



---

# **T H È S E**

pour l'obtention du titre de

**Docteur du Centre International d'Etudes Supérieures en  
Sciences Agronomiques de Montpellier**

**Ecole Doctorale : Sciences des Procédés Sciences des Aliments**

---

**Développement d'une méthode de comparaison de  
données asynchrones en vue de la formalisation d'un  
raisonnement par analogies. Application à l'aide à la  
décision en viticulture**

---

présentée et soutenue publiquement le 03 juillet 2012 par

**Séverine DUPIN**

## **JURY**

Président :	Dr. Christian GARY,	INRA Montpellier
Rapporteurs :	Pr. Gilbert GRENIER, Pr. J. Bernardo ROYO DIAZ,	Bordeaux Sciences Agro Univ. Publique de Navarre (Esp.)
Directeur de thèse :	Pr. Jean-Michel ROGER,	Irstea Montpellier
Examineurs :	Dr. Thibaut SCHOLASCH, Pr. Kees VAN LEEUWEN,	Fruition Sciences Inc. Bordeaux Sciences Agro
Invités :	Pr. Bruno TISSEYRE,	Montpellier SupAgro

*A toi qui m'a vue comme ton propre enfant  
A toi qui me rassurait petite  
A toi qui savait reprendre ta place pour laisser la sienne à mon père  
A toi qui m'a orientée vers la biologie  
A toi que je ne peux pas oublier  
Après 10 ans, me voilà au bout du chemin  
La route fut longue  
Mais ton enseignement m'a appris à ne jamais renoncer.*

*A toi avec qui j'ai partagé mon amour du vin  
Pour notre affection commune pour la Syrah  
Parce que tu m'appelais ton docteur en alcoologie  
Parce que je n'aurais pas eu le temps de te montrer mes résultats  
Apprécie de là haut  
Ce que je m'apprête à décrire.*

*A tous ceux qui m'ont quitté trop tôt  
Et que j'aurai voulu voir ici plutôt que dans mes souvenirs  
Pour la force que vous m'avez donnée  
Pour les tempêtes que vous avez essuyées  
Pour les sourires que vous m'avez extirpés  
C'est aussi votre thèse.*

*Il n'y a pas de réussite facile ni d'échec définitif. Marcel Proust*

*Il faut toujours viser la lune car même en cas d'échec, on atterrit dans les étoiles. Oscar Wilde*



# Remerciements

Quand on y pense, un thésard, après avoir rédigé plus de 100 pages sur sa thèse, ne devrait pas avoir de difficultés à rédiger les quelques lignes de remerciements qui vont suivre. Et pourtant ... Allez, je tente quand même l'expérience.

Je souhaite tout d'abord remercier l'ensemble du personnel de l'UMR ITAP que ce soit à Irstea ou sur le site de Montpellier SupAgro pour leur accueil et leur soutien scientifique au long de mes 4 années de présence à leur côté. Qu'ils aient travaillé avec moi ou non, ils ont toujours été intéressés par mon travail et m'ont permis d'avoir des réflexions de groupe qui ont abouti à des rectifications que quelques uns de mes résultats. Au sein de cette équipe, je remercie notamment son directeur actuel Tewfik Sari et sa directrice lors de mon arrivée dans l'équipe Véronique Bellon-Maurel. Je vous remercie de m'avoir permis de réaliser cette thèse au sein de votre équipe et de m'avoir donné accès à l'encadrement dont j'ai pu bénéficier.

Je remercie aussi particulièrement toute l'équipe administrative de l'unité pour leur aide à propos des différents dossiers à remplir notamment lors de mon départ en Californie. Michèle, Nicolas, Danielle, Joëlle, vous m'avez épargné bien souvent de longues heures perdues pour remplir ou faire remplir ces documents, je m'en souviendrais encore longuement.

Je pense aussi évidemment à ceux qui m'ont fait confiance en premier lieu pour m'offrir cette opportunité de thèse : Bruno Tisseyre, Alexia Gobrecht et Jean-Michel Roger.

Alexia m'a dit un jour : « Bruno avait confiance en toi pour cette thèse et vu qu'on fait confiance à Bruno les yeux fermés... alors te voilà. » Comment ne pas repenser à cette phrase au moment où je referme mon manuscrit. Merci Bruno d'avoir accepté de me confier ce sujet qui te tenait tant à cœur. Et puis merci pour toutes les étapes que tu m'as fait franchir. Je crois que je ne t'aurai épargné aucun maux, mais tu as toujours su me pousser pour que je continue d'avancer. Entre les jours où je perdais courage car je ne voyais pas le bout du tunnel, les jours où j'oubliais que je travaillais sur une thèse méthodologique et me remettait à faire de la viticulture « de base », les jours où il fallait me secouer pour que l'écriture d'article avance, les après-midi passés à me forcer à couper mes phrases « à la proust » pour qu'elles soient courtes et faciles à traduire... je me rends compte de la patience et des nerfs qu'il t'a fallu pour m'encadrer au long de tout ce travail. Je ne garde que des bons souvenirs de nos travaux. Et pourtant quand tu me remontais les bretelles, tu ne le faisais pas avec des pincettes. Ce que j'ai accompli, c'est grâce à l'opportunité que tu m'as confiée. Alors si aujourd'hui, je cherche à tout prix à faire du vin par ordinateur, tu y es forcément pour quelque chose ...

Bon mais c'est pas tout, mais j'avais deux directeurs de thèse... Alors oui bien sur, Jean-Michel, comment te remercier pour ces trois années. Tu as récupéré une Ingénieure

qui avait peur de mal faire devant toi, tu m'as transformée en docteur qui est fière de ce qu'elle a accompli sous ta tutelle. Ma thèse t'amusait car elle te permettait de revenir aux travaux que tu effectuais il y a quelques années. Pour moi, au départ, c'était moins amusant car il me fallait te suivre dans tes disgressions. Mais, tu as réussi à me faire aimer les analyses de données .... et surtout leurs résultats. Je me souviens de mes débuts chaotiques, du jour où d'épuisement j'ai craqué devant toi et de ta réaction calme et posée. Je ne sais pas si des thésards comme moi tu en as vu souvent ... Merci pour tes idées lumineuses, même si, pour toutes les enregistrer au cours de mes comités de thèse, il m'aurait fallu un magnétophone. Merci d'avoir réussi à mettre ta sauce chimométrie dans mes problèmes d'aide à la décision pour la viticulture...

Enfin, dans cette équipe de choc, il reste celle qui dirigeait le projet Vinnotec. Alexia, je crois que je ne pourrais pas compter sur mes doigts le nombre de fois où tu m'as dit que nous t'avions largué d'un point de vue scientifique et que si tu venais au comité de thèse c'était pour jouer le rôle de Candide. Justement grâce à ce rôle, tu m'as bien des fois coupé les ailes en plein vol, mais tu m'as permis ainsi d'éviter des erreurs ou de m'extasier de résultats pas très innovants. Merci encore.

Je ne peux pas oublier de remercier les autres encadrants de cette thèse, membres du comité de thèse et convoqués à des réunions pleines de rebondissements : Hernan Ojeda, Olivier Naud, Bernard Palagos, Kees van Leeuwen. Grâce aux conseils de chacun, j'ai pu avancer sur cette thèse. Merci Hernan pour ta connaissance en viticulture et de Pech Rouge... les conclusions des analyses auraient été bien moins efficaces sans toi. Merci Olivier pour nos séances du vendredi soir 19h dont je repartais toujours avec plein d'idées de développement. Merci Bernard pour les séances d'analyses de données et les explications sur les différentes analyses descriptives. Enfin merci Kees d'avoir accepté de continuer à m'encadrer après mon diplôme d'Ingénieur et de nous avoir apporté vos connaissances expertes sur la vigne et les cinétiques de maturation.

Un grand merci aussi à l'ensemble des personnes ayant accepté d'être membres du jury de cette thèse et notamment au président du jury Christian Gary. Evaluer une thèse multidisciplinaire n'est jamais de tout repos. Merci pour vos remarques pertinentes, pour les questions posées qui m'amènent encore aujourd'hui à réfléchir sur les perspectives d'utilisation de la thèse. Et merci d'avoir tendu l'oreille en ce 3 juillet pour m'écouter malgré mon extinction de voix.

Enfin, je tiens à remercier l'ensemble des employés de la station expérimentale INRA Pech Rouge pour le temps passé à regrouper les données qui m'ont servi pour « jouer ».

Partons sur la côte pacifique, merci à tous les employés de l'entreprise Fruition Sciences pour leur accueil durant l'été 2011 : Virgine, Sébastien, Thibaut, Ana et Benjamin. Merci pour l'opportunité et la confiance que vous avez eue lors de ma présence sur place. Merci aussi Thibaut pour toutes les discussions constructives que nous avons eues face à mon travail... et que nous continuerons d'avoir compte tenu des nombreuses questions que tu as posé à propos de mon manuscrit et auxquelles je n'aurais jamais fini de répondre si je veux être vraiment précise.

Sur un registre un peu moins officiel, je souhaite plus que tout remercier mes parents et mes frères et sœurs pour leur soutien pendant cette « longue traversé du désert ». Si jamais je perdais confiance en moi, vous ne l'aviez pas perdue et vous me poussiez à reprendre courage. Lulu, ma petite Lulu, je t'ai rarement vue aussi fière que ce 3 juillet... Et je me souviens de ta remarque à propos de mon résumé de thèse : « Waouh, c'est bien la première fois que je comprends ce que tu fais dans ta thèse... ». Ca fait sourire, mais ça remonte aussi beaucoup le moral. Bon j'ai vu la fierté dans les yeux de mes parents et de Guillaume aussi. Je me sentais sur un petit nuage de vous voir aussi fiers. C'est pas tous les jours que votre famille vous applaudit debout après que le verdict soit tombé... Merci d'avoir été présents ce jour là malgré vos emplois du temps compliqués et d'avoir fait le trajet depuis Paris pour m'écouter et me soutenir. Papa, Maman, Mylène, Guillaume, Lulu merci pour les nombreuses relectures et pour vos questions innocentes qui permettent de rectifier bien des erreurs.

Enfin, j'ai une pensée émue pour toi, Jean-Pierre. Tu savais que je devrais faire de la biologie, avais tu imaginé où j'en arriverais et ce que je m'apprêtais à faire de ma vie lorsque tu m'as parlé de la prépa bio de Janson il y a 10 ans? Bien sur j'ai toujours un pincement au cœur quand je me souviens que tu n'es plus là pour voir ce que j'ai accompli, mais j'ose espérer qu'il est un endroit d'où tu peux observer tout ça.

Merci aussi à mes deux grands frères du Cemagref (bon d'accord d'Irstea) : Florent et Bertrand. La petite, elle a réussi hein... Merci pour vos notions d'informatique qui m'ont aidée à m'en sortir et pour les délires passés au bureau, comme la traduction de ma problématique de thèse en jeu vidéo, qui nous permettaient d'avancer à tous les trois. Oui Flo, je pense forcément à toi. Le labo a été bien vide le jour où tu es parti pour le LIRMM. Mais nous continuions de discuter. Et bien que j'en ai peut-être plus appris sur la génération de colonnes que toi sur le vin, tu m'as apporté une aide essentielle avec des conseils....

Enfin, je ne pouvais pas conclure ces remerciements sans toi, miss Axelle. Imaginez vous une amie qui passe des heures avec vous à discuter de votre thèse, qui fait oui oui au téléphone car la seule chose dont vous avez besoin c'est de vous sentir écoutée pour réfléchir, qui débarque en urgence chez vous un week-end parce qu'elle sent que vous commencez à péter les plombs, qui relit toute votre thèse pour chasser ces malheureuses fautes d'orthographe, qui fait votre taxi quand vous vous cassez le pied et vous brulez le bras simultanément pour que vous puissiez continuer d'aller travailler, qui prend 15 jours de congés pour venir vous préparer à votre soutenance et y assister et qui sait aussi passer d'excellents moments de détente avec vous (d'ailleurs on a pas encore atteint la Lune hein!)... Une amie comme ça, je sais qu'il n'en existe qu'une... et j'ai eu la chance de la trouver. Alors merci Axelle pour cette thèse qui est aussi à moitié la tienne... Une très très grand merci !





# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Projet Vinnotec : développement des nouvelles technologies pour la viticulture et l'œnologie . . . . .	4
1.2	Evolution récente de la viticulture : une augmentation des données de suivi enregistrées . . . . .	6
1.3	L'utilisation des données de traçabilité . . . . .	8
1.3.1	Prise de décision par analogies entre parcelles et cycles de production	8
1.3.2	Définition de l'objectif pratique de la thèse : vers la création d'un outil d'aide à la décision facilitant les analogies entre parcelles . .	10
1.4	Organisation du mémoire de thèse . . . . .	11
<b>2</b>	<b>Analyse de la problématique</b>	<b>13</b>
2.1	Est-il possible de comparer les parcelles et les millésimes à partir des données brutes des bases de données de traçabilité ? . . . . .	14
2.1.1	Le système de production viticole : un système temporisé et cyclique	14
2.1.2	Comparaison de parcelles et de millésimes à partir des données brutes des bases de données de traçabilité . . . . .	15
2.2	Comment rendre comparables les données de traçabilité ? . . . . .	18
2.2.1	Etat de l'art . . . . .	18
2.2.2	Application et limites de l'utilisation de ces propositions pour la comparaison de données de traçabilité viticoles . . . . .	20
2.3	Collaboration entre données de traçabilité et expertise pour créer un espace commun pour la comparaison de parcelles et / ou de millésimes . .	21
<b>3</b>	<b>Approche théorique</b>	<b>23</b>
3.1	Définitions . . . . .	24
3.1.1	Mode d'étude d'un phénomène temporel . . . . .	24
3.1.2	Description d'une base de données spatio-temporelles . . . . .	25
3.2	Méthode de comparaison de plusieurs individus. . . . .	27
3.2.1	Rappel du problème soulevé par la comparaison d'individus. . . .	27
3.2.2	Approche méthodologique proposée . . . . .	27
3.2.3	Utilisation des vecteurs de paramètres pour faciliter les comparaisons	29
3.3	Bilan : Schéma récapitulatif des étapes de retraitement des données . . .	30

<b>4</b>	<b>Application 1 : Etude du climat d'un domaine sur plusieurs années entre la floraison et la véraison</b>	<b>33</b>
4.1	Introduction : Effet des paramètres climatiques sur l'évolution de la plante entre la floraison et la véraison . . . . .	35
4.1.1	Effet de la température sur la vigne et le raisin . . . . .	36
4.1.2	Effet du régime hydrique sur la vigne . . . . .	37
4.1.3	Effet du rayonnement solaire sur la vigne . . . . .	37
4.2	Matériel et Méthodes . . . . .	38
4.2.1	Données climatiques utilisées . . . . .	38
4.2.2	Définition de la base de données d'étude . . . . .	39
4.2.3	Grandeurs climatiques utilisées . . . . .	41
4.2.4	Calcul des paramètres . . . . .	41
4.2.5	Utilisation des paramètres . . . . .	45
4.3	Résultats . . . . .	48
4.3.1	Adéquation de l'exemple à la méthode proposée . . . . .	48
4.3.2	Résultats de l'ACP réalisée en utilisant les indices bioclimatiques non pondérés . . . . .	48
4.4	Etude comparative de l'ACP avec indices bioclimatiques pondérés . . . . .	55
4.4.1	Analyse comparative du plan constitué des deux premières composantes principales . . . . .	56
4.4.2	Analyse comparative du plan constitué de la troisième et de la quatrième composante principale . . . . .	57
4.5	Conclusion : Bilan de l'étude comparative du climat entre cépages et millésimes . . . . .	59
<b>5</b>	<b>Application 2 : Modélisation de l'augmentation du pH pendant la maturation</b>	<b>61</b>
5.1	Introduction : l'évolution du pH au cours de la maturation . . . . .	62
5.2	Matériel et méthodes . . . . .	63
5.2.1	Matériel expérimental . . . . .	63
5.2.2	Définition d'un temps thermique pour chaque jour . . . . .	64
5.2.3	Modélisation paramétrique de l'augmentation du pH dans les baies de raisin . . . . .	65
5.2.4	Utilisation des modèles . . . . .	66
5.3	Résultats et discussion . . . . .	66
5.3.1	Caractérisation du mésoclimat de chacun des trois domaines choisis pour l'utilisation du modèle en extensif . . . . .	66
5.3.2	Adéquation de l'exemple aux problèmes de comparaison des séquences temporelles . . . . .	67
5.3.3	Validation d'un modèle d'augmentation du pH sur les 53 individus de la base de données . . . . .	68
5.3.4	Comparaison graphique de la cinétique de pH des trois parcelles étudiées . . . . .	71

5.4	Conclusion : Un modèle avec du potentiels pour comparer les cinétiques d'augmentation du pH dans différentes parcelles . . . . .	75
<b>6</b>	<b>Application 3 : Modélisation de l'accumulation de sucres dans les baies de raisin pendant la maturation</b>	<b>77</b>
6.1	Introduction : Evolution des sucres dans les baies de raisin . . . . .	78
6.2	Matériel et Méthodes . . . . .	78
6.2.1	Matériel expérimental . . . . .	79
6.2.2	Modélisation paramétrique de l'accumulation de sucres dans les baies	83
6.2.3	Utilisation du modèle . . . . .	84
6.3	Résultats et Discussion . . . . .	85
6.3.1	Adéquation des exemples aux problèmes de comparaison de séquences temporelles . . . . .	85
6.3.2	Validation d'un modèle d'accumulation de sucres dans les baies de raisin pour chaque individu . . . . .	86
6.3.3	Résultats des analyses descriptives de la base de données 1 après modélisation . . . . .	95
6.3.4	Résultats des analyses descriptives de la base de données 2 après modélisation . . . . .	105
6.4	Conclusion : Une méthodologie générique qui permet de comparer des individus hétérogènes . . . . .	114
<b>7</b>	<b>Conclusion et perspectives</b>	<b>117</b>
7.1	Bilan des travaux réalisés . . . . .	118
7.1.1	Une méthode de modélisation paramétrique de séquences temporelles pour permettre les comparaisons . . . . .	118
7.1.2	Trois applications différentes d'une même méthode . . . . .	119
7.1.3	Atteinte de l'objectif de la thèse et généricité de la méthode . . . . .	121
7.2	Perspectives de développement du travail proposé . . . . .	122
7.2.1	Elargissement du nombre de grandeurs modélisées, du nombre de paramètres calculés . . . . .	122
7.2.2	Analyse multivariée des individus . . . . .	124
7.2.3	Modification des modèles pour représenter au mieux les cinétiques réelles des différentes grandeurs . . . . .	125
7.2.4	Utilisation d'autres méthodes d'analyses pour comparer les individus	126
7.2.5	Calcul de distance à partir des positions des individus sur les plans factoriels des analyses de données . . . . .	127
7.2.6	Mise en place d'un système à base de cas . . . . .	127
7.2.7	Evaluation de la qualité d'ajustement des modèles en fonction du type de séquence temporelle . . . . .	127
7.2.8	Simulation de la fin d'une cinétique . . . . .	129
7.3	Utilisation du travail de recherche réalisé pour aider au pilotage de la vigne	130
	<b>Bibliographie</b>	<b>135</b>

<b>Annexes</b>	<b>147</b>
<b>A Détail du choix des valeurs seuils pour les compteurs</b>	<b>147</b>
<b>B Calcul détaillé des indices bioclimatiques entre floraison et véraison</b>	<b>149</b>
B.1 Les indices basés sur des cumuls de températures . . . . .	149
B.1.1 Cumul de températures base 10 : Les degrés jour de Winkler (Win) [98] . . . . .	149
B.1.2 Cumul de températures base 10 et sans prise en compte des températures supérieures à 19 °C : Indice de Gladstones (Glad) [42] .	149
B.1.3 Cumul de températures moyennes et maximales : indice héliothermique de Huglin (IH) [49] . . . . .	150
B.1.4 Cumul de précipitations soustrait du cumul de températures : indice thermo-pluviométrique (ITP) de Ribéreau-Gayon et Peynaud, in [11] . . . . .	150
B.2 Les indices d'évaluation de l'état hydrique . . . . .	150
B.2.1 L'indice de sécheresse (IS) [92] . . . . .	151
<b>C Analyse détaillée du plan factoriel 3-4 de l'ACP réalisée à partir des paramètres climatiques sur les 11 cépages et les 7 millésimes</b>	<b>153</b>

# Liste des acronymes

<b>BD</b>	<b>B</b> ase de <b>D</b> onnées
<b>ACP</b>	<b>A</b> nalyse en <b>C</b> omposantes <b>P</b> incipales
<b>ACM</b>	<b>A</b> nalyse des <b>C</b> orrespondances <b>M</b> ultiples
<b>AFM</b>	<b>A</b> nalyse <b>F</b> actorielle <b>M</b> ultiple
<b>CAH</b>	<b>C</b> lassification <b>A</b> scendante <b>H</b> ierarchique
<b>RMSE</b>	<b>R</b> oot <b>M</b> ean <b>S</b> quare <b>E</b> rror - erreur quadratique moyenne
<b>ETP</b>	<b>E</b> vapotranspiration <b>P</b> otentielle
<b>GDD</b>	<b>G</b> rowing <b>D</b> egree <b>D</b> ays - temps thermique

# Liste des symboles

pH	potentiel hydrogène
AT	Acidité totale
$x_j$	mesures
$t_j$	dates de mesures
$s_j$	localisations des mesures
$e_a$	estimations de mesures
$T_i$	dates représentant des bornes d'intervalles
$\Phi$	phénomène temporel
$X_g(t)$ ou $g(t)$	Modèle d'une grandeur
W	Watts
J	Joules
m	mètres
cm	centimètres
°C	degrés celcius
°Br	degrés brix
°alc	degrés alcooliques
°j	degré jour
g	grammes
L	litres
j	jours

**Résumé** - L'objectif initial de ce travail de thèse est de valoriser les informations relatives au suivi temporel de la vigne, du raisin et de l'environnement de la plante et enregistrées dans des bases de données (BD) de traçabilité pour permettre la comparaison entre parcelles et millésimes, en vue de décisions par analogies.

Les travaux réalisés durant cette thèse ont permis de proposer une méthode de transformation qui permet de représenter des ensembles de données asynchrones dans un espace commun afin de les comparer. Cette méthode s'appuie sur l'expertise du système de production. Dans ce travail de thèse, cette méthode a été appliquée à la comparaison de couples parcelle  $\times$  millésime.

L'expertise du système de production viticole permet, dans une première phase, de définir (i) la forme générale de la cinétique d'évolution de grandeurs de mesures évaluées sur la vigne, le raisin ou l'environnement de la plante et (ii) l'effet du climat sur la plante. Cette expertise est utilisée, dans une seconde phase, pour proposer des modèles paramétriques de l'évolution de chaque grandeur. Les données de suivi de chaque couple parcelle  $\times$  millésime permettent d'ajuster les paramètres du modèle. Un vecteur de paramètres est défini par couple parcelle  $\times$  millésime. Ce vecteur représente l'espace commun qui rend les couples parcelle  $\times$  millésime comparables. Deux stratégies de comparaison sont alors possibles : (i) les comparaisons sont réalisées à partir des paramètres (méthode intensive), ou (ii) les comparaisons sont réalisées à partir de l'estimation de la valeur de la grandeur pour chaque couple parcelle  $\times$  millésime et chaque unité de temps, commune à tous les couples (méthode extensive).

Cette méthode a été appliquée à trois exemples différents.

Dans une première application, les climats de différents millésimes intervenus sur différents cépages, entre la floraison et la véraison, ont été comparés entre eux après modélisation des grandeurs de mesure climatiques, à l'aide de modèles très simples.

Dans une seconde puis une troisième application, la cinétique d'augmentation du pH et d'accumulation des sucres dans les baies de raisin pendant la maturation a été modélisée sous la forme d'une sigmoïde. Les comparaisons ont ensuite été réalisées en travaillant sur (i) la courbe représentative de chaque cinétique (pH), (ii) les paramètres du modèle (sucres) et (iii) une estimation journalière de la concentration en sucres dans les baies.

Les bases de données utilisées dans ces applications proviennent de deux régions très différentes. Des données issues du suivi de la station expérimentale INRA Pech Rouge en Languedoc-Roussillon, dans le sud de la France, ont été utilisées pour l'application 1 et une partie de l'application 3. Des données de suivi de différents domaines de la Napa Valley en Californie ont servi pour l'application 2 et une partie de l'application 3.

Le changement d'espace de représentation des données apporte une connaissance nouvelle pour décrire les individus et les phénomènes temporels de la vigne. Cette connaissance pourrait permettre de formaliser un raisonnement par analogies utilisant et valorisant l'expérience passée pour la gestion du millésime en cours.

**Mots-clés** - *Vitis Vinifera L.* - Modélisation - Expertise - Comparaisons - Analogies - Analyses de données

S verine DUPIN.

Irstea Montpellier - UMR Information, Technologies, Analyse environnementale, Proc d s agricoles - 361 rue Jean Fran ois Breton BP5095 - 34196 Montpellier Cedex 5 (France).



---

.

---

---

# Chapitre 1

## Introduction

---

### Sommaire

<b>1.1</b>	<b>Projet Vinnotec : développement des nouvelles technologies pour la viticulture et l'œnologie . . . . .</b>	<b>4</b>
<b>1.2</b>	<b>Evolution récente de la viticulture : une augmentation des données de suivi enregistrées . . . . .</b>	<b>6</b>
<b>1.3</b>	<b>L'utilisation des données de traçabilité . . . . .</b>	<b>8</b>
1.3.1	Prise de décision par analogies entre parcelles et cycles de production . . . . .	8
1.3.2	Définition de l'objectif pratique de la thèse : vers la création d'un outil d'aide à la décision facilitant les analogies entre parcelles	10
<b>1.4</b>	<b>Organisation du mémoire de thèse . . . . .</b>	<b>11</b>

---

Ce travail de thèse s'inscrit dans le cadre d'un projet de recherche et développement collaboratif visant à développer les technologies de l'information et de la communication pour la viticulture : le projet Vinnotec. L'objectif de ce projet est de fournir de nouveaux outils aux viticulteurs. Ces outils permettront à ceux-ci de faciliter leur travail au jour le jour et à terme de renforcer leur compétitivité sur le marché mondial.

## 1.1 Projet Vinnotec : développement des nouvelles technologies pour la viticulture et l'œnologie

Le projet Vinnotec est un projet de recherche et développement collaboratif. Il a débuté en juillet 2007 et se terminera en juin 2012. L'objectif global de ce projet est le développement des technologies de l'information et de la communication pour répondre aux nouveaux enjeux de la filière viti-vinicole. Le projet a été labellisé par le pôle de compétitivité Qualiméditerranée.

Il s'agit d'un projet de recherche et développement partiellement financé par la région Languedoc-Roussillon. L'essentiel des partenaires du projet est donc situé en Languedoc-Roussillon.

13 partenaires sont associés au projet. Le consortium du projet est constitué de (i) 4 centres de recherche et de formation : l'INRA (institut national de la recherche agronomique), l'IFV (institut français de la vigne et du vin), Irstea (institut de recherche en sciences et technologies pour l'environnement et l'agriculture) anciennement Cemagref et Montpellier SupAgro (centre international d'étude supérieures en sciences agronomiques de Montpellier), (ii) 7 industriels : le Groupe ICV (institut coopératif du vin), Ondalys, Pellenc SA, Neotic, Inter Rhône, Vivelys et Sodimel et (iii) 2 caves coopératives : La cave des vigneron du Mont Tauch et la cave de Laudun et Chusclan Vignerons.

Les objectifs opérationnels du projet sont : (i) de faciliter la collaboration entre les acteurs de la recherche et de l'industrie travaillant dans le domaine des nouvelles technologies appliquées à la viticulture en Languedoc-Roussillon, (ii) de permettre le développement d'outils innovants et de méthodes d'aide à la décision au vignoble et en cave basés sur les technologies de l'information et de la communication, les innovations devant rapidement être mises sur le marché et (iii) de permettre l'adaptation, grâce aux innovations développées, du vin au marché actuel et ainsi accroître la compétitivité des entreprises.

Pour atteindre ces objectifs, le projet a été divisé en différents axes de recherche. Les axes 1 et 2 sont organisés autour du développement d'outils innovants pour la gestion du vignoble (axe 1) et des vinifications (axe 2). Un troisième axe a été proposé autour de la création de systèmes d'aide à la décision pour la gestion du vignoble et / ou des vinifications. Enfin, un quatrième axe a pour but de développer une plateforme de stockage et d'échange de données de suivi d'un domaine. Ces données seront ainsi accessibles pour le test des outils mis en place.

### **Axe 1 : Développement d'outils innovants pour la gestion du vignoble**

Différents outils ont été développés grâce à la collaboration entre recherche et industrie dans le cadre du projet Vinnotec. Ces outils sont basés sur :

- l'imagerie multispectrale pour permettre la caractérisation des parcelles par télé-détection
- la vision numérique pour évaluer le rendement et la qualité potentiels du raisin
- la spectrométrie proche infrarouge pour suivre rapidement et de manière non destructive la maturation du raisin et son potentiel aromatique
- le développement des systèmes d'information pour proposer un système innovant ultra ergonomique de saisie de l'information à la parcelle

### **Axe 2 : Développement d'outils innovants pour la gestion des vinifications**

Les outils développés dans cet axe sont basés sur :

- la vision numérique pour évaluer le potentiel qualitatif et l'aspect sanitaire du raisin à l'entrée en cave
- la spectrométrie proche infrarouge intégrée dans un capteur permettant le suivi et le contrôle des fermentations alcooliques et de l'élevage du vin

### **Axe 3 : Mise en place de systèmes d'aide à la décision**

Dans cet axe, deux types de système d'aide à la décision ont été considérés.

Le premier système d'aide à la décision est basé sur le traitement des données de traçabilité des domaines dans le but d'analyser les performances techniques des structures de production.

Le second système d'aide à la décision a pour objectif de valoriser les données de traçabilité des domaines pour la création de nouvelles règles de décision spécifiques au contexte local. Pour ces travaux, l'expertise viticole et l'expérience des viticulteurs à propos de leurs domaines ont été utilisés pour aider à la valorisation des données de traçabilité.

A terme, ces travaux devraient permettre de proposer un outil d'aide à la décision au vignoble basé sur la réalisation d'analogies entre parcelles et millésimes. Ce travail donne les premiers fondements nécessaires pour cet outil d'aide à la décision.

Cette valorisation des données de traçabilité par mobilisation de l'expertise viticole constitue le travail effectué dans cette thèse. Elle sera donc détaillée dans ce mémoire.

### **Axe 4 : Création d'une plateforme de stockage et d'échange de données pour le domaine expérimental de l'INRA Pech Rouge**

Afin de faciliter le développement et le test des outils et des systèmes d'aide à la décision créés dans le cadre du projet, une plateforme de stockage et d'échange de données a été

mise en place sur un domaine expérimental de l'INRA : la station expérimentale INRA de Pech Rouge.

Cette plateforme contiendra à terme l'historique complet de traçabilité du domaine. Les nouvelles mesures de suivi effectuées au vignoble ou en cave y seront enregistrées en temps réel.

Cette plateforme permet aux industriels d'avoir accès à un grand nombre de données pour comparer leurs résultats au suivi classique du domaine.

Les outils développés ont aussi pu être testés dans le domaine. Pour faciliter l'interprétation de leurs résultats, les industriels ont accès via cette plateforme à de nombreuses données complémentaires.

Une partie des données historiques de traçabilité du domaine, enregistrées dans le cadre du suivi classique du domaine, a aussi servi au test de la méthodologie proposée dans la thèse.

Ce projet est donc un projet regroupant un grand nombre d'acteurs. Cette thèse n'est qu'une partie du projet mais fait partie intégrante de la recherche pour le développement des technologies de l'information et de la communication à destination de la viticulture et l'œnologie.

## 1.2 Evolution récente de la viticulture : une augmentation des données de suivi enregistrées

Depuis les années 80, l'agriculture a tiré profit de l'émergence, l'amélioration et de la convergence de plusieurs technologies comme des techniques de contrôle automatique, les télécommunications, les techniques de traitement de l'information et l'amélioration et la miniaturisation des ordinateurs [100].

Ces technologies ont permis un meilleur enregistrement des données de suivi dans des bases de données. Ces changements ont aussi provoqué une augmentation constante des mesures de suivi effectuées sur les systèmes de production agricole [4, 100]. Plus récemment, le développement de capteurs portatifs de mesures sur la plante ou de réseaux de capteurs automatisés entraîne une augmentation du nombre de mesures de suivi réalisées. Plus de mesures sont réalisées au sein d'une même parcelle et la fréquence des mesures augmente [32]. Les mesures sont de plus géolocalisées, c'est à dire que la position de chaque mesure est connue de façon précise. Grâce à cette géolocalisation, les mesures de traçabilité peuvent être répétées exactement au même point. Elles favorisent donc un suivi temporel des plantes. Elles favorisent aussi les comparaisons intra-parcellaires des mesures.

Les systèmes de production agricole sont suivis grâce à des mesures ponctuelles datées. Ces mesures sont une information locale et spécifique pour caractériser le système de production à un instant donné.

En viticulture, ces mesures peuvent être :

- des relevés réalisés sur la plante comme, par exemple, la mesure de la surface foliaire
- des relevés effectués au niveau des baies : mesure du poids de 100 baies, mesure de la concentration en sucres, en acides ou en anthocyanes dans les baies de raisin, par exemple
- des relevés permettant de caractériser l’environnement de la plante peuvent aussi être entrepris : mesure de la température de l’air, mesure de la teneur en eau des sols par exemple

Ces mesures sont réalisées de façon manuelle ou automatisée. Elles sont utilisées pour prendre une décision ou pour assurer une traçabilité du système de production.

En viticulture, des capteurs effectuant des mesures sur la plante, le raisin ou l’environnement de la plante ont été développés ou adaptés de l’agriculture pour être appliqués aux caractéristiques de la vigne. Ces nouvelles technologies ont principalement permis de réaliser plus de mesures d’une même grandeur au sein d’une même parcelle [4, 91]. Les mesures sont généralement présentées sous forme de cartes où est représentée la variabilité d’une grandeur de mesure au sein d’une ou de plusieurs parcelles. Les mesures réalisées sont enregistrées dans des bases de données de traçabilité.

De multiples capteurs sont en cours de développement ou de test. Le nombre de mesures, permettant de suivre le système de production viticole, issues de ces capteurs ne fera qu’augmenter avec le temps. De nombreux travaux de la littérature font référence à ces capteurs, parmi lesquels :

- Des caméras ont été utilisées dans les bandes spectrales du visible, du proche infrarouge ou de l’infrarouge thermique. Elles ont permis d’effectuer des estimations à haute résolution spatiale de l’état hydrique de la vigne [1, 45, 57, 67, 85, 88].
- Des caméras multispectrales ont aussi permis de déterminer la surface foliaire de la vigne [55, 54, 69] ou la vigueur de la plante [4]
- Des réseaux de capteurs sans fils ont été développés et installés dans des parcelles. Ils permettent des mesures de composantes climatiques : température de l’air, température du sol, humidité de l’air etc. Ces réseaux ont permis l’étude du microclimat au sein d’une parcelle et de définir le mésoclimat global d’une parcelle ou d’un ensemble de parcelles [3, 8, 65]
- Des capteurs installés sur les machines à vendanger permettent d’effectuer des mesures de poids de récolte directement lors de la vendange. Des estimations de rendement à un niveau intraparcellaire ont pu en être retirées [13, 89, 90]
- Des mesures de résistivité ont permis de créer des cartes représentant la composition des sols [2, 13]
- Des capteurs permettant la mesure de la concentration de certains composés chimiques dans les baies de raisin ont aussi été développés. Ils permettent d’estimer à tout moment la concentration en eau, sucres [35, 41, 46], anthocyanes, flavonols [18, 46] ou la concentration en acides aussi évaluée par le niveau de pH [23, 46]

dans les baies de raisin [44]

- Des capteurs permettant la mesure de la surface foliaire de chaque plante ont aussi été testés

Le développement de ces nouvelles technologies va entraîner l'augmentation du nombre de données de suivi (i) de la plante, (ii) de la composition chimique des baies ou (iii) de l'environnement de la plante. Ces données sont donc disponibles pour prendre des décisions ou sont enregistrées pour représenter la traçabilité du système de production au niveau parcellaire ou intraparcellaire.

Les données de suivi de la plante, de son environnement ou issues des mesures de capteurs décrits ci-dessus sont enregistrées dans des bases de données. Ces bases de données contenant toute la traçabilité d'un domaine pour différents cycles de production sont appelées bases de données de traçabilité [5].

Toutes ces données nouvellement disponibles constituent un fort potentiel pour caractériser le système de production viticole. Elles peuvent permettre de caractériser le système de production à plusieurs dates. Elles peuvent aussi permettre de caractériser plusieurs positions au sein d'une même parcelle. Ces données ont donc un important potentiel pour comparer en inter ou intraparcellaire l'état du système de production. Il faut donc valoriser ces données pour permettre ces comparaisons. Ce souhait de valorisation a été l'un des fondements pour proposer le sujet de ce travail de thèse.

## 1.3 L'utilisation des données de traçabilité

### 1.3.1 Prise de décision par analogies entre parcelles et cycles de production

Dans leur article sur la modélisation de la prise de décision en agriculture, Attonaty et al. [5] font un bilan des formalismes utilisés pour la prise de décision en agriculture. Ces formalismes sont, selon ces auteurs, principalement basés sur la comparaison d'ensembles homogènes de parcelles. La comparaison d'ensembles homogènes de parcelles peut prendre en compte un nombre variable de systèmes de productions : parcelles et cycles de production.

Les systèmes de production agricoles fonctionnent de façon cyclique et récurrente [6]. C'est pourquoi la comparaison du système de production en cours de cycle de production avec des cas déjà résolus pourra faciliter des prises de décisions.

La figure 1.1 présente comment la comparaison de différents systèmes de production peut amener à une prise de décision.

