

Vins blancs de Colombar et rosés de Négrette : création et validation de modèles de prévisions de leur potentiel aromatique par systèmes IRTF

Eric SERRANO¹, Thierry DUFOURCQ¹, Carole FEILHES¹, Sylvie ROUSSEL²

¹ Institut Français de la Vigne et du Vin – Pôle Sud-Ouest – Vinnopôle, BP22, 81310 Lisle Sur Tarn - France

² Ondalys, 385 Avenue des Baronnes 34 730 PRADES LE LEZ - France

Email: eric.serrano@vignevin.com

Résumé : Un programme de recherche mis en place par l'IFV Sud-ouest et ses partenaires (Syndicat des Côtes de Gascogne et Conseil Général de la Haute-Garonne) entre 2006 et 2008 s'est attaché à adapter et à optimiser les équipements spectroscopiques moyen infrarouge (IRTF) afin de fournir un outil permettant de définir le potentiel aromatique d'un moût à son arrivée au quai de réception. L'étude a porté sur le cépage Négrette dans un objectif d'élaboration de vin rosé, et sur le cépage blanc du Gers le Colombar. A l'issue de ces trois années d'expérimentation, deux modèles ont été élaborés par l'IFV et la société Ondalys permettant de discriminer le potentiel aromatique des raisins à la récolte selon 2 classes (aromatiques / non aromatiques). Les erreurs observées de prévision s'établissent aux environs de 25% dans les deux cas. Ces modèles ont ensuite été validés en 2010 sur un millésime n'ayant pas été utilisé pour leur construction.

Mots-Clés : IRTF, raisins, qualité aromatique, thiols, vins blancs, rosés, modélisation

Introduction

La mondialisation du marché des vins impose aux entreprises de la filière une démarche nouvelle d'industrialisation. En complément des marchés traditionnels des appellations d'origine, le développement d'une démarche, réactive, en grand volume est un axe stratégique souhaité par les grosses structures de production. Cette démarche nécessite le développement ou l'adaptation de nouveaux outils technologiques dans le schéma d'élaboration des vins pour limiter les coûts et assurer une qualité identifiée et constante du produit fini. A l'instar de l'industrie agro-alimentaire, elle nécessite des outils fiables d'identification qualitative de la matière première. De nouvelles technologies ont fait leur apparition permettant de réaliser des empreintes spectrales à l'aide du moyen infra-rouge de lots de vendange. Son exploitation en routine nécessite cependant une expertise des paramètres fournis pour répondre efficacement aux attentes précises de l'entreprise. L'enjeu est de pouvoir rapidement constituer des groupes d'apport de « qualités » homogènes, afin d'éviter un mélange entre des qualités extrêmes conduisant à un résultat moyen et une perte économique majeure. Il s'agit de mettre en relation une mesure immédiate d'un lot de vendange à son potentiel aromatique, pour orienter l'itinéraire technique de la vinification vers une production optimale de composés aromatiques. A terme, il s'agit d'améliorer la « productivité aromatique » de sa vendange grâce à une information analytique adaptée et rapide. L'IFV Sud-ouest a développé pour ce programme d'une période de 5 ans une méthodologie adaptée et un savoir-faire en matière de traitement de l'information issu de l'Infrarouge. L'étude s'appuie sur un partenariat étroit avec des laboratoires de recherche et d'analyses dans les domaines du potentiel aromatique des vins (INRA Montpellier, SA Nyséos) et de la Chimie (SA Ondalys). A l'issue de la création des modèles, le projet s'est attaché à :

1. Valider les modèles de prévision de la qualité avec des raisins qui n'ont pas servi à leur élaboration et, qui de plus sont issus d'un autre millésime (2010).
2. Transférer les bases de données historiques de l'appareil de mesure spectroscopique utilisé (CETIM type Multyspec) vers un spectromètre de la marque FOSS, et établir un modèle propre FOSS.
3. Analyser l'incidence de la date de prélèvement des raisins (jour des vendanges ou quelques jours avant) pour l'analyse spectrale sur les résultats des modèles.

Matériels et méthodes

Trois années d'expérimentation (2006-2007-2008) ont consisté à établir une base de données comportant les paramètres analytiques des raisins à la récolte, leur spectre infrarouge, les données analytiques et de dégustation des vins finis. A partir de 70 échantillons de raisins prélevés au moment de l'apport de la récolte en cave, et vinifiés en conditions standards expérimentales, une base de données comprenant des mesures spectroscopiques des raisins et les résultats des dégustations issues de collègues d'experts est élaborée. Le principe reposait ensuite sur l'analyse chimiométrique des données spectrales et sensorielles (classement des vins) pour la création des modèles. La modélisation appliquée dans ce projet est essentiellement basée sur une méthode de discrimination supervisée appelée la PLS-AFD (Régression aux Moindres Carrés Partiels suivie d'une Analyse Factorielle Discriminante). Il s'agit d'une méthode de discrimination multivariée capable de trouver les axes les plus discriminants dans des données spectrales, c'est-à-dire des variables nombreuses et corrélées.

