

Estimation précoce du rendement de la vigne : corrélation entre le volume de la grappe de *Vitis Vinifera* en cours de croissance et son poids à la récolte

Eric SERRANO¹, Laure GONTIER¹, Thierry DUFOURCQ², Sylvie ROUSSEL³

¹Institut Français de la Vigne et du Vin – Pôle Sud-Ouest - V'Innopôle - BP 22 - 81310 LISLE/TARN

²Institut Français de la Vigne et du Vin - Pôle Sud-Ouest - Château de Mons - 32100 CAUSSENS

³ONDALYS - 385 Avenue des Baronnes - 34730 PRADES-LE-LEZ

Email: eric.serrano@vignevin.com

Résumé : Cette étude s'insère dans le cadre d'une recherche de méthodologie simple et rapide d'estimation précoce des rendements d'une parcelle de vigne. La méthode proposée exploite, pour l'estimation du poids moyen de la grappe, l'existence de relations entre son volume à différents stades phénologiques et son poids à maturité. Le volume de la grappe a été calculé grâce à la formule de volume d'un cône. Plus de vingt mille données sont traitées statistiquement; elles portent sur quatre millésimes (2001-2004), trois stades phénologiques (nouaison, fermeture de grappe et début véraison), et 19 cépages du genre *Vitis vinifera*. Les modèles ont été construits par variété, à chaque stade. L'influence de la région, du stade phénologique, de la variété et du millésime a été étudiée. La performance des modèles est fortement influencée par le nombre de parcelles incluses dans le modèle. Les prédictions de poids moyens de grappes sont satisfaisantes à partir du stade fermeture de grappe, avec des erreurs moyennes en cross-validation inférieures à 10%. Les modèles ont également été validés grâce à une série de tests indépendants sur le millésime 2004. Lorsque plusieurs années sont incluses dans le modèle, la stabilité des prédictions augmente, permettant d'obtenir une performance moyenne proche de 9 %. Trois groupes homogènes de cépages ont été identifiés, permettant de bâtir des modèles sur ces groupes de cépages, et non pas séparément.

Mots-Clés : estimation, rendement, vigne, volume de grappe

Introduction

La production viticole a toujours été marquée par l'irrégularité de la quantité de la vendange d'une année à l'autre. Un des enjeux est de pouvoir fournir à l'ensemble des acteurs de la filière impliqués dans l'estimation d'un rendement une méthode fiable et rapide permettant des pronostics précoces.

Différentes méthodes ont tenté de pronostiquer de façon précoce le volume de récolte pour des secteurs géographiques plus ou moins étendus. Les premières tentatives d'estimation des rendements sont notifiées dans la bibliographie en 1955 (Wurgler, Leyvraz et Bolay).

A l'échelle de la parcelle, les études les plus abouties portent essentiellement sur des comptages d'inflorescences, de grappes par souche ou de baies par grappe (Murisier, 1986). Des études menées en Alsace (1957-1971, Huglin et Schneider) indiquent que le poids moyen des grappes est le facteur primordial de variations engendrant les sources d'erreur. Dufourcq et Serrano (2003) ont démontré que le coefficient de variation annuel du poids moyen des grappes sur une même parcelle de Négrette atteint 15%, confirmant les résultats d'Huglin et al. (1975) sur Syrah ou de Gerbier et Remois (1977) en Champagne.

Aussi, si certaines des méthodes proposées permettent d'obtenir des estimations satisfaisantes, elles ont été peu développées dans la pratique car elles nécessitent soit l'acquisition de données historiques sur une longue période (schneider, 1997), soit une mise en œuvre trop lourde (comptage du nombre de grains présents sur la grappe, nombre de grains par rameau, poids des grains) au vu de la qualité de l'information fournie (Booyesen et Al, 1978).

Cette étude pose comme condition l'affranchissement des variables nombres de baies par grappes et poids moyen indicatif d'une baie en raison de la lourdeur de mise en pratique de telles observations. L'approche repose sur l'existence d'une corrélation entre le volume de la grappe durant son développement et son poids à la vendange. L'étude s'attache à balayer la première phase de croissance des baies afin de déterminer une période optimale de mesure pour la mise au point d'un modèle mathématique de corrélation. En parallèle, elle étudie l'influence du millésime, du cépage et de la parcelle sur la précision des résultats. Elle définit enfin des règles d'échantillonnage à l'échelle de la parcelle viticole en prenant en considération le nombre de grappes par pieds.