Pour prendre une décision sur un système de production à une date  $t_j$ , correspondant à un stade clef du développement de la plante, l'état actuel du système est défini [17]. Cet état actuel prend en compte l'historique du système de production, de son environnement (climat, sol...) et de l'effet de l'environnement sur le système de production. Il est défini grâce aux données de suivi de la plante et de son environnement enregistrées tout au



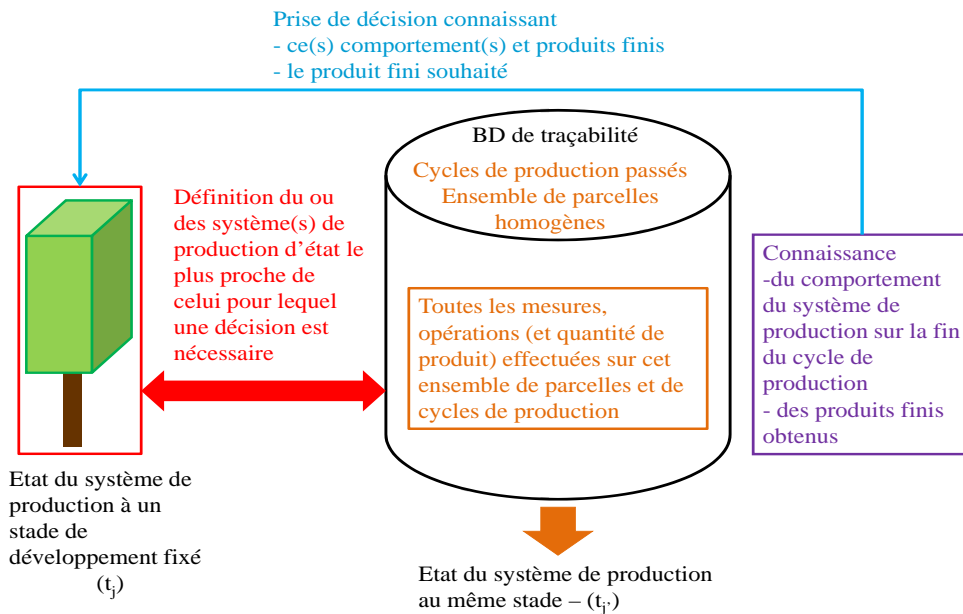


FIGURE 1.1 – Prise de décision par analogies en agriculture

long du cycle de production.

Pour une date  $t_{j+1}$  équivalente à la date  $t_j$ , l'état du système de production lors des cycles passés est défini. Ce travail est effectué pour différentes parcelles et différents cycles.

L'état du système de production est alors comparé à l'état des systèmes des couples parcelle  $\times$  cycle de production de la base de données de traçabilité. Les cas les plus proches sont alors déterminés.

Pour ces parcelles, le comportement de la plante et de son environnement sur la fin du cycle végétatif et la quantité et qualité du produit fini obtenu sont connus. Les décisions et opérations effectuées sur la plante ou son environnement pour obtenir le produit fini sont aussi connues.

En fonction du produit fini qu'il souhaite obtenir, l'agriculteur va prendre une décision, en faisant des hypothèses sur l'évolution du climat sur la fin du cycle de production. Cette décision sera le meilleur équilibre entre risques et meilleur traitement pour essayer d'obtenir le produit fini décidé en début de saison [17]. L'agriculteur peut donc décider soit (i) d'effectuer les mêmes actions que sur une des parcelles, pour un cycle de production, ressemblantes déterminées [17], soit (ii) de choisir des actions différentes de celles effectuées sur les parcelles ressemblantes pour tenter de modifier la quantité ou la composition du produit fini.

Un tel raisonnement est reproduit à chaque étape où une décision stratégique doit être prise. Ce type de raisonnement, fonctionnant sur des comparaisons avec des cas déjà résolus est appelé raisonnement par analogies.

Pour permettre la caractérisation d'un grand nombre d'états du système de production, de comportements et de produits finis différents, un tel raisonnement doit bénéficier d'une grande quantité de données et de références disponibles pour les comparaisons [5]. Des procédures d'apprentissage pourront alors être mises en place. Ces procédures ont pour objectif d'aider la prise de décision grâce aux comparaisons [32, 38].

La définition des états antérieurs du système de production des couples parcelle  $\times$  cycle de production utilisés pour les analogies et leur comportement sur la fin du cycle de production fait appel à la connaissance experte des cultures [28]. Elle fait aussi appel à la connaissance experte de leur domaine par les agriculteurs. Cette connaissance est empirique et mal formalisée par les agriculteurs eux-mêmes [63].

La formalisation des processus de prise de décision se heurte donc à cet empirisme [29]. Chaque agriculteur est le meilleur expert de son domaine et cette connaissance peut-être valorisable pour faciliter la prise de décision [29, 31].

Les bases de données de traçabilité contiennent l'historique de suivi d'un très grand nombre de systèmes de production. Les données de suivi d'un système de production peuvent, à tout moment, être mobilisées pour caractériser son état à un stade de développement donné.

Les données contenues dans les bases de données de traçabilité devraient donc permettre de mettre en place des raisonnements par analogie.

### **1.3.2 Définition de l'objectif pratique de la thèse : vers la création d'un outil d'aide à la décision facilitant les analogies entre parcelles**

L'objectif initial de la thèse est la valorisation des données contenues dans les bases de données de traçabilité des domaines viticoles pour faciliter la prise de décisions stratégiques sur le système en cours de production.

Connaissant le potentiel d'un raisonnement par analogies pour la prise de décision en agriculture, il a été décidé de proposer un formalisme qui permette de faciliter la comparaison de systèmes de production viticole et la réalisation d'analogies entre eux.

Les données de suivi du système de production, contenues dans les bases de données de traçabilité, ont un fort potentiel pour représenter à tout moment, pour un stade de développement de la plante choisi, l'état du système de production quel que soit le millésime.

L'objectif opérationnel du travail réalisé est donc de valoriser les données de suivi des bases de données de traçabilité pour faciliter la réalisation d'analogies entre différents systèmes de production viticoles, quels que soient les millésimes, ou cycles de production, durant lesquels ils ont été produits.

## 1.4 Organisation du mémoire de thèse

Après ce chapitre qui a permis de décrire le contexte et le problème opérationnel dans lequel se situe ce travail de thèse, la suite du manuscrit sera structurée en six chapitres.

Le chapitre 2 présentera l'analyse de la problématique. Dans ce chapitre, les potentialités de comparaisons de systèmes de production viticole en utilisant les informations brutes des bases de données de traçabilité seront évaluées. Puis, les outils déjà disponibles pour faciliter la comparaison de séquences temporelles seront présentés. Enfin, les besoins, propres au cas étudié, pour permettre la comparaison de couples parcelle  $\times$  millésime seront déterminés.

La méthodologie utilisée sera formalisée dans le chapitre 3. Les différents concepts utilisés seront définis dans ce chapitre. Une méthode de modélisation de la cinétique de grandeurs de mesure basée sur la collaboration entre les données de suivi et l'expertise du système de production étudié sera présentée.

Cette approche méthodologique sera ensuite appliquée à trois exemples.

La modélisation de l'évolution du climat entre floraison et véraison pour 7 millésimes et 11 cépages sera présentée dans le chapitre 4. L'apport de cette modélisation pour la réalisation d'analogies sur le comportement climatique des millésimes sera ensuite détaillé.

Puis, l'utilisation d'un modèle commun pour représenter l'évolution de différentes grandeurs mesurées dans les baies de raisin, la fonction logistique, sera détaillée. La modélisation de l'augmentation du pH pour 53 parcelles d'un même millésime et la comparaison de parcelles à partir d'une visualisation graphique de cette augmentation sera présentée dans le chapitre 5. La modélisation de l'accumulation de sucres dans les baies de raisin pour 7 millésimes d'une même parcelle puis 65 parcelles d'un même millésime et la comparaison de ces millésimes ou parcelles selon deux types d'utilisation des modèles sera ensuite présentée dans le chapitre 6.

Enfin, le chapitre 7 exposera les conclusions et les perspectives de ce travail de thèse. Des perspectives pour améliorer la méthode de modélisation et les comparaisons réalisées seront présentées.



# Chapitre 2

## Analyse de la problématique

---

### Sommaire

<b>2.1</b>	<b>Est-il possible de comparer les parcelles et les millésimes à partir des données brutes des bases de données de traçabilité ?</b>	<b>14</b>
2.1.1	Le système de production viticole : un système temporisé et cyclique . . . . .	14
2.1.2	Comparaison de parcelles et de millésimes à partir des données brutes des bases de données de traçabilité . . . . .	15
<b>2.2</b>	<b>Comment rendre comparables les données de traçabilité ?</b>	<b>18</b>
2.2.1	Etat de l'art . . . . .	18
2.2.2	Application et limites de l'utilisation de ces propositions pour la comparaison de données de traçabilité viticoles . . . . .	20
<b>2.3</b>	<b>Collaboration entre données de traçabilité et expertise pour créer un espace commun pour la comparaison de parcelles et / ou de millésimes</b>	<b>21</b>

---

L'objectif de ce travail de thèse est la valorisation des bases de données de traçabilité pour faciliter un raisonnement par analogies entre parcelles et / ou millésimes.

Afin de proposer une méthode de comparaison de parcelles et d'individus, il faut répondre aux questions suivantes :

1. Est-il possible de comparer des parcelles et / ou des millésimes à partir des données brutes des bases de données de traçabilité ?
2. Quels sont les outils déjà disponibles permettant de faciliter ces comparaisons ?
3. Quelles recherches sont nécessaires pour permettre la comparaison de parcelles et / ou millésimes à partir des informations contenues dans les bases de données de traçabilité viticoles ?

## 2.1 Est-il possible de comparer les parcelles et les millésimes à partir des données brutes des bases de données de traçabilité ?

### 2.1.1 Le système de production viticole : un système temporisé et cyclique

Chaque année, la vigne suit un cycle biologique fixé. Ce cycle est représenté dans la figure 2.1, pour l'hémisphère nord. Ce cycle peut être subdivisé en (i) un cycle végétatif et (ii) un cycle reproducteur. Le cycle végétatif représente le développement du feuillage depuis le débourrement jusqu'à la chute des feuilles en passant par le moment où la surface foliaire est maximale. Le cycle reproducteur représente le développement de la baie de raisin depuis sa formation au moment de la nouaison jusqu'à sa maturité, moment où elle pourra être vendangée.

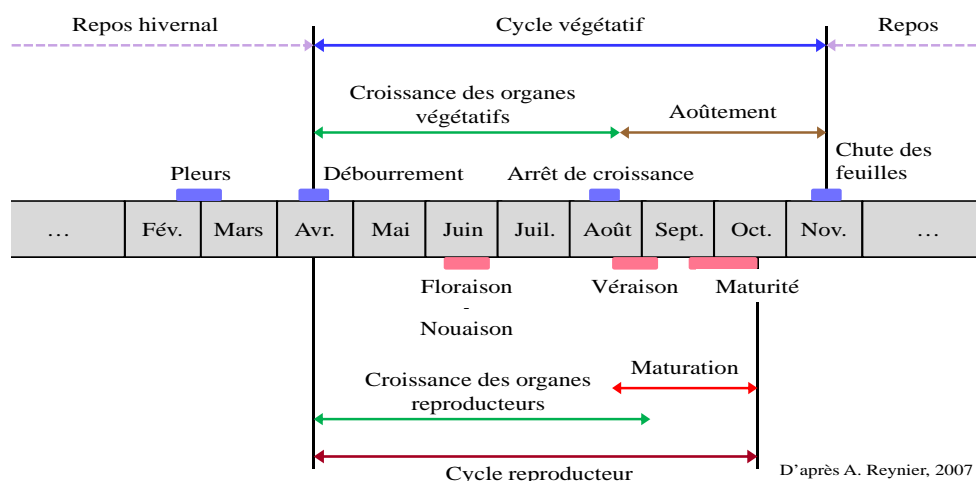


FIGURE 2.1 – Cycle biologique annuel de la vigne [79]

L'évolution de la plante suit donc les mêmes étapes quel que soit le millésime. Ces étapes sont ordonnées dans le temps. Une échelle de temps représentant le développement de la plante peut donc être définie. La date, dite en calendrier julien, à laquelle se fait chaque mesure permettra de faire référence à une étape du développement de la plante.

Les mesures qui vont être utilisées pour la comparaison de parcelles et / ou millésimes permettent l'étude de l'évolution de la plante ou de la baie de raisin. Elles ont un comportement dynamique. Comparer deux mesures provenant de stades de développement de la plante différents n'aura donc aucune utilité.

Pour la comparaison de parcelles et / ou de millésimes, le stade de développement auquel les mesures font référence doit donc être identique.

### **2.1.2 Comparaison de parcelles et de millésimes à partir des données brutes des bases de données de traçabilité**

Les parcelles doivent pouvoir être comparées entre elles qu'elles proviennent du même millésime ou non. Le système de production unitaire utilisé pour les comparaisons représente donc un couple parcelle  $\times$  millésime.

#### **Informations contenues dans les bases de données de traçabilité utilisées pour les comparaisons**

Dans les bases de données de traçabilité, les dates d'opération ainsi que, s'il y en a eu, les quantités de produit appliquées sont conservées. Ces bases de données contiennent aussi un grand nombre de données de suivi : (i) suivi de l'état hydrique de la plante, (ii) suivi climatique, (iii) suivi de la surface foliaire, (iv) suivi de la maturation du raisin et bien d'autres. Ce sont ces données de suivi qui vont être valorisées pour aider à la comparaison de couples parcelle  $\times$  millésime.

Chaque donnée de suivi représente l'évaluation d'une grandeur de mesure à laquelle est associée sa date de mesure. Pour chaque couple parcelle  $\times$  millésime, une grandeur de mesure est donc représentée par un ensemble de données de suivi associées à leur date de mesure. La fréquence des mesures dépend de chaque grandeur et de chaque couple parcelle  $\times$  millésime.

Deux grands types de grandeurs et de fréquence de mesure peuvent tout de même être considérés :

- Grandeurs mesurées à l'aide de capteurs automatiques, avec des mesures à haute fréquence et à pas de temps régulier. Ces grandeurs peuvent caractériser la plante et son environnement. Les capteurs associés peuvent être des capteurs climatiques (température de l'air, vitesse du vent...) ou des capteurs situés sur la plante (mesure de la température de feuillage, évaluation du flux de sève...)
- Grandeurs nécessitant la présence du viticulteur sur le terrain pour être mesurées. La fréquence des mesures de ces grandeurs est beaucoup plus faible que celle réalisée par des capteurs automatiques. Le pas de temps entre deux mesures est aussi très

hétérogène. Ces grandeurs sont toutes les grandeurs mesurées sur la plante et son environnement pour lesquelles les mesures ne sont pas automatisées. Ce peut être les mesures de poids des baies, de potentiel hydrique, de concentration en sucres ou en acides dans les baies pour n'en citer que quelques unes

Ces deux types de grandeurs sont très différents l'un de l'autre. Permettent-ils de comparer des couples parcelle  $\times$  millésime en tant que tels ?

**Les informations issues des bases de données de traçabilité permettent difficilement la comparaison des couples parcelle  $\times$  millésime**

Pour comparer des couples parcelle  $\times$  millésime à partir des données brutes les représentant, les données sont comparées deux à deux.

Dans cette partie, les deux types de grandeurs seront analysés séparément. Les grandeurs mesurées à l'aide de capteurs automatiques seront étudiées en premier. Puis, en second, les grandeurs mesurées grâce à des relevés sur le terrain seront analysées.

Les grandeurs de mesure évaluées à l'aide de capteurs automatiques sont représentées par un grand nombre de données.

Soit la température moyenne de l'air, mesurée par une station météorologique automatique à fréquence journalière. Soit la comparaison de la température moyenne de l'air entre le débourrement et la floraison pour deux millésimes d'une même parcelle de Carignan d'un domaine du Languedoc-Roussillon : la station expérimentale INRA Pech Rouge (figure 2.2).

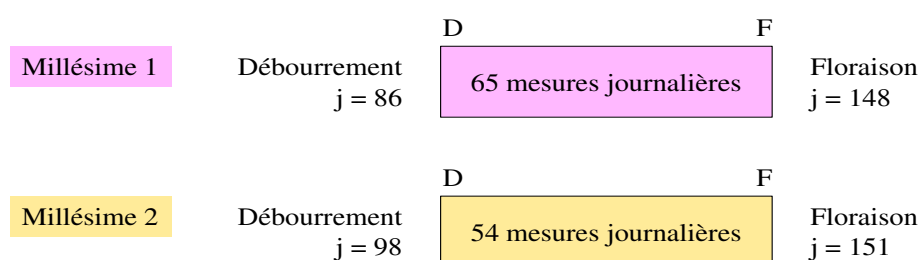


FIGURE 2.2 – Dates de débourrement et de floraison et nombre de mesures de la température journalière pour une même parcelle et 2 millésimes différents

Les dates de débourrement et de floraison sont différentes pour ces deux millésimes. Entre le 98ème et le 148ème jour, les températures moyennes journalières peuvent être comparées deux à deux. Mais les températures des jours compris entre le 86ème et le 97ème jour du millésime 1 ne peuvent pas être comparées avec le millésime 2. De même, les températures des jours compris entre le 149ème et le 151ème jour du millésime 2 ne peuvent pas être comparées avec le millésime 1. A partir des données brutes, les



deux millésimes ne pourront donc pas être comparés sur la totalité de la période de débourrement à floraison.

La température de l'air mesurée au 100<sup>ème</sup> jour de l'année ne représente pas de façon précise l'évolution de la température sur toute la période entre le débourrement et la floraison. La température de l'air peut varier considérablement d'un jour à l'autre. L'étude de la température de l'air entre débourrement et floraison doit donc prendre en compte un comportement moyen de celle-ci sur toute la période plutôt qu'une mesure ponctuelle.

Il est donc difficile de comparer ces deux millésimes en se basant sur les données brutes de température moyenne journalière. Un retraitement de ces données mènera à une comparaison plus précise.

La difficulté de comparaison de l'exemple donné dans la figure 2.2 est généralisé à tout type de comparaison de couples parcelle  $\times$  millésime quelle que soit la grandeur de mesure, mesurée automatiquement.

Dans le cas de la comparaison de deux couples parcelle  $\times$  millésime sur des intervalles avec des bornes identiques, la comparaison au jour le jour de données brutes pourra être réalisée. Chaque mesure journalière ne représente pas l'évolution de la grandeur sur toute la période étudiée. La comparaison au jour le jour de données brutes ne permet donc pas de réaliser une comparaison objective, représentant toute la période étudiée, de couples parcelle  $\times$  millésime pour la grandeur.

Pour réaliser des comparaisons qui seront ensuite utilisées pour faire des analogies entre le passé et le présent, il faut donc retraiter les données de suivi des grandeurs de mesures évaluées automatiquement.

La figure 2.3 présente l'exemple du suivi de l'augmentation de la concentration en sucres dans les baies de raisin pour une même parcelle et trois millésimes différents. Chaque point représente une mesure ponctuelle de cette concentration.

Pour chacun des trois millésimes présentés, les mesures sont réalisées à des dates différentes. Le nombre total de mesures varie d'un millésime à l'autre (3 pour 2007, 4 pour 2003 et 7 pour 2006). Les données de ces trois millésimes ne peuvent donc pas être comparées une à une. De plus, le cycle reproducteur de la vigne et la forme de la cinétique d'accumulation des sucres dans les baies de raisin impliquent qu'il n'y a aucune utilité à comparer des mesures de juillet avec des mesures réalisées au mois d'août ou au mois de septembre. Ces trois millésimes ne peuvent donc pas être comparés en se basant sur les informations brutes des bases de données de traçabilité.

De tels exemples pourraient être donnés pour toutes les grandeurs de mesures mesurées directement sur le terrain. La comparaison de couples parcelle  $\times$  millésime pourrait être possible si toutes les mesures de ces couples étaient synchrones. L'exemple présenté dans la figure 2.3 met donc en évidence qu'il est très difficile de comparer des couples parcelle  $\times$  millésime à partir des données brutes pour les grandeurs de mesures évaluées directement sur le terrain.

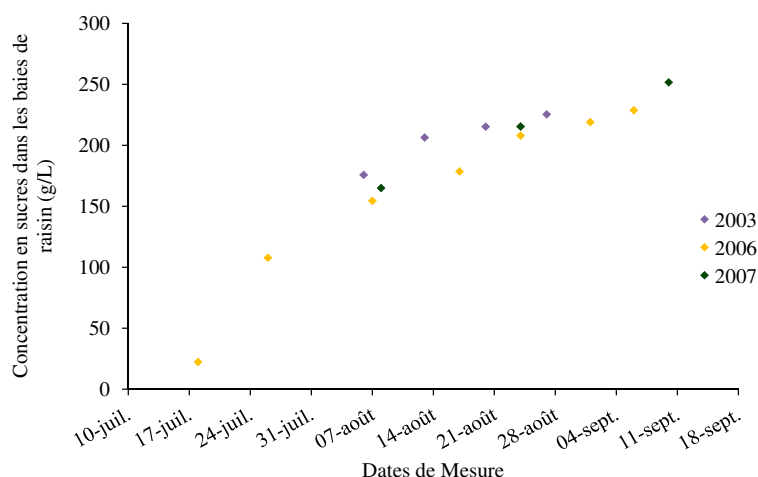


FIGURE 2.3 – Mesures ponctuelles de la concentration en sucres dans les baies de raisin pour une même parcelle et 3 millésimes différents

Quelle que soit la grandeur de mesure, il y a une hétérogénéité (i) du nombre total de mesures et (ii) des dates de mesures de cette grandeur. Les données sont dites asynchrones.

La quantité totale de mesures d'une grandeur pour une parcelle et un millésime fixé est appelée dimension de l'espace dans lequel ce couple parcelle  $\times$  millésime est représenté.

Les couples parcelle  $\times$  millésime sont donc représentés dans des espaces de dimension différente. La comparaison des couples parcelle  $\times$  millésime à partir des données brutes des bases de données de traçabilité n'est donc pas possible.

Afin de pouvoir comparer les couples parcelle  $\times$  millésime entre eux, il faut trouver un moyen de les représenter dans un espace commun. Cet espace commun devra permettre de représenter avec le moins de perte d'information possible la cinétique de la grandeur de mesure pour laquelle les couples parcelle  $\times$  millésime sont comparés.

L'objectif de ce travail de thèse est donc la définition d'un espace commun entre les couples parcelle  $\times$  millésime qui permette la comparaison de ces couples et limite le plus possible la perte de l'information initiale contenue dans les bases de données de traçabilité.

## 2.2 Comment rendre comparables les données de traçabilité ?

### 2.2.1 Etat de l'art

Dans d'autres domaines que l'agriculture, l'utilisation de bases de données spatio-temporelles pour enregistrer et valoriser le suivi de systèmes très différents, du plus simple au plus complexe (du processus industriel avec le suivi d'une seule grandeur de mesure

[86] au suivi de l'historique d'un malade en hôpital par exemple [20]) a été largement étudiée.

Les données à enregistrer dans ces bases de données étaient (i) mesurées sur des pas de temps différents, (ii) asynchrones et (iii) en nombre différent d'un cas à l'autre. Ces données présentaient donc les mêmes caractéristiques que les données viticoles décrites dans la partie 2.1.2.

Le premier objectif des travaux disponibles dans la littérature est de permettre un enregistrement homogène des données.

Pour permettre cet enregistrement, des intervalles de temps identiques d'un cas à l'autre ont été définis [20, 86, 97]. Des données faisant référence à un même intervalle de temps peuvent être enregistrées dans une même matrice et sont comparables entre elles.

Les solutions proposées pour enregistrer les données dans des intervalles de temps identiques sont principalement basées sur la discrétisation de l'axe temporel en modules temporels [43, 53, 84, 87]. Les modules temporels sont des intervalles de temps de taille identique définis pour une grandeur de mesure. L'ensemble des modules temporels définis pour une grandeur de mesure couvre toute la surface de l'intervalle initial d'étude de cette grandeur.

Chaque module temporel est représenté par une date de début, une date de fin et une durée [97]. Des relations de précédence et d'inclusion entre deux intervalles de temps peuvent alors être proposées [43, 87].

Le second objectif des travaux disponibles dans la littérature est de proposer un formalisme d'enregistrement des données qui facilite leur comparaison.

Pour pouvoir comparer différents cas entre eux, l'étape qui suit la discrétisation de l'axe de temps en modules temporels est d'enregistrer les données mesurées dans le module temporel auquel elles appartiennent. Pour ce faire, la date de mesure est alors comparée aux dates de début et de fin de chaque module [97].

La comparaison de différents cas se fait à partir de la comparaison d'une et une seule valeur représentative de la grandeur par module temporel. Les données brutes enregistrées dans les bases de données temporelles peuvent présenter différents cas :

- Une et une seule donnée pour chaque module temporel a été mesurée
- Plusieurs données ont été mesurées pour chaque module temporel
- Il y a au moins un module temporel pour lequel une mesure manque

Dans les deux derniers cas, pour avoir une et une seule valeur représentant la grandeur par intervalle, il faut convertir les données déjà existantes [86]. Des fonctions de conversion sont donc définies pour chaque grandeur de mesure [43]. Ces fonctions de conversion sont soit des fonctions d'interpolation s'il manque des valeurs pour certains modules temporels, soit des fonctions d'agrégation s'il y a un trop grand nombre des mesures par module temporel [53, 84].

La définition de la fonction de conversion est spécifique à chaque grandeur. Pour ne pas perdre d'information, la fonction de conversion de chaque grandeur doit être définie

à partir de la connaissance experte de cette grandeur [84, 97].

Pour enregistrer et valoriser des données hétérogènes dans des bases de données temporelles, les étapes suivantes ont donc été proposées :

- Discrétisation de l'axe temporel en modules temporels
- Conversion des données pour ne garder qu'une et une seule valeur par module temporel via une fonction de conversion qui fait appel à la connaissance experte de la grandeur de mesure

En viticulture, la connaissance experte du système de production doit permettre de proposer des fonctions de conversion représentant l'évolution de chaque grandeur de mesure.

## 2.2.2 Application et limites de l'utilisation de ces propositions pour la comparaison de données de traçabilité viticoles

Le système de production viticole a été largement étudié. Son cycle biologique annuel (figure 2.1) est connu. Mais surtout, la cinétique de chaque grandeur de mesure évaluée sur la plante ou le raisin a été étudiée. Un exemple de ce type de cinétique est donné dans la figure 2.4. Elle représente l'évolution entre la nouaison et la maturité du volume des baies. Sur le même graphique, les transformations globales d'autres grandeurs mesurées dans la baie sont données mais leur cinétique n'est pas tracée.

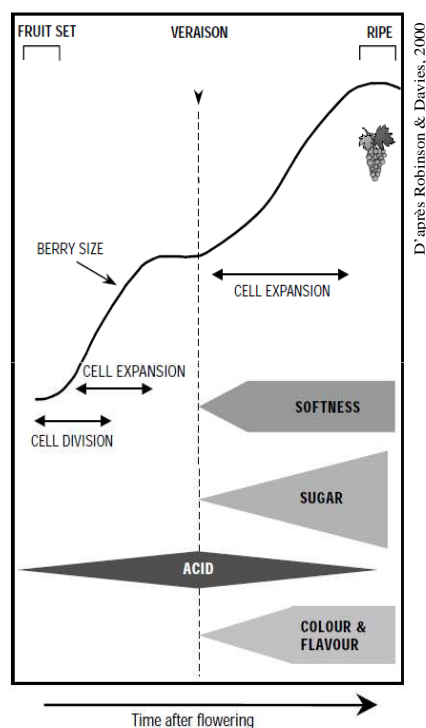


FIGURE 2.4 – Cinétique du volume des baies de raisin de la nouaison à la maturité et évolution globale des autres grandeurs mesurées dans les baies

L'effet de l'environnement sur la plante et le raisin a lui aussi été étudié. Ces connaissances ont été reportées dans la littérature et sont donc disponibles pour valoriser les données contenues dans les bases de données de traçabilité.

La connaissance experte de la vigne et de son cycle biologique doit donc permettre de définir des modules temporels utiles à l'étude de chaque grandeur de mesure. Ces modules temporels ont comme bornes des stades de développement de la plante. Pour chaque couple parcelle  $\times$  millésime, la date de ces stades est différente. Les modules temporels ont donc des bornes de dates différentes en fonction de chaque couple parcelle  $\times$  millésime. La longueur de chaque module sera elle aussi propre à chaque couple parcelle  $\times$  millésime. La définition des modules temporels telle que présentée dans les travaux cités dans la section 2.2.1 devra donc être adaptée aux spécificités du cas viticole. Les définitions fournies dans ces travaux ne peuvent donc pas être directement appliquées à la valorisation des informations des bases de données de traçabilité viticoles.

Enfin, la connaissance de la cinétique des grandeurs de mesure évaluées sur la plante et la baie de raisin et de l'effet des grandeurs de mesure évaluées sur l'environnement de la plante sur celle-ci et son raisin doit permettre de proposer des fonctions de conversion objectives pour extrapoler ou agréger les données brutes des bases de données de traçabilité. Ces fonctions de conversion sont spécifiques à chaque grandeur. Les fonctions de conversion proposées dans les travaux cités dans la section 2.2.1 ne pourront donc pas être utilisées pour la valorisation des informations contenues dans les bases de données de traçabilité viticole.

Pour proposer une nouvelle représentation de l'information contenue dans les bases de données de traçabilité, il faut donc proposer un formalisme de discrétisation de l'axe temporel propre aux spécificités du système de production viticole. Les fonctions de conversion doivent être définies pour chaque grandeur en mobilisant l'expertise du système de production viticole.

## **2.3 Collaboration entre données de traçabilité et expertise pour créer un espace commun pour la comparaison de parcelles et / ou de millésimes**

La figure 2.5 résume l'approche envisagée pour permettre la comparaison de parcelles et / ou de millésimes à partir des informations contenues dans les bases de données de traçabilité viticole.

Pour chaque couple parcelle  $\times$  millésime, un nombre identique de valeurs est estimé. Ces valeurs ne sont plus asynchrones. Les couples sont représentés dans un espace commun. Les couples parcelle  $\times$  millésime ont donc été rendus comparables.

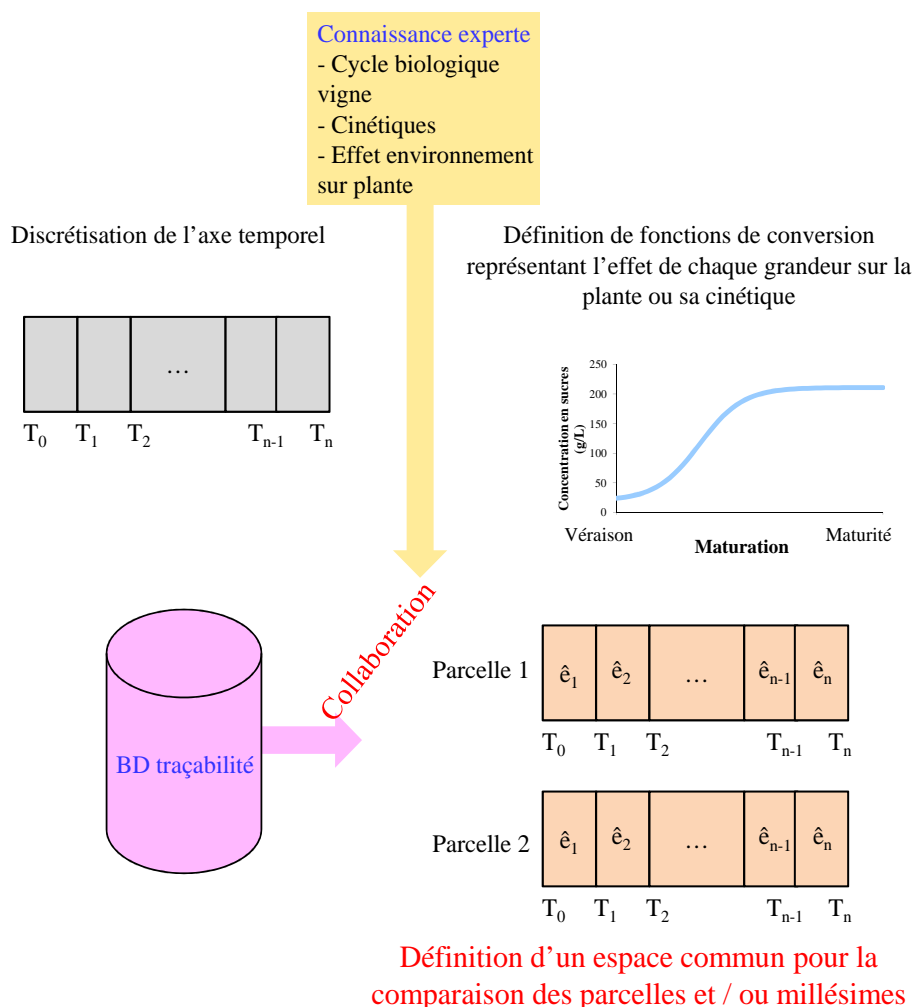


FIGURE 2.5 – Collaboration entre données de traçabilité et expertise pour comparer les parcelles et / ou millésimes

Pour définir la discrétisation temporelle servant pour l'étude de chaque grandeur et la fonction de conversion relative à cette grandeur, l'expertise du système de production viticole est essentielle. Cette expertise est disponible dans la littérature.

La proposition de ce travail de thèse est de faire collaborer les données des bases de données de traçabilité viticoles avec la connaissance experte de ce système pour créer un espace commun de comparaisons entre les couples parcelle  $\times$  millésime.

La méthodologie mise en place pour définir cet espace commun de comparaison est détaillée dans l'approche théorique qui est présentée en chapitre 3.

# Chapitre 3

## Approche théorique

---

### Sommaire

<b>3.1 Définitions</b> . . . . .	<b>24</b>
3.1.1 Mode d'étude d'un phénomène temporel . . . . .	24
3.1.2 Description d'une base de données spatio-temporelles . . . . .	25
<b>3.2 Méthode de comparaison de plusieurs individus.</b> . . . . .	<b>27</b>
3.2.1 Rappel du problème soulevé par la comparaison d'individus. . .	27
3.2.2 Approche méthodologique proposée . . . . .	27
3.2.3 Utilisation des vecteurs de paramètres pour faciliter les comparaisons . . . . .	29
<b>3.3 Bilan : Schéma récapitulatif des étapes de retraitement des données</b> . . . . .	<b>30</b>

---

## 3.1 Définitions

Afin de pouvoir valoriser les bases de données de traçabilité en utilisant les concepts présentés dans le chapitre 2, il est nécessaire de définir différents termes propres au cas en cours d'étude. Nous proposons donc une terminologie propre, adaptée des définitions issues des travaux sur les bases de données temporelles [43, 53, 84, 86, 87, 97] et élaborée à partir de la connaissance experte du système de production viticole (ou agricole) et des besoins relatifs à la définition d'un espace commun.

### 3.1.1 Mode d'étude d'un phénomène temporel

**Définition 3.1.1** *Systeme  $\Phi$*  : Un système est un ensemble d'éléments interagissant entre eux ou interdépendants qui forment un tout.

Soit l'étude d'un système dans le temps,

**Définition 3.1.2** *Phénomène temporel  $\Phi$*  : Un phénomène temporel est un ensemble de modifications d'un système qui suit une ou plusieurs lois générales applicables au système considéré ou à un système analogue.

Cette définition large s'applique aux évolutions internes du système étudié comme les phénomènes biologiques, biochimiques etc. Elle s'applique également aux évolutions liées à l'environnement du système considéré comme la température, l'humidité etc.

**Définition 3.1.3** *Grandeur  $g$*  : Une grandeur de mesure permet l'évaluation de l'évolution du phénomène temporel. Cette évaluation est directe ou indirecte via des effets induits sur le système par le phénomène étudié.

**Définition 3.1.4** *Base temporelle  $T_\Phi$*  : Un phénomène temporel est étudié sur un intervalle de temps donné appelé base temporelle.

**Définition 3.1.5** *Cycle temporel  $T$*  : Un cycle temporel est l'intervalle de temps regroupant toutes les bases temporelles d'étude d'un système.

**Définition 3.1.6** *Discrétisation temporelle  $\Delta$*  : Découpage de la base temporelle en une succession d'intervalles de temps contigus  $W_a$  nécessaires pour l'étude de  $\Phi$ .

La discrétisation est définie, en fonction du phénomène temporel en cours d'étude, selon une grille régulière ou irrégulière.

$\Delta$  est caractérisée par  $N$  intervalles contigus  $[W_1, \dots, W_N]$  et  $N+1$  bornes de ces intervalles telles que  $\forall a \in [1; N], W_a = ]T_{a-1}; T_a]$  et  $W_1 = [T_0; T_1]$ .

L'ensemble des fenêtres  $W_a$ , comme on peut le voir dans la figure 3.1, recouvre la totalité de la base temporelle  $T_\Phi$ . L'intervalle de temps du cycle temporel couvert par chaque base temporelle  $T_\Phi$  dépend du phénomène temporel étudié.



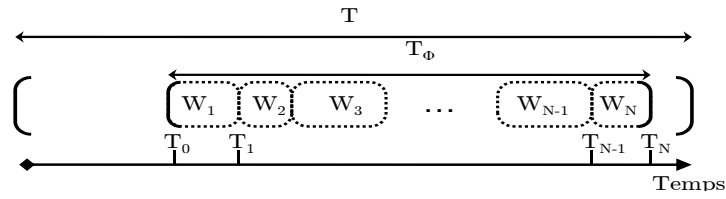


FIGURE 3.1 – Discrétisation de l'échelle temporelle pour l'étude d'un phénomène temporel

**Définition 3.1.7** *Mesure  $x$*  : Une mesure est l'évaluation de la valeur d'une grandeur à un moment donné.

**Exemples** : Soit le système pied de vigne. De nombreux phénomènes peuvent être étudiés sur celui-ci comme l'évolution de la surface foliaire de la plante ou celle de sa balance hormonale.

