

A Chenin decision help system to predict grape berries maturity and anticipate wine potentialities

Ilaria Brunetti¹, Mejean Perrot Nathalie², Picque Daniel², Guérin Laurence³

¹Institut National de la Recherche Agronomique d'Avignon, ilaria.brunetti@inra.fr

²Institut National de la Recherche Agronomique de Grignon, nathalie.mejean@inra.fr

³Institut Français de la Vigne et du Vin

L'objectif principal de notre travail est de développer un outil d'aide à la décision sur la maturation des baies pour le cépage Chenin blanc. Un modèle qui lie l'analyse sensorielle des baies au type de vin à élaborer sera également abordé. Nous nous intéressons à l'ensemble des sources d'information disponibles (les données expérimentales, la littérature scientifique mais aussi l'expertise humaine) pour caractériser les mécanismes de la maturation dans le but de modéliser l'évolution des indicateurs de maturité considérés durant cette période. Plusieurs étapes sont nécessaires pour conduire cette démarche : faire un état des lieux des connaissances concernant la maturation des baies de raisin, définir les indicateurs de maturité à prendre en compte et les variables qui influencent leur évolution. Puis, il faut définir le réseau d'interaction, et donc la structure du réseau bayésien et évaluer ses performances.

1 INTRODUCTION

Parmi les demandes des acteurs de la filière viticole, le besoin de méthodes d'évaluation de la maturité des baies de raisin et de leur potentiel à la vinification selon des « cibles vins » déterminées est clairement exprimé. Plusieurs types de maturité peuvent être mesurés, montrant bien la complexité du problème : maturité technologique, aromatique, sensorielle. Des outils et des capteurs portables sont continuellement développés ou améliorés et plusieurs équipements sont apparus ces dernières années mesurant la couleur des baies, la teneur en sucre, la maturité polyphénolique ou le potentiel aromatique (Ben Ghazlen, et al., 2010), (Geraudie, Roger, & J. M., 2010). Cependant, ces analyses sont pour la plupart des mesures en laboratoire, consommatrices de temps et généralement coûteuses pour un suivi rapproché de la maturation. Dans la pratique, des réseaux de suivis de maturité existent dans toutes les régions viticoles françaises, qui consistent à référencer les cépages et les parcelles associées les plus représentatives de la zone viticole donnée ; d'assurer des prélèvements de raisins, et la réalisation des analyses physico-chimiques. Outre le fait que la mise en œuvre est fortement mobilisatrice, les analyses réalisées sont représentatives de l'état de maturité des raisins, au temps t , et n'ont aucune valeur de prédiction de l'évolution de la maturité.

L'utilisation de modèles mathématiques est une piste très intéressante car elle peut permettre de comprendre les mécanismes gérant les processus, d'anticiper les évolutions et ainsi de s'y adapter. On peut citer l'outil « *épiculture* » développé par l'IFV qui permet de prédire les risques phytosanitaires pour établir le calendrier de traitement (Raynal, Debord, Guittard, & Vergnes, 2010), ou les travaux de l'UMR SPO de Montpellier sur la prédiction des cinétiques de fermentation (Goelzer, Charnomordic, Colombié, Fromion, & Sablayrolles, 2009). Pour le suivi de la maturation des baies, un modèle liant météo et maturité technologique a été développé, basé sur les Réseaux Bayésiens Dynamiques, pour les cépages Cabernet Franc et Gamay (Baudrit, et al., 2015), (Perrot, et al., 2015). Au-delà de ces travaux peu de modèles ont été développés et le caractère prédictif de ces modèles est généralement absent. Des approches par régression PLS (partial least square) ont été proposées (Claverie, et al., 2008) pour mettre

en lien des données climatiques et pédologiques avec des indicateurs de maturité des baies (sucre, acidité totale) pour les cépages Merlot et Cabernet Sauvignon. Cependant, la nature des variables retenues dans le modèle limite son utilisation en tant qu'outil d'aide à la décision. Les connaissances disponibles sur la maturation sont d'origines diverses et exprimées sous différentes formes (équations, expertise, bases de données, ...), dans différents formats (numérique, symbolique, linguistique) et à différentes échelles d'observation (microbiologique, physico-chimique, organoleptique). Pour développer un outil d'aide à la décision, il faut disposer de modèles mathématiques capables de prendre en compte ces différentes sources de connaissance exprimées en différents formats (N., Trelea, Baudrit, Trystram, & Bourguine, 2011), (Baudrit, Wullemin, & Perrot, 2013). Les réseaux bayésiens dynamiques RBD (Murphy, 2002) fournissent un formalisme mathématique pratique qui permet de décrire des systèmes complexes en exploitant ces différentes sources de connaissance. Nous avons donc développé des modèles mathématiques adaptés à la description de la maturation des baies de raisin et à la prédiction du potentiel qualitatif du vin. L'objectif principal est de proposer un outil d'aide à la décision robuste utilisable par la filière vini-viticole afin de prédire l'évolution de différentes variables représentatives de la maturation des baies de raisin. Cet outil est conçu sur les données de parcelles du cépage Chenin blanc en région Anjou et Touraine.

1 MATERIELS ET METHODES

1.1 DONNÉES EXPÉRIMENTALES

Les données expérimentales proviennent de parcelles situées en Val de Loire suivies par l'IFV sur plusieurs millésimes, avec des prélèvements hebdomadaires permettant d'avoir une évolution régulière des indicateurs de maturité. Pour chaque parcelle expérimentale, les données de la station météorologique la plus proche ont été récupérées.

Les données expérimentales couvrent deux zones géographiques : l'Anjou et la Touraine pour un total de 30 parcelles différentes et une période allant de 1989 à 2017. Selon la parcelle considérée, l'ensemble des millésimes n'est pas forcément couvert et le nombre de prélèvements par année peut différer.

1.2 INDICATEURS DE MATURITE ET VARIABLES METEOROLOGIQUES

Trois types de variables ont été utilisés pour développer le modèle mathématique.

- Les **variables climatiques** influençant l'évolution des indicateurs de maturité. Quatre variables climatiques sont utilisées : l'humidité relative **HR**, la température **T**, la pluviométrie **PI** et l'ensoleillement **Ins**.

Variable	Description	Calculation	Unit
HR	Relative humidity	$mean(\sum_{i=1}^7 \frac{HRmin_i + HRmax_i}{2})$	%
T	Temperature	$\sum_{i=1}^7 \frac{Tmin_i + Tmax_i}{2}$	°C
PI	Rainfall	$\sum_{i=1}^7 Pl_i$	mm
Ins	Insolation	$\sum_{i=1}^7 Ins_i$	h

- Les **variables physico-chimiques**, qui sont les indicateurs de maturité des baies de raisin obtenues après analyse en laboratoire. Six indicateurs de maturité sont fréquemment utilisés par les experts, faisant référence à la maturité technologique (alcool, pH, acidité), le potentiel phénolique des baies (maturité et extractibilité des polyphénols) et la maturité aromatique (arômes de la pulpe). Parmi ces indicateurs, trois indicateurs ont été retenus pour définir la maturité des baies de raisins : le **sucre** (s) en g/l, l'**acidité totale** (ac) en $g_{H_2SO_4}/L$ et l'**acide malique** (ac_m) en g/L, (Barbeau, 2003), (Riou, 1994). Pour les variables physico – chimiques nous disposons des données des millésimes **1989, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2016, 2017, 2018**. Les données du 2018 ont été récoltées mais pas encore traitées.
- **Les variables sensorielles raisin**, qui sont les indicateurs de maturité des baies de raisin obtenues après analyse sensorielle réalisée au sein de l'IFV. Un ensemble représentatif de ces indicateurs a été retenu avec l'aide de l'ESA et de l'IFV ; les indicateurs sensoriels choisis peuvent être groupés dans les trois familles suivantes.
 - Indicateurs relatifs à la **baie** : teinte, fermeté/élasticité, dilacération de la pellicule, force pour détacher le pédicelle, quantité de chair sur le pédicelle, croquant, gélatineux de la pulpe, agressivité de la pellicule, couleurs des pépins ;
 - Indicateurs relatifs au **jus** : teinte, sucre, acide
 - Indicateurs des **arômes** : intensité globale aromatique, fruit blanc, fruit jaunes, fruit exotique, fruité agrumes.