Matériels et méthodes

Plus de vingt mille données portant sur le volume de la grappe à différents stades phénologiques de la vigne et son poids à la récolte ont été collectées. Cette banque de données expérimentales a été constituée sur dix neuf cépages *Vitis vinifera* au cours de six millésimes (1999 à 2004). Les cépages étudiés sont le Cabernet Franc, Cabernet Sauvignon, Chardonnay, Colombard, Auxerois, Duras, Fer Servadou, Gamay, Grenache, Gros Manseng, Loin de l'œil, Mauzac, Merlot, Muscadelle, Négrette, Sauvignon B., Syrah, Tannat et Ugni B. Pour chacun d'eux, deux à trois parcelles sont identifiées en fonction de leur niveau de production ou de leur type de sol.

La forme d'une grappe est assimilée à un cône. Les mesures sur fruits portent sur la hauteur et la circonférence de la grappe à l'aide d'un outil artisanal permettant une mesure directe.

Les mesures de gabarit sont effectuées aux stades nouaison, début de fermeture de la grappe et début véraison sur un échantillon de 50 à 200 grappes par parcelle. Elles sont baguées et identifiées lors de la première mesure. Elles sont pesées individuellement à la récolte.

Estimation de l'erreur d'échantillonnage : Sur plusieurs parcelles de vigne, de superficies de 0,3 ha à 0,6 ha, est déterminé pour chacun des pieds le nombre de grappes présentes. L'enregistrement des données est réalisé dans le respect de la disposition des pieds au sein de la parcelle. Divers schémas d'échantillonnage sont simulés portant d'une part sur le nombre d'individus au sein de l'échantillon, et d'autre part sur la répartition géographique de l'échantillon.

Modèle de corrélation linéaire : Les modèles proposés sont basés sur des régressions linéaires, selon l'équation

$$\hat{w} = a_0 + a_1 * v$$

, où \hat{w} est l'estimation du poids.

Dans un premier temps, les modèles sont calculés pour chaque variété, millésime et stade phénologique. Ils sont ensuite validés par le traitement statistique en Validation Croisée (leave-one-out cross-validation). Cette méthode de validation consiste à générer N modèles par permutation, chacun d'eux étant étalonné sur N - 1 individus, et testé sur l'individu n'ayant pas participé au modèle. La pertinence

des modèles est évaluée à partir de l'Erreur Standard issue Validation Croisée (SECV – Standard Error of Cross-Validation). Cette erreur peut être exprimée en % par rapport au poids moyen de la grappe propre à la parcelle considérée, produisant alors un Coefficient de Variation.

Dans un deuxième temps, les modèles sont validés par une série de tests indépendants. Les modèles sont calculés à partir des données issues des millésimes 2001, 2002 et 2003 et évalués sur leur capacité à prédire les poids des grappes du millésime 2004.

Analyse en Composantes Principales (ACP) et Classification hiérarchique ascendante (CAH) : Les modèles de corrélation établis par variété et par millésime sont analysés par Analyse en Composantes Principales (ACP). A chaque stade phénologique, les ACP sont effectuées sur les coefficients de régression, afin de définir des groupes potentiels de variétés ou de millésimes.

La classification hiérarchique ascendante (CAH) a pour objet de regrouper les millésimes ou les variétés. La similitude entre les groupes est calculée sur les résultats de l'ACP, et est basée sur la distance euclidienne entre les barycentres des groupes.

Le logiciel de traitement des données utilisé est MATLAB (The Mathworks, NATIC, USA).

Résultats et discussions

Optimisation de l'échantillonnage : nombre de pieds à observer :

La figure 1 présente la part d'erreur induite par l'échantillonnage, aléatoire ou en bloc, d'un nombre croissant de pieds. Quelle que soit la configuration de la parcelle, l'échantillonnage est plus précis lorsqu'on effectue des prélèvements de façon aléatoire. L'échantillonnage en bloc engendre une erreur supérieure de 1% à 5%, en fonction des parcelles et en fonction du nombre de pieds pris en compte. Une précision intéressante de l'échantillonnage, définie par l'écart à la moyenne, est obtenue à partir d'un échantillon d'au moins 40 pieds. Cet écart varie de 14% (Duras) à 4% (Colombard). Il est fonction de la surface considérée, de sa densité de plantation et de son hétérogénéité (pieds manquants ou hétérogénéité ecophysio-logique). Au-delà de 60 pieds dénombrés, le gain en précision n'est pas augmenté. Un échantillonnage d'au moins 40 pieds apparaît nécessaire pour limiter l'erreur de mesure tout en conservant une mise en œuvre au vignoble réalisable. En deçà, le risque d'erreur est augmenté de façon importante ; au-delà, le gain en pertinence est limité pour le temps à y consacrer.