Pour étudier le développement de la baie de raisin au cours du millésime, un des phénomènes temporels,  $\Phi$ , qui peut être suivi est l'augmentation du poids des baies entre la formation de la baie et la vendange. La grandeur de mesure,  $g$ , est donc le poids de la baie. La base temporelle d'étude,  $T_\Phi$ , sera la période entre la formation de la baie, estimée au moment de la floraison, et la vendange.

Pour étudier l'évolution de la quantité de feuillage de la vigne au cours du millésime, un des phénomènes temporels,  $\Phi$ , qui peut être suivi est l'évolution de la surface foliaire exposée entre la formation des feuilles et la chute de celles-ci. La grandeur de mesure,  $g$ , est donc la surface foliaire exposée. La base temporelle,  $T_\Phi$ , est la période entre la formation des premières feuilles, estimée au moment du débourrement et la chute des feuilles.

Ces deux phénomènes temporels sont étudiés sur des bases temporelles différentes. D'autres phénomènes seraient étudiés sur d'autres bases temporelles. L'intervalle de temps regroupant toutes ces bases est le millésime. Le cycle temporel,  $T$ , est donc représenté par un millésime.

La discrétisation temporelle,  $\Delta$ , d'étude de chaque phénomène sera quand à elle définie en fonction des besoins exprimés pour chaque étude de ce phénomène.

### 3.1.2 Description d'une base de données spatio-temporelles

**Définition 3.1.8** *Base de données spatio-temporelles TDB* : Une base de données spatio-temporelles est constituée d'un ensemble de descripteurs  $d_j$ . Chaque descripteur  $d_j$  représente une donnée brute.

$$TDB = \{d_j, j \in [1; J]\}$$

**Définition 3.1.9 Descripteur :** Un descripteur  $d_j$  est une donnée datée et géoréférencée. Tout descripteur est donc considéré comme étant un triplet  $d_j = (t_j, s_j, x_j)$  où :

$x_j$  est la mesure

$t_j$  est la date à laquelle la mesure a été effectuée

$s_j$  est l'emplacement géographique de la mesure

**Définition 3.1.10 Individu  $I$  :** Un individu est un ensemble de descripteurs  $d_j$  de même localisation et dont toutes les dates de mesure sont incluses dans un seul cycle temporel.

Tout individu est donc caractérisé par  $D_I$ , un sous ensemble de descripteurs de la base de données  $TDB$ . L'individu  $I$  est tel que :

$$\forall d_j \in D_I, t_j \in T, s_j \text{ est unique}$$

**Définition 3.1.11 Séquence temporelle  $TS_{g,I}$  :** La séquence temporelle est l'ensemble des descripteurs de  $D_I$  relatifs à une même grandeur de mesure  $g$ .

A des fins de simplification, il sera sous-entendu dans la suite du document que la grandeur de mesure de chaque séquence temporelle est clairement définie. Elle sera donc explicitée clairement et la séquence temporelle sera notée  $TS_I$  dans la suite du document.

**Exemples :**

Soit à nouveau le système pied de vigne, la base de données spatio-temporelles,  $TDB$ , relative à l'étude de ce système contiendra toutes les mesures datées et géoréférencées de suivi de tous les phénomènes temporels possibles sur vigne. Un individu,  $I$ , sera le regroupement de toutes les mesures effectuées sur une même zone,  $s_j$ , souvent une parcelle, pour un millésime,  $T$ . Pour chaque individu,  $I$ , un sous ensemble  $D_I$  regroupant toutes les données de la base de données spatio-temporelles  $TDB$  correspondant au millésime  $T$  auquel l'individu  $I$  se rapporte et à sa localisation  $s_j$  pourra donc être défini.

Pour étudier la grandeur de mesure poids des baies, le descripteur,  $d_j$ , est la mesure à une date donnée du poids des baies. La séquence temporelle,  $TS_I$ , relative à l'étude de la grandeur de mesure poids des baies sera l'ensemble des mesures ponctuelles du poids des baies et leur date de mesure.

De même, pour étudier la grandeur de mesure surface foliaire exposée, le descripteur,  $d_j$ , est la mesure ponctuelle datée de cette grandeur. La séquence temporelle,  $TS_I$ , regroupera l'ensemble de ces descripteurs soit les mesures associées à leur date de mesure.

## 3.2 Méthode de comparaison de plusieurs individus.

### 3.2.1 Rappel du problème soulevé par la comparaison d'individus.

La comparaison d'individus entre eux est difficile compte tenu des propriétés suivantes :

1. Caractère asynchrone des données : Les descripteurs des séquences temporelles ne sont pas mesurés à des dates identiques. De plus, les phénomènes temporels étudiés,  $\Phi$ , ont un comportement dynamique. Il n'y a donc aucun intérêt à comparer deux données relevées à des dates différentes.
2. Différence de dimension des séquences temporelles. Le nombre de données contenues dans chaque séquence temporelle varie d'un individu à l'autre. Une comparaison terme à terme est alors impossible.

Ces contraintes mettent en évidence la difficulté de comparaison de séquences temporelles quels que soient les individus considérés. De ce fait, il devient nécessaire de développer des méthodes adaptées pour permettre la représentation de tous les individus dans un espace commun. Les individus pourront alors être comparés dans cet espace commun.

### 3.2.2 Approche méthodologique proposée

Dans cette partie, la méthode proposée pour comparer différents individus à partir des séquences temporelles représentant un phénomène temporel est détaillée. Notre approche est basée sur la définition de modèles paramétriques et leur ajustement à l'aide des séquences temporelles.

Soit  $(\Phi, \Delta)$  un couple phénomène temporel et discrétisation de la base temporelle associée à ce phénomène.

Soit  $TS_I$  la séquence temporelle qui permet de caractériser  $(\Phi, \Delta)$ . Selon la définition 3.1.11,  $TS_I$  peut s'écrire de la manière suivante :

$$TS_I = \begin{bmatrix} \vdots & \vdots \\ t_j & x_j \\ \vdots & \vdots \end{bmatrix}, t_j \in T_\Phi$$

Soit l'hypothèse qu'un phénomène temporel est correctement caractérisé dès lors qu'une mesure de la séquence temporelle  $TS_I$  est présente pour chaque fenêtre  $W_a$  de la discrétisation temporelle  $\Delta$ .

Quel que soit le descripteur, aucune information n'est fournie *a priori* à propos de la fenêtre à laquelle celui-ci appartient.

L'approche proposée dans ce travail de thèse est basée sur la projection des descripteurs des  $TS_I$  dans chaque fenêtre  $W_a$  et la conversion des données brutes pour ne conserver qu'une seule valeur par fenêtre  $W_a$ . Quel que soit  $d_j$  de  $TS_I$ , on dit que :  $d_j$  se projette dans  $W_a$  si et seulement si  $t_j \in ]T_{a-1}; T_a]$ .

**Définition 3.2.1** *Nombre de descripteurs projetés dans une fenêtre  $W_a$ ,  $b_{a,I}$  :* L'entier naturel  $b_{a,I}$  représente le nombre de descripteurs projetés dans la fenêtre  $W_a$  pour l'individu  $I$ .

**Typologie des modifications de séquences temporelles nécessaires en fonction des entiers  $b_{a,I}$**

- $\forall a \in [1; N]$  si  $b_{a,I} > 1$ , il faudra agréger la connaissance contenue dans  $TS_I$ .
- $\forall a \in [1; N]$  si  $b_{a,I} = 1$ , la connaissance contenue dans  $TS_I$  correspond aux besoins exprimés pour l'analyse de  $\Phi$
- $\forall a \in [1; N]$  si  $\exists a / b_{a,I} = 0$ , Certaines fenêtres ne présentent aucune information. Il faudra interpoler la connaissance contenue dans  $TS_I$ .

Pour obtenir une valeur par fenêtre  $W_a$  et représenter l'individu dans l'espace commun choisi pour la comparaison, l'approche proposée va *faire collaborer les données mesurées avec un modèle paramétrique*. Le modèle paramétrique représente l'évolution du phénomène temporel pour la grandeur considérée.

Cette approche reste la plus générale possible afin d'étudier l'ensemble des problèmes identifiés qu'ils soient de sur-échantillonnage ou de sous-échantillonnage des séries temporelles par rapport aux besoins exprimés pour l'analyse de  $\Phi$ .

**Définition 3.2.2** *Modèle  $M(g)$  :* Le modèle  $M(g)$  est une représentation simplifiée, sous forme mathématique, de la connaissance experte de l'évolution de la grandeur  $g$  servant à l'étude du phénomène  $\Phi$  sur la base temporelle  $T_\Phi$ .

Plusieurs modèles peuvent être définis pour la même grandeur. Un modèle peut être de type très simple comme une moyenne ou un écart-type d'une séquence temporelle. Des modèles plus complexes représentant une fonction de forme connue peuvent aussi être utilisés.

Soit maintenant  $\Theta = \{\theta_k\}$ ,  $k \in \mathbb{N}^*$ , un vecteur de paramètres qui définit  $M(g)$ . Chaque paramètre  $\theta_k$  est ajustable.

L'approche proposée consiste à ajuster le modèle grâce aux descripteurs de la séquence temporelle  $TS_I$ . Chaque paramètre  $\theta_k$  sera donc défini grâce aux valeurs expérimentales de  $TS_I$ .

Le modèle  $M(g)$  est défini grâce à l'expertise du système de production. L'ajustement de l'ensemble des paramètres de  $\Theta$  pour chaque individu permet la définition d'une et une seule formule de  $M(g)$  pour l'individu  $I$ . Pour chaque individu, un modèle sera donc défini grâce à la collaboration entre l'expertise du système de production et les descripteurs des séquences temporelles extraites des bases de données de traçabilité.

Le modèle étalonné pour l'individu  $I$  sera noté  $M_I(g)$  et les paramètres de ce modèle seront notés  $\Theta_I = \{\theta_{k,I}\}$ ,  $k \in \mathbb{N}^*$ .

### 3.2.3 Utilisation des vecteurs de paramètres pour faciliter les comparaisons

Soit un modèle  $M_I(g)$ , deux utilisations du modèle peuvent être envisagées (i) en extension et (ii) en intension.

La définition en extension d'un concept est le fait de formuler sa signification en explicitant l'ensemble des objets qui le représentent.

**Définition 3.2.3** *Utilisation extensive d'un modèle  $M_I(g)$  :* Un modèle  $M_I(g)$  est utilisé en extensif quand une valeur de la grandeur  $g$  pour l'individu  $I$  est estimée pour chaque fenêtre  $W_a$  de la discrétisation  $\Delta$  associée à l'étude de  $g$ . L'estimation de la grandeur  $g$  pour l'individu  $I$  et la fenêtre  $W_a$  est notée  $\hat{g}(I, W_a)$

La définition en intension d'un concept est le fait de spécifier toutes les propriétés nécessaires et suffisantes permettant de définir l'appartenance d'un terme au concept.

**Définition 3.2.4** *Utilisation intensive d'un modèle  $M_I(g)$  :* Le vecteur  $\Theta_I$  représente les informations nécessaires et suffisantes pour définir tout point du modèle. Ce vecteur représente donc l'utilisation intensive du modèle  $M_I(g)$ .

**Exemples :** Pour étudier l'évolution du poids de la baie entre la floraison et la récolte, la discrétisation temporelle d'étude peut être choisie comme homogène, chaque fenêtre  $W_a$  étant un jour.

L'utilisation en intension du modèle reviendra à travailler sur les valeurs des paramètres du modèle qui permet de représenter la cinétique du poids des baies de la floraison à la récolte.

L'utilisation en extension du modèle, qui permet de représenter la cinétique du poids des baies de la floraison à la récolte, revient à estimer une valeur du poids des baies pour chaque jour entre la floraison et la récolte.

Soit  $I_1$  et  $I_2$ , deux individus à comparer pour une grandeur  $g$  en se basant sur un unique modèle  $M(g)$ .

L'ajustement de  $M(g)$  avec les séquences temporelles  $TS_1$  et  $TS_2$ , correspondant respectivement aux individus  $I_1$  et  $I_2$ , permet d'obtenir les vecteurs  $\Theta_{I_1}$  et  $\Theta_{I_2}$ .

$I_1$  et  $I_2$  sont comparés selon l'utilisation intensive des modèles  $M_{I_1}(g)$  et  $M_{I_2}(g)$  si les paramètres de  $\Theta_{I_1}$  et  $\Theta_{I_2}$  servent pour la comparaison. A l'inverse,  $I_1$  et  $I_2$  sont comparés selon l'utilisation extensive des modèles  $M_{I_1}(g)$  et  $M_{I_2}(g)$  si les estimations de la valeur de  $g$  pour chaque fenêtre  $W_a$  de la discrétisation temporelle  $\Delta$  sont utilisées pour la comparaison.

### 3.3 Bilan : Schéma récapitulatif des étapes de retraitement des données

La figure 3.2 présente un schéma récapitulatif des différentes étapes mises en place pour retraiter les données brutes des bases de données de traçabilité viticoles. L'expertise est d'abord mobilisée pour définir des modèles paramétriques de l'évolution de chaque grandeur sur des bases temporelles jugées utiles pour l'analyse. Puis les modèles peuvent être utilisés en intensif ou en extensif. Les résultats des modèles, paramètres ou estimations, sont enregistrés dans de nouvelles bases de données qui seront utilisées pour les comparaisons.

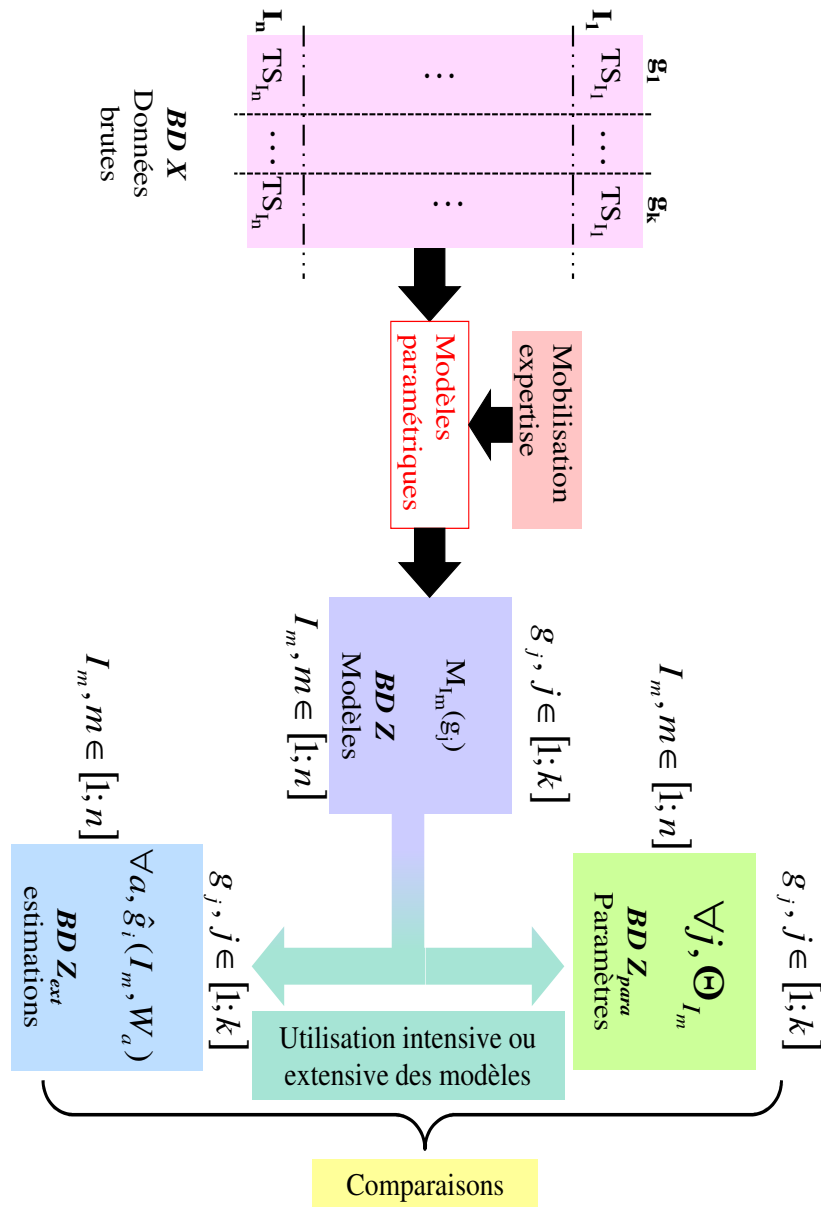


FIGURE 3.2 – Schéma récapitulatif des étapes de retraitement proposées

La base de données construite à partir des paramètres ajustés de chaque modèle d'un individu,  $I$ , est appelée base de données paramétrique,  $Z_{para}$ . La base de données construite avec les estimations de chaque grandeur  $g$  pour chaque individu  $I$  dans chacune des fenêtres  $W_a$  de la discrétisation temporelle  $\Delta$  choisie pour l'étude de la grandeur  $g$  est appelée base de données extensive,  $Z_{ext}$ .

A titre d'exemple, il a été proposé de comparer les individus en utilisant des techniques d'analyses descriptives classiques telles que l'analyse en composantes principales (ACP), l'analyse des correspondances multiples (ACM) ou la classification ascendante hiérarchique (CAH). Dans les applications présentées par la suite, les comparaisons se sont appuyées sur des ACP.





# Chapitre 4

## Application 1 : Etude du climat d'un domaine sur plusieurs années entre la floraison et la véraison

---

### Sommaire

<b>4.1</b>	<b>Introduction : Effet des paramètres climatiques sur l'évolution de la plante entre la floraison et la véraison . . . . .</b>	<b>35</b>
4.1.1	Effet de la température sur la vigne et le raisin . . . . .	36
4.1.2	Effet du régime hydrique sur la vigne . . . . .	37
4.1.3	Effet du rayonnement solaire sur la vigne . . . . .	37
<b>4.2</b>	<b>Matériel et Méthodes . . . . .</b>	<b>38</b>
4.2.1	Données climatiques utilisées . . . . .	38
4.2.2	Définition de la base de données d'étude . . . . .	39
4.2.3	Grandeurs climatiques utilisées . . . . .	41
4.2.4	Calcul des paramètres . . . . .	41
4.2.5	Utilisation des paramètres . . . . .	45
<b>4.3</b>	<b>Résultats . . . . .</b>	<b>48</b>
4.3.1	Adéquation de l'exemple à la méthode proposée . . . . .	48
4.3.2	Résultats de l'ACP réalisée en utilisant les indices bioclimatiques non pondérés . . . . .	48
<b>4.4</b>	<b>Etude comparative de l'ACP avec indices bioclimatiques pondérés . . . . .</b>	<b>55</b>
4.4.1	Analyse comparative du plan constitué des deux premières composantes principales . . . . .	56
4.4.2	Analyse comparative du plan constitué de la troisième et de la quatrième composante principale . . . . .	57
<b>4.5</b>	<b>Conclusion : Bilan de l'étude comparative du climat entre cépages et millésimes . . . . .</b>	<b>59</b>

Application 1 : Etude du climat d'un domaine sur plusieurs années entre la floraison et la véraison

---

Dans ce chapitre, les potentialités d'utilisation de la méthode pour caractériser le vécu climatique de chaque parcelle pour chaque millésime sont étudiées. Ce travail ne vise pas à fournir de nouvelles connaissances sur la climatologie du domaine dont les données proviennent mais à vérifier que la méthode proposée permet de retrouver une connaissance empirique déjà acquise sur le vécu climatique de chaque millésime.

## 4.1 Introduction : Effet des paramètres climatiques sur l'évolution de la plante entre la floraison et la véraison

Tout au long de son cycle de développement et de reproduction [79], la vigne est sensible au climat [19, 37]. La température, le régime hydrique et le rayonnement solaire sont les trois facteurs de climat les plus importants pour activer ou limiter la croissance de la vigne et le développement de la baie [16, 70, 80]. L'activité métabolique de la plante a une réponse continue en fonction de ces facteurs. Des niveaux trop faibles et éventuellement trop élevés de ces facteurs auront alors pour effet d'inhiber cette activité métabolique, et donc la croissance de la plante et le développement de la baie.

A la floraison, les fleurs sont pollinisées et vont se différencier en baies. La baie va alors se développer comme un organe photosynthétique jusqu'à la véraison [11]. Elle reçoit beaucoup de nutriments venant du feuillage ou des racines. Au moment de la véraison, la baie change de fonction et devient un organe puits qui servira à la reproduction de la plante.

Le climat facilite ou inhibe le développement de la plante et de la baie. La plante est beaucoup moins sensible aux effets du climat après la véraison qu'entre la floraison et la véraison [9, 16, 58]. De plus, le poids final de la baie, et donc le rendement final de la parcelle est quasiment fixé au moment de la véraison [72]. Des auteurs ont même proposé le poids des baies 4 semaines après la nouaison comme le meilleur estimateur du rendement final [11].

Le climat entre floraison et véraison a aussi un effet important sur l'accumulation de différents composés dans les baies et le feuillage [9, 52, 58, 70].

La période de floraison à véraison est donc la période où le climat, en plus de son effet sur la croissance de la plante, peut avoir l'effet limitant le plus marqué sur le raisin, son rendement et sa composition à la vendange. Le climat peut aussi influencer la composition de la baie entre la véraison et la récolte, mais cette période n'a pas été étudiée dans cette application. Il a donc été choisi d'appliquer l'approche théorique proposée à un ensemble de paramètres climatiques entre floraison et véraison. L'objectif de ce travail sur les paramètres climatiques est de permettre la comparaison d'individus sur la base d'une caractérisation du climat entre la date de floraison et la date de véraison.

Dans un premier temps, cette partie donnera un rapide état de l'art relatif à l'incidence des paramètres du climat sur la vigne et le raisin. Cet état de l'art permettra de justifier

et de présenter succinctement un certain nombre d'indices utilisés classiquement pour décrire l'incidence du climat sur la plante et le fruit. Enfin, l'intégration de ces indices comme expertise dans l'approche générale proposée et leur utilisation pour comparer des individus seront discutées.

#### 4.1.1 Effet de la température sur la vigne et le raisin

La chaleur est fondamentale pour le développement du végétal [19, 76]. La température est souvent considérée comme le facteur climatique majeur agissant sur la vigne [98]. La température a un effet cumulatif sur la plante [98]. Chaque stade phénologique est atteint après l'accumulation d'une certaine quantité de températures actives [76, 77, 95]. Dans le Bordelais, par exemple, il a été observé que la date de véraison est déterminée par (i) la date de floraison, (ii) le nombre de jours ayant dépassé 30°C entre la floraison et la véraison [56] et (iii) le cumul de températures actives depuis le début du cycle de la plante.

Connaissant l'effet cumulatif de la température sur la plante, de nombreux modèles basés sur l'effet cumulatif des températures ont été proposés pour estimer la date d'atteinte des stades phénologiques [11, 42, 49, 76, 98].

L'intensité de la réponse de la plante à la température dépend de la valeur de cette température. Ainsi, en dessous d'un certain seuil, estimé à 10°C, les températures n'ont pas d'action sur celle-ci [14]. De plus, au dessus d'un autre seuil, estimé entre 35 et 40°C, l'activité de la plante est inhibée [19, 37]. L'intensité de la photosynthèse suit une courbe de forme approximativement gaussienne en réponse à la température [15]. Une photosynthèse plus active permet un développement de la plante et de la baie ainsi qu'une maturation plus rapides.

Quelle que soit la période du cycle végétatif, la photosynthèse est maximale quand la température est proche de 25°C [15, 16]. Elle est presque inexistante avant 10°C et s'annule à partir de 40°C [15, 16]. Le développement de la baie et la maturation présentent une vitesse maximale quand la température est comprise entre 20 et 25°C [16, 19, 42, 58, 59, 72, 80].

Pour avoir un fonctionnement métabolique optimal, la vigne a donc besoin d'une chaleur relative mais pas excessive.

Le taux de nouaison (différenciation des fleurs en baies) est réduit aux températures basses (<15°C) et élevées (>35°C) [11, 16, 19, 36, 58]. Un climat trop froid ou trop chaud au moment de la floraison va donc favoriser les erreurs de différenciation de la fleur en baie telles que le millerandage et la coulure [80].

Enfin, les températures trop basses et trop élevées entre floraison et véraison limitent la taille finale des baies de raisin. Au sein de la plage de températures permettant un grandissement des baies efficace, une augmentation de la température va avoir un effet bénéfique sur le poids final de la baie [71].

La température entre floraison et véraison peut aussi avoir un effet sur l'accumulation des différents composés chimiques dans les baies de raisin. L'accumulation des sucres dans

les baies est plus sensible aux températures entre la floraison et la véraison et pendant la véraison que pendant la maturation [52].

### 4.1.2 Effet du régime hydrique sur la vigne

Le régime hydrique de la vigne est déterminé par le bilan entre l'eau qui lui est apportée via les précipitations et l'irrigation, si elle est autorisée, et les pertes d'eau par la plante et son environnement : la transpiration de la plante, perte d'eau par ruissellement et drainage [16, 30]. S'il y a eu plus de transpiration que d'eau apportée, le régime hydrique est alors déficitaire et l'on parle de contrainte hydrique.

Son niveau peut être évalué grâce à des indices climatiques et notamment l'indice de sécheresse [70, 92].

Ce facteur est, après la température, le facteur limitant le plus important pour le grandissement de la baie entre floraison et véraison.

Il faut 250 à 350 mm d'eau à la vigne entre le débourrement et la vendange pour que la contrainte hydrique ne soit pas trop sévère [37]. Dans les vignobles où la température est suffisante pour l'activité de la vigne, la contrainte hydrique devient le facteur limitant essentiel dans la détermination du poids final de la baie. C'est le cas des vignobles situés en zone méditerranéenne [16, 37, 82]. Cette contrainte va aussi conditionner la vigueur de la plante [16].

La photosynthèse est d'abord peu affectée par une contrainte hydrique modérée avant une valeur seuil après laquelle son taux diminue de façon significative [15].

Une contrainte hydrique sévère entre floraison et véraison va ainsi limiter de façon irréversible le poids final de la baie. Les précipitations vont quant à elles permettre d'augmenter le poids final de la baie [11, 16, 36, 66, 70, 80, 93].

Une contrainte hydrique sévère, notamment dans les climats chauds, au moment de la nouaison peut entraîner une coulure importante [70, 80]. Des pluies au même moment vont aussi favoriser la coulure et le millerandage [11, 36].

Une contrainte hydrique modérée après la nouaison peut avoir un effet positif sur la qualité du raisin en favorisant l'accumulation de polyphénols et d'arômes dans les baies de raisin [70, 74, 93].

L'alimentation hydrique doit donc, à l'instar de la température, être comprise dans une gamme de valeurs optimales, favorisant l'activité de la vigne [80]. Cette gamme de valeurs varie en fonction du stade phénologique [70]. Une contrainte hydrique sévère a des conséquences d'autant plus graves qu'elle apparaît tôt [19].

### 4.1.3 Effet du rayonnement solaire sur la vigne

Le rayonnement solaire sert pour la photosynthèse. Il va permettre l'activité des feuilles et donc la synthèse des nutriments qui vont être utilisés pour le développement des baies et de la plante [15]. Dans les zones viticoles traditionnelles, ce n'est en général pas un facteur limitant [80].

La photosynthèse et le poids de la baie sont d'autant plus importants que la radiation solaire est importante, et ce jusqu'à un seuil après lequel la réponse de la baie et de la plante au rayonnement stagne [16, 9].

Plus il y aura de rayonnement entre floraison et véraison et plus le poids et la taille des baies seront élevés [16, 52, 56].

La lumière est nécessaire à la synthèse des polyphénols [19, 37]. Une plus importante exposition des baies entre floraison et véraison va accentuer l'accumulation des sucres, des anthocyanes et des tanins dans les baies [9, 52, 58].

Hormis les trois facteurs détaillés ci-dessus, d'autres facteurs climatiques peuvent avoir des effets sur la vigne. Ces facteurs ne seront pas détaillés dans cette introduction mais on pourra citer par exemple (i) l'humidité de l'air [15] ou (ii) la vitesse du vent [52] qui agissent sur le micro-climat de la baie et influencent ainsi sa taille et sa composition [7].

Le climat change beaucoup les caractéristiques du raisin mûr [11]. L'effet du climat sur la plante est déterminant et complexe. Il y a donc un réel intérêt à étudier son évolution sur chaque millésime. Sur la période de floraison à véraison, connaître l'évolution du climat et ses caractéristiques comparées aux autres millésimes pourrait permettre de faire, par analogie, des hypothèses sur le rendement ou la composition du raisin attendu.

L'évolution du climat a été étudiée au travers de modèles simples : moyenne, écart-type ou comptage de jours ayant dépassé un seuil. Ces modèles seront présentés dans ce chapitre.

## 4.2 Matériel et Méthodes

### 4.2.1 Données climatiques utilisées

Cette étude a été réalisée à partir des données climatiques de la station expérimentale INRA Pech Rouge. Ce domaine est situé à Gruissan, en Languedoc-Roussillon, dans le sud de la France. Ses Coordonnées géographiques sont 43°10'N, 3°06'E dans le système géodésique WGS84.

Les données d'une station météorologique automatique ont été utilisées. Cette station était située au sein d'une des parcelles de la station expérimentale INRA Pech Rouge. Cette station effectuait des mesures horaires ou journalières systématiques de différentes grandeurs qui étaient ensuite archivées sur un serveur de l'INRA. Ces relevés étaient réalisés pour le suivi habituel du domaine. Pour cette application, les données journalières ont été utilisées.

## 4.2.2 Définition de la base de données d'étude

Le travail réalisé portait sur la comparaison entre le vécu climatique de différents individus sur la période allant de la floraison à la véraison. La base temporelle de travail ( $T_{\Phi}$ ) correspondait donc à l'intervalle entre ces deux stades phénologiques.

### Utilisation d'observations de dates phénologiques pour définir les bornes de la base temporelle

De 2003 à 2009, sur un réseau de 11 parcelles, les dates des principaux stades phénologiques ont été enregistrées dans la base de données de traçabilité représentant tout le domaine. 10 cépages étaient représentés dans ce réseau : Carignan, Chardonnay, Gamay, Grenache noir, Marselan, Mourvèdre, Petit Verdot, Portan, Syrah et Viognier. 9 parcelles correspondaient chacune à un cépage. Le nom du cépage était donc utilisé en référence à la parcelle. 2 parcelles de Syrah étaient comprises dans ce réseau de parcelles. Une de ces parcelles était cultivée dans la plaine (Syrah plaine), tandis que l'autre était cultivée sur le plateau calcaire du domaine (Syrah plateau).

Les dates de mi floraison et mi véraison des 11 cépages ou dénomination pour la Syrah ont été extraites de la base de données de traçabilité contenant l'ensemble des mesures et observations réalisées sur toutes les parcelles du domaine. Ces dates correspondaient à l'observation de l'apparition des fleurs ou de la véraison des baies sur la moitié des vignes ou des baies de chaque parcelle. Les dates de floraison et véraison, en jour julien, et le nombre de jours entre floraison et véraison pour chacun des cépages ou dénomination pour la Syrah ont été regroupées dans le tableau 4.1.

Chaque cépage, ou dénomination pour la Syrah, représentait une localisation  $s_j$  indépendante. Par mesure de simplification, les deux parcelles de Syrah ont été considérées comme des cépages indépendants et nommées cépage dans la suite de cette application. L'ensemble de ces localisations a été étudié sur 7 millésimes ( $T$ ). Un individu,  $I$ , était donc l'étude d'un cépage pour un millésime.

Tous les relevés journaliers des grandeurs climatiques ont été extraits du serveur de l'INRA.

Chaque descripteur était une mesure journalière d'une grandeur et sa date de mesure. Pour chaque cépage, la séquence temporelle ( $TS_I$ ) représentative de chaque grandeur contenait l'ensemble des descripteurs de cette grandeur dont les dates de mesures étaient comprises entre la floraison et la véraison.

Les mesures provenaient d'une station à relevés automatiques. Hormis dysfonctionnement de la station, une mesure par jour de chaque grandeur est donc réalisée. Le nombre de descripteurs d'une séquence temporelle correspondait donc au nombre de jours entre floraison et véraison pour chacun des individus.

Pour cette application, la discrétisation temporelle choisie ne comportait qu'une seule fenêtre  $W_a$  entre la floraison et la véraison. Cette unique fenêtre correspondait donc à

Application 1 : Etude du climat d'un domaine sur plusieurs années entre la floraison et la véraison

TABLE 4.1 – Dates en jours julien de la floraison et de la véraison et nombre de jours entre ces deux stades pour les 11 cépages et les 7 millésimes étudiés

Cépage	Date de floraison						
	2003	2004	2005	2006	2007	2008	2009
Carignan	148	154	154	151	142	147	144
Chardonnay	137	152	150	146	135	139	139
Gamay	136	151	147	147	134	135	141
Grenache noir	147	155	155	153	141	145	145
Marselan	146	154	154	154	140	145	145
Mourvèdre	150	159	158	155	148	151	148
Petit Verdot	145	155	153	150	139	145	145
Portan	139	150	150	145	136	136	143
Syrah plaine	148	156	154	151	140	146	144
Syrah plateau	148	157	157	149	141	145	145
Viognier	142	153	152	149	135	136	140

Cépage	Date de véraison						
	2003	2004	2005	2006	2007	2008	2009
Carignan	202	208	209	198	194	206	197
Chardonnay	196	202	203	198	197	203	196
Gamay	194	197	200	188	185	191	196
Grenache noir	202	209	209	196	205	206	200
Marselan	200	207	207	197	200	205	201
Mourvèdre	198	205	206	206	200	208	198
Petit Verdot	209	212	212	205	203	203	197
Portan	193	202	199	194	192	199	196
Syrah plaine	202	204	205	198	195	203	198
Syrah plateau	196	204	205	195	197	204	196
Viognier	198	203	204	206	201	200	197

Cépage	Nombre de jours floraison à véraison						
	2003	2004	2005	2006	2007	2008	2009
Carignan	53	53	54	46	51	58	52
Chardonnay	58	49	52	51	61	63	56
Gamay	57	45	52	40	50	55	54
Grenache noir	54	53	53	42	53	60	54
Marselan	53	52	52	42	59	59	55
Mourvèdre	47	45	47	50	51	56	49
Petit Verdot	63	56	58	54	63	57	51
Portan	53	51	48	48	55	62	52
Syrah plaine	53	47	50	46	54	56	53
Syrah plateau	47	46	47	45	55	58	50
Viognier	55	49	51	56	65	63	56



tout l'intervalle de la base temporelle  $T_{\Phi}$ .

### 4.2.3 Grandeurs climatiques utilisées

Différentes grandeurs ont été retenues de part leur influence sur la plante, son environnement et le raisin. Elles sont présentées dans le tableau 4.2.

Les grandeurs choisies étaient relatives aux facteurs mis en avant par l'étude bibliographique de l'effet du climat sur l'activité métabolique de la vigne : température de l'air et du sol, évapotranspiration et hauteur de précipitations rayonnement solaire toutes deux relatives au régime hydrique de la plante, rayonnement solaire, humidité de l'air et vitesse du vent.

### 4.2.4 Calcul des paramètres

La cinétique des grandeurs climatiques ne suit pas une forme fixe. Le climat varie d'un jour à l'autre d'année en année. La connaissance experte du climat ne peut donc pas se traduire sous la forme de cinétiques. L'évolution de chaque grandeur climatique étant aléatoire, la meilleure façon d'envisager le climat dans sa globalité est d'étudier son comportement moyen sur une période donnée.

La connaissance experte du climat se traduit dans la définition de fenêtres de températures, niveau hydrique ou rayonnement solaire qui activent plus ou moins ou même inhibent l'activité métabolique de la plante.

Entre les différents individus, le temps entre floraison et véraison varie de 40 à 65 jours (tableau 4.1). Chaque grandeur était donc caractérisée par 40 à 65 descripteurs dans la fenêtre  $W_a$ . Pour ne garder qu'une seule valeur de la séquence temporelle dans la fenêtre  $W_a$ , les modèles choisis étaient donc des modèles permettant d'agrèger les données.

### Moyenne, somme et écart-type de descripteurs

Pour agréger l'information contenue dans les séquences temporelles représentatives de chaque grandeur, la moyenne et / ou la somme des descripteurs ainsi que leur écart-type ont été calculés pour toute la période entre floraison et véraison. Les grandeurs pour lesquelles ces calculs ont été réalisés, le type de calcul et le code de paramètre employé sont récapitulés dans le tableau 4.3.

Pour la plupart des grandeurs, la moyenne des descripteurs a été calculée entre la floraison et la véraison. La somme des descripteurs pour la même période n'a été calculée que dans certains cas. Enfin, l'écart-type des descripteurs a été calculé pour toutes les grandeurs où une moyenne ou une somme ont été calculées.