Divers prétraitements spectroscopiques et différentes méthodes de discrimination ont été testés, afin d'optimiser les résultats de discrimination des modèles créés sur chacun des cépages. Pour la validation de ces modèles, l'IFV a réalisé sur le millésime 2010, des mesures avec deux spectromètres moyen infrarouge couvrant l'offre des équipements du commerce : CETIM type Multyspec et FOSS type Winescan

Comme pour la création des modèles, l'acquisition des spectres est effectuée sur des raisins, issus de parcelles préalablement sélectionnées, au moment de la récolte. Ce prélèvement est réalisé dans les bennes au quai de réception des caves coopératives. Les spectres ainsi disponibles pour chaque cépage et pour chaque spectromètre constituent la base de données sur laquelle porte l'étude de validation. En parallèle, des lots de raisins (70kg) de chaque parcelle sont prélevés au quai de réception et sont vinifiés en conditions standards expérimentales. Les vins issus de ces lots sont ensuite soumis à une dégustation par un jury expert IFV. Le principe de la dégustation pour les deux cépages repose sur le classement des vins en 2 catégories :

- classe A : aromatique
- classe B : non aromatique

Pour le cépage Colombar, la classe finale retenue correspond à un compromis entre la classe donnée en dégustation et la composition analytique des vins en molécules aromatiques reconnues influentes

sur la palette aromatique des vins de ce cépage : les thiols volatils (3MH, A3MH).

Résultats et discussions

Création des modèles : La performance des modèles de discrimination est évaluée par un taux d'erreur global de classification. Les études menées sur les trois millésimes (2006, 2007 et 2008) ont permis d'identifier des modèles de prédiction de la qualité aromatique des vins blancs de Colombar et rosés de Négrette. Deux classes qualitatives peuvent ainsi être prédites à l'arrivée des raisins au quai de réception. Sur Négrette Rosé, la performance des modèles de discrimination sur les 3 millésimes donne les résultats suivants :

- En étalonnage : 23.8% d'erreur
- En validation croisée (leave-one-out) : 25.4% d'erreur. Cette erreur est optimiste, car elle n'écarte qu'un seul échantillon par modèle.
- En validation croisée (leave-10-out / 10 au hasard) : Entre 27 et 30% selon la randomisation. Ces erreurs sont plus réalistes que la précédente.

Tableau I : Matrice de confusion Négrette Rosé pour la discrimination des 2 classes – modèles 2006+2007+2008

6LVs		Bons	Mauvais	Erreur 2nd ordre
		117	72	
Bons (A)	cal	96	23	19.3%
	cv	94	33	26.0%
Mauvais (B)	cal	21	49	30.0%
	cv	23	39	37.1%
Erreur 1er ordre	cal	17.9%	31.9%	
	cv	19.7%	45.8%	
Erreur totale				23.3%
				29.6%

Sur Colombar, le tableau suivant présente la synthèse des résultats obtenus sur les données issues du spectromètre CETIM lors des millésimes 2006, 2007 et 2008. Ces modèles ont été établis pour discriminer 2 classes de qualité : une ayant un potentiel aromatique élevé regroupant les échantillons « bons » + « très bons » et une ayant un potentiel qualitatif faible, regroupant les échantillons « mauvais » à « très mauvais ». Les performances se situent à 27% d'erreur. Des essais ont aussi été menés en basant les modèles uniquement sur les 2 classes extrêmes (Très bons / Très mauvais), améliorant les résultats jusqu'à 24% d'erreur.

Tableau II : Synthèse des performances des modèles spectroscopiques Colombar

Nom du modèle	Nombre de facteurs (PLS AFD)	Erreur en étalonnage (année)	Erreur en validation croisée (année)
Modèle 2006+2007+2008	7	T : 22% (39/178) B : 23% M : 21%	T : 27% (48/178) B : 27% M : 27%
Modèle 2006+2007+2008 Classes extrêmes	10	T : 14% (16/116) B : 13% M : 15%	T : 24% (28/116) B : 21% M : 28%

Validation des modèles 2006-2007-2008 : La validation des modèles consiste à effectuer une prédiction de classes grâce aux spectres 2010 acquis sur CETIM et de confronter ces résultats au classement obtenu par l'analyse sensorielle des vins. La performance des modèles de discrimination est ensuite évaluée par un taux d'erreur global de classification. Ce taux d'erreur est complété par les erreurs de discrimination par classe (erreur de 1er ordre) et les taux d'impuretés (erreur de 2nd ordre) pour chaque classe, au sein d'une « matrice de confusion ».