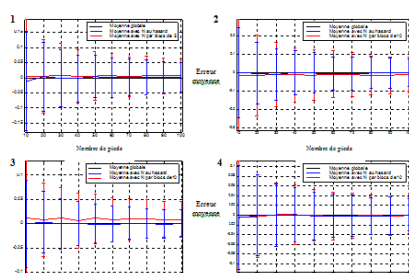


Figure 1 : écart par rapport à la moyenne du nombre de grappes par pied, en fonction d'un échantillon N de pieds dénombrés ; 1 : Négrette ; 2 : Duras ; 3 : Colombard ; 4 : Merlot ; millésime 2001.

Modèles de corrélation pour l'estimation des poids moyens des grappes :

Quel que soit le stade phénologique, le coefficient de corrélation calculé par étalonnage pour un cépage donné, un millésime donné et un stade phénologique donné est élevé. Les moyennes obtenues sur l'ensemble des cépages et des millésimes sont présentées dans le tableau 1. Les performances des modèles en validation croisée sont plus hétérogènes mais demeurent très satisfaisantes pour les stades les plus tardifs. A la nouaison, les erreurs sont moins acceptables, comprises entre 10 et 40 g (SECV). Quel que soit le millésime ou le cépage, la qualité de l'estimation est systématiquement plus précise lorsque la mesure est plus tardive. Il

Stade	r ² (étalonnage)	r ² (validation croisée)
Nouaison	0.91	0.42
Fermeture de la grappe	0.90	0.71
Véraison	0.91	0.76

Tableau 1 : coefficients de détermination (r²) moyen de la relation linéaire entre le volume de la grappe à différents stades phénologiques et son poids à la vendange (19 cépages, 4 millésimes)

existe notamment une grande différence entre les résultats obtenus à la nouaison et ceux issus des deux autres stades phénologiques (figure 2). La figure 3 montre les coefficients de variation obtenus en Validation Croisée pour l'ensemble des cépages au stade fermeture de la grappe en 2003 et 2004. Les performances des modèles présentent des erreurs moyennes proches de 10 %, l'ensemble d'entre-elles étant inférieur à 20%. Si le modèle est construit à partir de plusieurs parcelles par cépage, les performances de l'estimation baissent, quel que soit le stade phénologique. Les mêmes tendances sont observées à l'analyse des données obtenues à la véraison, avec des coefficients de variation légèrement meilleurs, inférieurs dans tous les cas à 13 %, pour l'estimation du poids moyen de la grappe par parcelle.

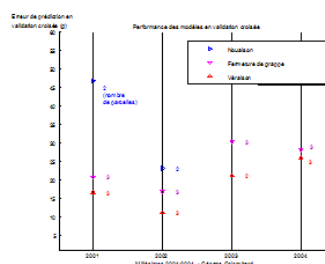


Figure 2 : performance du modèle en Validation Croisée sur Colombard selon le stade phénologique

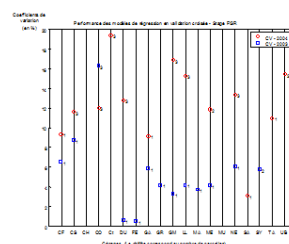


Figure 3 : coefficient de variation en Validation Croisée des 19 cépages en 2003 et 2004 au stade fermeture de la grappe

Validation des modèles par l'estimation du poids moyen pour l'année suivante :

Parmi les 19 variétés étudiées, seulement 4 ont fait l'objet de mesures durant 4 années (2001 à 2004) au stade phénologique fermeture de la grappe et véraison : Colombard, Duras, Merlot et Négrette. Une première analyse des modèles montre que ces derniers sont différents selon les millésimes (figure 4). Le millésime 2003 s'avère être atypique pour 3 des 4 variétés étudiées (exception du Colombard). L'étude est réalisée sur le

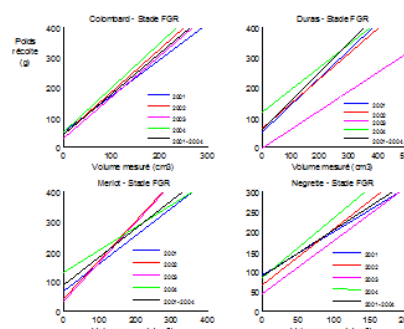


Figure 4 : modèles de corrélation de 4 cépages au stade fermeture de la grappe

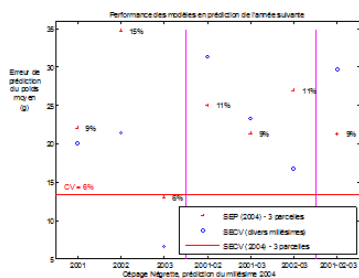


Figure 5 : performance du modèle pour l'estimation du millésime 2004 à partir de données antérieures

millésime 2004 à partir de modèles bâtis sur toutes les combinaisons possibles, issues des années précédentes (2001 à 2003).