Chaque indicateur est noté avec une valeur (entière) entre 1 et 5 (sauf les variables « Teinte de la baie » et « Teinte du jus », notées de 1 à 10) ; les notes sont ensuite moyennées pour toutes répétitions et tous juges. Nous avons utilisé les données sensoriels raisin des millésimes 2016 et 2017.

1.3 RECUEIL D'EXPERTISE

Afin de comprendre l'ensemble des éléments qui définissent la potentialité d'un vin, nous avons interviewé deux vignerons experts du chenin, Claude Papin du Château Pierre Bise à Beaulieu sur Layon (Anjou) ainsi que Christophe Vigneau, du **Domaine Vigneau-Chevreau**, Chançay (Touraine). La méthodologie de recueil s'appuie sur la méthode des **entretiens d'explicitations** développée par Pierre Vermersch (Vermersch, 2006). Deux entretiens ont eu lieu avec chacun dans lesquels nous avons d'abord défini ensemble la cible mentale idéale pour leur production et quels étaient les éléments qui permettaient d'atteindre cette cible : des conditions pedo-géologiques aux pratiques agricoles.

1.4 FORMALISMES MATHÉMATIQUES PROPOSÉS

Les modèles de prédiction proposés s'appuient sur deux formalismes particuliers qui permettent d'entrer en étroite interaction avec les experts du chenin afin de formaliser la connaissance implicite (contenue dans les données expérimentales à disposition) et explicite (verbalisation experte, savoirs faire).

1.4.1 Introduction aux outils de modélisation

Le premier formalisme est un formalisme qui repose sur une représentation graphique probabiliste Bayésienne : les **réseaux bayésiens dynamiques (DBN)**. Ainsi à partir d'une représentation graphique des liens entre variables sous forme de graphe orienté (défini, dans notre cas, grâce à l'expertise humaine) et d'une base de données, des probabilités conditionnelles sont calculées automatiquement permettant de relier ces variables ; Les variables doivent être discrétisées pour construire le modèle. Le réseau étant dynamique, il permet de calculer l'évolution des variables en plusieurs pas de temps, dans notre cas deux pas de temps équivalents à deux semaines de maturation.

Plus particulièrement, un réseau Bayésien Dynamique RBD (ou *Dynamic Bayesian Network*, DBN) est un modèle graphique d'une distribution de probabilité multivariée qui capture les propriétés

d'indépendance conditionnelle entre variables. D'une part, il s'agit d'une représentation graphique de la probabilité de distribution jointe, d'autre part, il représente la dépendance entre variables. De manière formelle, un réseau bayésien dynamique est un graphe acyclique orienté dont les nœuds représentent les variables et les arcs marquent la dépendance conditionnelle entre variables. Les réseaux bayésiens dynamiques sont une extension des réseaux bayésiens dans lesquels les nœuds représentant une variable discrète aléatoire sont indexés au temps et fournissent une représentation compacte de la probabilité de distribution jointe pour un intervalle de temps fini (Baudrit, et al., 2015) (Perrot, et al., 2015).

Le deuxième formalisme est la logique floue, qui, pour la partie que nous développons dans ce rapport, repose sur la théorie des ensembles flous. Cette théorie est une approche de formalisation mathématique passant par la notion de fonctions d'appartenance floues. Elles représentent une projection d'un espace de variables numériques sur un espace de variables symboliques manipulées par les experts du domaine. Cela peut être pour prendre une décision d'action ou fusionner des informations de différents formats et types pour prendre une décision. Ces variables numériques, une fois projetées sur l'espace symbolique, donnent un degré d'appartenance flou à chaque symbole défini pour une variable numérique. Les degrés d'appartenance sont ensuite fusionnés à l'aide d'opérateurs logiques (ie. : produit/somme bornée ou min/max) selon des règles établies avec l'expert, par exemple : si acidité ok ET sucre ok ET pH ok = VIN CIBLE. Pour plus de détails méthodologiques sur la mise en œuvre de DBN ou la logique floue se référer à (Perrot, et al., 2015).

1.4.2 Evaluation des modèles

Les modèles Bayésiens sont évalués par une **validation croisée** (*cross validation*) du type « *leave-one-out* » : le modèle utilise le jeu de données complet moins une observation pour apprendre les paramètres (*validation set*), et l'observation (*training set*) est utilisé pour évaluer le modèle. Ce processus est répété pour toutes les observations, pour obtenir des prédictions pour l'ensemble des prélèvements (Geisser, 1993).

Les réseaux bayésiens, en se basant sur des probabilités conditionnelles, exigent des variables discrètes (voir la sous-section 3.1.1) et les résultats obtenus sont en termes de classes discrètes. Il est donc nécessaire de définir une méthode pour comparer les résultats aux valeurs observés ; nous avons transformé la classe prédite en valeur continu de la façon suivante :

$$x = \sum_{i=1}^{n_x} \bar{x}_i P(i)$$

Où n_x est le nombre de classes pour la variable x ; \bar{x}_i et $P(i)$ sont respectivement la valeur moyenne et la probabilité de la i -ème classe.

Pour évaluer les résultats du modèle de prédiction des variables chimiques, les valeurs prédites sont comparées aux valeurs observées avec le calcul du **RMSE (Root Mean Square Error)** et du RMSE exprimé en pourcentage de la gamme de valeurs observées, le **RRMSE**.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^{prédit} - x_i^{observé})^2} \quad RRMSE = \frac{RMSE}{x_{max}^{observé} - x_{min}^{observé}} \times 100$$

Pour le modèle flou, le pourcentage des points correctement classifiés est comptabilisé aussi.

2 RESULTATS : MODELES DE PREDICTION POUR LE CHENIN

Trois modèles de prédiction de la maturité des baies ont été développés en s'appuyant sur le formalisme mathématique des réseaux Bayésiens dynamiques :

- Indicateurs chimiques en fonction des variables météo ;
- Indicateurs sensoriel raisin en fonction des variables météo ;
- Indicateurs sensoriel raisin en fonction des variables météo et chimiques ;

De plus, un quatrième modèle de prédiction de la qualité globale attendue du vin à partir des indicateurs des maturités chimiques a été développé, en se basant sur la théorie de la logique floue.

2.1 LES MODELES CHENINS BASES SUR LES RESEAUX BAYESIENS DYNAMIQUES (DBN)

Ces réseaux sont définis par i) une discrétisation des variables, ii) sa structure qui va définir les relations entre variables et iii) ses paramètres, un tableau de probabilités conditionnelles (CPT) qui définit quelle est l'influence des variables prédictives sur les variables prédites. La structure du réseau a été développée sur la base de précédents travaux qui ont donné lieu au développement du logiciel PREVIMAT (Brousset, 2009) (Baudrit, et al., 2015) et de la connaissance experte.

2.1.1 La discrétisation des variables

Le réseau bayésien dynamique, en se basant sur des probabilités conditionnelles, exige des variables discrètes : il est donc nécessaire de construire une discrétisation préliminaire des variables continues dont on dispose. La sensibilité des résultats du modèle à la discrétisation est importante et cette étape est cruciale pour bien prédire. Il existe différentes méthodes ayant une influence directe sur la qualité des calculs : la méthode des quantiles, la méthode des moyennes emboîtées, la méthode des amplitudes. C'est cette dernière qui a été choisie dans notre cas permettant ainsi des classes d'amplitude similaires pour ce qui concerne les variables que nous avons choisi de manipuler : variables climatiques, variables sensorielles raisin et les variables physicochimiques instantanées ou variations des variables physicochimiques d'un pas de temps à un autre. Le nombre de classes a été choisie (avec l'expertise humaine) en choisissant un nombre d'intervalles limité mais suffisamment représentatif.

Les intervalles définis pour les variables climatiques sont les suivants :

- $Ins = [[15,30], [30,40], [40,55], [55,60], [60,75]]$
- $Pl = [[0,10], [10,20], [20,30], [30,45], [45,70], [70,100]]$
- $T = [[0,11], [11,15], [15,17], [17,19.5], [19.5,22]]$
- $HR = [[60,70], [70,75], [75,80], [80,90], [90,100]]$

Pour les variables sensorielles raisin, la discrétisation est fixée aux 4 intervalles unitaires de 1 à 5 (10 dans le cas de la teinte de la baie et du jus), c'est-à-dire $[[1,2], [2,3], [3,4], [4,5],]$. Pour la discrétisation des variables physico-chimiques un outil innovant de **discrétisation semi-automatique** a été développé. Il s'appuie sur la notion de co-occurrence entre les variables et leurs variations. Pour cela nous utilisons un outil de visualisation mis en place au sein de l'équipe Malices du laboratoire (**EvographDice**). Couplé à cette approche de visualisation, nous définissons une approche d'optimisation basée sur les algorithmes évolutionnaires adaptés (Boukhelifa, Tonda, Trelea I.-C., Perrot, & Lutton, 2017) qui, à partir de la discrétisation donnée, détermine la discrétisation des variables correspondantes.