Pour réaliser la validation sur un deuxième spectromètre de marque FOSS, la base de données historique 2006-2007-2008 de CETIM a été transférée vers FOSS.

Validation des modèles de prédiction de la qualité des vins de Négrette rosés : En 2010, plus de 30 parcelles ont été suivies (prélèvements de raisins à la récolte, acquisition des spectres, vinifications et dégustations) pour la validation du modèle. La qualité gustative des vins 2010 donne un classement de 60% de bons (A) et 40% de mauvais (B). La classification par le modèle CETIM des vins 2010 aboutit à une erreur de 30% sur les 30 échantillons ayant une note définie (tableau suivant).

Tableau III : Matrice de confusion pour la discrimination des 2 classes – modèles 2006+2007+2008 CETIM

6 LVs		Dégustation Classe A	Dégustation Classe B	Erreur 2nd ordre
		18	12	
Modélisation (A)	test	14	5	26,3%
Modélisation (B)	test	4	7	36,4%
Erreur 1er ordre	test	22,2%	41,7%	
Erreur totale				30,0%

Les classes de qualité établies à partir des spectres FOSS conduisent à des résultats tout aussi corrects que ceux acquis avec les spectres historiques CETIM (30% d'erreur). L'analyse des matrices montre une très bonne sélectivité des échantillons classés A (10% d'erreur), mais une mauvaise sensibilité (50% des bons sont classés en mauvais).

Validation des modèles de prédiction de la qualité des vins blancs de Colombar :

La validation des modèles sur ce cépage a porté sur différents modes d'échantillonnage des raisins pour l'acquisition des spectres. L'effet du prélèvement sur la discrimination des classes de qualité est ainsi mesuré. Trois types de prélèvements sont réalisés pour chacune des 20 parcelles 2010 à l'étude :

- point V : prélèvement à la benne au quai de réception sur la vendange mécanique (type de prélèvement ayant servi pour l'acquisition de la base de données 2006, 2007, 2008 du modèle CETIM)
- point R : 200 baies prélevées à la parcelle juste avant la récolte
- point G : portions de grappes prélevées à la parcelle juste avant la récolte

Sur la base des spectres CETIM 2006-7-8 transférée de FOSS, plusieurs modèles de discrimination avec des prétraitements différents sont comparés. L'analyse a mis en évidence une certaine facilité de transfert des spectres CETIM vers les spectres FOSS. Sur la base des spectres FOSS 2006-2007-2008 transférés de CETIM, un modèle de discrimination a été établi et basé sur les mêmes paramètres de prétraitements que les modèles CETIM. La validation croisée est réalisée sur des blocs de 4 échantillons et le modèle est testé sur les spectres de FOSS 2010 aux stades V, R et G. Les résultats de validation sur la base des modèles CETIM ou FOSS 2006-7-8 ne sont pas satisfaisants, l'erreur moyenne étant de 50%. Les validations 2010 à partir des modèles 2006-7-8 ne donnant pas de résultats satisfaisants, un nouveau modèle a été construit sur les 4 millésimes

Tableau IV : Matrice de confusion pour le modèle FOSS 2006-2007-2008 et FOSS 2010 V et R sur Colombar

5LVs		Bons	Mauvais	Erreur 2nd ordre
		101	111	
Bons (A)	cal	68	39	36.4%
	cv	67	40	37.4%
Mauvais (B)	cal	33	72	31.4%
	cv	34	71	32.4%
Erreur 1er ordre	cal	32.7%	35.1%	
	cv	33.7%	36.0%	
Erreur totale				34.0%
				34.9%

(2006-7-8-10). Deux zones spectrales ont été sélectionnées. L'erreur globale est de 34% en étalonnage et de 35% en validation croisée.

Les erreurs ne sont pas également réparties entre les 4 millésimes. Les millésimes 2006 et 2007 ne sont pas vraiment satisfaisants (>40% d'erreur) alors que 2008 et 2010 sont bien prédits (25% d'erreur pour le millésime 2010 stade V). L'application du modèle sur les spectres du type de prélèvement G, donne une erreur de 37%, soit proche du prélèvement de type R (36%). On constate un accord des estimations de 71% entre ces deux types de prélèvement. Le modèle semble donc robuste à la méthode de prélèvement de l'échantillon.