Lorsque les modèles sont bâtis sur les données de 3 années, les coefficients de variation augmentent en moyenne d'un facteur 1,5 à 2, en comparaison aux meilleures performances obtenues par le modèle en Validation Croisée sur le même millésime.

La figure 5 montre les résultats des estimations pour le cépage Négrette au stade fermeture de la grappe. Quand le modèle est calculé sur une seule année, les performances sont très différentes d'une année à l'autre, selon la proximité entre l'année modélisée et l'année à prédire. Lorsque plusieurs années sont incluses dans le modèle, la stabilité des prédictions augmente, permettant d'obtenir une performance moyenne intéressante proche de 9 %.

Groupe de variétés : Certaines variétés peuvent être regroupées à l'intérieur d'un même modèle. Ce regroupement permet d'accroître la banque de données, et ainsi d'augmenter la stabilité des modèles. L'objectif final est de générer un seul et même modèle adapté à plusieurs variétés. Basés sur les deux derniers stades phénologiques étudiés, 3 groupes ont été identifiés comme pouvant dépendre d'un seul et même modèle de corrélation :

- Groupe 1 : Duras, Merlot, Mauzac, Syrah
- Groupe 2 : Gamay, Chenin, Tannat
- Groupe 3 : Cabernet Sauvignon, Cabernet Franc, Gros Manseng

Les autres cépages ne peuvent être assimilés à aucun de ces 3 groupes.

Groupes de millésimes : Un même type d'analyses a été réalisé sur les différents millésimes étudiés, afin d'identifier des groupes. L'objectif est de pouvoir établir des formules moyennes en fonction des critères climatiques (par exemple millésime sec et humide) et ainsi de pouvoir encadrer la prévision du volume moyen des grappes. L'analyse statistique montre qu'il n'existe pas de forte hétérogénéité entre les années. Seul le millésime 2003 apparaît significativement différent des autres millésimes.

L'analyse statistique identifie donc bien les conditions exceptionnelles de croissance des grappes observées en 2003.

Conclusion

Cette étude statistique met en évidence la pertinence de la mesure du volume d'une grappe durant sa phase herbacée pour en estimer son poids à la récolte.

Des modèles de corrélation stables ont pu être mis en évidence entre la mesure du volume à la fermeture de la grappe ou à la véraison, et son poids à la vendange.

Les modèles ont été validés par validation croisée sur plusieurs années, et testés sur lot totalement indépendant - le millésime 2004 -. Ces modèles ont été appliqués pour chacune des 19 variétés étudiées. L'ensemble des coefficients de variation obtenus est inférieur à 20 %.

Sur les cépages présentant une banque de données importante, des erreurs d'estimation inférieure à 10% sont obtenues. Un effet parcelle est mis en évidence mais celui-ci tend à être gommé par le regroupement de données issues de plusieurs millésimes.

Certains groupes de cépage et millésimes ont également pu être identifiés. Le regroupement de données issues de différents cépages et/ou millésimes laisse ainsi entrevoir la possibilité de construire des modèles de corrélation toujours plus stables et plus performants. L'année 2003 apparaît cependant atypique ; l'introduction de données de ce millésime dégrade les performances des modèles.

Enfin, dans le cadre d'une application de terrain, l'étude démontre qu'un échantillonnage minimum de 40 ceps permet de limiter de manière optimale l'erreur de mesure pour le calcul du nombre de grappes par cep.

Références bibliographiques

- BOOYSEN J.H., ORFFER C.J., BEUKMAN E.T., 1978. Crop forecasting for vine grapes in South-Africa. OVRI Stellenbosch, 78-94.
- GERBIER N., REMOIS P., 1977. Influence du climat sur la qualité et la production du vin de Champagne. Monographie n°106, Météorologie Nationale.
- HUGLIN P., BALTHAZARD J., 1975. Variabilité et fluctuation de la composition des inflorescences et des grappes chez quelques variétés de V. vinifera. Vitis, 6-13.
- MURISIER F., JEANGROS B., AERNY J., 1986. Maîtrise du rendement et maturité du raisin. Revue Suisse Vitic. Arboric. Hortic. Vol 18 (3) : 149-156.
- SCHNEIDER C., 1995. La prévision, un outil pour la maîtrise des fluctuations de rendement en viticulture. C.R. GESCO, 240-246.
- SERRANO E., DUFOURCQ T., CHABERT M., 2002. Recherche d'une méthode simple et fiable d'estimation des rendements à la parcelle. Actes de colloque Journée Technique Maîtrise des rendements en viticulture, dec. 2002. Station Régionale ITV Midi-Pyrénées.
- WURGLER W., LEYVAZ H. ET BOLAY A., 1955. Peut-on prévoir le rendement de la vigne avant le débournement ? Annuaire agr. D. I. Suisse, 766-783.