Les résultats de la discrétisation semi-automatique des variables physico-chimiques

Pour l'**acidité totale**, la discrétisation de la **variation Δac fixée** est la suivante :

- $[[-\infty, -1.5], [-1.5, -1], [-1, -0.6], [-0.6, 0]]$

La discrétisation de **ac déterminée par l'algorithme d'optimisation** est :

- $[-\infty, 5.47], [5.47, 6.33], [33, 7.94], [7.94, +\infty]$

Pour l'**acide malique**, la discrétisation de la **variation** Δac_m **fixée** est la suivante :

- $[-\infty, -2.5], [-2.5, -1.5], [-1.5, -0.75], [-0.75, -0.5], [-0.5, 0]$

La discrétisation de ac_m déterminée par l'algorithme d'optimisation est :

- $[-\infty, 3.66], [3.66, 4.6], [4.6, 5.68], [5.68, 6.88], [6.88, +\infty]$

Pour le **sucre**, la discrétisation de la **variation** Δs **fixée** est la suivante :

- $[0, 12], [12, 20], [20, 35], [35, +\infty]$

La discrétisation de s déterminée par l'algorithme d'optimisation est :

- $[-\infty, 156.9], [156.9, 182.86], [182.86, 201.8]$

On affine ensuite cette discrétisation en ajoutant les classes suivantes :

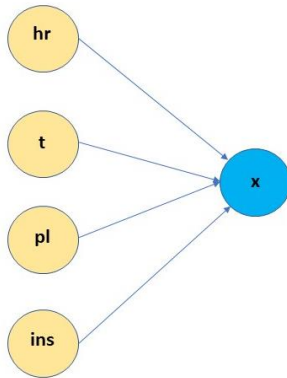
$[201.8, 210], [210, 220], [220, 230], [230, 240], [240, +\infty]$.

2.1.2 Modèle DBN 1 : Prédiction des indicateurs physico-chimiques en fonction des variables météo

La structure du réseau

Le premier modèle développé, inspiré par les précédents travaux sur le Cabernet-Franc et le Gamay (Baudrit et al. 2015) est celui qui permet de prédire les indicateurs physico-chimiques en fonction des conditions météorologiques. Pour chaque indicateur de maturité physico-chimique, les variables climatiques jouant un rôle clé dans leur cinétique ont été sélectionnées à partir des dires d'experts et de la littérature. En particulier, l'humidité relative n'impacte que les deux acidités, l'ensoleillement le sucre, alors que température et pluviométrie influencent les trois variables

La variation des variables et la dynamique du modèle



Pour le modèle de prédiction des variables physico-chimiques, nous avons introduit dans le réseau RBD des nouvelles **variables d'état intermédiaires** : les **variations** des variables physico-chimiques. En effet, nous regardons l'évolution des indicateurs dans les quatre dernières semaines avant la vendange et donc la trajectoire d'une variable peut être supposée stable et la variation constante à conditions météorologique identiques : à ce stade, on peut en effet imaginer que seule la météo peut causer une déviation significative de cette trajectoire.

Pour chaque indicateur physico-chimique $x \in \{ac, ac_m, s\}$, à temps $t + 1$, on a : $\mathbf{x}(t+1) = \Delta \mathbf{x}(t+1) + \mathbf{x}(t)$ et donc : $\Delta \mathbf{x}(t) := \mathbf{x}(t) - \mathbf{x}(t-1)$.

Figure 1: Relations entre les variables climatiques et chimiques

Comme déjà anticipé, la valeur (absolue) d'une variable nous sert d'indicateur du stade de maturation auquel on se trouve alors que la variation, prédite en fonction des variables climatiques indique la

trajectoire de maturation.

La dynamique du modèle

Nous définissons un RBD en trois pas de temps où chaque pas correspond à une semaine de suivi. Au temps $t = 0$, toutes les variables sont observées ; aux deux temps successifs, seules les variables météo sont observées, alors que les variables physico-chimiques et leurs variations sont prédites. La figure 5 montre la dynamique du réseau dans le temps : un indicateur physico-chimique au temps $t + 1$, $x(t + 1)$ est prédit en fonction de sa valeur à l'instant précédent $x(t)$, ainsi que de la valeur de la variation, $\Delta(t) = x(t) - x(t - 1)$. La variation $\Delta(t)$ est prédite quant à elle en fonction des variables climatiques entre $(t - 1)$ et t . À noter : les variables physico-chimiques sont indépendantes entre-elles. Le graphe dans son ensemble est présenté en Figure 3.

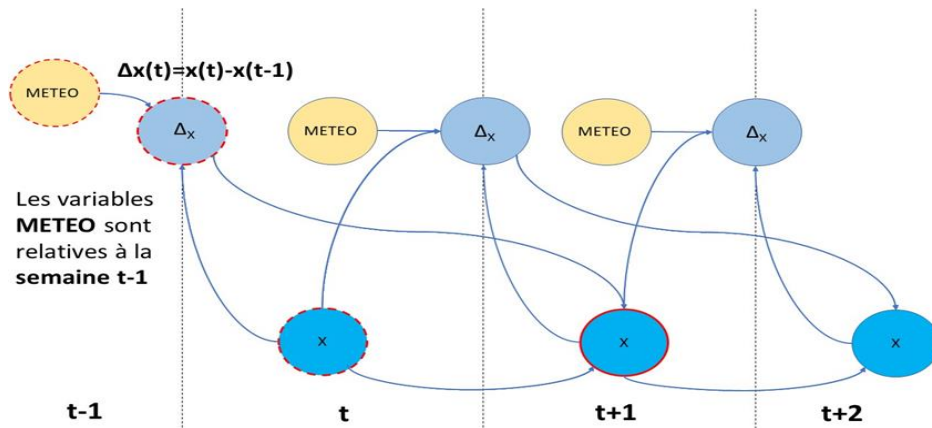


Figure 2: la dynamique temporelle du réseau bayésien pour les indicateurs physico-chimiques

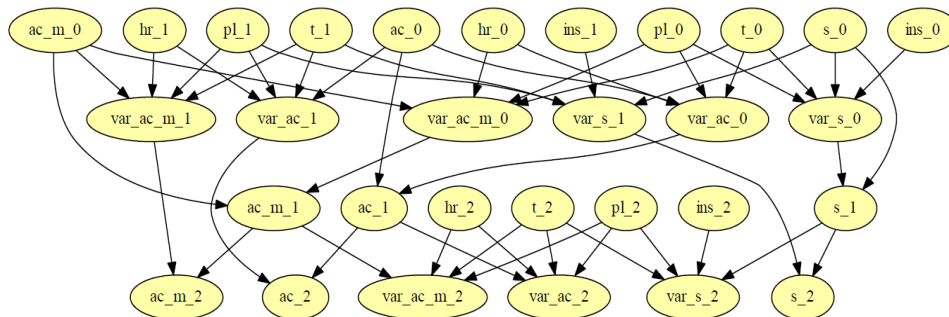


Figure 3: le réseau complet dans les trois pas de temps

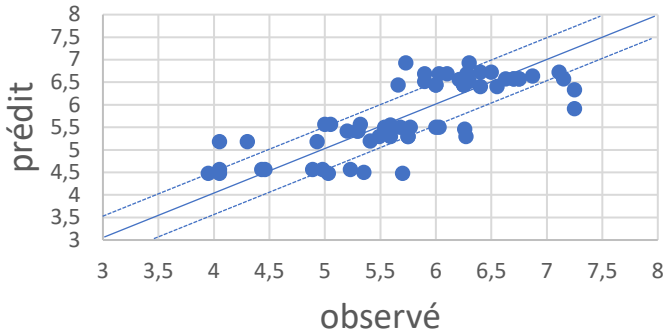
Résultats Les résultats sont résumés dans le tableau suivant. Les meilleurs résultats sont obtenus pour l'acidité totale ; sucre et acide malique