### Compteurs de jours où une grandeur a dépassé un seuil

La connaissance experte du climat sur la vigne permet de proposer des températures au niveau desquelles la croissance et la maturation sont de vitesse maximales, des niveaux

TABLE 4.2 – Grandeurs utilisées pour la comparaison du comportement climatique sur différents individus

Grandeur	Code	Unité
Amplitude thermique journalière	AMP	°C
Température minimale de l'air	TN	°C
Température maximale de l'air	TX	°C
Température moyenne de l'air	TM	°C
Température minimale du sol à 10 cm de profondeur	S1N	°C
Température maximale du sol à 10 cm de profondeur	S1X	°C
Température moyenne du sol à 50 cm de profondeur	S5M	°C
Rayonnement solaire global	RG	J.cm <sup>-2</sup>
Evapotranspiration potentielle (méthode de Penman-Monteith)	ETP	mm
Hauteur des précipitations	RR	mm
Intensité maximale des précipitations	RRX	mm
Humidité minimale de l'air	UN	%
Humidité maximale de l'air	UX	%
Humidité moyenne de l'air	UM	%
Nombre d'heures où l'humidité de l'air a été inférieure à 40%	U4	h
Nombre d'heures où l'humidité de l'air a été supérieure à 80%	U8	h
Nombre d'heures où l'humidité de l'air a été supérieure à 90%	U9	h
Vitesse moyenne du vent	V	m.s <sup>-1</sup>
Vitesse maximale du vent	VX	m.s <sup>-1</sup>

TABLE 4.3 – Type de calcul réalisé par grandeur et nom du paramètre extrait

Grandeur	Moyenne	Somme	Ecart-type
AMP	AMPM		EtAMP
ETP	ETPM	ETPS	EtETP
RG	RGM		EtRG
RR		RRS	EtRR
S1N	S1NM		EtS1N
S1X	S1XM		EtS1X
S5M	S5MM	S5MS	EtS5M
TM	TMM	TMS	EtTM
TN	TNM		EtTN
TX	TXM		EtTX
UM	UMM		EtUM
UN	UNM		EtUN
UX	UXM		EtUX
V	VM		EtV
VX	VXM		EtVX

de rayonnement global utiles à la photosynthèse etc. Cette connaissance a été utilisée pour définir des seuils de température, précipitations, rayonnement global et humidité de l'air qui ont une influence sur le comportement de la vigne.

Les connaissances disponibles dans la littérature et citées dans la partie 4.1 ont été utilisées pour proposer des seuils ou des intervalles de températures déterminants pour la croissance de la vigne.

Le détail des choix effectués est donné dans l'annexe A.

Pour chacun de ces seuils, si le descripteur répondait à l'inégalité ou était dans l'intervalle proposé un compteur était incrémenté de 1. Ainsi, le nombre de jours où une grandeur a franchi un seuil ou bien était restée dans un intervalle fixé a été comptabilisé.

Les valeurs de nombre de jours ainsi calculées ont été divisées par le nombre de jours entre floraison et véraison de chaque individu. Chaque paramètre représentait donc la fréquence d'apparition d'un certain seuil pour une grandeur fixée et par individu.

Les différents compteurs mis en place sont détaillés dans le tableau 4.4.

## Indices bioclimatiques

Différents indices bioclimatiques ont été proposés dans la littérature pour suivre l'évolution de la plante ou prédire l'atteinte des stades phénologiques et la maturité [11]. Ils sont calculés à l'aide de différentes grandeurs : température de l'air, hauteur des précipitations, insolation etc.

Ces indices sont considérés comme très utiles pour caractériser une situation viticole. Quelques uns des indices bioclimatiques disponibles dans la littérature ont été calculés

TABLE 4.4 – Compteurs du nombre de jours ayant dépassé un seuil et description des seuils

Paramètre	Grandeur d'origine	Calcul
TM20	TM	TM < 20
TM25	TM	TM > 25
TMMat	TM	TM ∈ [20; 25]
TMCR	TM	TM ∈ [20; 30]
TX25	TX	TX < 25
TX30	TX	TX > 30
TX35	TX	TX > 35
TXOPT	TX	TX ∈ [30; 35]
RR15	RR	RR ∈ [0; 15]
RR30	RR	RR > 30
RRX30	RRX	RRX > 30
RG22	RG	RG > 2200
RG11	RG	RG < 1100
U46	U4	U4 < 6
U4F	U4	U4 > 12
U86	U8	U8 < 6
U8F	U8	U8 > 12
U96	U9	U9 < 6
U9F	U9	U9 > 12

et insérés à la base de données paramétrique de cet exemple.

Les indices utilisés dans cette application sont de deux types : (i) indices basés sur des cumuls de températures, indices visant à caractériser la croissance de la plante et la vitesse de maturation ou (ii) évaluation de l'état hydrique des plantes. Les calculs ont été réalisés entre la floraison et la véraison. Les indices ont donc été adaptés à la base temporelle choisie.

Les indices calculés étaient : le cumul de températures base 10 ou indice de Winkler (Win), le cumul de températures base 10 où les températures supérieures à 19 °C ne comptent pas dans le calcul de la somme ou indice de Gladstones (Glad), l'indice héliothermique de Huglin (IH), l'indice thermopluviométrique de Ribéreau-Gayon et Peynaud (ITP) et l'indice de sécheresse (IS).

Le détail des calculs effectués est donné dans l'annexe B.

Les indices Win, Glad, IH et ITP, correspondant à des sommes de températures, avaient, par millésime, une valeur d'autant plus élevée que le temps passé entre la floraison et la véraison était grand. Afin de pouvoir aussi travailler avec des valeurs ne prenant pas en compte la précocité de chaque cépage et chaque millésime, pour chaque individu, les valeurs calculées ont été pondérées du nombre de jours compris dans chaque base temporelle. Ces valeurs pondérées ont été définies comme des paramètres à part entière.

#### 4.2.5 Utilisation des paramètres

Pour chaque individu  $I$ , représentant un cépage étudié sur un millésime, l'ensemble des paramètres présentés dans la section 4.2.4 a été enregistré dans une nouvelle base de données dite paramétrique avant d'être analysé.

Pour cette application, la base temporelle,  $T_\Phi$ , n'a été discrétisée qu'en une seule fenêtre  $W_a$ . Chaque paramètre correspondait donc à l'estimation d'une valeur de la grandeur pour chaque fenêtre de la discrétisation temporelle. L'utilisation intensive et extensive des modèles était donc identique.

Les données ont ensuite été analysées par analyse en composante principale (ACP) [64, 83]. Deux analyses ont été effectuées. La première prenait en compte les valeurs des indices bioclimatiques non pondérées. Elle permettait ainsi d'évaluer les potentiels de la méthode pour caractériser le climat à partir de calculs couramment réalisés par les viticulteurs. La seconde analyse, quant à elle prenait en compte les valeurs pondérées des indices bioclimatiques.

Ces analyses ont été réalisées avec le logiciel STATlab V3.0 (SLP Infoware, Ivry sur Seine, France).

La figure 4.1 donne un récapitulatif de tous les paramètres utilisés dans cette application et du type de calcul ayant permis leur modélisation. Un numéro et un symbole a été attribué à chacun des paramètres pour une lecture plus facile et rapide des plans d'ACP. Pour les indices bioclimatiques, les grandeurs de mesures utilisées pour le calcul de chaque indice sont rappelées. Les indices bioclimatiques pondérés ont la lettre P

Application 1 : Etude du climat d'un domaine sur plusieurs années entre la floraison et la véraison

---

indiquée à la fin de leur code, leur couleur et numéro sont différents des autres indices.

Paramètre	Numéro	Type calcul	Symbole	Paramètre	Numéro	Type calcul	Symbole
AMPM	1	Moy.	◆	ETPM	25	Moy.	■
EtAMP	2	ET	◆	ETPS	26	Somme	■
TMM	3	Moy.	◆	EtETP	27	ET	■
TMS	4	Somme	◆	RRS	28	Somme	■
EtTM	5	ET	◆	EtRR	29	ET	■
TNM	6	Moy.	◆	UMM	30	Moy.	■
EtTN	7	ET	◆	EtUM	31	ET	■
TXM	8	Moy.	◆	UNM	32	Moy.	■
EtTX	9	ET	◆	EtUN	33	ET	■
S5MM	10	Moy.	◆	UXM	34	Moy.	■
S5MS	11	Somme	◆	EtUX	35	ET	■
EtS5M	12	ET	◆	RGM	36	Moy.	●
S1N	13	Moy.	◆	EtRG	37	ET	●
EtS1N	14	ET	◆	RR15	38	Compt.	■
S1X	15	Moy.	◆	RR30	39	Compt.	■
EtS1X	16	ET	◆	RRX30	40	Compt.	■
TM20	17	Compt.	◆	RG22	41	Compt.	●
TM25	18	Compt.	◆	RG11	42	Compt.	●
TMMat	19	Compt.	◆	U46	43	Compt.	■
TMCR	20	Compt.	◆	U4F	44	Compt.	■
TX25	21	Compt.	◆	U86	45	Compt.	■
TX30	22	Compt.	◆	U8F	46	Compt.	■
TX35	23	Compt.	◆	U96	47	Compt.	■
TXOPT	24	Compt.	◆	U9F	48	Compt.	■
VM	49	Moy.	●	VXM	51	Moy.	●
EtV	50	ET	●	EtVX	52	ET	●
Win	53	Ind. - TM	▲	WinP	58	Ind. - TM	▲
Glad	54	Ind. -TM	▲	GladP	59	Ind. -TM	▲
IH	55	Ind. - TM et TX	▲	IHP	60	Ind. - TM et TX	▲
ITP	56	Ind. - TM et RR	▲	ITPP	61	Ind. - TM et RR	▲
IS	57	Ind. - ETP et RR	▲	ISP	62	Ind. - ETP et RR	▲

FIGURE 4.1 – Liste des paramètres utilisés dans l'application et symbole associé

## 4.3 Résultats

### 4.3.1 Adéquation de l'exemple à la méthode proposée

Les séquences temporelles sont constituées par le regroupement de tous les descripteurs d'une grandeur mesurés entre la date de floraison et la date de véraison.

Le nombre de jours entre floraison et véraison varie d'un individu à l'autre (tableau 4.1). Le nombre des descripteurs contenus dans chaque séquence temporelle varie donc d'un individu à l'autre. Les séquences temporelles sont donc de dimensions différentes.

Chaque individu a une date de floraison et une date de véraison propre (tableau 4.1). Soit deux individus de la base de données : le Portan (Individu  $I_1$ ) et la Syrah Plaine (Individu  $I_2$ ) pour le millésime 2003. Ces deux individus sont de dimension identique (53 jours - voir figure 4.2), l'individu  $I_1$  présente une date de floraison et de véraison 9 jours avant celles de l'individu  $I_2$ .

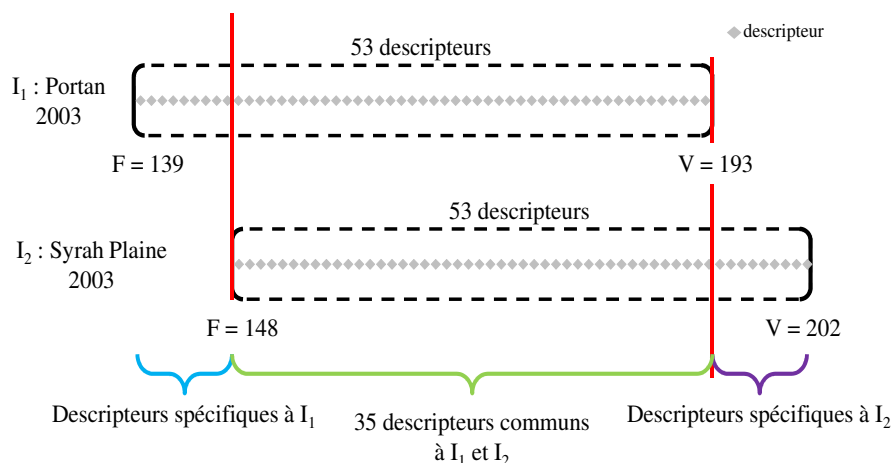


FIGURE 4.2 – Etude de deux individus de même dimension : TS asynchrones

Les 9 premiers descripteurs de la séquence temporelle de  $I_1$  ne renseignent pas l'individu  $I_2$ . Les 9 derniers descripteurs de l'individu  $I_2$  ne renseignent pas l'individu  $I_1$ . Les mesures de chacun des deux individus ne sont donc pas toutes effectuées à des dates identiques. Les mesures des séquences temporelles  $TS_1$  et  $TS_2$  sont donc asynchrones.

De telles observations pourraient être réalisées quels que soient les individus. Les données brutes sont donc toujours asynchrones.

Les séquences temporelles qui seront comparées dans cette application présentent donc les deux problèmes cités dans la section 3.2.1. L'étude de cette base de données constitue donc bien un exemple applicatif pour la méthodologie proposée.

### 4.3.2 Résultats de l'ACP réalisée en utilisant les indices bioclimatiques non pondérés

Pour cette analyse, 57 variables ont été prises en compte.



Dans cet exemple, les 4 premières composantes principales ont été conservées. Elles représentent respectivement 42,3 %, 19,13%, 11,86% et 10,54% de variance expliquée. L'ensemble des ces 4 axes permet donc d'expliquer 83,84% de la variance totale entre les 11 cépages pour les 7 millésimes étudiés. Le pourcentage de variance résiduel, soit 16.16%, est expliqué par les autres axes de l'ACP mais leur pourcentage de variance expliquée, toujours inférieur à 6%, a été jugé comme négligeable face à celui des autres axes.

L'analyse détaillée du plan 1-2 sera présentée. Les conclusions tirées de l'analyse du plan 3-4 seront ensuite données. L'analyse détaillée du plan 3-4 est disponible dans l'annexe C.

### Analyse du plan constitué des deux premières composantes principales

La figure 4.3 présente la position des individus dans le plan constitué des deux premières composantes principales. Le symbole de chaque individu est le millésime durant lequel cet individu a été produit.

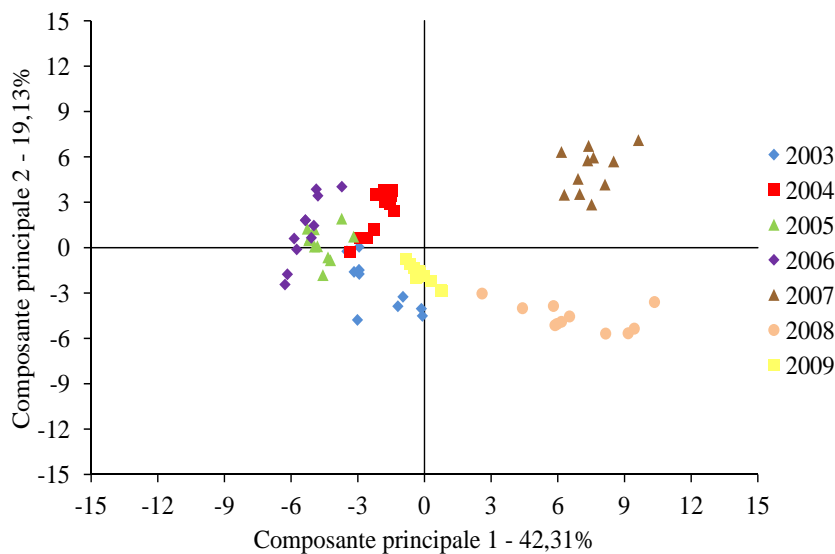


FIGURE 4.3 – Position des individus sur le plan constitué des deux premières composantes principales

Chaque millésime forme un groupe séparé. Pour la plupart, ils sont positionnés sur l'axe 1 et étendus sur l'axe 2. Sur la gauche de la figure, les millésimes 2006, 2005 et 2004 sont regroupés sur l'axe 1 et centrés sur l'axe 2. Les millésimes 2003 et 2009 sont proches de cet ensemble, plus proches du centre de l'axe 1 et négatifs sur l'axe 2. 2007 est positif sur les deux axes, nettement séparé des autres millésimes. Enfin, tous les individus du millésime 2008 sont dans le quart de plan positif sur l'axe 1 et négatif sur l'axe 2. Ce millésime est nettement séparé des 6 autres. Le nuage de points représentatif de ce millésime est, *a contrario* des autres millésimes, très étendu sur l'axe 1.

Ce plan met donc en avant un regroupement des individus selon le millésime pendant lequel ils ont été produits puis une répartition des cépages au sein du nuage de points représentant le millésime. Il y a donc probablement un type de climat par millésime qui,

Application 1 : Etude du climat d'un domaine sur plusieurs années entre la floraison et la véraison

TABLE 4.5 – Moyenne ou somme des variables climatiques principales entre le 01 juin (jour 152) et le 15 juillet (jour 196)

	$TM_m$ (°C)	$RR_s$ (mm)	$ETP_m$ (mm)	$RG_m$ (J.cm <sup>-2</sup> )	$UM_m$	$S5M_m$ (°C)
2003	23,9	5,5	6,0	2521,8	65,0	22,9
2004	21,9	10,5	6,2	2553,5	58,4	22,2
2005	22,7	22,5	6,1	2684,1	60,6	24,0
2006	22,9	3,0	6,0	2684,0	63,9	24,8
2007	20,8	49,0	5,1	2426,4	71,3	21,4
2008	20,8	45,0	5,8	2524,0	65,5	22,6
2009	22,3	45,0	5,9	2514,3	60,2	23,5

TM : température moyenne, RR : somme de précipitations, ETP : évapotranspiration potentielle, RG : rayonnement global, UM : humidité moyenne de l'air, S5M : température moyenne du sol à 50 cm de profondeur, m : moyenne, s : somme

pour la base de données étudiée, peut être très différent d'un millésime à l'autre. Cette typologie climatique annuelle détermine une position globale de chaque individu qui est ensuite ajustée en fonction de chaque cépage. Ces ajustements sont la représentation de l'atteinte des stades phénologiques à des dates différentes entre les cépages.

Le tableau 4.5 donne la moyenne (i) de la température moyenne de l'air (TM) et du sol à 50 cm de profondeur (S5M), (ii) de l'évapotranspiration potentielle (ETP), (iii) du rayonnement global (RG), (iv) de l'humidité moyenne de l'air (UM) et (v) la somme des précipitations (RR) entre le 01/06 (jour 152) et le 15/07 (jour 196). Ces dates sont proches des dates de floraison et véraison observées pour les différents cépages et les différents millésimes, données dans le tableau 4.1.

La température moyenne ( $TM_m$ ) de l'air des millésimes 2007 et 2008 est nettement plus faible que celle des autres millésimes. 2007, 2008 et 2009 présentent un cumul de précipitations ( $RR_s$ ) quasiment identique et supérieur à celui des autres millésimes. 2007 présente aussi la valeur de température du sol, de rayonnement global et d'ETP la plus basse et la valeur d'humidité de l'air la plus élevée. 2007 est donc un millésime plus frais, humide et moins ensoleillé que les autres millésimes sur cette période. 2008 est lui aussi plus frais et pluvieux mais le rayonnement global, la température du sol ou l'humidité de l'air sont proches des moyennes des autres millésimes. 2009 est pluvieux mais chaud et ensoleillé. Il présente une forte évapotranspiration ( $ETP_m$ ) et une faible humidité de l'air ( $UM_m$ ), comme les autres millésimes. Enfin, sur cette période, 2003, 2004, 2005 et 2006 présentent des moyennes climatiques semblables, air et sol chaud, peu de pluie et de forts évapotranspiration et rayonnement global.

Ces moyennes démontrent principalement une dualité au niveau de la température de l'air et les précipitations entre 2007 et 2008 et tous les autres millésimes. De plus, 2007 est le millésime avec le climat le plus différent des autres, plus frais, humide, pluvieux et moins ensoleillé. Cette dualité se retrouve sur l'axe 1. Cet axe permet donc de différencier les années de climat chaud, sec et ensoleillé des années de climat frais et pluvieux.

TABLE 4.6 – Temps entre floraison et véraison et position sur l'axe 2 par millésime

2003		2004		2005		2006		2007		2008		2009	
FV	CP2	FV	CP2	FV	CP2	FV	CP2	FV	CP2	FV	CP2	FV	CP2
47	-0,24	45	3,50	47	1,22	40	4,04	50	7,11	55	-3,60	49	-0,73
47	0,05	45	3,80	47	1,26	42	3,44	51	6,73	56	-3,03	50	-1,08
53	-1,48	46	3,77	48	1,92	42	3,87	51	6,33	56	-3,85	51	-1,35
53	-1,60	47	3,45	50	0,57	45	1,47	54	5,96	57	-4,55	52	-1,69
53	-1,48	49	3,06	51	0,49	46	1,82	55	5,78	58	-4,00	52	-1,56
53	-3,24	49	2,88	52	0,08	46	1,82	55	5,70	58	-4,90	53	-1,90
54	-1,74	51	2,41	52	0,73	48	0,66	59	4,55	59	-5,04	54	-1,93
55	-3,87	52	1,17	52	0,10	50	0,61	61	4,17	60	-5,13	54	-2,18
57	-4,03	53	0,67	53	-0,64	51	-0,11	63	3,50	62	-5,35	55	-2,01
58	-4,51	53	0,65	54	-0,83	54	-1,76	63	3,55	63	-5,68	56	-2,89
63	-4,77	56	-0,29	58	-1,81	56	-2,44	65	2,86	63	-5,66	56	-2,84

FV : nombre de jours entre floraison et véraison, CP2 : Position de l'individu sur la seconde composante principale

Sur l'axe 2, le millésime 2007 est séparé d'un groupe constitué des millésimes 2003, 2008 et 2009. Les valeurs fournies dans le tableau 4.5 ne permettent pas de séparer le millésime 2007 du groupe des trois millésimes. Aucune conclusion *a priori* à propos de la position des millésimes ne peut donc être tirée sur cet axe.

Le tableau 4.6 donne, pour chaque millésime, la correspondance pour chaque individu entre le nombre de jours entre la floraison et la véraison et la position sur l'axe 2.

Globalement, pour chaque millésime, la position de chaque cépage sur l'axe 2 est d'autant plus négative que le temps entre floraison et véraison est long. L'axe 2 pourrait donc être un axe qui permet, au sein d'un même millésime, de séparer les individus selon la longueur du temps passé entre la floraison et la véraison.

Pour un même nombre de jours entre floraison et véraison, les dates d'atteinte des stades peuvent varier de plusieurs jours. Par exemple, comme on peut le voir dans le tableau 4.1, en 2005, les cépages Chardonnay, Gamay et Marselan sont tous caractérisés par une durée de 52 jours entre floraison et véraison. Le Gamay a atteint la floraison le 147ème, le Chardonnay le 150ème et le Marselan le 154ème jour. Les dates de véraison sont respectivement le 200ème, le 203ème et le 207ème jour. Le Gamay a donc subi le même climat que le Marselan entre le 154ème jour et le 200ème jour mais un climat spécifique entre le 147ème et le 154ème jour. Le Marselan a quant à lui subi un climat spécifique entre le 200ème et le 207ème jour. Ces cépages n'ont donc pas vécu un climat totalement identique entre floraison et véraison. Cette différence peut donc expliquer que ces individus ne soient pas confondus.

La figure 4.4 présente le cercle des corrélations des paramètres étudiés dans l'ACP

## Application 1 : Etude du climat d'un domaine sur plusieurs années entre la floraison et la véraison

pour le plan constitué des deux premières composantes principales (61% de l'inertie du nuage). Les paramètres se répartissent principalement de part et d'autre de l'axe 1 et négativement sur l'axe 2.

Un grand nombre de paramètres est à proximité du cercle des corrélations. Les paramètres sont donc bien représentés dans ce plan.

La figure 4.4 propose également une interprétation des axes qui sera justifiée dans les paragraphes suivants.

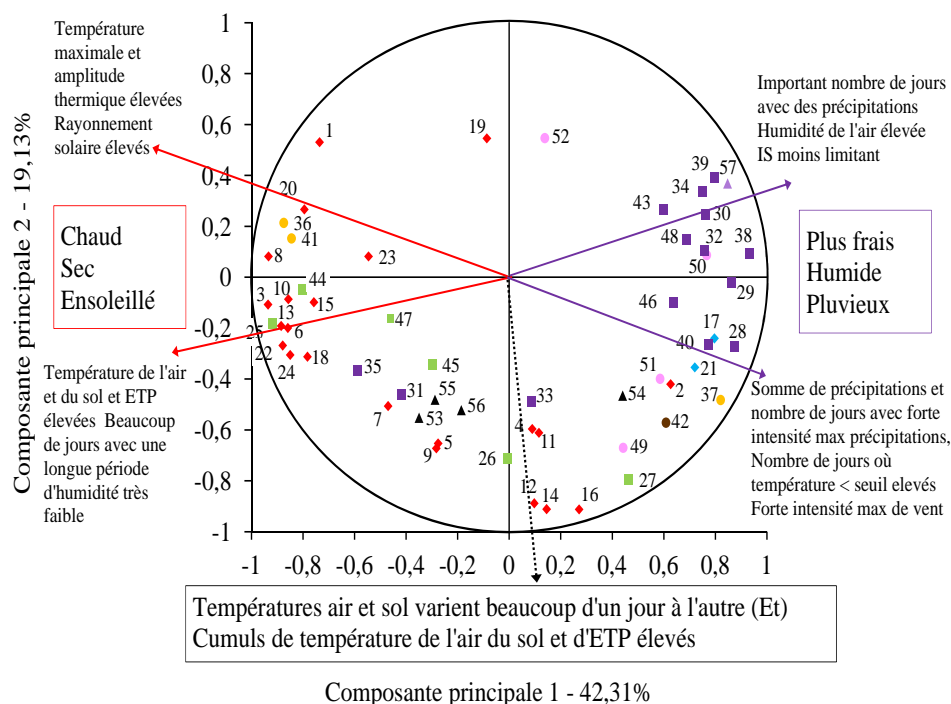


FIGURE 4.4 – Cercle de corrélation des variables - plan 1-2 - ACP réalisée sur les paramètres climatiques entre floraison et véraison

Sur l'axe 1, à chaque extrémité les paramètres se répartissent de chaque côté de l'axe.

Les paramètres corrélés négativement à l'axe 1 et légèrement positivement à l'axe 2 correspondent à des températures maximales de l'air élevées, des amplitudes thermiques journalières importantes, des températures moyennes optimales pour la croissance de la plante et un rayonnement solaire élevé. Cette direction, dans le sens négatif de l'axe 1 et légèrement positif de l'axe 2, représente donc des températures et rayonnement optimaux pour la croissance de la plante et la photosynthèse.

Les paramètres orientés négativement sur l'axe 1 et légèrement négativement sur l'axe 2 correspondent à des températures de l'air et du sol élevées, que ce soit en moyenne ou selon des compteurs de jours où la température a dépassé un seuil fixé, les paramètres de transpiration élevée de la plante et un air très sec. L'ensemble de ces paramètres permettent de caractériser une direction qui représente un climat chaud et sec, favorisant l'évapotranspiration.

Une simplification possible du sens négatif de l'axe 1 est un climat chaud, sec et

ensoleillé entre la floraison et la véraison, favorable à la croissance de la plante et de la baie et à la photosynthèse.

Les paramètres corrélés positivement à l'axe 1 et légèrement positivement à l'axe 2 correspondent à une humidité de l'air élevée, un grand nombre de jours pluvieux et un indice de sécheresse moins limitant. Cette direction représente donc un climat humide et pluvieux entre floraison et véraison.

Les paramètres corrélés positivement à l'axe 1 et légèrement négativement à l'axe 2 représentent des précipitations élevées, des températures fraîches et une forte intensité maximale du vent. Cette direction caractérise donc un climat plus frais et pluvieux.

L'axe 1 est donc un axe qui permet d'opposer les climats chauds, secs et ensoleillés et les climats plus frais, humides et pluvieux.

L'axe 2 est principalement déterminé par les paramètres se projetant négativement sur celui-ci.

Cet axe est corrélé négativement avec les paramètres correspondants à des cumuls de température, indices bioclimatiques ou sommes de températures, le cumul d'évapotranspiration et les paramètres mettant en avant une importante variabilité des températures entre les jours.

Il est donc possible que cet axe permette de différencier, pour des températures, des sommes de précipitations ou des rayonnements solaires semblables, les individus pour lesquels la période de floraison à véraison est plus longue des individus pour lesquels elle est plus courte.

La figure 4.5 présente à nouveau la position des individus dans la carte factorielle constituée des deux premières composantes principales de l'ACP en cours d'étude. Les directions mises en avant grâce à l'analyse du cercle des corrélations sont représentées sur la carte.

Les millésimes 2006, 2005 et 2004 puis 2003 et 2009 sont négatifs sur l'axe 1. Ils sont donc caractérisés par un climat plus chaud, sec et ensoleillé entre la floraison et la véraison que les millésimes 2007 et 2008. Dans le sens opposé, 2007 et 2008, positifs sur l'axe 1, présentent un climat plus frais, humide, pluvieux et moins ensoleillé. Lors de l'analyse du tableau 4.5, il avait été noté une dualité entre les millésimes 2007 et 2008 et tous les autres millésimes. Les millésimes 2007 et 2008 étaient effectivement plus pluvieux et frais que les autres. La position des millésimes sur l'axe 1 est donc cohérente avec les moyennes du tableau 4.5.

Les millésimes 2006, 2005 et 2004 sont très peu étalés sur l'axe 1 et plus étendus sur l'axe 2. Sur ces millésimes, les cépages ont subi un type unique de climat, chaud, sec et ensoleillé. Les individus sont ensuite étalés selon le nombre de jours entre floraison et véraison, nombre qui fait varier le cumul des températures. Comme il avait été noté grâce au tableau 4.6, les cépages à l'extrémité basse des nuages de points seront ceux pour lesquels le temps entre floraison et véraison est le plus long. Il y aura ensuite un

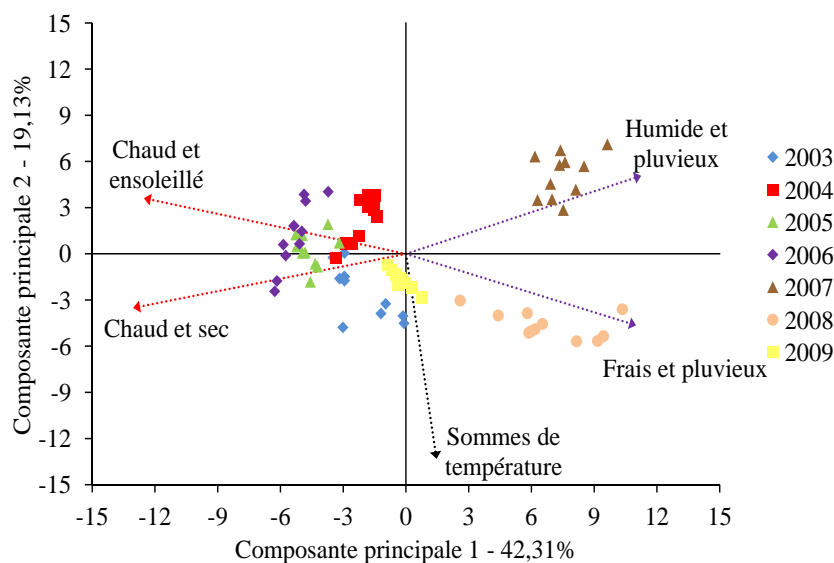


FIGURE 4.5 – Carte factorielle des individus analysés dans l'ACP avec les directions mises en avant

gradient de temps entre floraison et véraison, avec des individus positionnés de plus en plus haut dans le nuage de points.

Le millésime 2003 est séparé en deux nuages de points distincts. Le premier est, comme les millésimes 2006, 2005 et 2004, positionné sur l'axe 1 et étendu sur l'axe 2. Le second est positionné, négativement, sur l'axe 2 avec peu de variabilité sur celui-ci et plus étendu sur l'axe 1. Cette partie du nuage de points peut donc être constituée d'un ensemble d'individus avec des cumuls de température semblables mais des températures, rayonnement solaire et ETP différents.

Le millésime 2009 est centré sur l'axe 1. C'est un millésime chaud et ensoleillé mais aussi pluvieux (Tableau 4.5) entre floraison et véraison. Sa position est donc compensée entre les deux actions antagonistes des deux extrémités de l'axe 1. Sur l'axe 2, les individus sont à nouveau répartis en fonction du temps passé entre floraison et véraison.

En positif sur l'axe 1, le millésime 2007 est orienté par la direction représentant un temps humide et pluvieux tandis que le millésime 2008 est orienté par la direction caractérisant un temps plus frais et pluvieux. Ces deux millésimes, de comportement d'abord opposé aux 5 autres millésimes, ont donc chacun un comportement spécifique. A nouveau, sur l'axe 2, les individus se séparent en fonction du temps entre floraison et véraison.

Chaque millésime a donc un comportement propre sur ce plan. Pour la période allant de la floraison à la véraison, une typologie de chaque millésime, comparée aux autres millésimes, et ajustée à chaque cépage, peut facilement être proposée à partir de ces observations. Puis une typologie climatique propre à chaque individu peut être définie pour un millésime ou entre les millésimes. Ces typologies peuvent ensuite être comparées entre elles et des analogies entre individus quels que soient le cépage et le millésime peuvent être proposées.

Les deux premières composantes principales peuvent facilement être décrites grâce à un ensemble de paramètres. Elles permettent de séparer (i) les millésimes entre eux pour l'axe 1 et (ii) les cépages au sein d'un millésime pour l'axe 2.

L'insertion d'un nouveau millésime, en conservant les même 11 cépages, dans une telle analyse permettrait de connaître rapidement ses chaleur, pluviométrie et ensoleillement relatifs par rapport aux millésimes déjà présents dans la base de données.

### **Conclusions de l'analyse du plan constitué de la troisième et de la quatrième composante principale**

Une analyse similaire à celle présentée dans la section 4.3.2 a été effectuée sur le plan constitué de la troisième et de la quatrième composante principale. Ce plan représente 22,4% de variance totale expliquée.

Corrélé à des cumuls de températures et des indices bioclimatiques, l'axe 3 permet de différencier, au sein de chaque millésime, les cépages en fonction de la longueur de l'intervalle entre floraison et véraison. Les nuages de points représentatifs de chaque millésime sont très étendus sur cet axe.

L'axe 4 est déterminé par l'opposition entre (i) des températures maximales du sol et de l'air élevées associées à de fortes variations de la température d'un jour à l'autre ainsi qu'à un grand nombre de jours où la température moyenne a été raisonnable et (ii) un nombre de jours où l'air a rarement été extrêmement sec et d'importantes variations d'humidité de l'air d'un jour à l'autre. Cet axe permet de différencier le millésime 2004 (températures raisonnables) et le millésime 2006 (humidité variable). Les autres millésimes sont confondus sur cet axe.

Ce plan ne permet pas de différencier les millésimes entre eux, hormis 2004 et 2006 mais présente une forte différenciation des cépages au sein d'un millésime. Ce plan sera donc utilisé principalement pour différencier le vécu climatique des cépages au sein de chaque millésime et de comparer ces vécus pour comprendre des différences spécifiques entre cépages pour un millésime donné.

## **4.4 Etude comparative de l'ACP avec indices bioclimatiques pondérés**

Dans une seconde analyse, 54 variables ont été prises en compte. Les indices bioclimatiques ayant permis cette analyse sont les indices pondérés. Les trois paramètres correspondant à des sommes de descripteurs ont été retirés pour l'analyse. Leur valeur pondérée, soit la moyenne de la séquence temporelle sur la base temporelle, était déjà prise en compte dans l'analyse.

A nouveau, les 4 premières composantes principales ont été conservées. Elles représentent respectivement 50.3%, 18.25%, 11.24% et 8.36% de variance expliquée. L'ensemble de ces 4 axes permet donc d'expliquer 88.15% de la variance totale entre les 11 cépages pour les 7 millésimes étudiés. Les pourcentages de variance expliquée par chaque

axe sont du même ordre de grandeur que ceux de l'ACP réalisée avec les paramètres prenant en compte la précocité des cépages.

Dans cette partie, seulement la comparaison des résultats entre les plans de chacune des ACP sera présentée. L'analyse en détail de l'ACP, très semblable à celle de l'ACP réalisée avec les indices bioclimatiques non pondérés sera occultée.

#### 4.4.1 Analyse comparative du plan constitué des deux premières composantes principales

L'analyse de ce plan lors de la première analyse avait mis en avant :

- Une différenciation des millésimes sur l'axe 1 en fonction de la chaleur, de l'ensoleillement et de la sécheresse de chacun.
- Une différenciation des cépages au sein de chaque millésime sur l'axe 2 principalement imposée par les sommes de descripteurs et les indices bioclimatiques

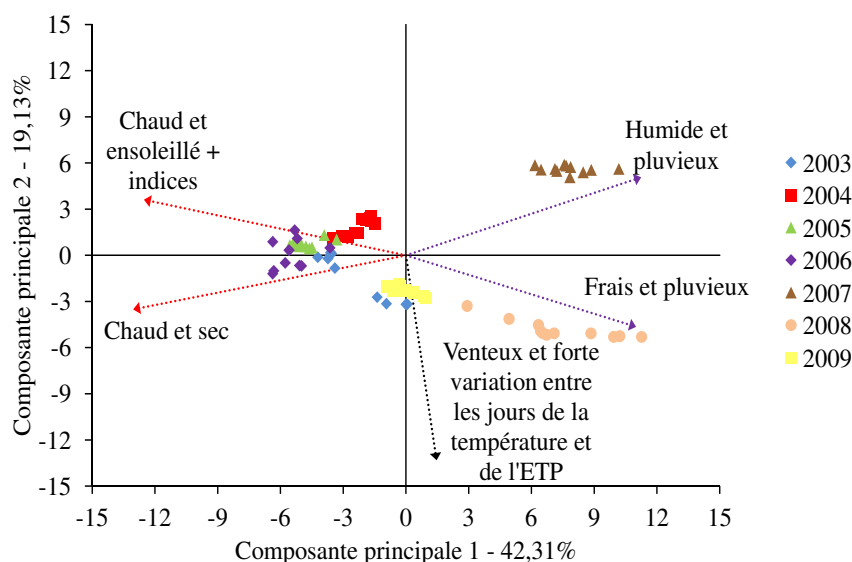


FIGURE 4.6 – Carte factorielle des individus analysés dans l'ACP sans paramètres prenant en compte la précocité des cépages avec les directions déterminées par les paramètres mises en avant

La figure 4.6 présente le plan factoriel issu de la seconde analyse. Les axes déterminés par l'analyse du cercle de corrélations sont aussi présentés sur cette figure.

Sur ce plan, les millésimes se différencient à nouveau sur l'axe 1 mais sont plus étalés. Au sein d'un millésime, il n'est plus possible de différencier les cépages sur l'axe 2.

Sur l'axe 1, les mêmes paramètres sont retrouvés de part et d'autre de celui-ci. Les indices bioclimatiques issus de sommes de températures se confondent aux paramètres représentant des températures chaudes.

