet de tenir compte de phénomènes non linéaires tels qu'ils peuvent se présenter dans l'élaboration du vin.

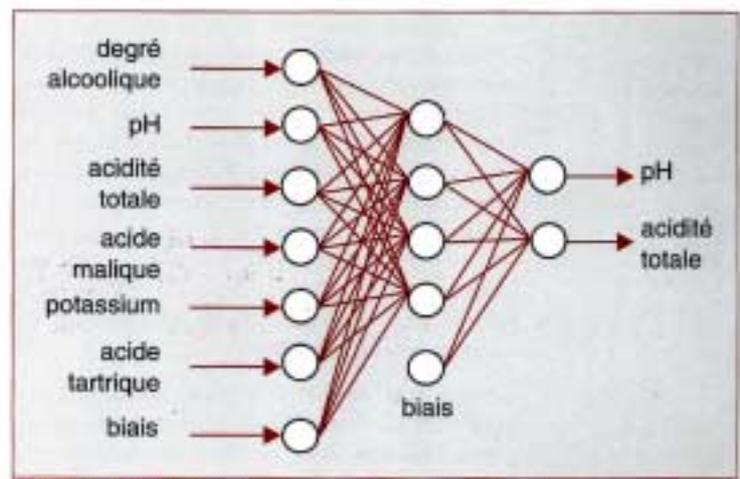
L'étude a débuté dans un premier temps dans le cadre du module « Intelligence artificielle » des enseignements de première année de l'Engref, pour évaluer l'intérêt des réseaux de neurones pour ce type d'application. Suite à cette préétude, une collaboration associée maintenant le Cemagref et le CIVC. *In fine*, l'objectif de ces modèles est de fournir aux œnologues un outil d'aide à la décision qui permette, dès la réception des moûts à la vendange, de choisir la filière de vinification (avec ou sans FML) à partir des caractéristiques initiales du moût et du niveau d'acidité final souhaité sur le vin.

Tableau 1. – Composition du jeu de données par filière de vinification, cépage et année de récolte ▼

Vinification ¹	Cépage	Années								Total
		1987	1988	1989	1990	1991	1992	1993	1994	
Avec FML	Chardonnay	4	7	13	5	17	10	13	15	84
	Pinot Noir	4	5	37	15	21	9	17	15	123
	Meunier	4	5	12	8	10	13	16	15	83
	Total	12	17	62	28	48	32	46	45	290
Sans FML	Chardonnay	1	1	7	3	16	10	13	15	66
	Pinot Noir	1	1	22	11	19	9	17	15	55
	Meunier	1	1	1	4	8	13	16	15	59
	Total	3	3	30	18	43	32	46	45	220
Total		15	20	92	56	91	64	92	90	510

Modélisation de l'acidité avec les réseaux de neurones

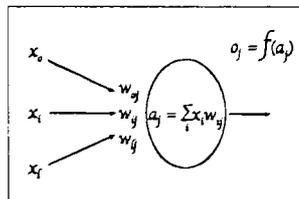
Un réseau de neurones formé de trois couches successives (figure 2) a été utilisé pour prédire le pH et l'acidité totale du vin. La couche d'entrée est formée par les paramètres physico-chimiques du moût. La couche cachée est formée d'un groupe de neurones connectés à la couche d'entrée et à la couche de sortie. La couche de sortie fournit les paramètres physico-chimiques du vin. Le nombre de neurones de la couche d'entrée est fixé par le nombre de paramètres analytiques mesurés au stade moût et, de manière similaire, le nombre de neurones de la couche de sortie est fixé par le nombre de paramètres à prédire pour le vin. Le nombre de neurones dans la couche cachée est établi empiriquement en fonction de la complexité