Variable	RMSE	RRMSE % ac = [3.4,12.5]; ac_m = [1.7,10] ; s = [144,271.8]
ac_1 (g/l)	0,536	6%
ac_2 (g/l)	0,648	7%
ac_m_1 (g/l)	0,825	9%
ac_m_2 (g/l)	0,867	10%
s_1 (g/l)	11,37	8%
s_2 (g/l)	12,87	10%

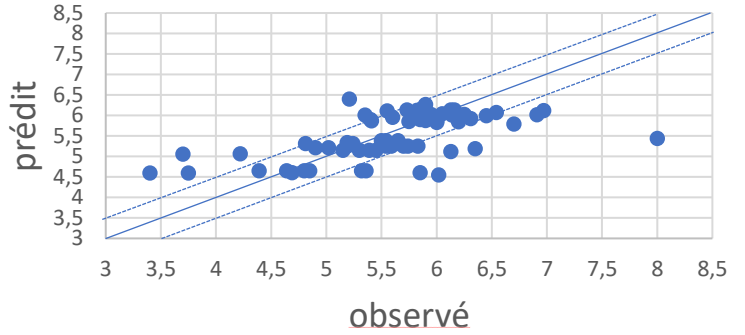
Dans les figures suivantes, les prédictions sont montrées avec des graphs du type observé-prédit, pour chaque variable et pour les deux pas de temps prédits, où le plus les points sont distribués long la

bissectrice (ligne continue en figure 4), les meilleures sont les prédictions. Pour l'acidité totale et l'acide malique, les parallèles à la bissectrice sont tracées à $\pm 0.5g/l$, pour le sucre à $\pm 10g$.

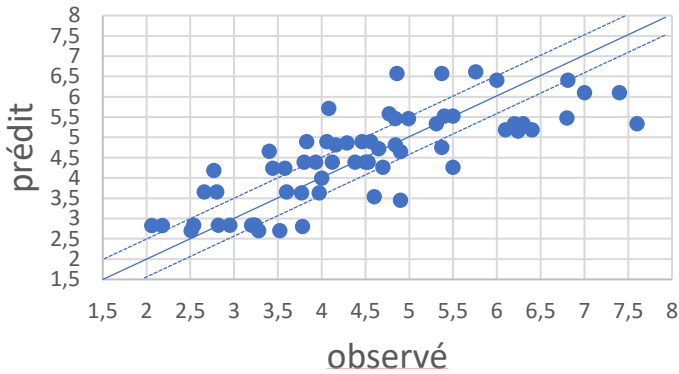
Acidité totale à t=1



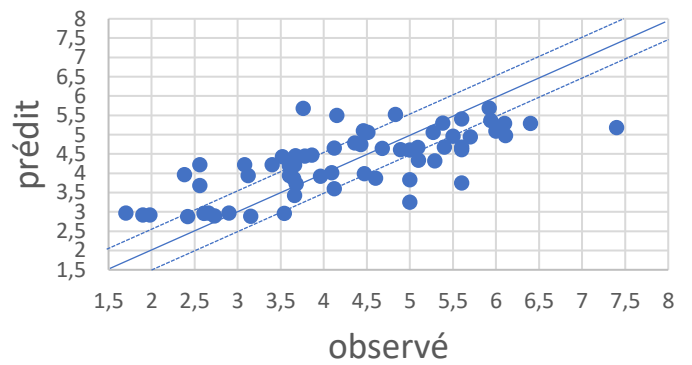
Acidité totale à t=2



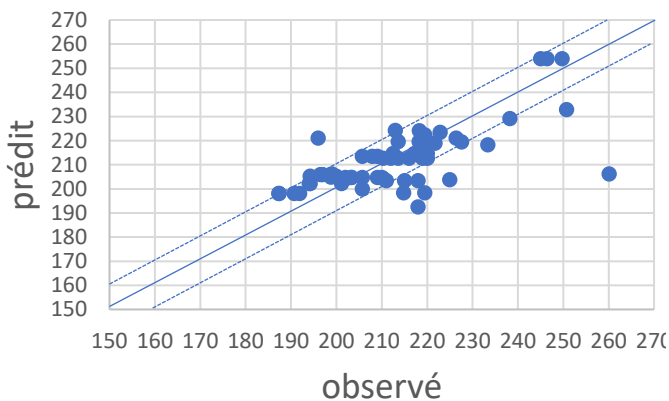
Acide malique à t=1



Acide malique à t=2



Sucre à t=1



Sucre à t=2

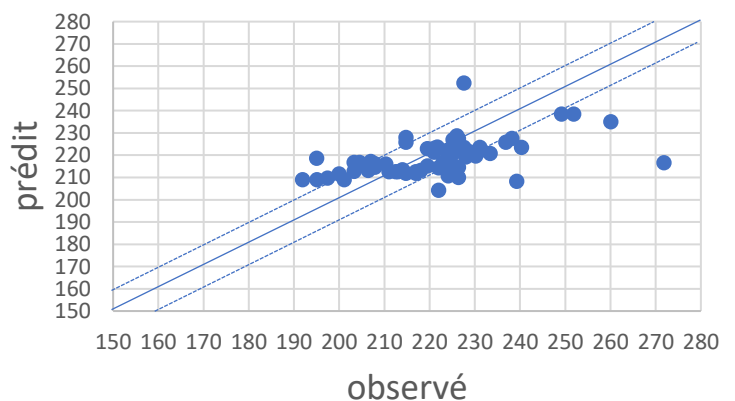


Figure 4: les résultats prédits pour les trois variables p-c dans les deux pas du temps. Sur la bissectrice les résultats bien prédits ; les lignes pointillées indiquent les prédictions à $\pm 0.5g/L$ pour les acidités, $\pm 10g/L$ pour le sucre

2.1.3 Modèle DBN 2 : les indicateurs sensoriels en fonction des variables météo

La structure du réseau Parmi les variables sensorielles analysées par l'IFV, un sous-ensemble représentatif de ces indicateurs a été retenu avec l'aide de l'ESA et l'IFV pour notre modèle :

- Indicateurs relatifs à la **baie** : teinte, dilacération de la pellicule, force pour détacher le pédicelle, agressivité de la pellicule, couleurs des pépins ;
- Indicateurs relatifs au **jus** : teinte, sucre, acide
- Indicateurs des **arômes** : intensité globale aromatique, fruit jaunes, fruité agrumes.

Dans ce cas, la variation des variables prédites n'est pas informative : les variables sont notées sur une échelle discrète de 1 à 5 (10 dans le cas de la teinte de la baie et du jus) et les notes sont moyennées pour les différentes répétitions et les différents juges. Il n'y a pas donc la même relation entre leur variation et leur valeur comme dans le cas précédent. Le RBD est donc construit uniquement avec les variables climatiques et sensorielles indiqués ; à différence du premier modèle, chaque variable est prédite en fonction de toutes les quatre variables météo. Les pas de temps sont toujours trois, avec une semaine d'observation et deux semaines de prédiction. Le modèle complet est présenté figure 5.

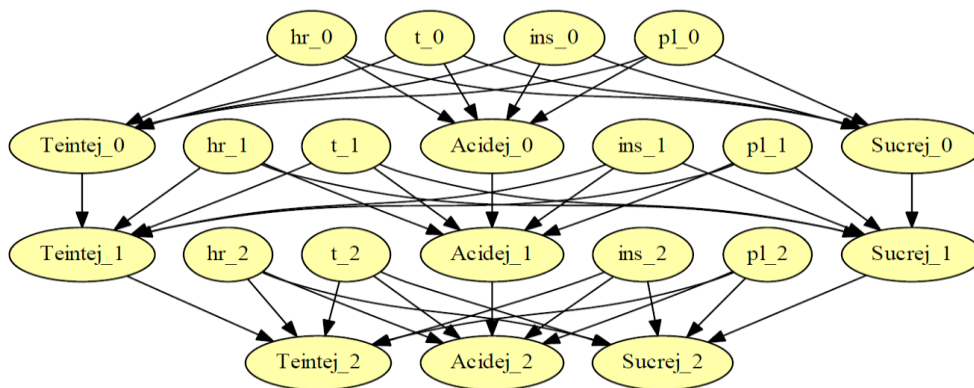


Figure 5: le réseau pour les variables sensorielles- jus dans les trois pas de temps (« unrolled »).