Sur l'axe 2; les paramètres représentant des sommes de descripteurs ou les indices bioclimatiques de sommes de températures ne sont plus représentés. Par ailleurs, les autres paramètres projetés négativement sur cet axe lors de l'analyse de l'ACP réalisée avec les indices bioclimatiques non pondérés sont à nouveau projetés négativement.



Sur ce plan, on voit donc que les paramètres prenant en compte la précocité des cépages n'ont pas créé un axe artificiel sur lequel les cépages se différencieraient les uns des autres. La disposition des nuages de points sur cette seconde analyse indique par contre que ces paramètres ont permis une meilleure différenciation des cépages, différenciation qui n'est pas due à des différences climatiques mais à une précocité des stades phénologiques changeant entre les cépages.

Grâce à ce plan, le vécu climatique global de chaque millésime peut donc être comparé à celui des six autres millésimes. De plus, pour un même cépage, en travaillant sur la projection de chaque individu, ce vécu climatique de chaque millésime peut à nouveau être comparé. Enfin, au sein d'un millésime, le vécu climatique de chaque cépage peut être comparé à celui des autres cépages. Pour chaque millésime, il faudrait maintenant analyser en profondeur le vécu climatique de chaque cépage en fonction des différents paramètres disponibles afin de comprendre la position de chacun au sein du nuage de points représentant le millésime.

Ce plan ne contredisant pas les informations observées lors de la première analyse de données, travailler avec les indices bioclimatiques non pondérés ne semblent pas influencer un artefact mathématique non prévu dans le cas présent.

#### **4.4.2 Analyse comparative du plan constitué de la troisième et de la quatrième composante principale**

L'analyse de ce plan lors de la première ACP avait mis en avant :

- Une différenciation des cépages au sein de chaque millésime sur l'axe 3, cet axe étant déterminé par des sommes de températures et des indices bioclimatiques issus de sommes de températures
- Une opposition entre (i) des températures maximales du sol et de l'air élevées avec de fortes variations de températures entre les jours et des températures moyennes raisonnables et (ii) une humidité de l'air très variable mais rarement extrêmement sec. Sur cet axe, 2004 était opposé à 2006. 2006 était orienté par la direction des paramètres représentant l'humidité de l'air variable tandis que 2004 était caractérisé par des températures moyennes raisonnables et des températures élevées chaudes.

La figure 4.7 présente le plan factoriel issu de cette seconde analyse pour la troisième et la quatrième composante principale. Les axes déterminés par l'analyse du cercle des corrélations sont indiqués sur ce plan.

Sur ce second plan, l'axe 3 est déterminé par une opposition entre des températures chaudes mais modérées, c'est à dire n'ayant pas atteint des valeurs où l'activité photosynthétique de la vigne est inhibée, et des jours peu secs. L'axe 4 est quant à lui déterminé par une opposition entre des jours dans forte humidité et les paramètres correspondant à une humidité de l'air très variable d'un jour à l'autre.

Les paramètres cités lors de la première analyse se retrouvent donc à nouveau. Ils sont par contre répartis homogènement sur ce plan. Le troisième axe de l'ACP réalisée

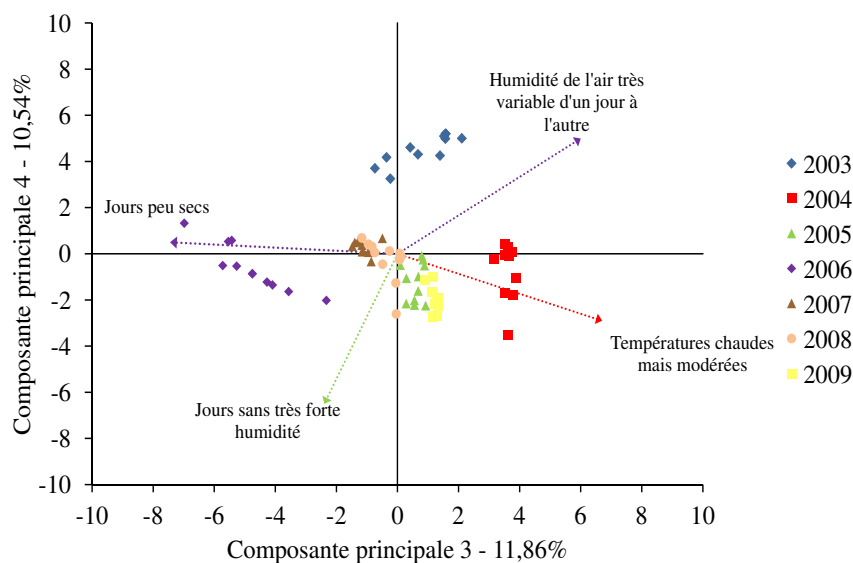


FIGURE 4.7 – Carte factorielle des individus analysés dans l'ACP sans paramètres prenant en compte la précocité des cépages avec les directions déterminées par les paramètres mises en avant pour la troisième et la quatrième composante principale

avec les paramètres prenant en compte la précocité de chaque cépage était exclusivement déterminé par ces paramètres. Cet axe était donc un axe artificiel. Le fait d'avoir oté ces paramètres a permis de mettre en avant une variabilité climatique entre les individus qui était auparavant présente uniquement sur l'axe 4.

2004 est toujours orienté par l'axe correspondant à des températures chaudes mais modérées. 2006 est quant à lui orienté par la direction déterminée par les paramètres représentant des jours avec peu de temps très sec. 2003, se différencie sur ce plan, ce qui n'était pas le cas lors de la première analyse, et est orienté par la direction déterminée par les paramètres représentant une humidité variable d'un jour à l'autre. Une nouvelle information est donc disponible sur ce plan. Des informations nouvelles sur les paramètres ayant eu une influence globale sur le climat de chaque millésime sont donc disponibles. L'absence des variables prenant en compte la précocité de chaque cépage a donc permis de mettre en avant des différences de vécu climatique de chaque millésime et chaque cépage.

Sur ce plan, l'utilisation de paramètres prenant en compte la précocité de chaque cépage a donc impliqué la création d'un axe artificiel, artefact mathématique des choix réalisés. Pourtant la position des millésimes, toujours cohérente entre les deux analyses permet de conclure sur l'importance des paramètres climatiques choisis et l'intérêt de l'analyse réalisée.

Sur ce plan, le vécu climatique de chaque millésime peut être comparé à celui des autres millésime et des informations nouvelles sur l'humidité de l'air sont disponibles. Ces informations sont des informations supplémentaires à la connaissances empirique apportée par les producteurs. A nouveau, il faudrait analyser chaque millésime pour comprendre le vécu climatique spécifique à chaque cépage puis analyser le vécu clima-

tique de chaque millésime pour chaque cépage pour prendre des décisions à partir de ces analyses.

Compte tenu de la proximité des résultats obtenus, l'utilisation de paramètres prenant en compte la précocité des cépages ne semble pas altérer les informations mises en avant par ces analyses. Les deux types d'analyse pourraient donc être entrepris pour donner une caractérisation globale du climat de chaque cépage pour chaque millésime entre la floraison et la véraison.

## 4.5 Conclusion : Bilan de l'étude comparative du climat entre cépages et millésimes

Dans cette application, le comportement climatique entre floraison et véraison subi par 11 cépages pour 7 millésimes d'un domaine du Languedoc-Roussillon a été étudié.

Plusieurs modèles de chaque grandeur de mesure ont été proposés. Ces modèles visent à agréger les informations contenues dans les séquences temporelles représentatives de chaque individu en une et une seule valeur sur toute la base temporelle d'étude, soit entre floraison et véraison. Les différents modèles permettent d'évaluer le comportement moyen de chaque grandeur et sa variabilité.

Les 77 individus ont ensuite été comparés grâce à des analyses en composantes principales utilisant pour variables les paramètres climatiques calculés en agrégeant les séquences temporelles. Grâce à ces ACP, entre floraison et véraison, le climat global de chaque millésime en comparaison aux autres millésimes, peut être analysé. Puis le climat spécifique vécu par chaque cépage pour un millésime ou entre millésimes peut être étudié. De telles analyses nécessitent d'étudier en profondeur chaque nuage de points. L'objectif du travail présenté dans cette thèse étant de mettre en avant les potentialités d'une méthode de comparaison de données asynchrones en vue de la réalisation d'analogies entre individus, ces analyses n'ont pas été réalisées.

A partir des plans factoriels obtenus, plusieurs niveaux d'analogies ont pu être identifiés :

- Analogies entre millésimes au niveau du climat global de chacun
- Analogies entre millésimes pour un cépage donné
- Analogies entre cépages pour un millésime donné

Les millésimes les plus proches, tous cépages confondus ou par cépage peuvent être différents.

Mais, il faudrait pouvoir réaliser ces analogies sur la base des 4 composantes principales identifiées.

En extrayant la position de chaque individu sur chacune des quatre composantes principales, 83% de la variabilité climatique sur la période floraison à véraison peut être représentée dans un espace de dimension 4.

Chaque composante principale est décrite par une combinaison de paramètres propre

aux spécificités de la base de données étudiée. Ces composantes principales représentent donc un méta-paramètre expliquant une part de la variabilité climatique dans un ensemble de parcelles et pour un ensemble de millésimes fixe. Suivant la base de données, chaque méta-paramètre a sa définition propre. Un tel ensemble de méta-paramètres pourrait être défini pour chaque période entre deux stades de développement clefs de la vigne.

A partir du calcul de ces méta-paramètres, il est possible de définir :

- Les millésimes de climat les plus proches sur la base des 4 composantes principales
- Les millésimes de climat les plus proches pour un cépage donné sur la base des 4 composantes principales
- Les cépages de climat les plus proches pour un même millésime sur la base des 4 composantes principales
- Les cépages de climat les plus proches tous millésimes confondus sur la base des 4 composantes principales

Dans le cadre de la mise en place d'un outil d'aide à la décision, une telle analyse du climat pourrait permettre de caractériser le climat de chaque millésime pour chaque cépage en utilisant des indices habituellement utilisés par les viticulteurs. Une telle caractérisation permettrait de confirmer les connaissances empiriques du producteur. De plus, une telle analyse pourrait permettre de rapprocher le millésime en cours du ou des millésimes dont il est le plus proche d'un point de vue climatique. Ces rapprochements pourraient être effectués en séparant les cépages ou non. Ces rapprochements sont utilisés par les viticulteurs de façon empirique pour prendre des décisions en connaissant le vécu climatique passé du millésime en cours et en présageant la réponse de la vigne au climat futur et à l'environnement. L'utilisation de ces travaux pourrait donc permettre, de façon automatisée, au viticulteur de caractériser rapidement le millésime sur lequel il est en train de travailler comparativement aux autres millésimes.

Ces rapprochements permettent donc d'envisager très rapidement la réaction de chaque cépage pour chaque type de millésime.

A partir de ces analogies climatiques, des hypothèses sur le comportement de la plante pourraient être envisagées. Ces hypothèses prendraient en compte la réponse de la vigne et de son raisin aux spécificités du millésime en cours. Un calendrier prévisionnel de travaux pourrait être envisagé en supposant que le climat du millésime reste dans la lignée de celui dont il est le plus proche à un moment donné.

# Chapitre 5

## Application 2 : Modélisation de l'augmentation du pH pendant la maturation

---

### Sommaire

<b>5.1</b>	<b>Introduction : l'évolution du pH au cours de la maturation</b>	<b>62</b>
<b>5.2</b>	<b>Matériel et méthodes</b>	<b>63</b>
5.2.1	Matériel expérimental	63
5.2.2	Définition d'un temps thermique pour chaque jour	64
5.2.3	Modélisation paramétrique de l'augmentation du pH dans les baies de raisin	65
5.2.4	Utilisation des modèles	66
<b>5.3</b>	<b>Résultats et discussion</b>	<b>66</b>
5.3.1	Caractérisation du mésoclimat de chacun des trois domaines choisis pour l'utilisation du modèle en extensif	66
5.3.2	Adéquation de l'exemple aux problèmes de comparaison des séquences temporelles	67
5.3.3	Validation d'un modèle d'augmentation du pH sur les 53 individus de la base de données	68
5.3.4	Comparaison graphique de la cinétique de pH des trois parcelles étudiées	71
<b>5.4</b>	<b>Conclusion : Un modèle avec du potentiels pour comparer les cinétiques d'augmentation du pH dans différentes parcelles</b>	<b>75</b>

---

Dans ce chapitre, les potentialités d'utilisation de tout type de séquence temporelle avec la méthode proposée pour comparer l'évolution du pH entre parcelles et millésimes sont évaluées, même dans les cas où la séquence temporelle ne couvrirait pas parfaitement l'ensemble de la cinétique : absence possible de quelques mesures en début et / ou en fin de cinétique.

## 5.1 Introduction : l'évolution du pH au cours de la maturation

Le potentiel hydrogène (pH) mesure l'acidité d'une solution. Ses valeurs sont comprises entre 0 et 14. Il est adimensionnel. Plus une solution est acide et plus sa valeur de pH est basse [22] et inversement.

Dans les baies de raisin, le pH dépend d'une relation complexe entre la concentration de chaque acide, sa (ou ses) constantes de dissociation et la concentration en cations dans le moût [19]. Il n'est donc pas toujours corrélé avec l'abondance des acides dans les baies. 90 % de l'acidité totale dans les baies est due à deux acides : l'acide malique et l'acide tartrique [11, 19]. Pourtant, le pH d'une baie dépend surtout des teneurs en acide tartrique et en potassium [80].

La teneur en acide tartrique de chaque baie augmente au cours de la phase de croissance herbacée des raisins (de la floraison à la véraison) puis reste constante pendant la maturation. Par phénomène de dilution, la concentration en acide tartrique va donc légèrement diminuer pendant la maturation [11]. Sur la même période, le raisin s'enrichit en potassium. La cinétique d'augmentation de la teneur en potassium dans les baies suit une courbe semblable à celle d'accumulation des sucres dans les baies [80]. La cinétique d'accumulation des sucres dans les baies prend la forme d'une sigmoïde entre la véraison et la récolte [19, 25, 40, 75]. De la véraison à la récolte, le bilan acido-basique entre l'acide tartrique et le cation potassium est donc de moins en moins acide. Le pH augmente donc au cours de la maturation. L'augmentation du pH suit une courbe de forme semblable à celle représentant l'accumulation des sucres, soit une courbe de forme sigmoïdale [18, 52].

La fonction logistique permet de représenter l'évolution d'une population au sein d'un milieu de taille non extensible [96]. Elle est de forme sigmoïdale. En viticulture, elle a déjà été utilisée pour modéliser l'accumulation de sucres dans les baies de raisin entre la floraison et la récolte [11, 25, 40, 68]. Il va donc être possible d'essayer de modéliser l'augmentation du pH au cours de la maturation en utilisant la fonction logistique et en adaptant celle-ci aux spécificités de cette grandeur. A notre connaissance, une telle approche n'a jamais été proposée dans la littérature.

## 5.2 Matériel et méthodes

### 5.2.1 Matériel expérimental

#### Sites d'étude

Cette étude a été réalisée sur les données provenant d'un ensemble de domaines de la région viticole de la Napa Valley en Californie. La Napa Valley est une vallée de production longue de 40 kilomètres et large de 10 kilomètres environ dont la pointe sud est située à une vingtaine de kilomètres au nord de la baie de San Francisco. La géographie et la géologie particulières de cette vallée impliquent que plusieurs mésoclimats sont présents dans la région, malgré sa surface limitée [33]

Cette étude a été réalisée à partir de 53 parcelles plantées du cépage Cabernet-Sauvignon. Elles ont été analysées uniquement sur le millésime 2011. Les parcelles choisies étaient situées dans des domaines de production différents de la Napa Valley. Les raisins étaient donc produits sur des sols très différents et dans des zones présentant des mésoclimats très différents. Les parcelles étaient irriguées sur tout le cycle végétatif. La densité de plantation variait entre 1500 et 7500 pieds par hectare. Les plantes étaient palissées, conduites en cordon de royat ou guyot. Les parcelles ont été plantées entre 1985 et 2008.

#### Données utilisées

Pour chaque individu, les mesures ont débuté avant ou au début de la véraison et se sont échelonnées jusqu'à la fin de la maturation. Un seul millésime,  $T$ , 2011, a été considéré. Chaque parcelle représentait une localisation unique,  $s_j$  (avec  $j \in [1; 53]$ ). D'après les producteurs dont les bases de données étaient utilisées, les mesures étaient réalisées pour représenter l'ensemble de la parcelle et évaluer son évolution globale au cours de la maturation. Chaque échantillon était donc réalisé en prenant des baies de différents pieds, à différents endroits de la grappe et vérées ou non vérées.

Pour le millésime 2011 et les 53 parcelles étudiées, la mesure,  $x_j$ , de pH dans les baies associée à sa date de mesure,  $t_j$ , et à son temps thermique de mesure dont le calcul est donné en section 5.2.2,  $GDD(t_j)$ , a été extraite de la base de données de traçabilité contenant l'ensemble des valeurs de relevés. Une base de données a été créée en regroupant les données selon les définitions de la partie 3.1. Pour chaque individu, la base temporelle,  $T_\Phi$ , a été définie comme l'intervalle entre la première mesure et la récolte. La base temporelle a ensuite été divisée en unités de temps calendaire,  $W_a$ , ou thermique,  $W'_a$ , depuis la première mesure jusqu'à la récolte.

Les données de description relatives à la base de données ont été regroupées dans le tableau 5.1. Le nombre de mesures des séquences temporelles représentant les différentes parcelles analysées a été indiqué pour chaque domaine. Au maximum 1 mesure par jour était réalisée. L'intervalle de temps entre deux mesures dépendait de chaque parcelle étudiée. La valeur d'indice héliothermique (IH) a été calculée selon la formule de Huglin [49]. La valeur d'indice de fraîcheur des nuits (IF) a été calculée selon la formule disponible dans l'article de Tonietto et Carbonneau [92].

TABLE 5.1 – Description des individus de la base de données - pH

Domaine	Nombre de parcelles	Nombre de mesures	IH	IF
2	3	14	2420,6	9,9
6	4	9 - 10	2333,7	10,3
9	7	13 - 16	2425,0	11,8
10	6	5 - 14	2320,9	10,4
11	1	12	2501,0	10,9
17	5	11 - 12	2336,8	9,6
19	4	11 - 12	2072,0	12,1
20	2	10 - 11	2415,3	11,8
23	6	10 - 12	2205,2	10,9
26	2	8	2034,6	10,6
27	1	6	2417,6	9,8
31	12	6 - 16	2137,5	14,1

## 5.2.2 Définition d'un temps thermique pour chaque jour

Pour chaque jour, une valeur de temps thermique a été définie en utilisant le modèle proposé par Parker et al. [76]. Le calcul du temps thermique faisait intervenir le calcul de la température moyenne journalière  $T_{moy}$  tel que défini dans la relation 5.1

$$T_{moy} = \frac{TN + TX}{2} \quad (5.1)$$

Avec  $TN$  : température minimale de l'air et  $TX$  : température maximale de l'air  
Le degré jour,  $^{\circ}j$ , de chaque jour était défini selon la relation 5.2.

$$^{\circ}j = \begin{cases} 0, & \text{si } T_{moy} \leq 0 \\ T_{moy}, & \text{si } T_{moy} > 0 \end{cases} \quad (5.2)$$

Enfin, le temps thermique, GDD (Growing Degree Day), de chaque jour a été calculé suivant la relation 5.3

$$\forall \text{ le jour } j, GDD(j) = \sum_{i=60}^{i=j} ^{\circ}j(i) \quad (5.3)$$

Dans le modèle proposé par Parker et al [76], le calcul des GDD est initialisé au jour 60 de l'année, soit le 01/03 pour les années non bisextiles comme 2011.

Pour chaque date de mesure,  $t_j$ , la valeur en GDD,  $GDD(t_j)$ , a donc été calculée et intégrée à la base de données de traçabilité contenant toutes les valeurs de relevés réalisés.



### 5.2.3 Modélisation paramétrique de l'augmentation du pH dans les baies de raisin

La fonction logistique a été adaptée pour modéliser l'augmentation du pH dans les baies de raisin depuis la véraison jusqu'à la récolte. Le modèle proposé est présenté dans l'équation 5.4.

$$x_{pH}(t) = x_{pH}(t_0) + \frac{pH_{max} - x_{pH}(t_0)}{1 + e^{\frac{Infl - t}{\tau}}} \quad (5.4)$$

où  $x_{pH}(t)$  représentait le niveau de pH dans les baies de raisin au temps  $t$ ,  $x_{pH}(t_0)$  correspondait au niveau de pH au moment du début de la véraison.

Le modèle paramétrique de l'équation 5.4 contenait 3 paramètres :  $pH_{max}$  qui représentait le niveau de pH maximal dans les baies,  $Infl$  qui correspondait à l'abscisse du point d'inflexion et  $\tau$  qui permettait d'ajuster la pente de la courbe au point d'inflexion. En plus des trois paramètres, la valeur réelle de la pente au point d'inflexion,  $Sl$ , était dérivée de l'équation 5.4 et calculée grâce à la formule 5.5.

$$Sl = \frac{pH_{max} - x_{pH}(t_0)}{4\tau} \quad (5.5)$$

La valeur d'initialisation de la cinétique,  $x_{pH}(t_0)$ , a été estimée en moyennant les niveaux de pH mesurés juste avant ou au moment du début de la véraison sur toutes les parcelles de Cabernet-Sauvignon de la base de données de traçabilité étudiée pour le millésime 2011. Cette moyenne était de 2,54.

Pour chaque individu, l'ajustement de la courbe définie équation 5.4 a été réalisée par la méthode des moindres carrés. L'algorithme d'optimisation utilisé était celui de Levenberg-Marquardt. La méthode a été implémentée avec la fonction *nlinfit* du logiciel MatLab v7.0 (The MathWorks Inc., Natick, MA).

Pour chaque individu, l'erreur quadratique moyenne, RMSE, d'ajustement du modèle aux descripteurs de la séquence temporelle a été définie grâce au calcul donné dans la formule 5.6.

$$\text{Soit } k \text{ le nombre total de descripteurs de } TS_I, \quad RMSE = \sqrt{\sum_{j=1}^{j=k} \frac{(x_j - \hat{x}_j)^2}{k}} \quad (5.6)$$

La RMSE représente l'écart moyen entre chaque descripteur de la séquence temporelle et l'estimation de ce descripteur grâce au modèle. Plus sa valeur est basse et plus la qualité d'ajustement du modèle est bonne.

En complément de cette valeur de RMSE, l'erreur moyenne d'ajustement du modèle a été calculée grâce à la formule donnée dans l'équation 5.7. Cette erreur correspondait à une évaluation pondérée du coefficient de variation entre la séquence temporelle et le

modèle.

$$\text{Soit } k \text{ le nombre total de descripteurs de } TS_I, \%_{\text{erreur}} = \frac{RMSE}{\sum_{j=1}^{j=k} \frac{x_j}{k}} \quad (5.7)$$

## 5.2.4 Utilisation des modèles

### Approche intensive

Pour chaque individu, les paramètres  $X_{s_{max}}$ ,  $Infl$  et  $Sl$  ont été estimés puis enregistrés dans une nouvelle base de données dite paramétrique ( $Z_{para}$ ).

### Approche extensive

Une base temporelle thermique,  $T'_\Phi$  unique a été définie. Les mesures de pH commencent avant la véraison. La base temporelle choisie était donc comprise entre 600 GDD avant le début de la véraison (-600) et 1500 GDD après le début de la véraison. Une discrétisation temporelle,  $\Delta'$ , homogène a été choisie. Chaque fenêtre,  $W'_a$ , correspondait à 20 GDD.

Pour chaque individu, la valeur de pH a donc été recalculée tous les 20 GDD. A partir de ces estimations, la cinétique d'augmentation du pH de chaque parcelle, en temps thermique, a été tracée. Les comparaisons ont été réalisées entre les courbes.

Pour cet exemple, uniquement 3 parcelles ont été étudiées. Elles provenaient chacune d'un domaine. Ces domaines étaient situés dans des zones différentes de la Napa Valley. Le sol et le mésoclimat de chacun des trois domaines était très différent de celui des deux autres. Leur numéro étaient le 7, le 16 et le 40.

La parcelle 7 appartenait au domaine 6, la parcelle 16 au domaine 10 et la parcelle 40 au domaine 19. Ces parcelles avaient été choisies aléatoirement. Un autre ensemble de trois parcelles aurait pu être choisi.

## 5.3 Résultats et discussion

Dans cette partie, l'équation du modèle proposée sera validée sur un grand nombre d'individus. Puis, la comparaison de trois parcelles sera réalisée à titre d'exemple du potentiel de la méthode proposée.

### 5.3.1 Caractérisation du mésoclimat de chacun des trois domaines choisis pour l'utilisation du modèle en extensif

La valeur de IH dans le domaine 19 (tableau 5.1) est nettement plus basse que celle calculée pour les domaines 6 et 10. La valeur de IF dans le domaine 19 (tableau 5.1) est nettement plus élevée que celle calculée pour les domaines 6 et 10. Le domaine 19 présente donc un mésoclimat nettement différent des deux autres domaines. Comparer l'augmentation du pH dans une parcelle de ce domaine à celle d'une parcelle des deux

autres domaines peut donc permettre d'envisager si l'effet du climat est important sur les différences d'augmentation du pH pour le millésime et les parcelles étudiées.

Le domaine 6 est situé dans la plaine de la vallée, au sud de celle-ci. Le domaine 10 est situé dans la plaine de la vallée plus au nord que le domaine 6. Les domaines 6 et 10, bien que présentant des valeurs d'indices bioclimatiques identiques, ne sont donc pas cultivés sur des sols identiques. Comparer l'accumulation du pH entre deux parcelles, chacune provenant d'un des deux domaines, permettra donc d'étudier si le type de sol est influant sur les différences d'augmentation du pH pour le millésime et les parcelles étudiées.

### 5.3.2 Adéquation de l'exemple aux problèmes de comparaison des séquences temporelles

La figure 5.1 présente la position des dates de mesures et les mesures effectuées sur chacune des trois parcelles utilisées pour l'étude comparative d'individus.

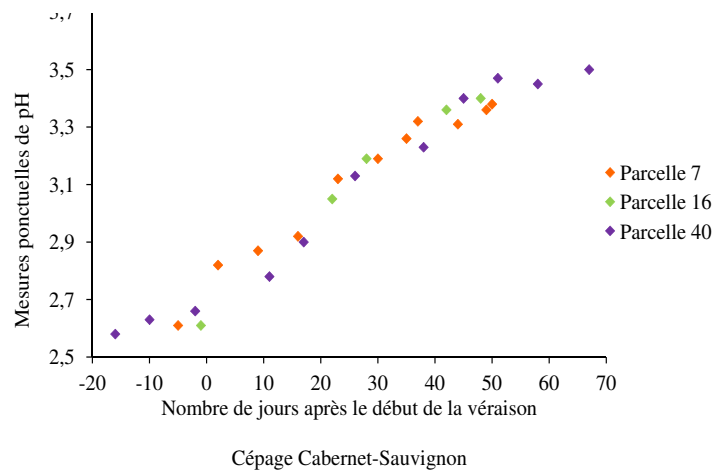


FIGURE 5.1 – Dates de mesures du pH des trois parcelles choisies

La parcelle 7 a fait l'objet de 11 mesures entre le 10ème jour avant véraison et le 50ème jour après véraison. La parcelle 16 n'a fait l'objet que de 5 mesures. Une mesure a été effectuée juste avant la véraison. les autres mesures se sont échelonnées entre le 20ème et le 50ème jour après la véraison. Enfin, la parcelle 40 a fait l'objet de 11 mesures selon une grille quasiment régulière entre le 20ème jour avant la véraison et le 70ème jour après la véraison.

La dimension de la séquence temporelle représentant la parcelle 16 est différente de celle des séquences temporelles représentant les parcelles 7 et 40. Sur la figure 5.1, il est aussi possible de voir que les descripteurs des trois séquences temporelles sont asynchrones. Les trois parcelles étudiées présentent donc l'ensemble des problèmes soulevés pour la comparaison d'individus présentés en partie 3.2.1.

Les trois domaines d'où proviennent les parcelles sont situés dans des mésoclimats différents (section 5.3.1). La température de l'air dans ces parcelles était donc très différente d'une parcelle à l'autre de part les mésoclimats de chaque domaine. Dû à ces

différences, pour un même jour après le début de la véraison, le cumul de températures (GDD) de chaque domaine sera donc différent. Les acides vont être détériorés et l'eau accumulée dans les baies de raisin en réponse à la température [9, 11, 16, 19, 52, 56, 72]. La vitesse d'augmentation du pH sera donc différente en fonction de la quantité de température accumulée dans chaque parcelle. Il faut donc vérifier si les relevés sont toujours asynchrones si l'échelle de temps est le calendrier thermique.

La figure 5.2 donne la position des mesures de chacune des trois parcelles étudiées après passage de l'échelle de temps en calendrier thermique. Malgré le changement d'échelle de temps, les données sont toujours asynchrones.

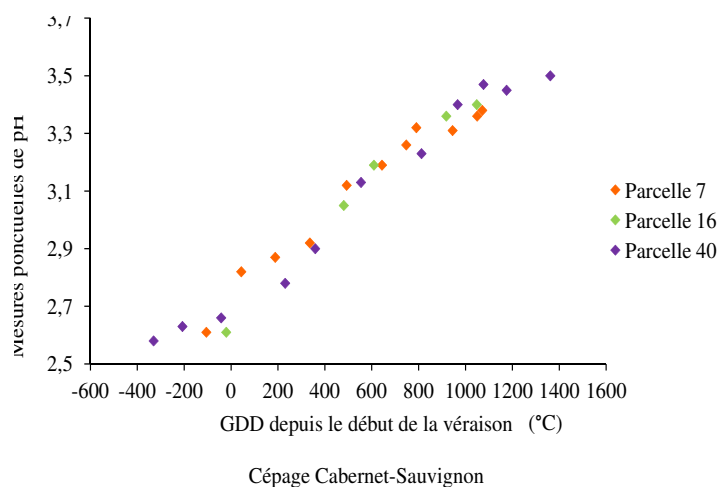


FIGURE 5.2 – GDD de mesures du pH des trois parcelles choisies

Quelle que soit l'échelle de temps considérée, les trois individus étudiés ont des séquences temporelles asynchrones. Deux séquences temporelles sont de dimension égale et la dernière est de dimension différente. Ces séquences temporelles constituent donc un bon exemple pour l'étude des problèmes de comparaison de données précitées. Cet exemple va donc permettre d'évaluer le potentiel de la méthode proposée pour leur résolution.

### 5.3.3 Validation d'un modèle d'augmentation du pH sur les 53 individus de la base de données

Le tableau 5.2 donne les statistiques de répartition des valeurs ajustées des trois paramètres définis dans l'équation 5.4 :  $pH_{max}$ ,  $Inf$  et  $\tau$ , de la pente réelle,  $Sl$ , de la courbe au point d'inflexion et de la valeur de RMSE et du pourcentage d'erreur entre la cinétique ajustée et les descripteurs de la séquence temporelle représentant chacun des 53 individus de la base de données.

#### Validation de la qualité globale de l'ajustement du modèle

Les valeurs de RMSE représentant la qualité d'ajustement du modèle sont comprises entre  $4,22 \cdot 10^{-3}$  et  $2,65 \cdot 10^{-1}$ .

Les valeurs de  $\%_{erreur}$  sont comprises entre 0.15 et 9.1 %. Ces valeurs indiquent que lors de l'ajustement du modèle, il y a au plus 10% d'écart entre les descripteurs de chaque

TABLE 5.2 – Répartition des paramètres ajustés lors de la modélisation du pH - Base de données pH - 53 individus

	$pH_{max}$	$Infl$ (j)	$\tau$ ( $j^{-1}$ )	$Sl$ ( $j^{-1}$ )	RMSE	$\%_{erreur}$
Minimum	3,23	138,57	7,04	$5,37 \cdot 10^{-4}$	$4,22 \cdot 10^{-3}$	0,13
Premier quartile	3,39	453,62	206,86	$8,47 \cdot 10^{-4}$	$4,75 \cdot 10^{-2}$	1,56
Médiane	3,47	576,05	241,76	$9,33 \cdot 10^{-4}$	$6,34 \cdot 10^{-2}$	2,15
Troisième quartile	3,59	660,86	288,01	$1,11 \cdot 10^{-3}$	$8,45 \cdot 10^{-2}$	2,82
Maximum	3,93	789,69	469,98	$3,33 \cdot 10^{-2}$	$2,65 \cdot 10^{-1}$	9,06

séquence temporelle et les estimations de ces descripteurs. La fonction logistique utilisée comme modèle s'ajuste donc avec un pourcentage d'erreur faible aux données de chaque séquence temporelle.

La formule du modèle donnée dans la formule 5.4 permet de représenter avec peu d'erreur l'augmentation du pH dans les baies de raisin. Ce modèle est donc bien un modèle représentatif de cette augmentation.

### Validation de la valeur de $pH_{max}$

Les valeurs de  $pH_{max}$  sont comprises entre 3,23 et 3,93.

Auparavant, de nombreux auteurs ont relaté des valeurs de pH à vendange pour des raisins de Cabernet-Sauvignon cultivés en Australie, France et Californie. En France, entre 1952 et 1958, sur le terroir de Pauillac, le pH à vendange de baies de Cabernet-Sauvignon a été observé à des valeurs comprises entre 2,9 et 3,4 [19]. Sur le même terroir, en 2000, un pH moyen des baies de 3,6 à la vendange a été relevé [11]. En 1998, un pH à vendange compris entre 3,35 et 3,61 a été observé [73]. Enfin, sur le terroir de Saint Emilion, un pH moyen de 3,4 a été relevé [94]. En Australie, en 2001 et 2002, un pH à maturité compris entre 3,66 et 4,26 a été évalué [23] puis sur les millésimes 2003 à 2005, un pH compris entre 3,43 et 3,61 a été observé [47]. Finalement, en 1967, en Californie, la valeur moyenne du pH des raisins de Cabernet-Sauvignon à vendange est évaluée à 3,81 [19]. En 2001, des valeurs comprises entre 3,6 et 4,1 sont notées [9]. Tous ces auteurs ont donc noté des valeurs du pH dans les baies de Cabernet-Sauvignon à la vendange comprises entre 2,9 et 4,26.

Les valeurs ajustées sur les individus de la base de données en cours d'étude sont cohérentes par rapport aux valeurs observées par tous ces auteurs. Il est donc possible de valider les valeurs ajustées de  $pH_{max}$  et de confirmer la qualité de l'estimation de ce paramètre.

### Validation de la valeur de $Infl$

Les valeurs de  $Infl$  sont comprises entre 139 et 790 GDD. Sur la région de production de la Napa Valley, le millésime 2011 a été très frais et caractérisé par une maturation languissante en début de celle-ci pour certaines parcelles (source personnelle). Cette atypicité pourrait donc expliquer la forte variabilité des valeurs du paramètre  $Infl$ .

En moyenne, entre véraison et vendange, les parcelles de la base de données en cours d'étude ont accumulé 1606 GDD. Au point d'inflexion, le pH n'a augmenté que de la moitié de son augmentation totale. Sur la partie de la courbe comprise entre *Infl* et la vendange, le pH devra donc augmenter autant que l'augmentation déjà réalisée de véraison à *Infl*. *Infl* est toujours inférieur à la moitié des GDD accumulés entre véraison et vendange. Il y a donc une augmentation de GDD au moins égale à l'accumulation de GDD ayant déjà eu lieu avant *Infl* pour permettre une augmentation de pH de même valeur que celle effectuée de véraison à *Infl*. Les valeurs de *Infl* ajustées sont donc logiques.

### Validation de la valeur de $\tau$ et de $Sl$

Le pH moyen à maturité des raisins des parcelles de la base de données étudiée est de 3,49. Le pH moyen des baies à véraison est de 2,54. Il y a donc une augmentation moyenne du pH de 0,95 au cours de la maturation. Les GDD moyens entre véraison et vendange pour les parcelles de la base de données sont de 1606 GDD. Si l'augmentation de pH dans les baies de raisin était linéaire entre la véraison et la vendange, ces moyennes reviendraient à une élévation de  $5,92 \cdot 10^{-4}$  de pH par GDD.

La pente au point d'inflexion est comprise entre  $5,37 \cdot 10^{-4}$  et  $4,45 \cdot 10^{-3}$ . Le premier quartile de répartition des valeurs de la pente  $Sl$ , donné dans le tableau 5.2, permet de montrer que 75% des valeurs sont supérieures à  $8,47 \cdot 10^{-4}$ . Les valeurs de pente au point d'inflexion sont donc en majorité supérieures à la valeur calculée si l'augmentation de pH était linéaire.

La cinétique d'augmentation du pH dans les baies de raisin est de forme sigmoïdale. L'augmentation du pH dans les baies de raisin va donc d'abord être lente avant d'entrer dans une phase d'accumulation linéaire rapide. Puis, en fin de maturation, cette augmentation va ralentir et atteindre un plateau qui correspondra à la valeur de  $pH_{max}$ . Par rapport à une augmentation totalement linéaire entre véraison et vendange, comme celle présentée dans le calcul ci-dessus, le pH dans les baies va donc augmenter de façon moins importante en début et fin de maturation. Par contre, en milieu de maturation et donc au niveau du point d'inflexion, où la pente est calculée, cette augmentation est plus rapide que si l'accumulation avait été exclusivement linéaire. Les valeurs de pente calculées sont supérieures à la valeur d'accumulation linéaire. Cette différence correspond aux observations réalisées ci-dessus.

Les valeurs de  $Sl$  sont de même ordre de grandeur que la valeur d'accumulation linéaire calculée. Les valeurs de  $Sl$  et donc celles de  $\tau$  sont donc logiques.

La pourcentage d'erreur d'ajustement du modèle aux descripteurs des séquences temporelles de chacun des 53 individus de la base de données est faible. De plus, les paramètres ajustés peuvent être validés à l'aide de données expertes ou en utilisant les connaissances contenues dans la base de données. Cette analyse permet donc de valider l'utilisation de la fonction logistique pour représenter l'augmentation du pH des baies de raisin au cours de la maturation, et ce sur un grand nombre d'individus.