▲ Figure 2. - Structure du réseau multicouches pour la prédiction de l'acidité

Les réseaux de neurones

Sous le terme de réseaux de neurones, on regroupe aujourd'hui des modèles dont l'intention est d'imiter certaines fonctions du cerveau humain en reproduisant quelques-unes de ses structures de base. Historiquement, c'est en 1943 que McCulloch et Pitts étudièrent un ensemble de neurones formels interconnectés, et montrèrent leurs capacités à calculer certaines fonctions logiques. S'inspirant de leurs travaux sur les neurones biologiques, ils ont proposé le neurone formel décrit dans la figure 5. Bien qu'ayant connu un fort intérêt dans les années 50, par l'intermédiaire de Hebb et Rosenblatt notamment, il faudra attendre le début des années 80 (Hopfield, Rumelhart et McClelland, Kohonen) pour voir la communauté scientifique se pencher à nouveau sur les réseaux de neurones. Les raisons de cette renaissance sont dues principalement au désenchantement vis-à-vis des performances de l'approche symbolique, qui avait monopolisé les recherches dans les années 60 et 70. Parmi les différents types de réseaux de neurones développés (Hopfield, pour les réseaux complètement connectés et à entrées binaires, Kohonen pour les réseaux topologiques auto-adaptatifs, Rumelhart et McClelland pour les réseaux à couches), les réseaux à couches sont ceux, pour des raisons encore mal identifiées, qui connaissent le plus grand succès.



choix initial aléatoire des connections synaptiques (w_i), on définit une fonction d'erreur $J_o(x, y, w)$. L'objectif consiste alors à trouver les valeurs w_i permettant de minimiser J_o pour les couples (x, y) utilisés dans la base d'apprentissage. Une des méthodes utilisées est l'algorithme de rétropropagation du gradient, basée essentiellement sur les règles de dérivation composée (Abdi, 1994).

Le processus est donc itératif, avec une présentation multiple des couples entrée-sortie qui vont contribuer à la modification

des connections synaptiques, jusqu'à atteindre un minimum de la fonction d'erreur. Ceci fait, on considère (ou non) que l'apprentissage est terminé, et que le modèle du réseau peut être utilisé. L'apprentissage plus ou moins rapide du réseau peut être réglé par des paramètres relevant de l'expertise.

Un des points clés du succès des réseaux de neurones est qu'il n'est pas explicitement nécessaire de fournir l'ensemble des règles de fonctionnement d'un système, mais qu'il est possible de faire apprendre ces règles à partir d'exemples. Cette approche est particulièrement intéressante pour les systèmes complexes, à nombreuses composantes, dont on connaît rarement l'ensemble des interactions. L'inconvénient des réseaux réside dans leur incapacité à fournir une explication des phénomènes appris. Toutefois, l'intérêt des réseaux dans la modélisation de phénomènes non linéaires, leur comportement relativement stable en présence de données inattendues ou fortement bruitées, et l'association avec des outils statistiques classiques en font des outils de modélisation séduisants, pour lesquels d'immenses projets de recherches sont en cours (400 millions de dollars pour un programme de 6 ans par le Pentagone américain, et 250 millions de dollars pour un programme de 10 ans au Japon, *The Economist*, 1995).

Ces derniers sont constitués de neurones organisés en couches. Les neurones des couches d'entrée et de sortie correspondent aux paramètres d'entrée et de sortie. Les paramètres de sortie peuvent correspondre à la prédiction d'un jeu de valeurs bien précises, ou à une classification. Les neurones sont reliés entre eux par des connections synaptiques, symbolisés par des valeurs numériques qui sont déterminées lors d'une phase d'apprentissage. A partir d'un ensemble connu de couples entrée-sortie (x, y) , et après un

du problème. Une fois la structure du réseau définie, il reste à trouver les valeurs des connexions entre les neurones qui permettent de réaliser l'apprentissage du réseau, c'est-à-dire de faire correspondre aux entrées présentées au réseau les sorties désirées. Les valeurs des connexions reliant les neurones sont modifiées en fonction des couples (entrées-sorties) présentés au réseau afin d'obtenir les valeurs désirées de pH et d'acidité totale. Le lecteur se reportera à l'encadré et à la bibliographie pour le détail du mécanisme de modification des connexions entre neurones.

Les deux types de filière d'élaboration des vins (avec et sans FML) correspondent à des processus biologiques différents. Deux modèles basés sur les réseaux de neurones sont donc développés pour prédire l'acidité des vins correspondant à chacune de ces filières. Toutefois, la méthodologie est la même dans les deux cas, seules varient les données utilisées (avec FML ou sans FML, comme indiqué dans le tableau 1) pour l'apprentissage de chaque réseau.