Résultats Pour les variables sensorielles, nous calculons le RMSE, et nous évaluons aussi le pourcentage de prédictions dont la distance de la valeur observée est inférieure respectivement à 0.5 et 0.25, Selon le dire des experts les résultats à 0.5 sont les plus représentatifs compte tenu de la sensibilité de la mesure sensorielle uniquement attendu à 0.5 près. Les résultats sont montrés dans les trois tableaux suivants (baie-jus-aromatique).

Variables BAIE	RMSE	% pred-obs < 0.5	% pred-obs < 0.25
Couleurpepin_1	0.4	75	50
Agrespelli_1	0.38	82	63.63
Dilapelli_1	0.4	77.3	47
Teinteb_1	0.91	56.8	25
Forcepedi_1	0.38	79.5	61.36
'Couleurpepin_2	0.48	70.5	43.2
Agrespelli_2	0.36	77.3	45
Dilapelli_2	0.35	86.36	59
Teinteb_2	1.07	43.18	27
Forcepedi_2	0.37	79.5	54

Pour les variables concernant la baie, la teinte est la variable plus difficile à prédire (Teinteb) et l'erreur est encore trop importante. Pour les autres variables on obtient des résultats prometteurs. À noter que la prédiction à deux semaines reste bonne et, pour la « dilacération de la pellicule » (Dilapelli), même meilleure que la prédiction de la première semaine. Pour les variables concernant le jus, on obtient une bonne prédiction pour le sucre (Sucrej), alors que pour les autres variables les résultats sont moins satisfaisants.

Variable JUS	RMSE	% pred-obs < 0.5	% pred-obs < 0.25
Aciditéj_1	0.69	52.3	29.5
Sucrej_1	0.44	79.6	43.2
Teintej_1	0.49	63.6	45.5
Aciditéj_2	0.8	52.3	22.7
Sucrej_2	0.55	72.7	40.9
Teintej_2	0.54	53.6	29.5

Les variables aromatiques sont également difficiles à prédire, mais on obtient des résultats prometteurs pour la variable « intensité globale aromatique » (IntGloAro) et « fruité agrumes » (FAgrume) à la deuxième semaine de prédiction.

Variable AROMATIQUE	RMSE	% pred-obs < 0.5	% pred-obs < 0.25
IntGloAro_1	0.73	47.5	30
FAgrume_1	0.68	62.5	32.5
Fjaune_1	0.74	50	32.5
IntGloAro_2	0.52	72.5	50
FAgrume_2	0.54	75	45
Fjaune_2	0.53	55	47.5

2.1.4 Modèle DBN 3 : les indicateurs sensoriels en fonction des variables météo et chimiques

La structure Le troisième modèle a été construit intégrant au précédent les variables physico-chimiques pour la prédiction des variables sensorielles. Ces dernières sont donc prédites en fonction de la météo et des valeurs observées de sucre, acidité totale et acide malique. Les indicateurs sensoriels retenus dans ce cas sont :

- Baie : teinte, agressivité pellicule, couleur pepin, dilacération pellicule.
- Jus : acide, sucre.
- Arômes : intensité globale aromatique, fruits jaunes, fruité agrumes.

Pour établir les liens entre variables sensorielles et physico – chimiques, nous nous sommes basés sur le tableau de corrélations établi par l’ESA, en figure 8 (à gauche). À ces liens (illustrés en figure 7, à droite), ils s’ajoutent ceux avec les variables climatiques, où, comme dans le modèle précédent, toutes les quatre

	AT	PH	AC_MAL	SUCRE_RED
Acidej	0,72	-0,48	0,61	-0,66
Agrespelli	0,49	0,02	0,58	-0,13
Couleurpepin	-0,57	-0,18	-0,69	0,05
Croquant	0,4	-0,27	0,31	-0,42
Dilapelli	-0,47	0,38	-0,37	0,52
FAgrume	0,61	-0,5	0,44	-0,69
Fermeteb	0,25	-0,2	0,19	-0,27
Fjaune	-0,81	0,46	-0,71	0,77
Forcepedi	-0,23	0,09	-0,23	0,17
IntGloAro	-0,84	0,51	-0,72	0,7
Nconfite	-0,59	0,45	-0,46	0,6
Qtitechairpe d	0,26	-0,05	0,33	-0,04
Sucrej	-0,69	0,28	-0,69	0,36
Teinteb	-0,78	0,14	-0,83	0,33

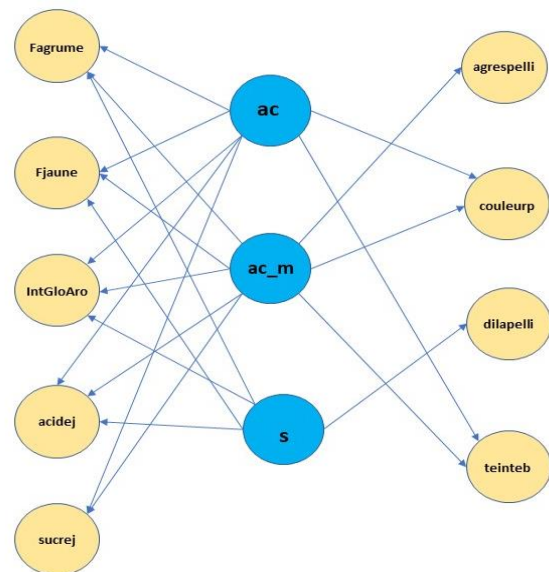


Figure 6: le tableau de corrélations entre indicateurs physico-chimiques et sensoriels et le réseau correspondant.

variables météo sont liées à chaque variable sensorielle. Le réseau complet est montré en Figure 9.

Les résultats Comme dans le cas précédent, nous calculons le RMSE et le pourcentage de prédictions dont la distance de la valeur observée est inférieure respectivement à 0.5 et 0.25. Comme dans le cas précédent, les résultats à 0.5 sont plus représentatifs. Le meilleur résultat est obtenu pour la variable « dilacération de la pellicule » (Dilapelli), pour laquelle plus du 90% des résultats ont une erreur inférieure à 0.5 dans les deux pas de temps. La prédiction de IntGloAro à t=1 est meilleure que dans le cas précédent, ainsi que pour l’acidité du jus (Acidej). Dans les autres cas, le fait d’ajouter les liens avec les indicateurs physico-chimiques n’améliore pas les résultats du modèle et, au contraire, on observe de très mauvais résultats. Les liens sont donc à revoir, mais le phénomène observé peut aussi être du au fait que nous ne disposons pas de suffisamment de données pour un tel tableau de probabilités

Variable	RMSE	% pred-obs < 0.5	% pred-obs < 0.25
Couleurpepin_1	0.57	56.8	27.3
Dilapelli_1	0.3	90.9	63.6
Teinteb_1	1.37	25	9.1
Acidej_1	0.54	63.6	29.5
Sucrej_1	0.74	36.4	20.5
IntGloAro_1	0.45	75	38.6
Fjaune_1	0.99	20.4	6.8
FAgrume_1	0.91	31.8	9.1
Couleurpepin_2	0.78	38.6	9.1
Dilapelli_2	0.31	90.9	56
Teinteb_2	1.52	22.7	9.1
Acidej_2	0.68	59.1	36.4
Sucrej_2	0.92	18.2	9.1
IntGloAro_2	0.51	68.2	36.4
Fjaune_2	0.76	34.09	13.4
FAgrume_2	0.92	15.9	2.3

2.2 MODELE DE LOGIQUE FLOUE

L'objectif du travail de recueil d'expertise mené lors du projet était de construire un algorithme permettant d'apporter une aide aux viticulteurs dans leur décision de vendanger ou non. Cet algorithme a été développé pour prédire, à partir des conditions climatologiques le moment à partir duquel le raisin pourrait être vendangé pour conduire à des vins cibles « idéaux » pour chaque producteur. Nous avons travaillé avec deux viticulteurs pour mettre au point la méthode et les approches mathématiques et algorithmiques associées (figure 7). Ainsi à partir des prédictions proposées par le modèle de réseau Bayésien Dynamique développé pour le chenin (DBN, climat vers variables physicochimiques et sensorielles), certaines variables prédites sont prises en compte par un système expert qui va calculer si à date prévue de vendange, compte tenu de leurs valeurs, le vin risque d'être dans la cible définie par l'expert (intégrée au Système Expert LF). Selon la réponse du système, le viticulteur peut avoir une alerte d'une vigilance potentielle par rapport à la date prévue de la vendange ou si la prédiction concerne des pas de temps plus en amont de la vendange, des correctifs à appliquer. Les correctifs ne sont pas développés dans notre approche pour l'instant mais pourraient l'être à terme.