### Potentiel de comparaison des individus à partir des valeurs des paramètres

Après ajustement, les paramètres de chaque individu peuvent être comparés aux paramètres des autres individus. En comparant deux individus, il est possible de connaître, pour chacun des paramètres, la position relative de l'un par rapport à l'autre.

Ainsi, si l'on compare un individu  $I_1$  à un individu  $I_2$ , il sera possible de dire si :

- le  $pH_{max}$  de  $I_1$  est plus ou moins élevé que celui de  $I_2$ , ce qui traduit une acidité plus ou moins importante dans les baies de l'un des deux individus
- $Infl$  est atteint plus tôt ou plus tard pour l'individu  $I_1$  par rapport à l'individu  $I_2$ . La différence de valeur de  $Infl$  traduit une précocité plus importante de l'augmentation du pH pour l'un des deux individus
- $Sl$  de  $I_1$  est plus élevée ou plus faible que celle de  $I_2$ . Cette différence représentera une plus forte rapidité d'augmentation du pH pour l'un des deux individus.

De telles comparaisons peuvent être faites en prenant en compte un plus grand nombre d'individus.

Mais, la comparaison un à un des paramètres ne prend pas en compte la forme de la cinétique complète d'accumulation. La comparaison d'individus en utilisant la représentation graphique des cinétiques modélisées permet de travailler sur toute la cinétique. Elle va être présentée dans la partie suivante.

### 5.3.4 Comparaison graphique de la cinétique de pH des trois parcelles étudiées

Les trois parcelles provenaient de trois domaines différents. Comme montré dans la partie 5.3.1, le climat et / ou le type de situation et de sol des 3 domaines choisis est différent d'un domaine à l'autre. Puisque le climat entre les domaines et donc entre les parcelles est différent, pour chaque jour, le  $\sum_j$  (section 5.2.2) dans chaque domaine est différent. La valeur de GDD sera donc différente d'une parcelle à l'autre. L'accumulation des températures tout au long de l'année, et donc a fortiori après véraison, peut être plus rapide dans un domaine par rapport à celle des autres domaines.

C'est pourquoi, pour ne pas insérer une variabilité supplémentaire dans la comparaison des cinétiques, l'échelle de temps choisie était l'échelle de temps thermique.

Le tableau 5.3 donne les valeurs ajustées des paramètres  $pH_{max}$ ,  $Infl$  et  $\tau$  définis dans l'équation 5.4 ainsi que de la pente réelle de la courbe au point d'inflexion,  $Sl$ . La RMSE et le pourcentage d'erreur de l'ajustement du modèle aux 3 séquences temporelles est aussi fourni.

Les trois parcelles choisies présentent des pourcentages d'erreur d'ajustement compris entre 1 et 3%. L'ajustement de la courbe aux descripteurs de la séquence temporelle,  $TSI$ , de chaque individu présente donc peu d'erreurs.

La valeur de  $pH_{max}$  dans la parcelle 40 est plus élevée que celle des deux autres parcelles. L'équilibre acido-basique dans la parcelle 40 est donc moins acide à la vendange que celui des deux autres parcelles.

TABLE 5.3 – Valeur des paramètres ajustés lors de la modélisation de l'augmentation de pH de trois parcelles de la base de données étudiées

	$pH_{max}$	$Infl$ (°C)	$\tau$ (°C <sup>-1</sup> )	$Sl$ (°C <sup>-1</sup> )	RMSE	% <i>erreur</i> (%)
Parcelle 7	3,46	363,55	312,23	$0,739 \cdot 10^{-3}$	$2,95 \cdot 10^{-2}$	0,93
Parcelle 16	3,42	422,94	182,32	$1,21 \cdot 10^{-3}$	$0,422 \cdot 10^{-2}$	3,12
Parcelle 40	3,55	516,41	277,55	$0,911 \cdot 10^{-3}$	$2,93 \cdot 10^{-2}$	3,01

Les GDD des points d'inflexion suivent un ordre. La parcelle 7 présente un *Infl* atteint plus tôt que celui de la parcelle 16, lui même atteint plus tôt que celui de la parcelle 40. L'augmentation du pH est donc la plus précoce dans la parcelle 7 et la plus tardive dans la parcelle 40.

Enfin, la pente au point d'inflexion est plus élevée pour la parcelle 16 que pour la parcelle 40, elle même ayant une pente plus élevée que la parcelle 7. L'augmentation du pH au point d'inflexion est donc plus rapide pour la parcelle 16 que pour la parcelle 40 puis pour la parcelle 7.

Les valeurs ajustées des paramètres pour les 3 parcelles étudiées présentent donc d'importantes différences. La comparaison des cinétiques d'augmentation devrait permettre de comprendre le comportement réel de chacune.

Les cinétiques du pH pendant la maturation des trois parcelles en cours d'étude sont représentées figure 5.3. Les valeurs des paramètres  $pH_{max}$  et *Infl* sont indiquées pour chacune d'elle.

Les courbes permettent de comparer l'évolution du pH de plusieurs parcelles en permettant l'analyse des courbes complètes ou de la position des paramètres pour chacune d'elle.

La courbe représentative de la parcelle 16 présente des plateaux marqués en début et fin de maturation et une pente dans la phase linéaire d'accumulation élevée. L'augmentation du pH commence donc lentement puis est rapide et atteint tôt le plateau de  $pH_{max}$ .

De comportement différent, la cinétique de la parcelle 7 présente des plateaux moins marqués en début et fin de maturation. L'augmentation du pH dans les baies des raisin de la parcelle 7 est moins rapide dans la phase linéaire d'accumulation que celle dans les baies des raisins de la parcelle 16. L'augmentation de pH dans les baies des raisins de la parcelle 7 est donc plus rapide que dans les baies de raisins de la parcelle 16 au début de la maturation puis plus lente et atteint moins facilement le plateau de  $pH_{max}$ .

La cinétique représentative de l'augmentation du pH dans les baies des raisins de la parcelle 40 est médiane entre les 2 autres cinétiques avant d'atteindre un  $pH_{max}$  plus élevé. En début de maturation (GDD < 200), l'augmentation de pH est donc plus rapide dans les baies des raisins de la parcelle 7 puis dans celles des raisins de la parcelle 40 puis dans celles des raisins de la parcelle 16. Cette tendance s'inverse dès 400 GDD après le début de la véraison.



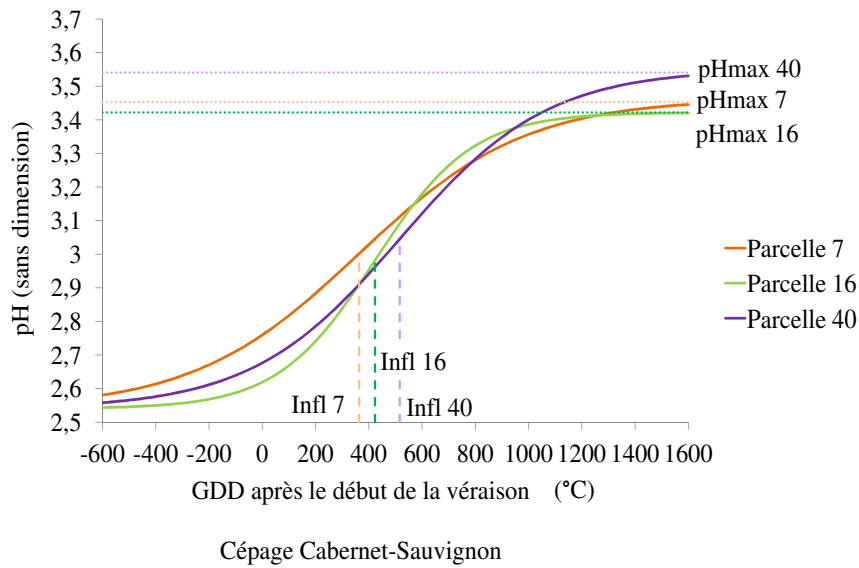


FIGURE 5.3 – Modèles de la cinétique du pH pendant la maturation de trois parcelles de Cabernet-Sauvignon

En fin de maturation, les trois cinétiques présentent un ralentissement plus ou moins marqué de l'augmentation du pH dans les baies. Le plateau de  $pH_{max}$  est atteint à partir de 900 GDD pour la parcelle 16, de 1200 GDD pour la parcelle 40 et est difficile à déterminer pour la parcelle 7. L'arrêt d'augmentation du pH est donc plus net dans les baies des parcelles 16 et 40 que dans les baies de la parcelle 7.

A tout moment, grâce à la figure 5.3, il est possible de définir la parcelle dans laquelle le niveau de pH est le plus élevé.

Jusqu'à 400 GDD après le début de la véraison, le niveau de pH est plus élevé dans les baies des raisins de la parcelle 40 que dans celles des raisins de la parcelle 16. Puis, après 600 GDD, le niveau de pH dans les baies de la parcelle 7 est plus faible que dans celle de la parcelle 16. Enfin, le niveau de pH dans les baies de la parcelle 40 devient plus élevé que dans les baies des deux autres parcelles à partir de 900 GDD. Ces comparaisons ne sont que des exemples de toutes les positions entre les trois courbes visibles que la figure 5.3.

La vitesse maximale d'accumulation va donc influencer le niveau de pH sur toute la période de maturation. Pour les trois parcelles étudiées, cette pente a influencé la position relative de la valeur de pH de chaque parcelle par rapport aux autres sur l'ensemble de la cinétique.

Comme le montre la figure 5.3, les baies de la parcelle 40 sont caractérisées par un  $pH$  plus élevé que celui des baies des deux autres parcelles à partir de 1000 GDD et sur toute la fin de la cinétique. Les mesures se sont prolongées plus longtemps pour la parcelle 40 que pour les deux autres parcelles. Nous supposons que la dernière mesure de chaque parcelle a été réalisée à proximité de la date de vendange et qu'il n'y a donc pas un problème d'ajustement du modèle.

Il a été rappelé dans la section 5.1 que le pH dans les baies de raisin est principalement dû à la concentration en acide tartrique et en cation potassium dans les baies. La teneur en acide tartrique des baies est fixée à la véraison et pendant la maturation, sa concentration diminue uniquement par phénomène de dilution. La concentration en cation potassium peut être augmentée par une forte alimentation hydrique du sol [19, 52].

Un pH élevé peut donc être dû à (i) une teneur en acide faible à la véraison, (ii) une concentration en acide tartrique diminuée par une forte dilution due à des baies plus grosses ou (iii) une concentration en cation potassium élevée.

Le pH plus élevé de la parcelle 40 peut donc être dû à (i) une teneur moins importante d'acide tartrique dans les baies de cette parcelle que dans les baies des deux autres parcelles à véraison, à (ii) des baies plus grosses pour la parcelle 40 que pour les deux autres parcelles ou (iii) à une concentration en cation potassium plus importante dans les baies de la parcelle 40 que dans les baies des deux autres parcelles.

La parcelle 40 pourrait donc avoir été irriguée avec de plus grandes quantités d'eau que les parcelles 7 et 16. Cette irrigation aurait alors permis que la concentration en cation potassium dans les baies de raisin soit plus importante. Ce point n'est qu'une supposition permettant d'expliquer le niveau plus élevé du pH dans les baies des raisins de la parcelle 40 par rapport à celui des deux autres parcelles et n'a pas pu être vérifié. Afin de pouvoir confirmer cette hypothèse, il faudrait connaître la cinétique d'accumulation dans les baies de chaque parcelle du cation potassium ainsi que l'évolution de la concentration en acide tartrique pour les 3 parcelles. Comprendre les différences dans ces cinétiques d'accumulation nécessite donc des mesures complémentaires.

Le point d'inflexion est atteint d'abord par la cinétique représentant l'augmentation de pH dans les baies de la parcelle 7 puis par celle représentant l'augmentation du pH dans les baies de la parcelle 16 et enfin par la cinétique représentant l'augmentation de pH dans les baies de la parcelle 40. L'augmentation de pH a donc été plus précoce dans les baies de la parcelle 7 que dans celle de la parcelle 16 ou de la parcelle 40. la valeur du paramètre  $Infl$  n'a pas de lien avec la pente de la cinétique et donc avec la notion de rapidité de l'augmentation du pH dans les baies de raisin.

En conclusion, l'étude de la cinétique de l'augmentation du pH dans différentes parcelles permet de mettre en avant les parcelles où cette augmentation a été rapide ou lente. Dans un objectif d'aide à la décision, une séquence temporelle en cours de construction pourrait être comparée à différents individus des millésimes passés dès les premières mesures. Cette comparaison pourrait permettre d'envisager si l'augmentation du pH dans les baies de raisin de la parcelle en cours d'étude est plus rapide ou plus lente et si la valeur de  $pH_{max}$  a des chances d'être plus élevée ou plus basse que celle des individus auxquels la cinétique en cours de construction est comparée.

## 5.4 Conclusion : Un modèle avec du potentiels pour comparer les cinétiques d'augmentation du pH dans différentes parcelles

L'augmentation du pH dans les baies de raisin a été modélisée grâce à une sigmoïde adaptée de la fonction logistique. Ce modèle a été validé sur un grand nombre de données et l'erreur d'ajustement du modèle s'est avérée faible.

Les paramètres du modèle peuvent être comparés entre eux. Les individus dont les  $pH_{max}$ ,  $Infl$  ou  $Sl$  sont proches peuvent être déterminés. Mais, la comparaison des paramètres ne permet pas la comparaison entre individus de la cinétique complète d'augmentation du pH.

Les cinétiques d'augmentation du pH dans trois parcelles de Cabernet-Sauvignon de la Napa Valley pour le millésime 2011 ont été comparées sur une base graphique. Il a été montré que le niveau du pH peut être ordonné entre les parcelles à tout moment de la cinétique. Cet ordre ne reste pas constant.

La comparaison graphique des cinétiques d'accumulation a permis de comparer les courbes dans leur ensemble et les paramètres des courbes. La compréhension des facteurs climatiques ou cultureux influençant les différences dans ces cinétiques reste encore à améliorer.

Si le pH des baies de raisin au moment de la vendange est trop élevé ( $> 3,6$ ), la vinification du raisin sera compliquée [52]. Suivre la cinétique du pH est donc essentielle pour s'assurer d'un bon potentiel de vinification. Comparer les cinétiques de pH pour tenter de comprendre dans quels cas (climat, pratiques culturelles) le pH sera trop élevé permettra donc de produire des règles de décisions pour limiter une trop grande augmentation non désirée du pH.

Afin de pouvoir évaluer l'effet du climat, des pratiques culturelles ou même du type de sol sur le cinétique d'augmentation du pH dans les baies de raisin, des analyses descriptives comme des analyses en composantes principales sur un plus grand nombre d'individus pourraient être réalisées. A partir de la carte factorielle de répartition des individus, des groupes d'individus en fonction du type de sol de chaque parcelle ou de la quantité d'eau apportée par irrigation pourraient être recherchés. De tels groupes, déterminés pour chaque base de données, permettraient par exemple d'apporter une connaissance sur les pratiques culturelles à éviter pour limiter les pH trop élevés. Cette connaissance est un exemple de règles de décisions qui pourraient être extraites de la mise en place à grande échelle de la méthode objet de cette thèse.

De plus, à partir des groupes mis en place, des règles de comportement indiquant le climat et les actions entreprises sur une parcelle et le type de raisin obtenu pourraient être mises en place. Ces règles serviraient alors de base à la mise en place d'un raisonnement par analogies.

Ce travail pourrait ensuite être couplé à la modélisation du climat telle que celle

présentée dans le chapitre 4. Une information globale sur le vécu climatique et la cinétique de pH de chaque individu pourrait alors être comparée avec la même information pour d'autres individus. Des analogies prenant en compte le vécu global de chaque individu pourraient alors être mises en place.

Ce travail d'application n'a pas été entrepris dans cette thèse. Les enjeux d'un tel travail sont présentés dans le chapitre 7 de ce manuscrit.

# Chapitre 6

## Application 3 : Modélisation de l'accumulation de sucres dans les baies de raisin pendant la maturation

---

### Sommaire

<b>6.1</b>	<b>Introduction : Evolution des sucres dans les baies de raisin</b>	<b>78</b>
<b>6.2</b>	<b>Matériel et Méthodes</b>	<b>78</b>
6.2.1	Matériel expérimental	79
6.2.2	Modélisation paramétrique de l'accumulation de sucres dans les baies	83
6.2.3	Utilisation du modèle	84
<b>6.3</b>	<b>Résultats et Discussion</b>	<b>85</b>
6.3.1	Adéquation des exemples aux problèmes de comparaison de séquences temporelles	85
6.3.2	Validation d'un modèle d'accumulation de sucres dans les baies de raisin pour chaque individu	86
6.3.3	Résultats des analyses descriptives de la base de données 1 après modélisation	95
6.3.4	Résultats des analyses descriptives de la base de données 2 après modélisation	105
<b>6.4</b>	<b>Conclusion : Une méthodologie générique qui permet de comparer des individus hétérogènes</b>	<b>114</b>

---

Dans ce chapitre, comme cela l'a été fait dans le chapitre précédent pour le pH, les potentialités d'utilisation de tout type de séquence temporelle avec la méthode proposée pour comparer l'accumulation de sucres entre parcelles et millésime sont évaluées, même dans les cas où les séquences temporelles ne couvriraient pas parfaitement l'ensemble de la cinétique : absence possible de quelques mesures en début et / ou en fin de cinétique.

## 6.1 Introduction : Evolution des sucres dans les baies de raisin

Les sucres sont des produits de la photosynthèse [11, 19, 25, 79, 80]. Ils sont principalement synthétisés dans les feuilles sous forme de saccharose [27]. Le saccharose est ensuite utilisé pour le métabolisme de la feuille, stocké sous forme d'amidon ou transporté vers les autres organes tels que les baies. Puis, le saccharose est hydrolysé en glucose et fructose, appelés sucres réducteurs, dont la concentration sera dosée dans les baies de raisin [27].

Les sucres réducteurs sont transformés en alcool pendant la vinification via le processus de fermentation alcoolique [12].

Les sucres sont essentiels dans la détermination de la qualité du raisin et leur évolution dans les baies a été largement étudiée.

De la floraison à la véraison, la concentration en sucres dans les baies reste assez faible. Cette concentration est de quelques dizaines de grammes par litre. Puis, à partir de la véraison, les sucres vont être accumulés dans les baies pour atteindre, à la vendange, des concentrations souvent supérieures à 200 grammes par litre [11, 19, 21, 72, 80].

La cinétique de l'accumulation des sucres dans les baies de raisin a la forme d'une sigmoïde. Cela signifie que l'accumulation est lente au début de la maturation puis rapide dans une phase appelée phase d'accumulation linéaire. Enfin, l'accumulation ralentit en fin de maturation [11, 19, 21, 72, 80].

La cinétique d'accumulation des sucres dans les baies de raisin a auparavant été modélisée grâce à la fonction logistique [11, 25, 40, 68, 75]. Dans cette application, la fonction logistique a donc été à nouveau adaptée à la modélisation de l'accumulation des sucres dans les baies de raisin.

## 6.2 Matériel et Méthodes

Une seule grandeur de mesure est considérée dans cet exemple : la concentration en sucres dans les baies. Deux cas d'étude différents ont été pris en compte.

## 6.2.1 Matériel expérimental

### Premier cas d'étude : base de données 1

**Site d'étude** Une première étude a été réalisée en utilisant des données provenant de la station expérimentale INRA Pech Rouge. Ce domaine est situé à Gruissan dans le sud de la France, dans la région de production du Languedoc-Roussillon ( $43^{\circ}10'N$ ,  $3^{\circ}06'E$ , dans le système géodésique WGS84). Les données utilisées pour l'application présentée ci-dessous provenaient d'une parcelle de Mourvèdre du domaine, mesurant 0,77 ha et plantée en 1991. La vigne était palissée et conduite en cordon de royat. La densité de plantation était de 4000 pieds par hectare.

**Données utilisées** Les mesures de la concentration en sucres ont été réalisées en grammes par litre de moût ( $g.L^{-1}$ ). Elles ont été effectuées entre la véraison et la récolte. Au maximum une mesure était effectuée pour une journée. Le nombre de mesures correspondait donc au nombre de jours où une mesure avait été effectuée. L'intervalle de temps entre deux mesures dépendait de chaque individu et le temps maximum entre deux mesures n'était pas fixé. Sept millésimes,  $T$ , de 2003 à 2009 ont été pris en compte dans cette base de données.

Les échantillons ayant permis l'évaluation de la concentration en sucres dans les baies étaient supposés comme représentatifs de l'ensemble de la parcelle étudiée. Ces échantillons étaient donc supposés constitués de baies à différents niveaux de maturation, correspondant ainsi aux différences de maturité entre les pieds ou au sein d'une grappe. A priori les échantillons étaient constitués d'un ensemble de baies et ne présentaient pas de grappes entières.

Les données de description relatives à la base de données 1 ont été regroupées dans le tableau 6.1. Les classes d'indices agroclimatiques ont été définies par classification ascendante hiérarchique (CAH) [83] des millésimes en fonction des trois indices bioclimatiques servant à la classification multicritères de Tonietto et Carbonneau [92].

Pour la parcelle et les millésimes étudiés, la mesure de la concentration en sucres dans les baies de raisin associée à sa date de mesure a été extraite de la base de données de traçabilité représentant l'ensemble des données relevées. Une base de données a été créée en regroupant les données selon les définitions données en partie 3.1. Une seule localisation  $s_j$ , correspondant à la parcelle de Mourvèdre choisie, a été prise en compte. Chaque millésime représentait donc un individu,  $I$ , unique. La base de données 1 était donc constituée de 7 individus. La base temporelle,  $T_{\Phi}$  a été définie comme l'intervalle entre véraison et récolte. Elle a ensuite été divisée en jours. Chaque fenêtre de temps,  $W_a$ , de la discrétisation temporelle,  $\Delta$ , était donc une journée.

### Second cas d'étude : base de données 2

**Site d'étude** Une seconde étude a été réalisée en utilisant des données provenant de plusieurs domaines de la région viticole de Napa Valley en Californie. La Napa Valley

TABLE 6.1 – Description des individus de la base de données 1 - INRA Pech Rouge, France

Millesime	Début de la véraison		Vendange		Nombre de mesures	Indices agroclimatiques			Classe agroclim.	Signification Classe agroclimatique
	Date	Jour julien	Date	Jour julien		IH	IS	IF		
2003	17/7	198	19/9	262	6	2569,3	-104,72	15,3	1	Chaud, modérément à très sec, nuits tempérées
2004	23/7	205	23/9	267	5	2286,1	-29,13	15,8	3	Chaud tempéré, modérément sec, nuits tempérées
2005	25/7	206	6/10	279	6	2369,3	-3,5	15,6	4	Atypique
2006	25/7	206	21/9	264	5	2516	-78,65	16,3	1	Chaud, modérément à très sec, nuits tempérées
2007	19/7	200	17/9	260	5	2333,3	-27,06	14,7	3	Chaud tempéré, modérément sec, nuits tempérées
2008	26/7	208	5/10	279	5	2098,1	-66,41	13,5	2	Chaud tempéré, modérément à très sec, nuits fraîches à tempérées
2009	17/7	198	10/9	253	5	2284,4	-88,9	15	2	Chaud tempéré, modérément à très sec, nuits fraîches à tempérées

IH : Indice héliothermique de Huglin, IS : Indice de sécheresse, IF : Indice de fraîcheur des nuits (Tonietto et Carbonneau, 2004)



est une vallée longue d'une quarantaine et large d'une dizaine de kilomètres située au nord de la baie de San Francisco. Elle est comprise entre  $38^{\circ}38'$  et  $38^{\circ}14'$  N de latitude et  $122^{\circ}27'$  et  $122^{\circ}17'$  O de longitude dans le système géodésique WGS84.

Un total de 65 parcelles a servi à l'analyse. Ces parcelles étaient toutes plantées du cépage Cabernet-Sauvignon. Elles étaient irriguées durant tout le cycle végétatif. Les densités de plantation variaient entre 1500 et 7500 pieds par hectare. Toutes les parcelles étudiées étaient palissées, conduites en cordon de royat ou guyot. Elles ont été plantées entre 1985 et 2008.

**Données utilisées** Les mesures de la concentration en sucres ont été réalisées en degrés brix ( $^{\circ}\text{Br}$ ). L'abaque de conversion des données en degrés Brix vers des valeurs de concentration en grammes par litre de moût ne permet de convertir les degrés brix qu'à partir de 10 degrés. L'ensemble des séquences temporelles considérées dans la base de données 2 ayant des valeurs de degrés brix inférieures à 10, les mesures ne pouvaient donc pas être transcrites en grammes par litre de moût. Les mesures ont commencé avant la véraison et se sont poursuivies jusqu'à la récolte. Au maximum une mesure par jour a été effectuée. Le nombre de mesures correspondait donc au nombre de jours où une mesure avait été effectuée. L'intervalle de temps entre deux mesures dépendait de chaque individu et le temps maximum entre deux mesures n'était pas fixé. Un seul millésime ( $T$ ) a été considéré. Les 65 parcelles représentaient chacune une localisation,  $s_j$ , unique.

Les échantillons ayant permis l'évaluation de la concentration en sucres dans les baies étaient supposés comme représentatifs de l'ensemble de la parcelle étudiée. Ces échantillons étaient donc supposés constitués de baies à différents niveaux de maturation, correspondant ainsi aux différences de maturité entre les pieds ou au sein d'une grappe. A priori les échantillons étaient constitués d'un ensemble de baies et ne présentaient pas de grappes entières.

Les données de description relatives à la base de données 2 ont été regroupées dans le tableau 6.2. La valeur des indices bioclimatiques utilisés par Tonietto et Carbonneau dans leur classification multicritère des climats viticoles [92] a été calculée grâce aux données météorologiques de chaque domaine. Les classes relatives à ces indices ont ensuite été définies par classification ascendante hiérarchique (CAH) [83].

Pour le millésime et les parcelles étudiées, la mesure de la concentration en sucres dans les baies de raisin associée à sa date de mesure a été extraite de la base de données de traçabilité représentant l'ensemble des données relevées. Une base de données a été créée en regroupant les données selon les définitions données en partie 3.1.

Un seul cycle temporel ( $T$ ) a été considéré. Chaque parcelle,  $s_j$ , représentait donc un individu  $I$ . Pour chaque individu, la base temporelle,  $T_{\Phi}$  a été définie comme l'intervalle entre la première mesure et la récolte. Le pas de temps minimum de mesure étant le jour, la base temporelle a donc été divisée en jour depuis la première mesure à la récolte. Chaque fenêtre  $W_a$  de la discrétisation temporelle  $\Delta$  était donc un jour.

TABLE 6.2 – Description des individus de la base de données 2 - Napa Valley, Californie

Domaine	Nombre parcelles	Nombre mesures	Indices agroclimatiques			Classe agroclim.	Signification classe agroclimatique
			IH	IS	IF		
2	3	15	2420,6	-122,70	9,9	2	Chaud tempéré, très sec, nuits très fraîches
6	6	12-13	2333,7	-101,45	10,3	2	Chaud tempéré, très sec, nuits très fraîches
7	3	11	2118,9	-113,45	14,2	4	Chaud tempéré, très sec, nuits tempérées
9	7	14-18	2425,0	-83,82	11,8	1	Chaud, modérément sec, nuit très fraîches
10	7	10-15	2320,9	-64,63	10,4	3	Chaud tempéré, modérément sec, nuits très fraîches
11	1	13	2501,0	-84,36	10,9	1	Chaud, modérément sec, nuit très fraîches
13	2	10	2277,7	-105,20	9,6	2	Chaud tempéré, très sec, nuits très fraîches
14	2	8-9	2420,0	-34,97	12,3	1	Chaud, modérément sec, nuit très fraîches
17	5	12-13	2336,8	-114,36	9,6	2	Chaud tempéré, très sec, nuits très fraîches
19	4	12-13	2072,0	-67,47	12,1	3	Chaud tempéré, modérément sec, nuits très fraîches
20	2	11-12	2415,3	-32,74	11,8	1	Chaud, modérément sec, nuit très fraîches
23	6	11-13	2205,2	-53,07	10,9	3	Chaud tempéré, modérément sec, nuits très fraîches
26	2	9	2034,6	-72,23	10,6	3	Chaud tempéré, modérément sec, nuits très fraîches
27	1	7	2417,6	-128,83	9,8	2	Chaud tempéré, très sec, nuits très fraîches
30	2	8	2415,3	-32,74	11,8	1	Chaud, modérément sec, nuit très fraîches
31	12	9-17	2137,5	-106,75	14,1	4	Chaud tempéré, très sec, nuits tempérées

IH : Indice héliothermique de Huglin, IS : Indice de sécheresse, IF : Indice de fraîcheur des nuits (Tonietto et Carbonneau, 2004)

## 6.2.2 Modélisation paramétrique de l'accumulation de sucres dans les baies

Dans différents travaux déjà publiés, sur des bases de données provenant de différentes régions viticoles, la sigmoïde représentant l'accumulation des sucres dans les baies de raisin a été modélisée depuis la floraison grâce à la fonction logistique, en adaptant cette fonction aux spécificités de l'accumulation de sucres dans les baies de raisin [11, 25, 40, 68].

Pour les deux cas d'étude considérés, la fonction logistique a été adaptée pour modéliser l'accumulation des sucres depuis la véraison. L'accumulation de sucres dans les baies de raisin a donc été modélisée grâce à l'équation 6.1

$$x_s(t) = x_s(t_0) + \frac{X_{smax} - x_s(t_0)}{1 + e^{\frac{Infl - t}{\tau}}} \quad (6.1)$$

où,  $x_s(t)$  représentait la concentration en sucres dans la baie de raisin au temps  $t$ ,  $t_0$  représentait le jour au niveau duquel le modèle était initialisé, soit la date correspondant au début de la véraison ( $t_0 = 0$ ).

Le modèle paramétrique présenté équation 6.1 contenait 3 paramètres :  $X_{smax}$  qui représentait le chargement maximum en sucres, soit le degré alcoolique potentiel du vin produit à partir de ce raisin,  $Infl$  qui représentait l'abscisse du point d'inflexion et  $\tau$  qui permettait d'ajuster la pente au point d'inflexion.

La pente,  $Sl$ , de la courbe au point d'inflexion a été calculée grâce à la formule donnée en équation 6.2. Sa valeur a été extraite pour l'utilisation du modèle en intensif en remplacement du paramètre  $\tau$ , car plus explicite.

$$Sl = \frac{X_{smax} - x_s(t_0)}{4\tau} \quad (6.2)$$

Pour la base de données 1, la valeur d'initialisation a été définie en se basant sur la connaissance de la concentration en sucres dans les baies de raisin non encore vérees ([19] p.100). Sa valeur a été fixée à  $x_s(t_0) = 20\text{g.L}^{-1}$ .

Pour la base de données 2, la valeur d'initialisation a été calculée en moyennant, pour le millésime 2011, les teneurs mesurées avant véraison sur toutes les parcelles de Cabernet-Sauvignon du système d'information étudié. Cette moyenne était de  $5.5^\circ\text{Br}$ .

L'ajustement du modèle à chaque séquence temporelle a été réalisé grâce à la méthode des moindres carrés. L'algorithme utilisé pour cet ajustement était celui de Levenberg-Marquardt. La méthode a été implémentée sous MatLab V7.0 (The MathWorks Inc, Natick, MA) avec la fonction de *curve fitting* non linéaire *nlinfit*. Pour chaque séquence temporelle  $TS_I$ , les résultats étaient extraits sous la forme du vecteur  $\Theta_I$ .

Pour chaque individu, l'erreur quadratique moyenne, RMSE, d'ajustement du modèle aux descripteurs de la séquence temporelle a été définie grâce au calcul donné dans la

formule 6.3.

$$\text{Soit } k \text{ le nombre total de descripteurs de } TS_I, \text{ } RMSE = \sqrt{\sum_{j=1}^{j=k} \frac{(x_j - \hat{x}_j)^2}{k}} \quad (6.3)$$

La RMSE représente l'écart moyen entre chaque descripteur de la séquence temporelle et l'estimation de ce descripteur grâce au modèle. Plus sa valeur est basse et plus la qualité d'ajustement du modèle est bonne.

En complément de cette valeur de RMSE, l'erreur moyenne d'ajustement du modèle a été calculée grâce à la formule donnée dans l'équation 6.4. Cette erreur correspondait à une évaluation pondérée du coefficient de variation entre la séquence temporelle et le modèle.

$$\text{Soit } k \text{ le nombre total de descripteurs de } TS_I, \%_{\text{erreur}} = \frac{RMSE}{\sum_{j=1}^{j=k} \frac{x_j}{k}} \quad (6.4)$$

### 6.2.3 Utilisation du modèle

Pour chacun des deux cas d'étude, les deux formes d'utilisation des modèles, présentées dans la partie 3.2.3, ont été étudiées.

#### Approche intensive

Pour chaque individu  $I$ , les paramètres  $X_{s_{max}}$ ,  $Infl$  et  $Sl$  ont été enregistrés dans une nouvelle base de données dite paramétrique ( $Z_{para}$ ) avant d'être analysés. Deux bases de données ont été créées, une pour chaque base de données initiale.

Les données ont ensuite été analysées par analyse en composante principale (ACP) [64, 83]. L'ACP a été réalisée avec le logiciel STATlab V 3.0 (SLP Infoware, Ivry sur Seine, France).

#### Approche extensive

Pour cette approche, une base temporelle,  $T_{\Phi}$ , unique, quelle que soit la base de données initiale, a été définie. Cette base temporelle couvrirait une fenêtre de temps comprise entre le jour de la véraison et le 65ème jour après la véraison. Une discrétisation temporelle homogène, chaque fenêtre  $W_a$  étant un jour, a été choisie. Pour chaque individu  $I$ , la concentration en sucres dans le moût a donc été estimée à partir du modèle ajusté pour chaque jour entre la véraison et le 65ème jour après la véraison. Ces estimations ont été enregistrées dans une nouvelle base de données dite extensive ( $Z_{ext}$ ), une par base de données initiale.

Les données ont ensuite été analysées par analyse en composante principale (ACP). Les variables à analyser par cette ACP étant en très grand nombre, mais faisant référence à l'évaluation du même phénomène, les composantes principales ont été étudiées sous

forme de loading [64]. L'ACP a été réalisée avec le logiciel MatLab V 7.0 (The MathWorks Inc., Natick, MA).

## 6.3 Résultats et Discussion

### 6.3.1 Adéquation des exemples aux problèmes de comparaison de séquences temporelles

#### Cas des Individus de la base de données 1

Dans la base de données 1, 5 à 7 mesures d'évaluation de la maturité ont été effectuées suivant le millésime (7 en 2003 et 2005, 5 en 2004, 6 les autres années). La figure 6.1 présente, pour chaque séquence temporelle, la position en nombre de jours après véraison de ces mesures.

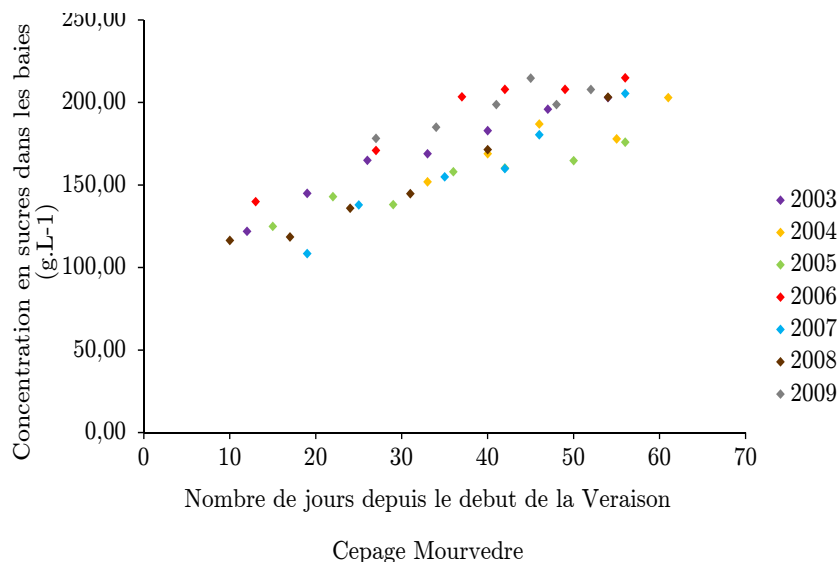


FIGURE 6.1 – Date des relevés effectués pour chacun des 7 individus de la base de données 1

Cette figure met en évidence des données très asynchrones entre les millésimes. Ainsi, pour le millésime 2004, les suivis de maturation ont commencé après le 30ème jour après la véraison alors que, pour le millésime 2008, ils ont débuté dès le 10ème jour après la véraison. Les concentrations en sucre étant très variables et les courbes des cinétiques étant difficiles à définir à l'œil nu, cette figure met bien en avant l'impossibilité de comparer ces 7 millésimes à partir des séquences temporelles brutes.

Ces observations confirment les problèmes de comparaison présentés en partie 3.2.1. Le cas d'étude de la base de données 1 permettra donc de travailler sur l'ensemble des problèmes de comparaison soulevés précédemment.

#### Cas des Individus de la base de données 2

Pour les parcelles étudiées dans la base de données 2, le nombre de mesures effectuées

par parcelle varie entre 7 et 18. Les mesures ont commencé avant véraison et se sont étendues jusqu'à la récolte.

Les dates de mesures ont été définies en fonction des besoins et des possibilités de chaque domaine. Elles sont réfléchies pour chaque domaine parcelle par parcelle. Toutes les parcelles d'un même domaine ne peuvent pas être évaluées le même jour. Les mesures sont donc asynchrones.

Une représentation semblable à la figure 6.1 pourrait être réalisée. Ce second cas d'étude, tout comme le premier, permettra donc d'étudier l'ensemble des problèmes de comparaison soulevés précédemment.