L'ensemble des valeurs mesurées de 1987 à 1993 a servi à l'apprentissage et à la détermination des valeurs de connexion du réseau, qui a ensuite été validé sur l'année 94. L'évaluation de la performance du réseau est réalisée en comparant les valeurs prédites avec les valeurs de pH et d'acidité effectivement mesurées. On exprime l'erreur commise par le réseau par l'erreur standard de prédiction (SEP), comme indiqué dans l'équation (1) où $Y_{i\text{prédit}}$ représente la sortie fournie par le réseau pour l'échantillon i et $Y_{i\text{mesure}}$ les valeurs fournies par les appareils de mesures, n étant le nombre d'échantillons mesurés et p le nombre de variables explicatives utilisées dans le modèle (dans le cas présent, $p = 6$, et n est égal en moyenne à 40).

$$\text{Eq. (1) } SEP = \sqrt{\frac{\sum_{i=1}^n (Y_{i\text{mesure}} - Y_{i\text{prédit}})^2}{n - p - 1}}$$

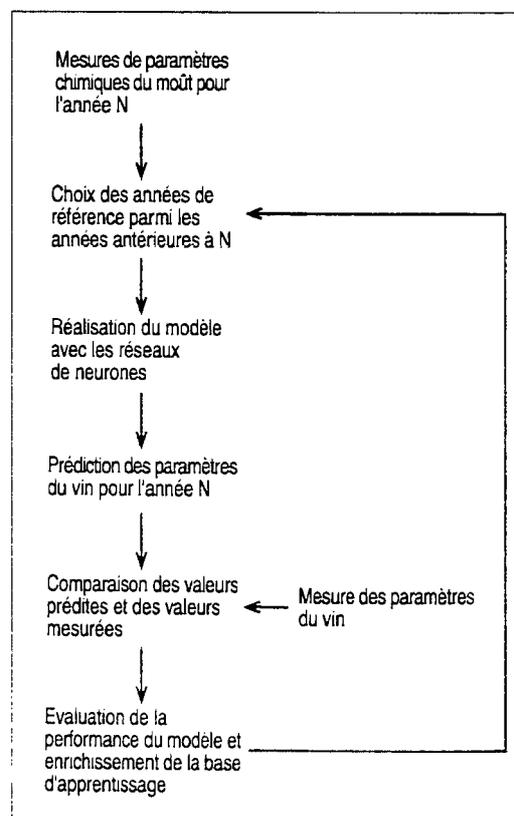
Pour l'année 95, les mesures sur le moût seront disponibles vers octobre, et les mesures sur vin vers novembre. L'évaluation se fera donc *a posteriori*. Le modèle réalisé va donc évoluer et sera enrichi chaque année selon le processus décrit sur la figure 3.

Figure 3. – Utilisation et évolution du modèle ►

Année de test	SEP du pH	SEP de l'acidité totale
1988	0.12	0.26
1989	0.06	0.26
1990	0.07	0.34
1991	0.05	0.33
1992	0.06	0.29
1993	0.04	0.22
1994	0.06	0.45

▲ Tableau 2. – Erreurs standard de prédiction (SEP) avec toutes les années sauf l'année de test comme base d'apprentissage.

Le tableau 2 montre les résultats obtenus dans le cas de la filière avec FML, quand on teste la performance du réseau sur l'année non incluse dans la base d'apprentissage. Pour les besoins de l'étude, on ne tient pas compte de l'antériorité des années. Ces résultats obtenus avec l'utilisation des réseaux de neurones montrent une amélioration par rapport au modèle quadratique, tout particulièrement pour la prédiction de l'acidité totale. Par exemple, pour l'année 89, on obtient



une précision de 0,60 g H₂SO₄/l avec un modèle basé sur une régression quadratique, et de 0,26 g H₂SO₄/l avec un modèle basé sur les réseaux de neurones.