Conclusion

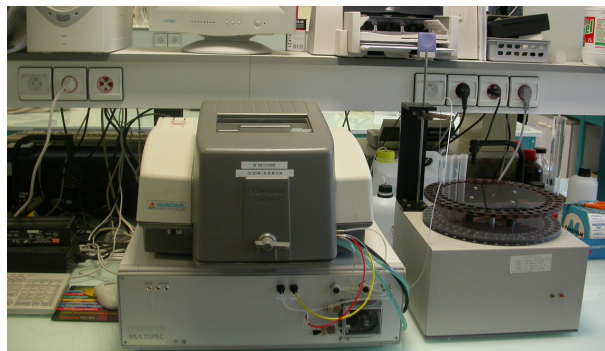
La validation du modèle de discrimination du potentiel aromatique de la Négrette établis sur les millésimes 2006-7-8 sur les spectres CETIM, a nécessité la réalisation d'un modèle moins sélectif, utilisant l'ensemble du spectre. La performance sur l'estimation de la qualité des vins de 2010 est alors de 30% d'erreur. Le transfert des bases de données historiques CETIM à un spectromètre FOSS a permis d'établir des modèles compatibles avec le spectromètre FOSS. En utilisant le même paramètre que pour le modèle CETIM (ensemble du spectre), le modèle aboutit à des classifications correctes pour FOSS 2010 de 30%. Sur Colombar, les résultats n'ont pas été satisfaisants avec le modèle historique. Si en validation croisée, les erreurs obtenues sont encourageantes (27%), la validation 2010 met en défaut le modèle : les résultats de la discrimination ne sont pas satisfaisants (proche d'un classement aléatoire, 50%) si on les compare aux classes réelles réalisées à partir des dégustations et des quantités de thiols mesurées dans les vins finis. Ces performances dégradées ont 3 sources possibles : l'effet du millésime, l'effet du transfert, la non pertinence du modèle. Pour compenser les sources d'erreurs possibles, les spectres 2006-2007-2008 et 2010 (stades R et V) ont été regroupés afin d'établir un modèle sur les 4 millésimes. Le modèle obtenu donne une erreur en validation croisée de 35%. Les performances en 2010 sont de 25% pour le stade V et de 36% pour le stade R en validation croisée. La comparaison des types de prélèvements R (raisins à la parcelle), G (grappes entières à la parcelle) et V (raisins prélevés à la benne après vendange mécanique) montre que les trois types de spectres sont assez différents. La différence entre les deux types de prélèvement R et G reste la plus faible. Cependant, les prédictions effectuées à l'aide du modèle sont satisfaisantes pour les trois types de prélèvements (25%, 36%, 37% d'erreur en validation croisée respectivement pour V, R et G). Parallèlement sur Négrette, un test réalisé sur des échantillons complémentaires permet de valider que les prédictions sont tout aussi correctes lorsque l'échantillon analysé est prélevé environ 5 jours avant la vendange.

Références bibliographiques

Dufourcq T., Davaux F., Serrano E., Roussel S., 2009, Prediction of a white wine aromatic potential by infrared spectroscopic measurements on grapes at harvest. 16th International GIESCO Symposium, Davis (USA) p207-211.

E. Serrano, T. Dufourcq, C. Feilhes, S. Roussel – Outil d'aide à la décision pour caractériser le potentiel olfactif et gustatif des raisins par spectroscopie infrarouge, 2èmes rencontres Qualiméditerranée : les TIC pour une agriculture performante et durable, 2010.

Dubernet M., Dubernet V., Lerch M., Coulomb S. et Traineau I., 2000. «Analyse objective de la qualité des vendanges par spectroscopie infrarouge à transformée de Fourier et réseaux de neurones». Revue Française d'Oenologie 185: pp. 18-21



Ce qu'il faut retenir

Sur la base de banques de données conséquentes (dégustations et spectres infrarouges), des modèles de corrélation multilinéaire ont été établis entre des mesures spectroscopiques dans le moyen infra-rouge (IRTf) sur moût de Colombar et de Négrette et deux classes aromatiques qualitatives par des méthodes de classification ad hoc.

La validation de ces modèles a abouti à des résultats satisfaisants sur Négrette en vue de la production de rosés aromatiques. En revanche, sur Colombar, du fait de nombreux transferts entre spectromètres, les résultats sont plus mitigés. Un nouveau modèle a été calculé en alimentant la banque de données d'une année supplémentaire.

Dans les deux cas, les erreurs sont de 30%.

Les modèles établis pour prédire le potentiel aromatique d'un moût de Colombar et de Négrette doivent à présent être transférés vers les utilisateurs finaux, laboratoires d'analyse et caves coopératives. Les résultats (performance de la validation et adaptation des modèles aux deux spectromètres du secteur viticole FOSS et CETIM) permettront d'envisager une phase de pré-industrialisation.