Quelle que soit la base de données étudiée, les données brutes sont donc asynchrones entre les individus. De plus, les séquences temporelles représentatives de chaque individu n'ont pas la même dimension. Pour une même grandeur, il n'est donc pas possible de comparer les individus en utilisant les séquences temporelles. L'approche de modélisation est donc un prérequis possible à toute comparaison d'individus. Les deux cas d'étude choisis sont donc de bons exemples pour le test des méthodes proposées dans l'approche théorique précédente.

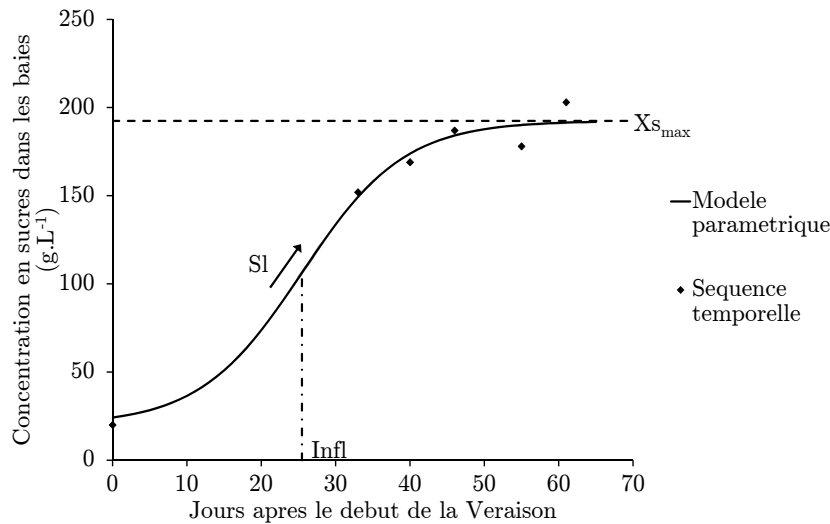
### 6.3.2 Validation d'un modèle d'accumulation de sucres dans les baies de raisin pour chaque individu

#### Construction d'une cinétique complète d'accumulation depuis la véraison jusqu'à la récolte

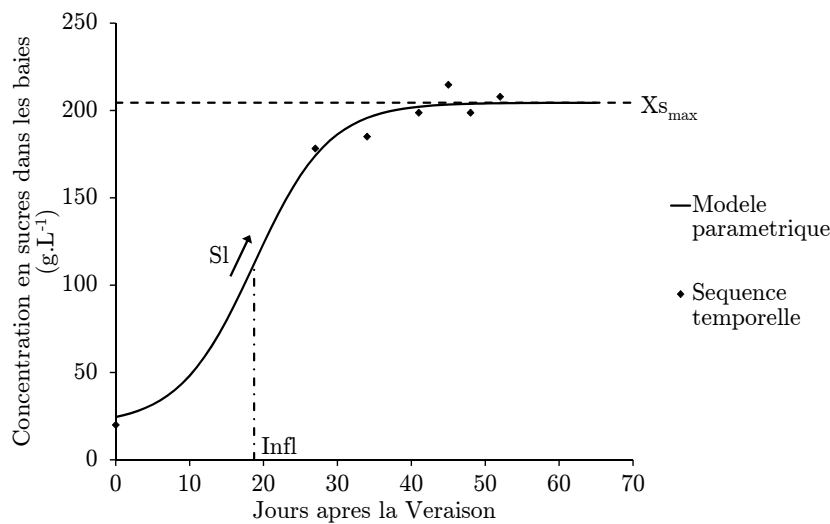
Les figures 6.2 et 6.3 présentent la forme du modèle après ajustement du vecteur  $\Theta_I$  et donc de la valeur des trois paramètres présentés dans l'équation 6.1,  $X_{s_{max}}$ ,  $Infl$  et  $\tau$ . Les courbes représentées correspondent respectivement aux millésimes 2004 et 2009 (base de données 1) puis aux parcelles 32 et 36 (base de données 2). Les trois paramètres extraits pour l'application en intensif du modèle sont présentés sur chacune des courbes.

Pour une même forme de courbe, les trois paramètres prennent des valeurs différentes en fonction de la séquence temporelle de l'individu. Par exemple, figure 6.2, le point d'inflexion de 2004 (figure 6.2(a)) est plus tardif que celui de 2009 (figure 6.2(b)), respectivement 25 et 19 jours après le début de la véraison. Figure 6.3, la valeur de concentration maximale de la parcelle 32 (figure 6.3(a)) est plus élevée que celle de la parcelle 36 (figure 6.3(b)). Grâce à la séquence temporelle représentative de chaque individu, l'ajustement des trois paramètres du modèle choisi pour représenter l'accumulation de sucres dans les baies de raisin permet donc de prendre en compte les spécificités de chaque individu et en particulier de ses conditions de maturation.

Les deux figures mettent aussi en évidence que le modèle choisi permet d'estimer l'évolution en continu de l'accumulation de sucres dans les baies de raisin de la véraison (ou de la première mesure si elle est réalisée avant la véraison) à la récolte. Il est donc possible d'estimer, à toute date de la période de maturation, la concentration en sucres dans les baies.



(a)



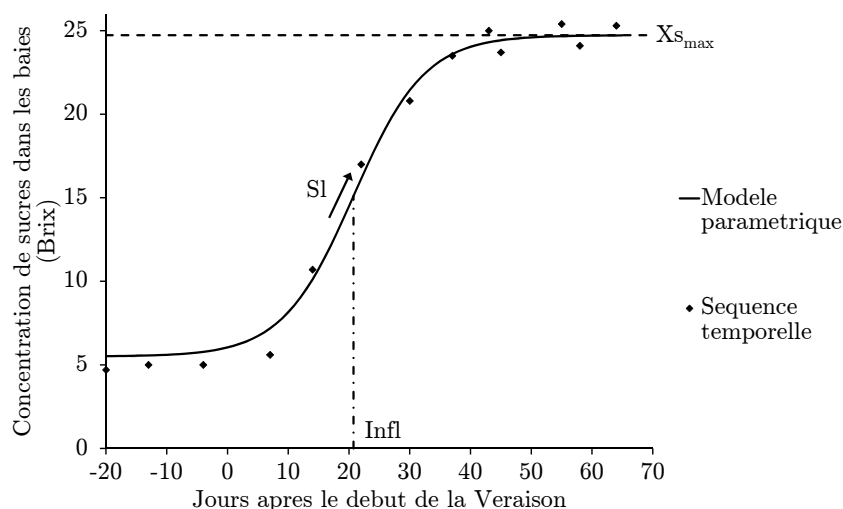
(b)

FIGURE 6.2 – Deux exemples de modèle paramétrique ajusté et sa séquence temporelle pour la base de données 1 : (a) cinétique du millésime 2004, (b) cinétique du millésime 2009

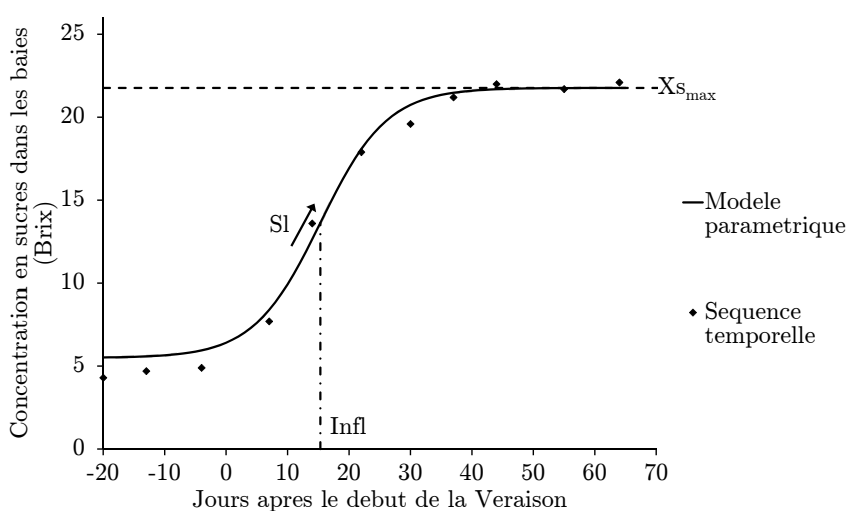
### Validation des paramètres calculés lors de l'ajustement du modèle à chaque séquence temporelle

Le tableau 6.3 présente les valeurs ajustées du vecteur  $\Theta_I$  pour le modèle représentant l'accumulation des sucres dans les baies de raisin défini en équation 6.1 pour chacun des 7 individus de la base de données 1. Dans ce modèle, il y a trois paramètres,  $X_{s_{max}}$ ,  $Infl$  et  $\tau$ . Les valeurs de la pente,  $Sl$ , au point d'inflexion sont aussi présentées. Enfin, la RMSE et le pourcentage d'erreur d'ajustement du modèle aux descripteurs de la base de données sont donnés.

Les données provenant de la Napa Valley étant plus nombreuses, le tableau 6.4 donne



(a)



(b)

FIGURE 6.3 – Deux exemples de modèle paramétrique ajusté et sa séquence temporelle pour la base de données 2 : (a) cinétique de la parcelle 32, (b) cinétique de la parcelle 36

les statistiques de répartition des paramètres ajustés lors de la modélisation de l'accumulation de sucres dans les baies de raisin pour les individus de la base de données 2. Les statistiques de répartition de la valeur de la pente au point d'inflexion, de la RMSE d'ajustement du modèle aux descripteurs de chaque séquence temporelle et du pourcentage d'erreur d'ajustement du modèle sont aussi renseignées.

### Qualité d'ajustement du modèle

Pour la base de données 1, les pourcentages d'erreur d'ajustement du modèle aux descripteurs de chaque séquence temporelle sont compris entre 3,5% et 13%. Le pourcentage d'erreur de 13% a été obtenu pour un seul individu. Pour les 6 autres individus,



Application 3 : Modélisation de l'accumulation de sucres dans les baies de raisin  
pendant la maturation

TABLE 6.3 – Valeurs ajustées des paramètres pour chaque individu de la base de données 1

Millésime	$X_{s_{max}}$ (g.L <sup>-1</sup> )	Infl (j)	$\tau$ (j <sup>-1</sup> )	$Sl$ (g.L <sup>-1</sup> .j <sup>-1</sup> )	RMSE (g.L <sup>-1</sup> )	% <sub>erreur</sub> (%)
2003	187,17	11,65	5,68	7,36	12,75	7,54
2004	192,43	25,46	6,88	6,27	7,49	4,21
2005	159,28	11,3	4,02	8,66	10,20	6,70
2006	201,77	10,91	3,45	13,17	12,88	6,75
2007	188,99	19,56	8,62	4,90	12,61	7,99
2008	195,93	15,8	11,08	3,97	19,34	13,03
2009	204,45	18,72	5,09	9,06	6,88	3,49

TABLE 6.4 – Statistiques de répartition des paramètres pour les individus de la base de données 2

	$X_{s_{max}}$ (°Br)	Infl (j)	$\tau$ (j <sup>-1</sup> )	$Sl$ (°Br.j <sup>-1</sup> )	RMSE (°Br)	% <sub>erreur</sub> (%)
Minimum	18,57	4,17	4,71	0,39	0,54	2,73
Premier quartile	22,18	11,32	5,89	0,59	0,90	5,76
Médiane	23,03	14,22	6,63	0,65	1,08	6,51
Troisième quartile	23,44	17,32	7,55	0,70	1,21	7,42
Maximum	24,73	21,40	12,11	0,88	1,80	12,88

le pourcentage d'erreur d'ajustement du modèle est inférieur à 8%. Il y a donc globalement moins de 10% d'erreur entre les descripteurs de chaque séquence temporelle et leur estimation par le modèle.

Pour la base de données 2, les pourcentages d'erreur calculés pour la base de données 2 sont compris entre 2,73% et 12,88%. 75% des individus présentent un pourcentage d'erreur inférieur à 7,5%. Il y a donc une erreur limitée entre les descripteurs de chaque séquence temporelle et leur estimation par le modèle.

L'évaluation de la concentration en sucres dans les baies de raisin est sensible à l'ensemble de baies utilisé pour les relevés. Les baies d'une même grappe ne mûrissent pas à la même vitesse. La maturation d'un pied peut aussi être plus rapide que celle du pied d'à côté. Il peut donc rapidement y avoir une importante différence de concentration dans les baies de raisin de différents endroits de la grappe ou de différents pieds d'une même parcelle. Pour avoir une estimation réelle de la concentration moyenne en sucres dans les baies de raisin d'une parcelle, il faut donc effectuer des relevés homogènes, en échantillonnant les mêmes cepes et les mêmes parties de la grappe. Si les relevés ne sont pas effectués de façon homogène, un bruit dans la cinétique réelle d'accumulation des sucres dans les baies de raisin peut être inséré. De plus, en fin de maturation, les baies sont très sensibles à la température et à l'humidité de l'air. Un phénomène de déshydratation et dilatation des baies peut être observé. Dans la fin de la maturation, il y a donc un bruit important entre les mesures de concentration.

Ces observations peuvent permettre d'expliquer les valeurs élevées (> 10%) de certains pourcentages d'erreur d'ajustement du modèle aux descripteurs des séquences tem-

porelles.

Malgré cela, il est possible de dire que le modèle s'ajuste bien aux séquences temporelles.

### Validation des valeurs ajustées pour le paramètre $X_{s_{max}}$

Pour la base de données 1, la valeur de  $X_{s_{max}}$  est comprise entre 160 (17,1°Br et 9,5°alc) et 205 g.L<sup>-1</sup> (21,1°Br et 12,2°alc). Dans la Napa Valley, le paramètre  $X_{s_{max}}$  varie entre 18,5 (175 g.L<sup>-1</sup> et 10,4°alc) et 25°Br (250 g.L<sup>-1</sup> et 14,9°alc).

L'effet (i) de différentes pratiques culturales (effeuillage, quantité de bourgeons laissées au moment de la taille ...) et (ii) de l'action du climat sur la vigne et sur la composition des baies a été évaluée dans différents pays et pour de nombreux cépages.

Un échantillon de ces travaux a été réalisé en Australie [21, 47], Chili [10], Espagne [82], Etats-Unis d'Amérique (Californie et Oregon) [9, 48, 52, 60], France [11, 18, 56, 94], Italie [50], Roumanie [51], Slovénie [24], et Suisse [61] pour les cépages Cabernet franc, Cabernet-Sauvignon, Chardonnay, Chenin, Grenache Noir, Merlot, Pinot Gris, Pinot Noir, Thompson seedless, et Zweiglet. Des mesures pour confirmer des résultats de modélisation de l'accumulation de sucres dans les baies de raisin en France, notamment en Champagne [68] ou dans le Bordelais [26] ont aussi été effectuées.

A la vendange, la concentration en sucres a été indiquée comme variant entre 16°Br (150g.L<sup>-1</sup> et 8,9°alc) et 29°Br (300 g.L<sup>-1</sup> et 17,8°alc). Une majorité des valeurs est comprise entre 17°Br (158 g.L<sup>-1</sup> et 9,4 °alc) et 25°Br (250 g.L<sup>-1</sup> et 14,8°alc).

Les valeurs calculées lors de l'ajustement du modèle à chaque séquence temporelle sont comprises dans les fenêtres de variations présentées ci-dessus. De plus, l'absence d'essai particulier sur les vignes étudiées permet d'expliquer des fenêtres de valeurs moins étendues que celles données par tous ces auteurs.

Finalement, les valeurs sont, au dire des producteurs, proches des concentrations observées à l'entrée en cave. Ces différents points permettent donc de valider un ajustement logique du paramètre  $X_{s_{max}}$ .

### Validation des valeurs ajustées pour le paramètre $Infl$

Le point d'inflexion est atteint entre 11 et 25 jours après le début de la véraison pour la base de données 1 et entre 4 et 21 jours après la véraison pour la base de données 2. La précocité et la vitesse de maturation dans les baies de raisin sont différentes d'un cépage à l'autre. Le cépage de la base de données 1 et de la base de données 2 étant différent, il n'est pas possible de comparer les valeurs du paramètres  $Infl$  dans ces deux bases de données.

Le point d'inflexion représente une concentration en sucres dans les baies égale à la concentration à véraison plus la moitié de la quantité totale accumulée ( $X_{s_{max}} - x_s(t_0)$ ).

La fonction logistique est symétrique. Le point d'inflexion se situe donc au milieu de l'axe de temps. L'atteinte du niveau maximal de cette fonction nécessite donc un temps double de celui attendu pour atteindre le point d'inflexion.

Les modèles ajustés ont une forme rappelant cette symétrie. Il doit donc falloir attendre un temps au moins approximativement double au temps nécessaire pour atteindre le point d'inflexion pour obtenir une concentration en sucres dans les baies proche de la valeur de  $X_{s_{max}}$ . Les baies doivent être vendangées après un temps au minimum double de la valeur du paramètre  $Infl$ .

Le temps le plus faible entre véraison et récolte pour les 7 individus de la base de données 1 est de 55 jours. Les valeurs de  $Infl$  ajustées pour la base de données 1 sont toujours inférieures à 27 jours après le début de la véraison, ce qui correspond à la moitié du temps minimum entre véraison et vendange. Pour la base de données 1, ces valeurs toujours inférieures à la moitié du temps entre véraison et vendange permettent de valider un ajustement logique de ce paramètre.

Le temps moyen entre véraison et récolte observé sur les parcelles de Cabernet-Sauvignon de la base de données 2 a été de 76 jours en 2011. Les valeurs de  $Infl$  ajustées pour la base de données 2 sont toujours inférieures à 38 jours après le début de la véraison, ce qui correspond à la moitié du temps minimum entre véraison et vendange. Pour la base de données 2, ces valeurs toujours inférieures à la moitié du temps entre véraison et vendange permettent donc à nouveau de valider un ajustement logique de ce paramètre.

### **Validation des valeurs ajustées du paramètre $\tau$ et de la pente, $Sl$ , de la courbe au point d'inflexion**

La pente représente l'accumulation maximum de sucres en une journée pour l'individu considéré. Pour la base de données 1, sa valeur est comprise entre 4 et 13  $\text{g.L}^{-1}.\text{j}^{-1}$ .

Pour le cépage Mourvèdre sur le domaine étudié, la moyenne de la concentration en sucres à maturité est de 197,3  $\text{g.L}^{-1}$ , soit une accumulation de 177,3  $\text{g.L}^{-1}$  depuis la véraison, pour un temps moyen entre véraison et vendange de 64 jours. Si l'accumulation de sucres est considérée comme linéaire, ce temps entre véraison et vendange et cette accumulation moyenne correspondraient à une augmentation de 2,8  $\text{g.L}^{-1}.\text{j}^{-1}$  de la concentration en sucres dans les baies.

L'accumulation de sucres dans les baies a la forme d'une sigmoïde. Elle est d'abord lente en début de maturation, puis elle s'accélère pour entrer dans une phase d'accumulation dite linéaire. En fin de maturation, elle ralentit pour atteindre un plateau maximum. La pente au point d'inflexion de la courbe représentant cette accumulation est donc supérieure à la pente d'une droite représentant la même accumulation de sucres.

Les valeurs de pentes calculées sont supérieures à cette valeur d'accumulation moyenne tout en restant dans le même ordre de grandeur. Ceci permet de valider un ajustement cohérent de ce paramètre.

Pour la base de données 2, une accumulation moyenne de 17,3°Brix (la moyenne de  $X_{s_{max}}$  est de 22,8 et la concentration en sucres avant le début de l'accumulation est de 5,5 °Brix) sur une période de 76 jours a été observée. Si l'accumulation est considérée comme linéaire, il y aura donc une augmentation de 0,23 °Br.j<sup>-1</sup>.

La pente de la courbe au point d'inflexion est calculée entre 0,39 et 0,88 °Br.j<sup>-1</sup>.

Ces valeurs sont supérieures à la valeur précédemment calculée tout en étant dans un

même ordre de grandeur. Les valeurs du paramètre  $Sl$  et donc du paramètre  $\tau$  dont elle dérive sont donc logiques.

### Retour sur la signification des paramètres du modèle

L'importation de sucres dans les baies de raisin se fait sous la forme de saccharose. Avant la véraison, le saccharose est consommé pour le métabolisme de la baie [80, 81]. Après la véraison, il est hydrolysé en fructose et glucose par une enzyme appelée invertase [27]. Ces sucres, appelés sucres réducteurs, sont accumulés dans la vacuole [78]. L'activité de l'invertase augmente de la floraison à la véraison puis reste constante sur toute la maturation. Ceci explique que l'accumulation des sucres dans les baies ne commence qu'à partir de la véraison.

L'importation de saccharose dans les baies se fait conjointement à l'eau et aux acides aminés via le phloème [21, 72]. Le phloème est inhibé à partir d'une concentration seuil en saccharose dans les baies dépendante du cépage. A titre d'exemple, cette concentration est de 20°Br dans les baies de Syrah [21, 81]. Le poids des baies est alors à son maximum et il peut diminuer par déshydratation des baies. L'action de la température porte à la fois sur l'activité photosynthétique, le métabolisme en général et l'intensité des migrations chez la vigne [80]. Une augmentation de la température favorise ces activités jusqu'à un niveau seuil à partir duquel l'activité de la feuille et le transport des nutriments est inhibé. Une alimentation hydrique suffisante est nécessaire au bon déroulement des processus biochimiques de la maturation. Mais, la circulation de la sève élaborée via le phloème n'est pas directement liée à l'état hydrique de la plante [80]. Une contrainte hydrique modérée n'a donc pas d'effet sur le potentiel de maturation de la baie tant que le métabolisme de création des sucres peut avoir lieu. Une contrainte hydrique trop forte, par contre, à partir de -0,6 MPa si l'on estime le potentiel de base [70], limite l'activité de la feuille et a donc un effet négatif sur la maturation. L'accumulation de sucres dans les baies sera donc plus rapide lorsque la température sera plus haute sauf si des températures ou des contraintes hydriques extrêmes sont atteintes.

Un point d'inflexion précoce peut donc être induit par un millésime chaud où l'afflux de sucres dans les baies est très rapide. La concentration inhibitrice de l'activité du phloème sera alors atteinte plus rapidement. La cinétique d'accumulation des sucres dans les baies de raisin sera donc précoce et rapide (pente élevée). Inversement, il sera supposé que les années ayant des points d'inflexion tardifs et avec une pente plus faible peuvent traduire un climat plus frais ou une contrainte hydrique très élevée.

L'activité du phloème s'arrête à partir d'une concentration en sucres seuil dépendante du cépage quel que soit le climat du millésime. Pour les millésimes à climat atypique, elle peut aussi se stopper plus tôt. Cela pourrait expliquer la concentration très faible pour les raisins de 2005 et une concentration assez faible dans les baies représentant le millésime 2003, pour les individus de la base de données 1. Les valeurs de  $X_{s_{max}}$  seront donc dues à une déshydratation des baies plus ou moins importante selon le climat. Les individus pour lesquels la fin de maturation est très chaude et sèche pourraient donc présenter des valeurs de  $X_{s_{max}}$  très élevées. De même, les individus avec les valeurs de  $X_{s_{max}}$  les plus importantes seront ceux avec le plus grand temps de latence entre la fin

de l'accumulation réelle et la vendange, à condition qu'aucune pluie ne vienne empêcher la déshydratation. La valeur de  $X_{S_{max}}$  n'est par contre pas déterminée par la précocité ou la rapidité de maturation.

La figure 6.4 montre la représentation graphique des 7 cinétiques ajustées grâce aux données de la base de données 1.

La figure 6.5, quant à elle, montre quelques exemples de courbes déterminées grâce aux données de la base de données 2. L'ensemble des courbes de cette base de données ne pouvant être tracé, la valeur minimale ou maximale de chacun des trois paramètres,  $X_{S_{max}}$ ,  $Infl$  et  $Sl$ , a été extraite de la base de données paramétrique relative à la base de données 2. Ces valeurs sont appelées valeurs extrêmes. Pour chacune de ces valeurs extrêmes, la valeur ajustée des deux autres paramètres, relatifs au modèle paramétrique ajusté dont est issue la valeur extrême, a permis de représenter la courbe réellement ajustée avec ce paramètre extrême. Ces 6 courbes permettent donc de représenter la variabilité des formes de courbes possibles avec les résultats d'ajustement de cette base de données.

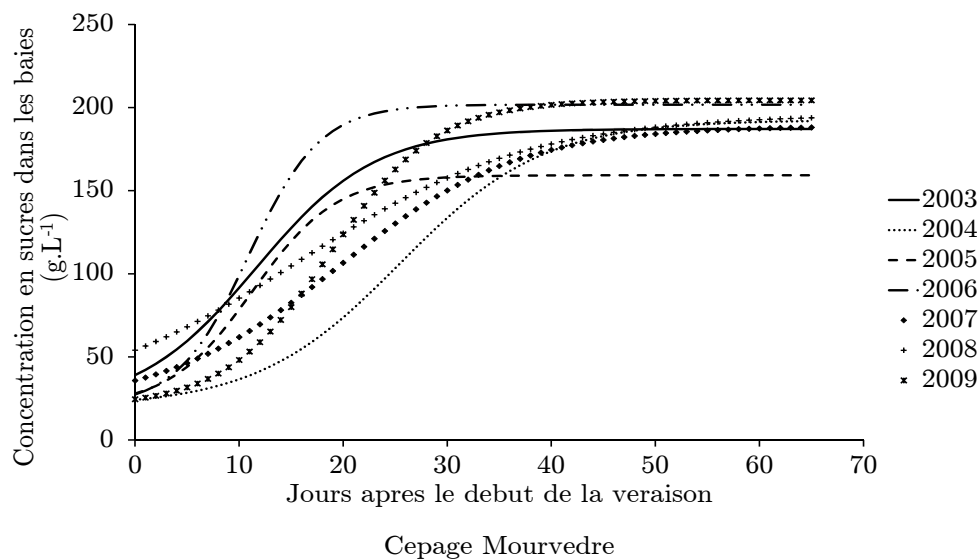


FIGURE 6.4 – Cinétiques d'accumulation des sucres dans les baies de raisin modélisées pour les 7 individus de la base de données 1

Quelle que soit la figure, chaque individu a un comportement spécifique.

Ainsi, pour la figure 6.4, 2006 est un millésime qui ressort comme ayant des baies très vite très concentrées en sucres, ce qui correspond selon les observations précédentes à un millésime chaud donc précoce et avec une accumulation rapide. A l'inverse, 2004 apparaît comme un individu avec des baies dans lesquelles la concentration reste longtemps faible, ce qui indique un millésime tardif et / ou lent et donc avec un climat plutôt plus frais.

Dans la figure 6.5, les principales différences entre les courbes s'observent au moment de la véraison ou peu après. Ainsi, les cinétiques caractéristiques du point d'inflexion

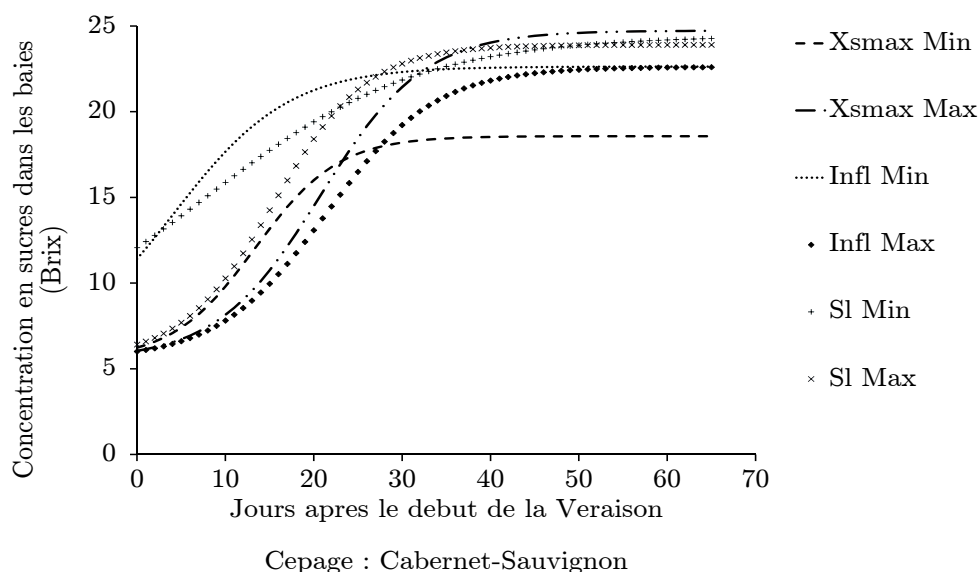


FIGURE 6.5 – Quelques cinétiques d'accumulation des sucres dans les baies de raisin pour la base de données 2

le plus précoce et de la pente la plus faible démontrent une accumulation des sucres nettement commencée à la date de véraison. La maturation a donc commencé avant cette date. Le début de la véraison est le moment où la première baie change de couleur ou devient translucide. La date du début de la véraison est positionnée grâce à des observations sur le terrain. Elle est difficile à évaluer, car il faut voir le changement de couleur de la première baie de la parcelle, qui peut être plus ou moins visible. Un retard dans la position de cette date peut donc être observé. Dans le cas de la base de données 2, les séquences temporelles couvrent toute la maturation. Pour des séquences temporelles ne couvrant que la fin de la maturation, l'ajustement de la courbe nécessite un point d'initialisation forçant le début de la cinétique à présenter des concentrations proches de la valeur experte fixée pour la véraison. Une date avancée ou retardée aura donc pour effet de modifier la valeur des paramètres mais l'ajustement restera toujours de forme identique et la valeur à véraison fixe. Pour des séquences temporelles démarrées avant véraison et allant jusqu'à la vendange, la courbe ajustée est affectée par la valeur d'initialisation estimée pour les premiers points de la séquence temporelle. Mais dès que l'accumulation linéaire commence, la cinétique s'ajuste aux valeurs mesurées. Une erreur dans la date de véraison pourra impliquer son estimation alors que la phase d'accumulation linéaire a déjà commencé. Le concentration en sucres dans les baies, au moment où la véraison commence, sera donc surestimée.

Il y a donc une variabilité très importante entre les cinétiques représentatives de chaque individu. Sur les courbes, il n'est pas possible de dire de façon certaine quels individus ont eu une cinétique semblable et quels individus ont eu une cinétique différente au regard de l'accumulation de sucres dans les baies.

Deux ACP ont été réalisées sur chacune des bases de données pour faciliter les comparaisons. La première utilise en variables les paramètres du modèle (utilisation intensive). La seconde a été réalisée en prenant chaque estimation journalière comme une variable indépendante (utilisation extensive).

### 6.3.3 Résultats des analyses descriptives de la base de données 1 après modélisation

#### Résultats de l'approche intensive

Pour cette application, les deux premières composantes principales ont été conservées. Elles représentent respectivement 51,49 et 36,85 % de variance totale expliquée. Le plan 1-2 correspond donc à 88,34 % de variance expliquée.

La figure 6.6 donne la position des 7 millésimes étudiés grâce à la base de données 1 sur le plan factoriel 1-2. Le millésime 2006 se situe en haut à gauche de la carte tandis que 2005 est en bas à gauche de celle-ci. 2004 est à droite, positionné sur l'axe 1. 2007 et 2008 sont plutôt à droite de la carte, un peu négatifs sur l'axe 2. 2003 est moins nettement à gauche que 2006 et 2005 mais toujours dans cette moitié de la carte factorielle et un peu négatif sur l'axe 2. Enfin, 2009 est centré sur l'axe 1 et nettement positif sur l'axe 2.

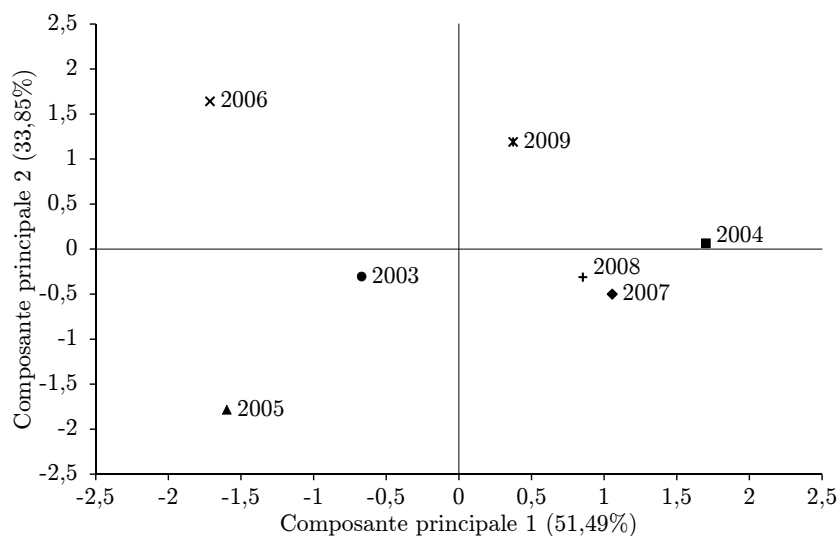


FIGURE 6.6 – Carte factorielle du plan 1-2 de l'ACP réalisée sur les 7 individus de la base de données 1 en prenant comme variables les 3 paramètres ajustés pour chaque modèle

Lors de l'analyse de la figure 6.4 et des valeurs des paramètres (tableau 6.3), 2003, 2005 et 2006 ont été observés comme les trois millésimes les plus précoces. Ils ont tous trois une pente au point d'inflexion assez élevée. L'accumulation a donc été précoce et rapide pour ces trois millésimes.

2004, 2007 et 2009 sont les millésimes les plus tardifs. La valeur de la pente au point d'inflexion de 2009 est plus élevée que celle de 2004 et 2007. Bien que tardive,

l'accumulation de sucres dans les baies de raisin pour le millésime 2009 a donc été rapide. 2004 et 2007 sont donc des millésimes tardifs avec une accumulation lente.

Enfin, 2008 est un millésime avec une pente très faible.

Il y a donc à gauche de la carte factorielle, les millésimes les plus précoces et de vitesse d'accumulation rapide, alors qu'à droite se trouvent les millésimes les plus tardifs et / ou de vitesse d'accumulation lente. Il est donc possible d'envisager que l'axe 1 est un axe de précocité et de rapidité de la maturation.

Les millésimes avec les  $X_{s_{max}}$  les plus élevés, 2006 et 2009 sont en haut de la carte, positifs sur l'axe 2. A l'autre extrémité, 2005, dont le  $X_{s_{max}}$  est très faible est en bas de celle-ci, négatif sur l'axe 2. Enfin, les 4 autres millésimes, de  $X_{s_{max}}$  proches et médians sont mal projetés sur la seconde composante principale. L'axe 2 pourrait donc être un axe représentant le chargement maximal en sucres dans les baies.

La figure 6.7(a) présente le cercle de corrélations des variables de l'ACP en cours d'étude. Chacun des trois paramètres est représenté à proximité du cercle des corrélations.

La première composante principale est bien corrélée positivement au paramètre  $Infl$  et moyennement corrélée négativement à la pente de la courbe au point d'inflexion. C'est donc un axe de précocité et rapidité de la cinétique d'augmentation de la concentration en sucres dans les baies de raisin. Plus un individu sera projeté négativement sur cet axe et plus il sera précoce et/ou avec une pente d'accumulation maximale rapide. La corrélation négative entre  $Sl$  et  $Infl$  permet de supposer que plus l'accumulation de sucres dans les baies présentera un point d'inflexion tardif et plus la vitesse maximale d'accumulation de sucres dans les baies sera faible. La corrélation moyenne et l'angle entre  $Sl$  et  $infl$  indiquent néanmoins que des exceptions peuvent exister à la relation citée juste avant. Ces conclusions recourent les observations réalisées lors de l'étude de la carte factorielle du plan 1-2. La première composante principale est donc bien un axe de précocité et rapidité de la maturation.

La seconde composante principale est bien corrélée positivement au paramètre  $X_{s_{max}}$  et légèrement corrélée positivement à la pente. Cet axe peut donc bien être décrit comme un axe représentant le chargement maximum en sucres dans les moûts.

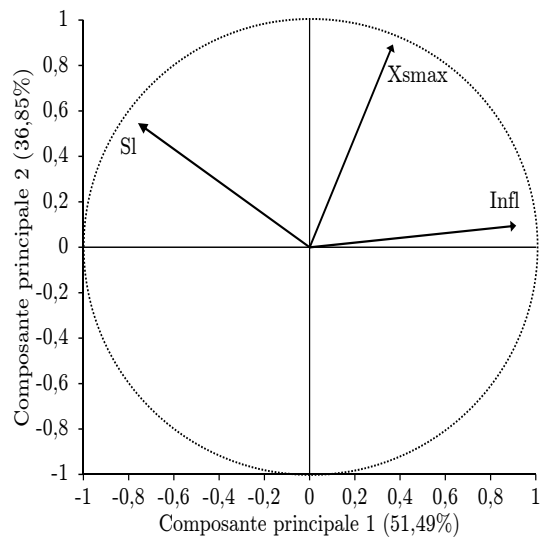
Ces deux composantes principales mettent en avant qu'il y a peu de relation entre le paramètre  $Infl$  et le paramètre  $X_{s_{max}}$ . Pour la base de données 1, la charge maximale en sucres dans les moûts n'est donc pas déterminée par la précocité de la parcelle.

La pente et le paramètre  $X_{s_{max}}$  sont projetés orthogonalement l'un à l'autre. Ces deux paramètres ne présentent donc aucune corrélation. La valeur de  $X_{s_{max}}$  ne sera donc en aucun cas déterminée par la valeur de  $Sl$  et inversement.