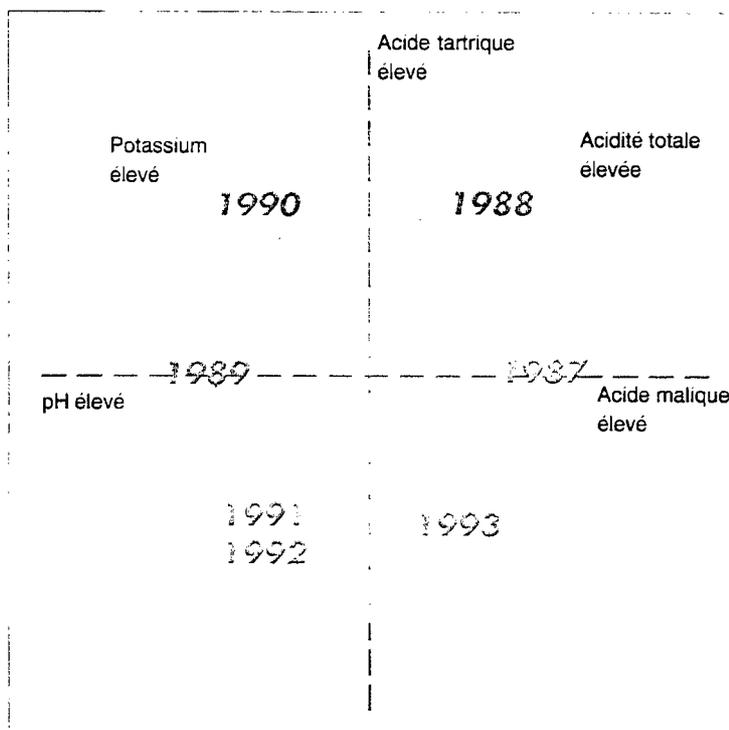
Un des points importants de la méthode consiste à identifier clairement les variables qui sont utilisées en entrées du réseau, et à effectuer ou non un prétraitement statistique avant l'entrée du réseau. Par exemple, on peut chercher à reproduire le comportement de l'œnologue qui, en présence d'une nouvelle année, s'inspire de l'ensemble des années passées. En effet, une étude statistique préalable, l'analyse en composante principale (ACP), a montré qu'il existe un « effet » millésime identifiable à partir des mesures physico-chimiques des moûts. L'ACP est une méthode linéaire de réduction de paramètres, qui permet de représenter géométriquement les individus et leur caractère. Dans le cas présent, la figure 4 représente les relations entre les moûts des années et leur composition physico-chimique. Par exemple, l'année 89 correspond à une année à pH élevé (sur la partie négative de l'axe des abscisses), alors que l'année 87 correspond à une année à acide malique élevé (sur la partie positive de l'axe des ordonnées), et 91, 92 et 93, par un niveau d'acide tartrique faible.

A partir de ces constats, on cherche à déterminer la base d'apprentissage la plus appropriée, pour décider quelles sont les années les plus « proches », chimiquement parlant, de l'année 94. Cela consiste à rechercher des caractéristiques chimiques similaires entre les moûts. L'utilisation des réseaux de neurones montre que certaines années sont mieux adaptées que d'autres pour réaliser l'apprentissage, mais n'apporte pas une amélioration significative de la prédiction.

Le modèle réalisé à partir des réseaux de neurones a été testé sur des données industrielles issues d'une maison de champagne. Il est tout d'abord important de remarquer que les données sont moins complètes et précises que celles réalisées au laboratoire. En effet, on ne dispose pas, par exemple, de l'acide tartrique, de l'acide malique et du potassium. Ces mesures sont alors remplacées par les valeurs moyennes de l'année correspondante relevées en laboratoire. Aussi, malgré les données incomplètes, le modèle basé sur les réseaux de

Cuvées	Année	pH			Acidité totale en gH ₂ SO ₄ /l		
		Industriel	Prédit	Erreur	Industrielle	Prédite	Erreur
B58	88	3.12	3.11	- 0.01	5.6	5.4	- 0.2
B18	88	3.04	3.03	- 0.01	5.8	5.5	- 0.3
B60	90	2.91	2.99	+ 0.08	6.1	6.7	+ 0.6
B10	90	2.88	2.97	+ 0.09	6.5	6.6	+ 0.1
B11	91	2.92	2.96	+ 0.04	6.4	6.7	+ 0.3
B82	92	2.96	2.96	0.00	6.5	6.4	- 0.1
B12	92	2.98	2.95	- 0.03	6.6	6.5	- 0.1
B13	93	2.91	2.96	+ 0.05	5.9	6.2	+ 0.3
B63	93	2.92	2.94	+ 0.02	5.8	6.4	+ 0.6