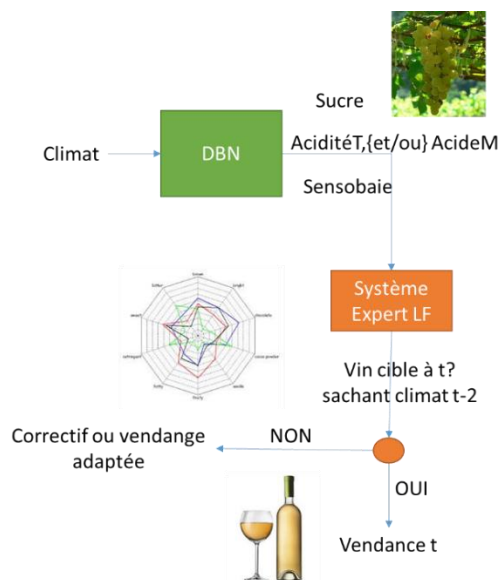


Figure 7: Algorithme de prédiction de la potentialité vin à partir d'information lors de la maturation des baies de raisin. Couplage DBN et Système expert Flou

Après avoir formalisé les savoirs élicités avec nos deux experts du Val de Loire, nous avons développé et testé la prédiction de l'algorithme flou (SE) pour des mesures réalisées sur le raisin la semaine de la vendange à partir 1/des données physicochimiques mesurées et 2/prédites par le DBN. Les parcelles utilisées pour la validation sont réparties entre l'Anjou et la Touraine avec 12 parcelles vendangées en 2017 et 8 parcelles vendangées en 2016 et donc des conditions climatologiques différentes.

Des prédictions pourraient être également proposées tout au long des mesures réalisées sur le terrain depuis la véraison afin de comparer la trajectoire potentielle du vin dans le temps de maturation à celle qui semble idéale pour la propriété concernée (ce que nous appelons cible mentale de l'expert). Ces prédictions sont réalisées pour conditions standard par ailleurs pour la propriété concernée (fermentation, mélanges, etc...)

2.2.1 Structure

Plus particulièrement deux modèles intégrant ces savoirs au travers de fonctions floues ont été testés. Un modèle représentant l'élicitation et formalisation des connaissances réalisée avec l'expert 1c (figure 8) qui prend en compte le Sucre, l'Acidité Totale (Acidité T) et l'Acide Malique (AcideM) mesuré sur le

raisin au moment de la vendange ou en cours de maturation. Un deuxième modèle (figure 9) formalisant des savoirs de l'expert 2, prenant en compte le Sucre, l'Acidité totale (AciditéT) et la Force pour détacher le pédicelle (ForcePedi).



Figure 8: Entrées/sorties du modèle SE Expert 1



Figure 9: Entrées/sorties du modèle SE Expert 2

2.2.2 Résultats

Validation de l'approche sur données mesurées

Les algorithmes sont testés indépendamment du DBN dans un premier temps sur 20 points qui correspondent aux vins évalués par l'expert 1, pour des parcelles contenues dans la base de données pour lesquelles nous avons toutes les données disponibles pour les vendanges 2016 et 2017. Ainsi sur la base des données concernant les variables d'entrées manipulées pour chaque modèle expert (par exemple Sucre, AciditéT et ForcePedi pour le SE Expert 2) une prédiction de la classe potentielle de vin est proposée par le système expert SE expert 1 et SE expert 2 (resp. figure 10 et 11) concernant la classe potentielle de vin prédite sachant les variables d'entrée mesurées au temps t juste avant vendange pour différents points d'expériences (20 points parcelles/années). Les variables d'entrées sont celles mesurées au temps t juste avant la vendange pour ces prédictions. Les classes considérées sont : Classe 1= loin cible ; classe 2= moyen cible et classe 3=cible. On compare la classe experte mesurée sur le vin mis en bouteille et dégusté. En fixant l'erreur absolue inférieure ou égale à 0.5 (seuil de perception sensorielle minimum) on obtient des pourcentages de bonne classification de 45% pour le SE EXPERT 1 et 55% pour le SE EXPERT 2.

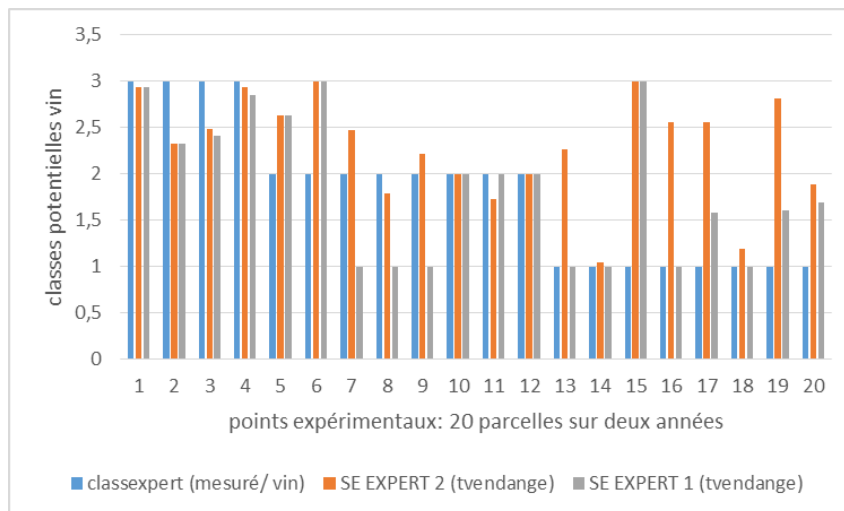


Figure 10: Prédiction des SE expert 1 et SE expert 2 concernant la classe potentielle de vin prédite sachant les variables d'entrée mesurées au temps t juste avant vendange pour différents points d'expériences (20 points parcelles/années).

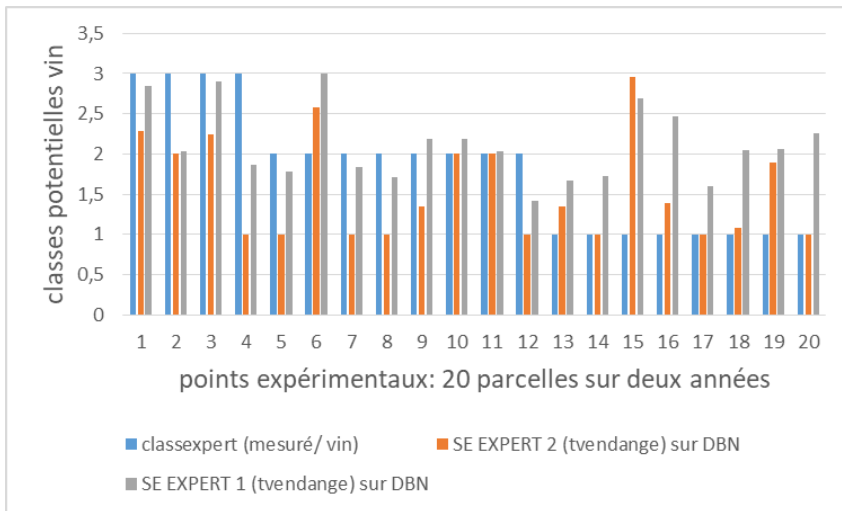


Figure 11: Prédiction des SE expert 1 et SE expert 2 concernant la classe potentielle de vin prédite sachant les variables d'entrée **prédites** au temps t juste avant vendange pour différents points d'expériences (20 points parcelles/années).

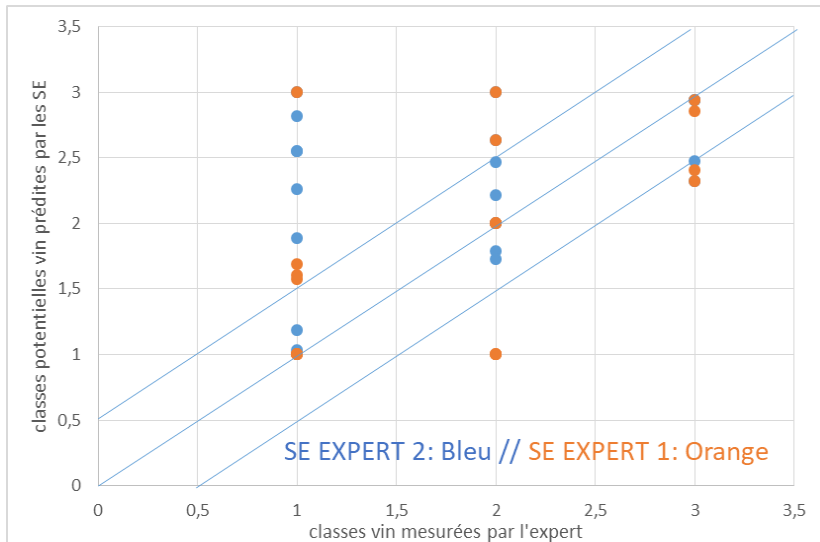


Figure 12: Prédiction des SE EXPERT 1 et SE EXPERT 2 concernant la potentialité du vin sachant les variables d'entrée **mesurées** au temps t juste avant vendange pour différents points d'expériences ainsi que les potentialités du vin sachant les variables d'entrée du SE prédites par le DBN. Nuage de points représentant la classe experte mesurée versus les classes expertes prédites pour les deux Systèmes Experts. Classe 1= loin cible ; classe 2= moyen cible et classe 3=cible. La cible est une cible mentale définie par l'EXPERT 1.

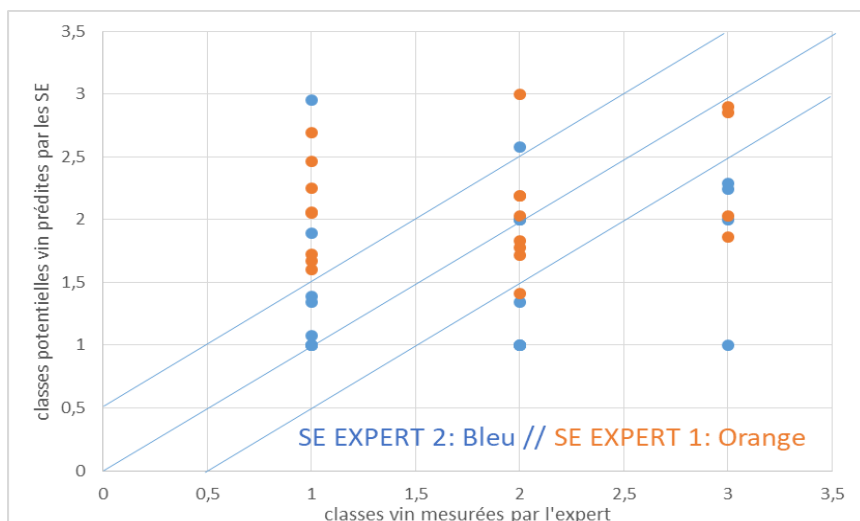


Figure 13: Prédiction des SE EXPERT 1 et SE EXPERT 2 concernant la potentialité du vin sachant les variables d'entrée **prédites par le DBN** au temps t juste avant vendange pour différents points d'expériences. Nuage de points représentant la classe experte mesurée versus les classes expertes prédites pour les deux Systèmes Experts. Classe 1= loin cible ; classe 2= moyen cible et classe 3=cible. La cible est une cible mentale définie par l'EXPERT 1.

Si on regarde plus en détails les prédictions qui ont été faite par le SE EXPERT 1 (figure 12) pattern décrit par l'Expert 1, on voit que les résultats sont globalement très pertinents avec seulement 5 points sur 20 qui ont un décalage d'évaluation supérieur ou égal à une classe et pour toutes les autres prédictions elles sont correctes dans un tube d'incertitude maximum de 0.69 c'est-à-dire un écart maximum de prédiction avec la classe réelle de 0.69 de valeur absolue (par exemple 1.69 au lieu de 1), ce qui est bon compte tenu de l'imprécision de mesure. Ce qui fait un taux de bonne classification de 75%. Les erreurs sont en majorité des erreurs de sous classements en 1 de vins qualifiés par l'expert en 2 (3 points sur les 5 points d'erreur), sachant que cette mesure sensorielle était influencée par le millésime et l'ensemble de la qualité des vins mesurés au cours de chacun des millésimes.

Pour les prédictions faites avec le SE EXPERT 2, dans lequel les patterns étaient décrits par l'expert 2, pour des vins évalués selon la cible mentale de l'expert 1, les différences sont un peu plus marquées (70% de bonne classification à 0.69, soit 6 erreurs) essentiellement sur la qualification des vins « loin de la cible ». Il est à noter que les cibles des vins sont différentes ainsi que les fermentations qui ne sont pas faites dans les mêmes conditions pour cet expert. Des réglages devraient donc être réalisés pour intégrer la cible mentale de l'expert 2. Ainsi plus de points pour le SE EXPERT 1 se retrouvent entre 0.5 et 0.69 d'erreur que pour le SE EXPERT 2 dans les points mal classifiés ce qui explique ce passage de 45% à 75% alors que pour le SE EXPERT 2, l'amélioration des résultats est moins forte (passage de 55% à 70%). Cependant les résultats restent très pertinents et cohérents entre les experts, malgré des cibles symboliques manipulées mentalement par chaque expert, qui diffèrent en fonction du résultat attendu pour chaque propriété et typicité.

Validation de l'approche sur données prédites par le DBN Dans cette configuration, Systèmes experts et DBN sont couplés. Ainsi les prédictions du DBN (sucre, acidité totale, acide malique, Force pour détacher le pédicelle) pour le temps t de maturation juste avant vendange sont les entrées des systèmes experts (cf. figure 7). Les prédictions du DBN sont réalisées à partir de données climatologiques mesurées à l'instant $t-1$ (t étant la vendange), c'est-à-dire 1 semaine avant la vendange. Les résultats sont présentés (figures 11 et 13). Dans cette configuration, le taux de bonne prédiction pour les deux SE est plus faible et de 40% pour un seuil inférieur ou égal à 0.5. Les erreurs au-delà du seuil de 0.5 sont contrairement à précédemment plus étalées et généralement supérieures à 0.69. Dans ce cas en effet deux erreurs de prédiction se cumulent : l'erreur lié au modèle et l'erreur liée au système expert. Une

optimisation des modèles couplés pourrait être réalisée pour diminuer cette erreur globale de prédiction.

Proposition d'un outil d'aide à la décision qui s'appuie sur les SE proposés

Ce modèle une fois optimisé est conçu pour pouvoir apporter une aide à la décision dès le début des mesures réalisées en cours de maturation. Les simulations proposées dans cette partie sont réalisées avec les valeurs des variables d'entrée mesurées. Ainsi nous présentons figure 14 les prédictions du modèle SE EXPERT 1 appliquées au suivi de 4 parcelles en cours de maturation. On peut observer la trajectoire que chacune de ces parcelles a dans le temps au regard de la potentialité vin attendue par l'expert 1. On peut voir notamment que du prélèvement 1 (début de maturation) au prélèvement 6 (prélèvement avant vendange), l'évolution de la trajectoire du raisin en termes de potentialité vin se rapproche pour 3 parcelles sur 4 de la cible mentale considérée comme optimale (valeur 3) par l'EXPERT 1. Par exemple, pour la parcelle CAD 57, on voit qu'elle développe ses potentialités du premier jour de suivi ou le potentiel vin serait loin de la cible si vendangé compte tenu de raisins non mûres, jusqu'à développer toutes ses potentialités pour au dernier prélèvement (le numéro 6) être dans le bon état pour développer un vin optimal, toutes choses égales par ailleurs. C'est une indication de bonne évolution et de moment idéal pour la vendange. Pour la parcelle ROC42 on voit que la trajectoire est assez différente avec un passage en classe 2 plus tardif (prélèvement 5 et non 4) et un rattrapage de la cible en fin de maturation (prélèvements 5 et 6) dû à des actions entreprises sur le terrain pour éviter une dérive de la maturation et donc une vendange plus tardive ou un vin plus loin de la cible attendu. Pour NOC52, la trajectoire semble idéale et proche de CAD57 mais en fin de maturation des éléments extérieurs ou de conduite de la vigne ont dû défavoriser l'atteinte de la potentialité attendue par l'expert 1. Il se peut également que cette parcelle n'appartenant pas à cet expert soit conduite pour atteindre une cible mentale experte différente. Enfin la dernière parcelle testée est une parcelle qui n'exprime pas sa potentialité vin au sens attendu par l'expert 1. Cela ensuite peut s'expliquer par différents facteurs qui pourraient être proposés par le système expert dans un deuxième temps.