▲ Tableau 3. – Prédiction de l'acidité sur les données industrielles



▲ Figure 4. – Regroupement des moûts par ACP

neurones présente l'avantage d'une certaine robustesse (tableau 3). On note la qualité de la prédiction, en particulier pour les valeurs de pH, avec toutefois des erreurs importantes pour les cuvées de l'année 90. Ceci peut s'expliquer par l'arypie de l'année 90, pour laquelle l'acide malique est extrêmement faible, et l'acide tartrique relativement élevé. Ce constat débouche sur la nécessité d'associer une mesure de confiance à la prédiction fournie par le réseau. Ceci fera l'objet de recherches ultérieures.

Conclusion

L'utilisation des réseaux de neurones montre une amélioration conséquente de la prédiction de l'acidité totale, avec une erreur divisée par deux par rapport aux modèles linéaires ou quadratiques. Le modèle réalisé à partir de données issues de laboratoire présente d'autre part un comportement robuste et satisfaisant lors de son application à des données industrielles. Toutefois, aucune amélioration n'est apportée dans la connaissance des processus de vinification. Ceci constitue l'une des limitations de ces modèles de type « boîte noire ».

Les travaux en cours permettent d'envisager une utilisation directe de l'outil par les œnologues. En effet, un programme informatique pourra fournir une prédiction de l'acidité pour un moût donné. Le vinificateur disposera d'un outil d'optimisation et d'une aide à la décision pour la réalisation des assemblages de vins en fonction du niveau d'acidité recherché. Cet outil pourra lui permettre d'orienter précisément la stratégie de vinification à adapter lors de l'élaboration des cuvées de champagne.

Résumé

L'acidité est un facteur important de la qualité des vins de Champagne. La prédiction du niveau d'acidité est indispensable au choix de la filière d'élaboration des vins dès la réception des moûts. Les tests effectués montrent que les modèles basés sur les réseaux de neurones permettent d'obtenir une meilleure précision que des modèles linéaires ou quadratiques, même si ce degré de précision est encore inférieur à celui des méthodes analytiques. Le modèle réalisé à partir de données de laboratoire issues de mini-vinifications a été testé avec succès sur des données industrielles, démontrant en cela sa robustesse. Cet outil d'aide à la décision pourrait être utilisé à l'avenir par les œnologues pour l'optimisation des assemblages des vins, contribuant à l'amélioration de la qualité des vins de Champagne.

Abstract

Acidity is an important factor in Champagne wine quality. Predicting acidity levels is therefore an indispensable requirement when determining which wine production process to use once the must has been supplied. Various tests which have been undertaken show that those models based on neural network programming are more accurate than either linear or quadratic function models, although they still prove to be less accurate than the more conventional analytical methods. The model generated using laboratory data from small-scale wine-making processes, has been successfully tested on data from industrial-scale processes, indicating the model's robustness. This decision-making tool could in future be used by wine-makers to optimise blending and thus improve the quality of the Champagne region's wines.

Bibliographie

- DAVALO, E., NAIM, P., 1980. *Des réseaux de neurones*, Eyrolles, 232 pages.
- STEINMETZ, V., DELWICHE, M.J.D., 1993. Neural network analysis of rose straightness, *Proceedings of the International Conference on Artificial Intelligence for Food and Agriculture*, Nîmes, France, p. 253-263.
- BOCHEREAU, L., SABLAYROLLES, J.M., 1993. Modelling alcoholic fermentation with neural networks, *Proceedings of the International Conference on Artificial Intelligence for Food and Agriculture*, Nîmes, France, p. 285-293.
- VALADE, M., LAURENT, M., 1993. La fermentation valolactique : réactivation des bactéries sur moût, *Le Vigneron champenois*, numéro 6, p. 5-17.
- ABDI, H., 1994. A neural network primer, *Biological systems. Special issue on neural networks*, volume 2, number 3, p. 247-281.
- The Economist*, April 15th, 1995, More in a cockroach's brain than your computers dream of.