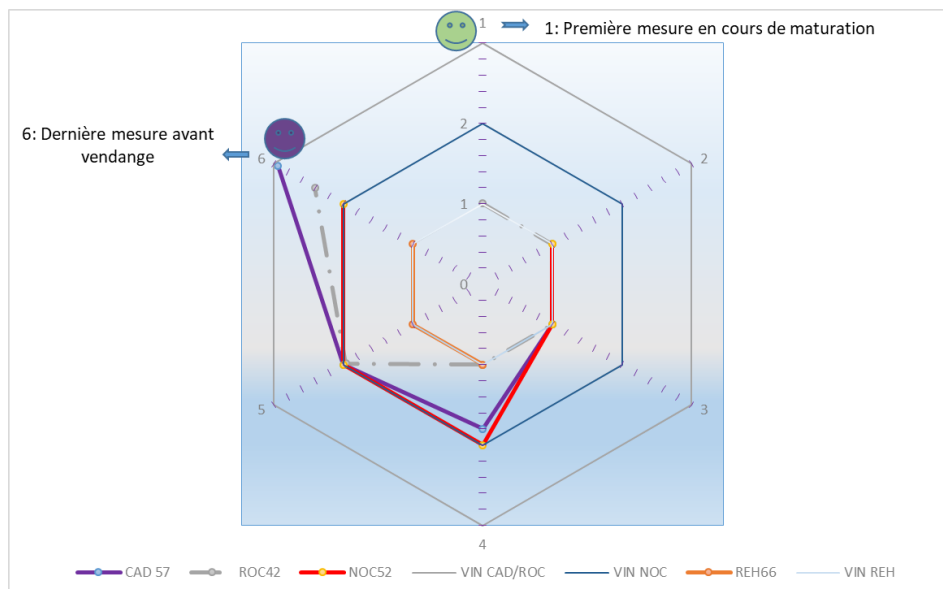


Figure 14: Trajectoires de maturation depuis le début de maturation noté 1 jusqu'à la fin de la maturation (mesure avant vendange) noté 6, évaluées par le système expert SE EXPERT 1 en termes de potentialité vin. Analyse sur 4 parcelles faisant parties des expériences analysées dans cette étude: CAD57, ROC42, REH66, NOC52. L'évaluation des vins de ces parcelles réalisées par l'expert 1, donc dépendante de la cible mentale attendue par cet expert, est présentée également (VIN NOC, VIN CAD/ROC, VIN REH). L'échelle du radar est de 0 à 3 avec 1 pour un vin loin de la cible attendue par l'expert 1 jusqu'à 3 pour un vin à la cible attendu par l'expert 1.

2.2.3 Discussion

Nous avons proposé dans cette étude un système expert d'aide à la décision qui présente sous la forme de « potentialité vin », adapté à la cible mentale d'un expert vigneron, une trajectoire potentielle vin en fonction de l'état de maturation du raisin, lui-même potentiellement prédit par un modèle DBN à partir des conditions climatologiques. Ce système pourrait être utilisé de manière prédictive. Il indiquerait quelques semaines en avance si un écart entre trajectoire réelle et trajectoire idéale se dessine. Des actions sur le terrain à certains moments de la maturation pourraient alors être proposées afin d'accompagner le vigneron dans sa conduite par anticipation de la vigne et partager les pratiques correctives de chacun dans ses vignes. Un travail doit se poursuivre pour intégrer les cibles mentales des différents acteurs et optimiser le couplage entre le modèle DBN et les systèmes experts.

3 CONCLUSIONS

Nous avons d'abord développé un outil de discrétisation semi-automatique pour les variables physico-chimiques considérées, intégrant expertise humaine et optimisation algorithmique afin d'améliorer le processus de discrétisation préliminaire au DBN. Nous avons ensuite construit le DBN pour ces variables, qui a donné des résultats prometteurs. Nous avons ensuite élargi l'étude aux indicateurs sensoriels du raisin et nous avons développé deux types de DBN : un réseau qui lie météo et variables sensorielles et un deuxième qui intègre également les variables physico-chimiques, en les utilisant comme des variables observées pour la prédiction des variables sensorielles. Nous avons observé que, ajouter les variables physico-chimiques n'améliore les résultats que dans certains cas. Nous avons aussi récolté l'expertise humaine, grâce au travail avec Monsieur Papin et Monsieur Vigneau, ce qui nous a permis de développer un ultérieur modèle SE qui lie les indicateurs de maturité au potentiel qualitatif du vin, lié à la cible de chaque vigneron. Le SE, jusqu'à présent testé sur les valeurs observées a été conçu de façon à pouvoir le lier directement aux prédictions du DBN, ce qui permettrait à terme d'avoir un modèle enchainé de la météo au potentiel du vin.

4 REMERCIEMENTS

Ce projet a été financé par le projet CASDAR 2015 porté par l'IFV sur le "Développement et transfert d'un outil d'aide à la décision appliqué à la maturation des baies de raisins rouge et blanc".

Nous remercions l'IFV, InterLoire et l'ESA d'Angers USC1422 GRAPPE et tous les partenaires du projet CASDAR pour les données qui nous ont été transmises pour réaliser notre travail de modélisation.

Nous remercions nos experts Messieurs Papin et Vigneau-Chevreau pour leur participation active et leur disponibilité.

5 BIBLIOGRAPHIE

- Barbeau, G. B.-H. (2003). Comportement de quatre cépages rouges du Val de Loire en fonction des variables climatiques. *J Int Sci Vigne Vin*, 38, 35-40.
- Baudrit, C. W.-H. (2013). Parameter elicitation in probabilistic graphical models for modelling multi-scale food complex systems. *Journal of Food Engineering*, 115(1), 1-10.
- Baudrit, C., Perrot, N., Brousset, J.-M., Guillemain, H., Perret, B., Picque, D., . . . Barbeau, G. (2015). A probabilistic graphical model for describing the grape berry maturity. *Computers and Electronics in Agriculture*, 118, 124-135.
- Ben Ghozlen, N. N. (2010). Assessment of grapevine maturity using a new portable sensor: Non-destructive quantification of anthocyanins. *Journal International des Sciences de la Vigne et du Vin*, Vol. 44, pp. 1-8.
- Boukhelifa, N., Tonda, A., Trelea I.-C., Perrot, N., & Lutton, E. (2017). Interactive Knowledge Integration in Modelling for Food Sustainability: Challenges and Prospects. *ACM CHI Workshop on Designing Sustainable Food Systems*.
- Brousset, J. (2009). *Caractérisation multifactorielle et modélisation de la maturité de baies de Cabernet Franc en moyenne vallée de Loire*. Rapport InterLoire.
- Claverie, M. P. (2008). Modélisation statistique de la qualité en viticulture par la méthode PLS Spline. *VIIe Congrès International des terroirs viticoles*.
- Geisser, S. (1993). *Predictive Inference*. New York: Chapman and Hall.
- Geraudie, V. R. (2010). Développement d'un appareil permettant de prédire la maturité du raisin par spectroscopie proche infra-rouge(PIR). *Revue Française d'Oenologie*, 240, 2-8.
- Goelzer, A. C. (2009). Simulation and optimization software for alcoholic fermentation in winemaking conditions. *Food Control*, 20(7), 635-642.
- Murphy, K. P. (2002). *Dynamic bayesian networks: representation, inference and learning*. PhD Dissertation. Berkeley: University of California.
- Perrot N., T. I. (2011). Modelling and analysis of complex food systems: state of the art and new trends. *Trends in Food Science & Technology*, 22(6), 304-314.
- Perrot, N. B. (2015). A Decision support system coupling fuzzy logic and probabilistic graphical approaches for the agri-food industry: prediction of grape berry maturity. *PLoS one*, 10(7), e0134373.
- Raynal, M. D. (2010). Epicure, a geographic information decision support system risk assessment of downy and powdery mildew epidemics in Bordeaux vineyards. . *Sixth international workshop on the grapevine downy and powdery mildew* (p. 144-146). Bordeaux: INRA-ISV.
- Riou, C. (1994). *Le déterminisme climatique de la maturation du raisin: application au zonage de la teneur en sucre dans la Communauté Européenne*. Office des Publications Officielles des Communautés Européennes.
- Vermersh, P. (2006). *L'entretien d'explicitation*. Paris: ESF.