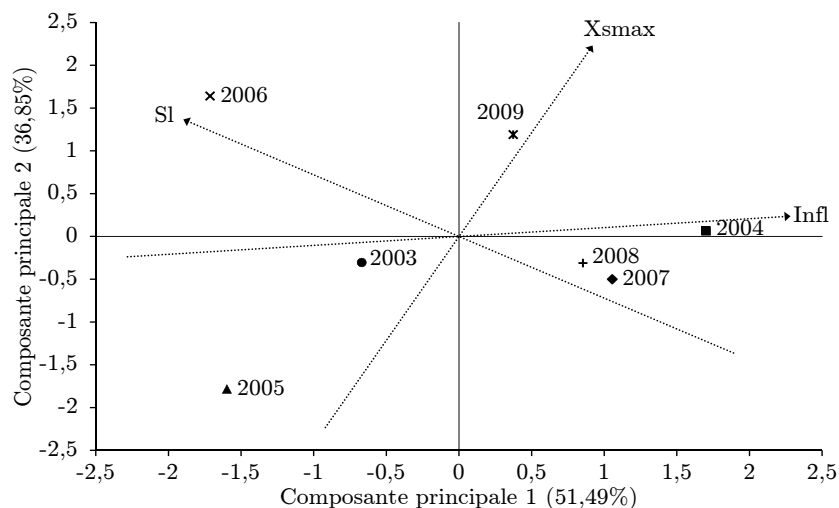
Le paramètre  $X_{s_{max}}$  est donc plus indépendant des deux autres paramètres que ceux-ci entre eux. Pour les individus de la base de données 1, la concentration maximale en sucres d'un individu n'est donc pas déterminée par sa précocité ou sa vitesse maximale d'accumulation.

Différents travaux antérieurs ont montré que la valeur de  $X_{s_{max}}$  est principalement déterminée par une déshydratation plus ou moins importante après l'arrêt d'apport actif





(a)



(b)

FIGURE 6.7 – Analyse en composante principale des 7 millésimes de la base de données 1 par les trois paramètres de chaque modèle : (a) cercle des corrélations, (b) carte factorielle du plan 1-2

de sucres dans les baies par le phloème [21, 80, 81]. La valeur de  $X_{smax}$  n'est donc effectivement pas liée à une plus grande précocité de la maturation ou à une vitesse d'accumulation plus élevée.

Dans la figure 6.7(b), les directions des paramètres, déterminées par l'ACP, ont été insérées à la carte factorielle étudiée.

La position des individus rappelle les résultats de l'analyse du tableau 6.3. Les individus qui avaient été observés comme précoces, 2003, 2005 et 2006, se projettent tous négativement sur l'axe représentatif du paramètre *Infl*. Les individus avec une valeur de  $X_{smax}$  élevée, 2004, 2006 et 2009, se projettent positivement sur la direction déterminée par ce paramètre. Enfin, les individus avec une pente faible, 2007 et 2008, se projettent

négativement sur l'axe déterminé par le paramètre  $Sl$ .

Grâce à ce plan factoriel, le paramètre le plus déterminant pour différencier un individu des 6 autres est mis en avant.

Le millésime 2006 est caractérisé par une pente très élevée. La position de 2005 et 2009 est plutôt déterminée par leur valeur de  $X_{s_{max}}$ . La valeur faible de la pente au point d'inflexion semble influencer le positionnement de 2007 et 2008 plus que celui du millésime 2004. Enfin, le millésime 2004 est plutôt marqué par un caractère tardif du millésime.

En 2004, Tonietto et Carbonneau ont proposé une classification multicritère des zones viticoles se basant sur trois indices bioclimatiques : l'indice héliothermique de Huglin (IH), l'indice de sécheresse (IS) et l'indice de fraîcheur des nuits (IF) [92]. Le tableau 6.1 présente les valeurs de ces indices pour chacun des 7 millésimes de la base de données et un regroupement par classification ascendante hiérarchique des millésimes en fonction de ces indices [83]. Ces classes présentent des liens avec la position des individus décrite ci-dessus :

- 2003 et 2006 sont des millésimes chauds et secs. La chaleur peut avoir induit une accumulation très rapide des sucres qui est ensuite stoppée par une contrainte hydrique trop importante ou même des températures trop élevées [9] (cas de 2003).
- 2004 et 2007 sont plus frais et présentent un bilan hydrique moins limitant. Ce contexte hydrique moins limitant a favorisé une accumulation plus lente qui n'est pas stoppée.
- 2008 a une cinétique très semblable à 2007. Pour ce millésime, les températures plus fraîches peuvent avoir compensé l'effet de la contrainte hydrique.
- 2009 est une année de vendange très précoce et très concentrée en sucres sur tout le domaine. Malgré un bilan climatique ne présentant pas tant de chaleur, la maturation a été extrêmement rapide. Compte tenu des informations disponibles, aucune interprétation rigoureuse n'a pu être formulée.
- 2005 est une année atypique. Un orage de 124 mm étant intervenu le 5 septembre a impliqué une confusion dans le calcul de l'indice de sécheresse. La saturation en eau des sols après l'orage peut avoir eu pour effet de stopper la maturation de ce millésime. Cette supposition a été validée par des employés de la station expérimentale de l'INRA Pech Rouge. Ce millésime précoce et rapide, qui aurait pu être aussi concentré que 2006 a donc pu être artificiellement bloqué à une concentration maximale en sucres très basse.

Pour la base de données 1, il est possible d'observer que la teneur maximale en sucres dans les baies de raisin n'est pas limitée par des températures trop basses. En effet, les concentrations maximales en sucres les plus basses ne sont pas liées aux millésimes où l'indice de Huglin est le plus bas. L'accumulation de sucres va être limitée par une contrainte autre que le besoin de chaleur. Une contrainte hydrique trop importante et des températures très élevées (cas de 2003) ou un événement climatique particulier (cas de 2005) ont par exemple limité la teneur maximale de sucres dans les baies.

Cette ACP met en avant l'intérêt de travailler sur les trois paramètres représentant la cinétique complète d'accumulation de sucres dans les baies de raisin entre la véraison et la vendange. En effet, la concentration maximale de sucres dans les baies, la précocité et la rapidité de maturation peuvent être reliées à des typologies climatiques et ainsi fournir une typologie globale du millésime.

Cette analyse met aussi en avant la nécessité d'avoir des connaissances sur le climat ou l'environnement de la plante représentative de chaque individu pour expliquer et pondérer les observations possibles lors de l'analyse de la maturation par exemple.

Pourtant, comme cela a été montré dans la figure 6.4, les paramètres ajustés et la forme de la courbe qui en résulte peuvent être très différents. L'approche intensive présentée ici permet de travailler sur la forme globale de chaque cinétique. L'étude de l'accumulation au jour le jour pourrait de ce fait apporter d'autres informations. C'est le sujet de la partie suivante sur l'utilisation du modèle en extension.

### Résultats de l'approche extensive

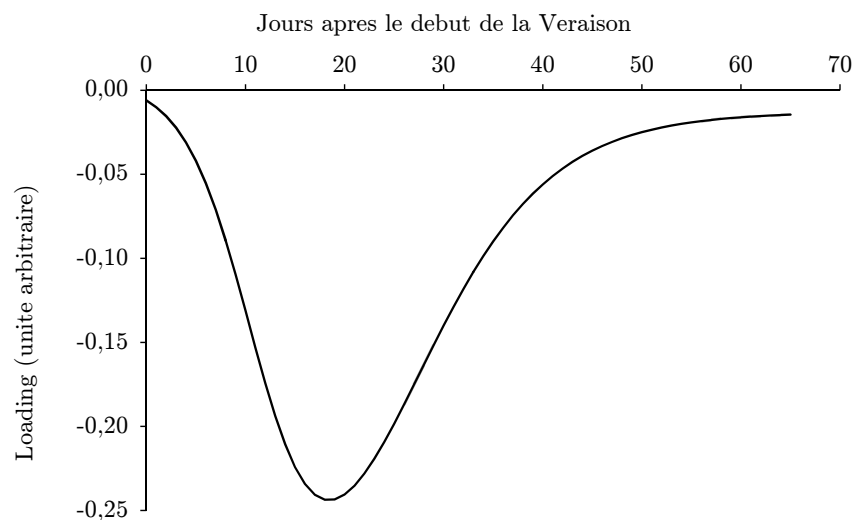
Cette seconde application considère que chaque estimation journalière de la concentration en sucres dans les baies de raisin est un descripteur indépendant pour l'ACP. La concentration en sucres dans les baies de chaque jour est donc une variable qui sera prise en compte indépendamment des autres dans l'ACP. Pour chaque individu, le nombre de variables est donc très important (66 dans le cas présent). Chacune des variables est dépendante des autres puisque faisant partie d'un même processus chimique. De telles analyses sont très courantes en chimométrie [39, 64, 99]. Les variables étant toutes dépendantes, les auteurs précités ont proposé d'étudier les composantes principales sous forme de loading plutôt que de travailler sur le cercle des corrélations. Un loading représente, pour l'ensemble des variables, l'évolution du poids de chacune d'entre elle dans la détermination de la composante principale.

Pour cet exemple, les deux premières composantes principales, correspondant respectivement à 70,77% et 23% de variabilité expliquée ont été conservées. La figure 6.8 présente le loading de la première composante principale (a) et celui de la seconde composante principale (b). Le jour après le début de la véraison est l'abscisse de ces courbes. En ordonnée, à chaque jour après le début de la véraison est associé le poids de celui-ci dans la détermination de la composante principale.

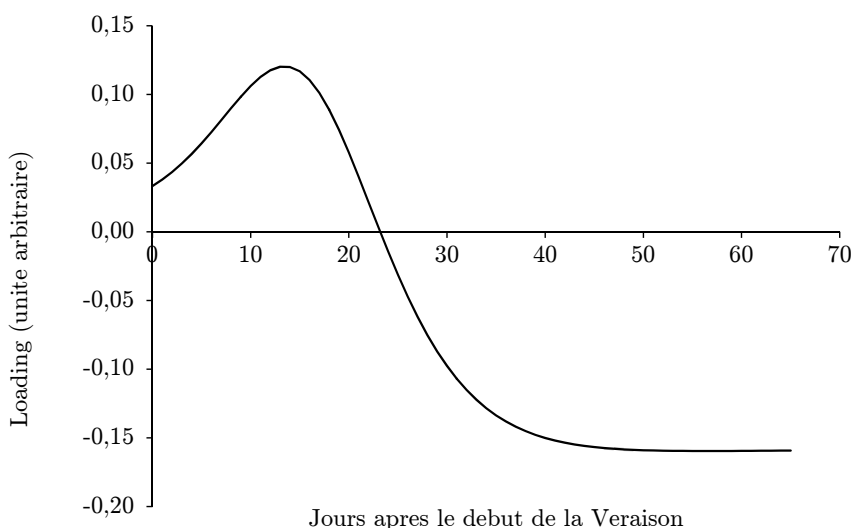
Le loading de la première composante principale représente 70,77% de variabilité expliquée. Il va donc permettre d'identifier les courbes qui sont les plus différentes, ainsi que les fenêtres qui impliquent cette différence.

Le loading de la première composante suit une courbe de forme approximativement gaussienne. Les dates avant le 10ème et après le 30ème jour après le début de la véraison ont un poids très faible dans la détermination de cette composante principale.

En première approximation, cette composante peut donc être simplifiée en un créneau entre le 10ème et le 30ème jour après la véraison. Dans ce créneau, les variables proches



(a) % de variance expliquée : 70,77%



(b) % de variance expliquée : 23,00%

FIGURE 6.8 – Loadings des deux premières composantes principales de l'ACP réalisée sur l'utilisation extensive des modèles ajustés à partir de la base de données 1 : (a) loading de la première composante principale, (b) loading de la seocnde composante principale

du 20ème jour après la véraison ont un poids plus important dans la détermination de la composante principale que les autres. Ce créneau représente la variabilité des courbes dans la phase d'accumulation linéaire.

Cette composante principale permet de différencier les cinétiques en fonction de l'intégrale de leur concentration en sucres entre le 10ème et le 30ème jour après la véraison. L'intégrale du taux de sucres entre deux dates sera d'autant plus importante que la concentration en sucres dans les baies est élevée et qu'elle augmente dans cette phase. Les millésimes vont donc se différencier, sur cette composante principale, en fonction de leur concentration en sucres relativement à celle des autres millésimes sur ce créneau.

Presque 71% de la variabilité entre les 7 courbes en cours d'étude est donc approximativement comprise dans un créneau limité et tôt après la véraison. La variabilité entre les sept cinétiques au moment de la véraison et en fin de maturation ne représente donc au plus que 30% de la variabilité totale.

Le loading de cette composante principale est entièrement négatif. Les individus ayant les concentrations en sucres les plus faibles entre le 10ème et le 30ème jour après la véraison seront donc projetés dans le sens positif et inversement.

La figure 6.9 présente les 7 cinétiques ajustées sur les individus de la base de données 1. Les parties des courbes grisées représentent les fenêtres de dates qui n'ont pas présenté de poids important dans la détermination du loading.

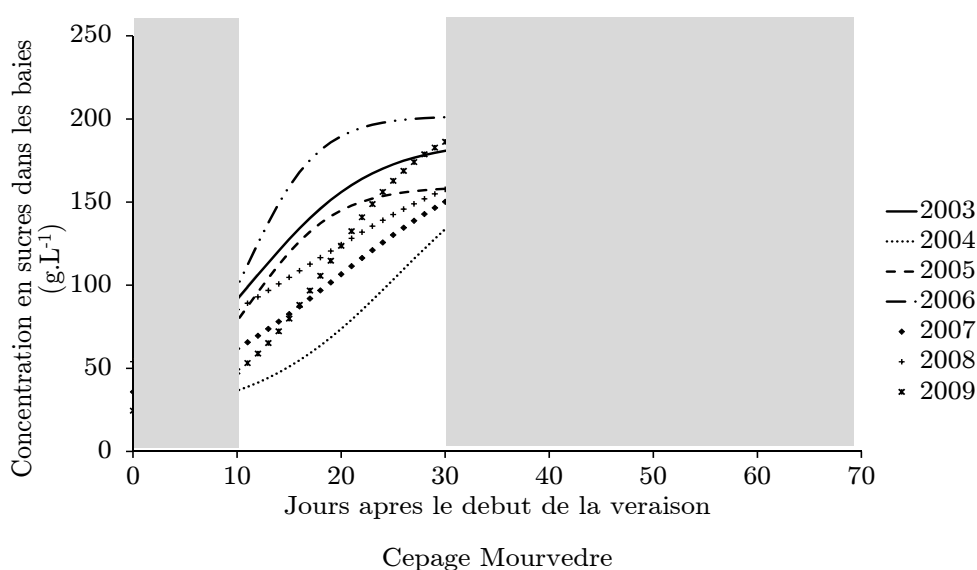


FIGURE 6.9 – Variabilité des 7 cinétiques étudiées dans le créneau déterminé par la première composante principale

Sur les courbes présentées dans la figure 6.9, les individus les plus concentrés dans le créneau contenu entre le 10ème et le 30ème jour après le début de la véraison sont le millésime 2003 et le millésime 2006. Ces individus ont auparavant été décrits comme précoces et présentant une accumulation rapide. Les individus les moins concentrés sont le millésime 2004 et le millésime 2007. Ces millésimes ont été caractérisés comme tardifs. Un individu précoce et rapide sera très rapidement fortement chargé en sucres. A l'inverse, un individu tardif et qui aurait une pente faible aura des concentrations en sucres faibles pour un même jour après la véraison. Cette composante principale peut donc être décrite comme un axe combinant précocité et rapidité d'accumulation soit la pente de la courbe dans la phase d'accumulation linéaire.

Le loading de la seconde composante principale est positif puis négatif. Les jours avant le 10ème jour après la véraison puis compris entre le 20ème et le 30ème jour après le

début de la véraison ont un poids très faible dans la détermination de cette composante principale. Les jours compris entre le 10ème et le 20ème jour après le début de la véraison ont un poids positif dans la détermination de cette composante. Puis, les jours postérieurs au 30ème jour après la véraison ont un poids négatif dans la détermination de celle-ci.

En première approximation, cet axe est donc une combinaison entre une intégrale de la concentration en sucres en début de phase d'accumulation linéaire et une intégrale de la concentration en sucres en fin de maturation.

Compte tenu de la base temporelle  $T_{\Phi}$  choisie, la période après le 30ème jour après le début de la véraison couvre une période de 35 jours. Le poids de la période après le 30ème jour après le début de la véraison est donc très important dans la détermination de ce loading. La concentration en sucres dans les baies en fin de maturation aura donc un effet plus important dans la détermination de la position d'un individu sur cette composante principale que sa concentration entre le 10ème et le 20ème jour après le début de la véraison.

La première partie du loading de la seconde composante principale peut être étudiée comme un créneau entre le 10ème et le 20ème jour après le début de la véraison. Les individus vont donc d'abord se séparer en fonction de l'intégrale de la concentration en sucres dans les baies de chacun dans ce créneau. La seconde partie de ce loading permet de différencier les individus en fonction de l'intégrale de la concentration de sucres dans leurs baies après le 30ème jour après le début de la véraison. La position de chaque individu sur cette composante principale sera donc une combinaison entre ces deux intégrales.

Puisque cette composante principale est déterminée par une première partie de loading positive puis une seconde partie de loading négative, les individus qui vont ressortir sur cette composante sont soit (i) les individus qui auront eu un début de maturation avec des concentrations en sucres plus faibles que les autres individus et une fin de maturation avec des concentrations en sucres plus élevées que les autres individus, soit (ii) les individus qui auront eu un début de maturation avec des concentrations en sucres plus élevées que les autres individus et une fin de maturation avec des concentrations en sucres plus faibles que les autres individus.

Si les deux valeurs d'intégrales sont élevées, alors la position de l'individu sur la seconde composante principale sera compensée entre la fenêtre comprise entre le 10ème et le 20ème jour après le début de la véraison et la fenêtre après le 30ème jour après le début de la véraison. De même, si les deux valeurs d'intégrales sont faibles, la position de l'individu en cours d'étude sur la deuxième composante principale sera à nouveau compensée.

Ce loading permet donc de mettre en avant les cinétiques les plus amples, soit avec le plus d'écart de concentration, entre la période comprise entre le 10ème et le 20ème jour après le début de la véraison et la fin de la maturation et les cinétiques les moins amples entre ces même périodes. Cette notion d'amplitude est explicitée dans la figure 6.10. Dans cette figure, on observe que la cinétique 1 est une cinétique avec une forte concentration en sucres par rapport à la cinétique 2 dans le créneau compris entre le 10ème et le 20ème jour après le début de la véraison et une faible concentration en sucres par rapport à la cinétique 2 après le 30ème jour après le début de la véraison. L'écart

entre les concentrations en sucres dans les baies des raisin représentées par la cinétique 1 est faible entre les deux fenêtres étudiées, tandis que celui entre les concentrations en sucres dans les baies de raisin représentées par la cinétique 2 est important entre les deux fenêtres étudiées. La cinétique 1 est représentative d'une cinétique peu ample. La cinétique 2 est représentative d'une cinétique ample.

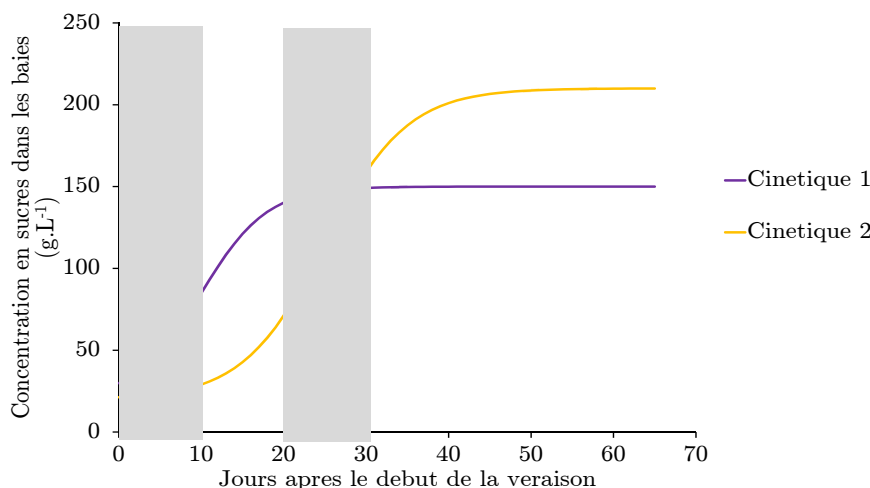


FIGURE 6.10 – Mise en avant de la notion d'amplitude de courbe pour la différenciation des individus sur la seconde composante principale

Les individus que cette composante fait donc ressortir sont les individus ayant un  $X_{s_{max}}$  élevé mais une concentration en sucres dans les baies faible peu de temps après le début de la véraison, d'un côté. Et, de l'autre côté, les individus ayant un  $X_{s_{max}}$  faible mais une concentration en sucres dans les baies élevée peu de temps après le début de la véraison.

La figure 6.11 présente les cinétiques ajustées pour chacun des 7 individus de la base de données 1. Les fenêtres de dates ayant un poids non déterminant pour la seconde composante principale sont grisées.

2005 est un millésime caractérisé par une cinétique très peu ample entre le créneau [10; 20] et le créneau postérieur au 30ème jour après le début de la véraison. Ce millésime devrait donc ressortir nettement positif sur cet axe. A l'inverse, 2009 est un millésime pour lequel la concentration en sucres dans les baies dans le créneau [10; 20] est un peu plus faible que la moyenne tandis que la concentration maximale en sucres dans les baies pour ce millésime est très élevée. Ce millésime devrait donc ressortir négativement sur cet axe.

A l'oeil nu, il n'est ensuite pas possible d'observer d'autres courbes pour lesquelles la faiblesse ou l'importance de l'amplitude entre les concentrations dans les baies de raisin sur les deux fenêtres de temps identifiées serait caractéristique.

La figure 6.12 présente la position des 7 individus de la base de données étudiée dans le plan défini par les deux composantes principales décrites ci dessus.

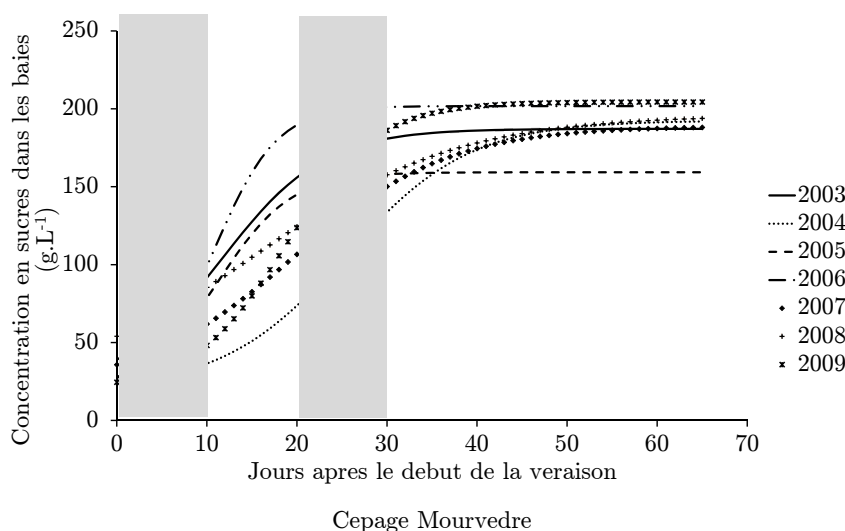


FIGURE 6.11 – Variabilité des 7 cinétiques étudiées dans les fenêtres déterminées par la seconde composante principale

Sur la première composante principale, les individus se projettent dans l'ordre suivant : 2004 puis 2007 puis un groupe médian composé de 2005, 2008 et 2009 enfin 2003 puis 2006. Sur la seconde composante principale, 2005 est positif, 2009 est négatif et les autres millésimes sont centraux. Sur ce plan, un classement des millésimes sur la première composante principale peut donc être observé et une différenciation en fonction de la seconde composante principale s'ajoute.

Comme la figure 6.9 le montre, 2004 est le millésime avec les baies les moins concentrées dans le créneau compris entre le 10<sup>ème</sup> et le 30<sup>ème</sup> jour après le début de la véraison, mis en avant par la première composante principale. Opposé à 2004, 2006 est le millésime avec les baies les plus concentrées dans ce même créneau. Dans l'analyse précédente, 2004 a été indiqué comme un millésime tardif et donc la maturation est plutôt lente. 2006 est quant à lui le millésime le plus précoce avec une pente maximale d'accumulation élevée. L'axe 1 permet donc effectivement de différencier les millésimes précoces des millésimes tardifs. Le créneau étudié faisant une intégrale entre le 10<sup>ème</sup> et le 30<sup>ème</sup> jour après la véraison, un millésime précoce mais dont l'accumulation s'arrête tôt (cas de 2003 et 2005) sera donc projeté de façon moins négative qu'un millésime précoce et dont l'accumulation n'est pas encore stoppée au 30<sup>ème</sup> jour après la véraison. C'est pourquoi 2003 et 2005 sont plus centraux que 2006 sur cet axe. 2007 est le second millésime le moins concentré sur le créneau [10; 30]. C'est un millésime tardif et d'accumulation lente mais sa différence de concentration dans les baies avec 2004 implique une position de valeur positive plus faible que 2004. Les autres millésimes sont centraux sur la fenêtre étudiée, ce qui explique leur position médiane sur l'axe 1.

Sur l'axe 2, seuls les millésimes 2005 et 2009 ressortent. Comme il a été dit précédemment, 2005 présente très peu de différence de concentration entre le créneau [10; 20] et le créneau postérieur au 30<sup>ème</sup> jour après le début de la véraison. Il est donc logiquement



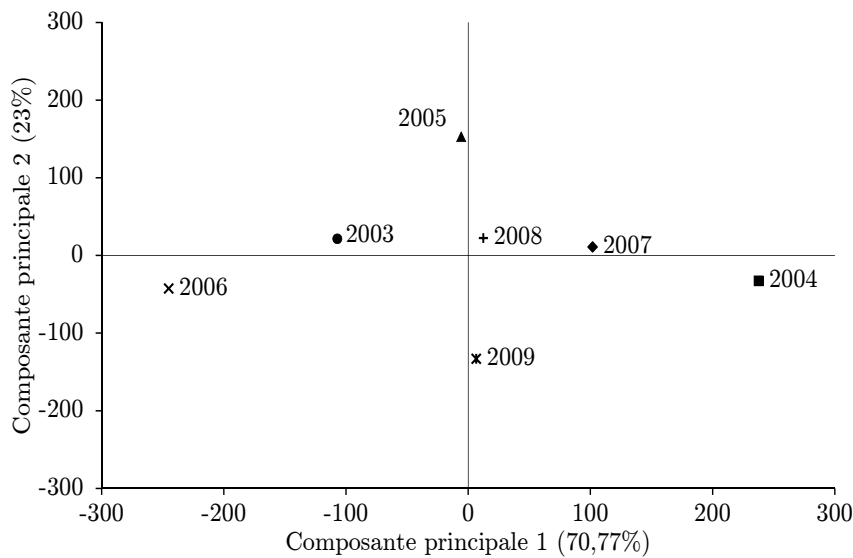


FIGURE 6.12 – Carte factorielle du plan 1-2 de l'ACP réalisée sur l'utilisation extensive des modèles ajustés à partir de la base de données 1

projeté positivement sur cet axe. 2005 est un millésime précoce, dont l'accumulation a été rapide puis stoppée à des concentrations très basses suite à un très fort orage. C'est pourquoi il présente aussi peu de différence de concentration entre les deux fenêtres considérées. 2009 est le second millésime le moins concentré entre le 10ème et le 20ème jour après le début de la véraison. C'est ensuite le second millésime le plus concentré après le 30ème jour après le début de la véraison. La pente au point d'inflexion de la cinétique de ce millésime est élevée. C'est un millésime qui a commencé sa maturation tard mais pour lequel celle-ci a ensuite été très rapide. C'est donc un millésime qui présente une grande différence de concentration entre le créneau [10; 20] et la fenêtre après le 30ème jour après le début de la véraison. Il est donc normal qu'il soit projeté négativement sur l'axe 2.

Dans ces deux exemples, 7 individus, initialement non comparables car étant caractérisés par des séquences temporelles asynchrones et de dimension différentes, ont pu être comparés à l'aide d'une ACP. Leur comportement a pu être expliqué par des connaissances sur le climat et l'analyse des cinétiques d'accumulation de chacun. Dans ce cas, la méthode de traitement des données proposée dans le chapitre 3 répond donc aux besoins exprimés de comparaison d'individus.

### 6.3.4 Résultats des analyses descriptives de la base de données 2 après modélisation

A partir des modèles ajustés pour les individus de la base de données 2, deux ACP ont été réalisées. La première est effectuée en prenant pour variables les paramètres ajustés pour chaque individu. La seconde utilise comme variables l'estimation journalière de la

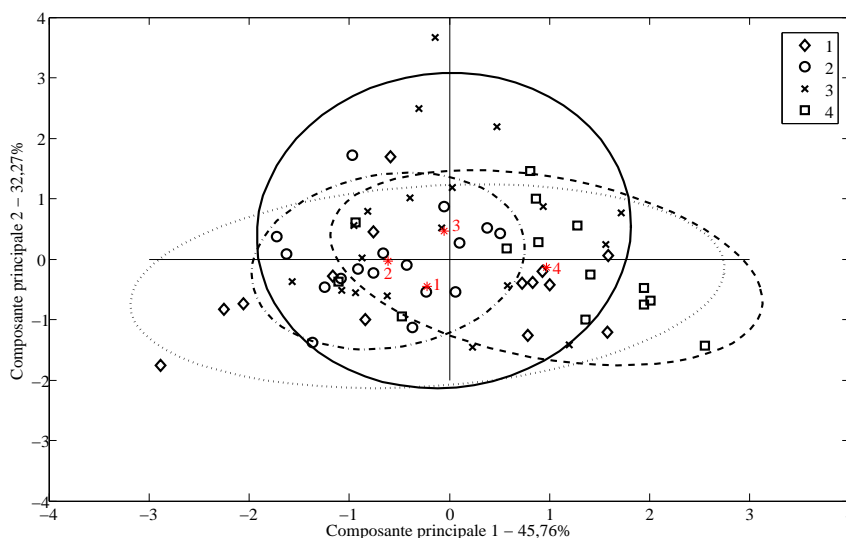


FIGURE 6.13 – Carte factorielle du plan 1-2 de l'ACP réalisée sur les 65 individus de la base de données 2. La classe 1 est représentée par l'ellipse en pointillés, la classe 2 par l'ellipse en pointillés mixtes, la classe 3 par l'ellipse en trait continu puis la classe 4 par l'ellipse en tirets.

concentration en sucres dans les baies de raisin entre le jour où la véraison commence et le 65ème jour après le début de la véraison. Ces deux analyses sont donc identiques aux analyses présentées pour l'étude de la base de données 1. L'objectif de ces analyses est de tester la possibilité à partir d'une méthodologie commune d'obtenir des paramètres de cinétiques ou leur représentation graphique utiles pour la comparaison d'individus. Cette analyse a donc pour objectif de tester la généralité de l'approche proposée.

Dans cette approche, un seul millésime,  $T$ , 2011 a été pris en compte. 65 parcelles  $s_j$  sont présentes dans la base de données 2. Ces parcelles représentent chacune un individu  $I$ .

### Résultats de l'approche intensive

Pour cette approche, uniquement les deux premières composantes principales ont été conservées. Elle représentent respectivement 45,76 et 33,27 % de variance expliquée. Le plan 1-2 représente donc 79,03 % de variance totale expliquée.

La figure 6.13 montre la position des individus de la base données 2 sur la carte factorielle constituée des deux premières composantes principales de l'ACP. le symbole choisi pour chaque individu correspond à la classe agroclimatique dans laquelle se situe le domaine auquel la parcelle appartient. Comme précédemment, cette classe a été déterminée grâce à la classification proposée par Tonietto et Carbonneau en 2004 [92].

Chaque zone est aussi représentée par une ellipse. L'ellipse de chaque classe représente la surface dans laquelle 95% des individus de cette classe devraient être présents si la distribution de ceux-ci était gaussienne. Le centre de chaque ellipse est le centre de gravité du nuage de points correspondant à la classe agroclimatique. Pour chacune des classes, il est mis en avant sur la carte factorielle par un point rouge.

Le tableau 6.5 donne les moyennes et écart-types des valeurs des paramètres ajustés pour chaque classe agroclimatique.

TABLE 6.5 – Moyenne et écart-type de chaque paramètre par zone agroclimatique

Zone	Brix Max ( $^{\circ}\text{Br}$ )	Infl (j)	Pente ( $^{\circ}\text{Br}.\text{j}^{-1}$ )
1	23,2 +/- 0,9	14,4 +/- 5,2	0,61 +/- 0,11
2	22,8 +/- 0,8	10,9 +/- 3,1	0,64 +/- 0,08
3	22,2 +/- 1,5	15,0 +/- 3,8	0,63 +/- 0,07
4	23,1 +/- 1,0	16,2 +/- 2,9	0,73 +/- 0,10

Le paramètre *Infl* a une valeur moyenne d'approximativement 11 jours après le début de la véraison pour la classe 2. C'est la valeur la plus basse entre les 4 zones agroclimatiques. Ce même paramètre a une valeur moyenne d'approximativement 16 jours après le début de la véraison pour la classe 4. C'est la valeur la plus haute entre les 4 classes. La classe 1 et la classe 3 présentent respectivement une valeur de *Infl* d'approximativement 14,5 puis 15 jours après le début de la véraison. La classe 2 est donc la classe la plus précoce alors que la classe 4 est la classe la plus tardive. Les deux autres classes présentent une précocité moyenne, plus proche de celle de la classe 4 que de celle de la classe 2.

La classe 3 présente une valeur de  $X_{s_{max}}$  d'approximativement  $22^{\circ}\text{Br}$  tandis que les trois autres classes présentent une valeur de  $X_{s_{max}}$  proche de  $23^{\circ}\text{Br}$ . La classe 3 a donc des parcelles qui ont présenté une concentration maximale en sucres dans les baies un peu plus basse que les trois autres classes.

Enfin, la classe 4 présente une valeur de pente proche de  $0,7^{\circ}\text{Br}.\text{j}^{-1}$  alors que la pente est calculée avec une valeur proche de  $0,6^{\circ}\text{Br}.\text{j}^{-1}$  pour les 3 autres classes. La pente des individus de la classe 4 est donc un peu plus élevée que celle des individus des trois autres classes. C'est donc une classe où l'accumulation de sucres dans les baies des individus y appartenant a été un peu plus rapide que dans les baies des individus des trois autres classes.

Sur la figure 6.13, les classes sont assez confondues. Des tendances peuvent tout de même être mises en avant. Ainsi, la classe 2 est la plus à droite du plan. La classe 1 possède les individus les plus à gauche du plan. Enfin, la classe 3 présente les individus les plus en haut du plan.

Sur la direction de la première composante principale, il y a donc un ordonnancement des classes, depuis la classe 1 vers la classe 2. La classe 4 a très peu d'individus sur la droite du plan et est globalement la plus à gauche. La classe la plus précoce ressort donc nettement sur cet axe. La zone la plus tardive est légèrement mise en avant. La première composante principale pourrait donc être à nouveau déterminée par le paramètre *Infl*.

Sur la direction de la seconde composante principale, uniquement la classe 3 se différencie des trois autres classes. La classe 3 présente des individus avec un  $X_{s_{max}}$  moyen plus faible que dans les autres classes. La seconde composante principale pourrait donc être à nouveau déterminée par le paramètre  $X_{s_{max}}$ .

La figure 6.14 présente le cercle des corrélations des trois variables de cette ACP.

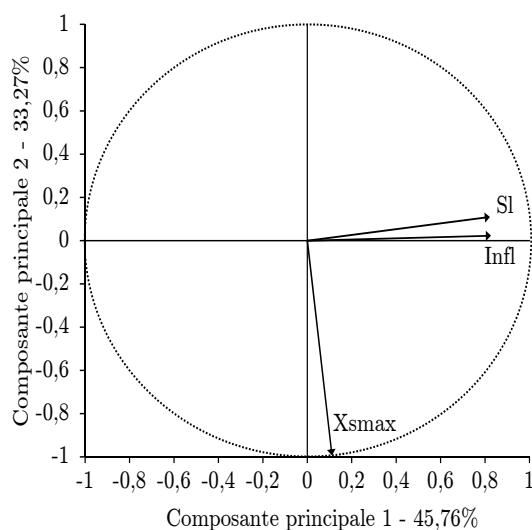


FIGURE 6.14 – Cercle des corrélations des variables de l'ACP réalisée sur les 65 parcelles de la base de données 2 en prenant en variables les trois paramètres ajustés pour chaque modèle

La première composante principale est déterminée par les paramètres *Infl* et *Sl*. Ils se projettent tous les deux dans le sens positif. Le paramètre *Infl* et le paramètre *Sl* sont donc corrélés. La classe 4, avec la valeur de *Infl* la plus tardive, est en effet la classe présentant la valeur de pente au point d'inflexion la plus élevée. Cet axe est donc un axe de tardiveté et de rapidité maximale d'accumulation. Plus un individu sera projeté en négatif sur cet axe et plus il sera précoce et / ou de vitesse maximale faible. Inversement, un individu projeté en positif sera tardif et de vitesse maximale d'accumulation des sucres dans les baies importante. Contrairement aux individus de la base de données 1, pour la base de données 2, la précocité d'accumulation, soit une valeur de *Infl* faible, ne préjuge pas de la rapidité de cette même accumulation de sucre dans les baies de raisin. Un individu plus tardif aura même tendance à présenter une vitesse maximale d'accumulation plus élevée.

La seconde composante principale est déterminée par le paramètre  $X_{s_{max}}$ . Il se projette dans le sens négatif. Cet axe est donc un axe de chargement maximal des baies. Plus la concentration maximale dans les baies d'un individu sera élevée et plus il sera projeté dans le sens négatif sur cet axe.

Comme lors de l'étude de la base de données 1, les paramètres  $X_{s_{max}}$  et *Sl* sont orthogonaux. Ils ne présentent donc pas de corrélation. La vitesse maximale d'accumulation n'a donc, pour cet ensemble d'individus, aucun effet sur la concentration maximale en sucres dans les baies. Les paramètres  $X_{s_{max}}$  et *Infl* sont quasiment orthogonaux. De la même façon que cela avait été observé lors de l'étude de la base de données 1, pour la base de données 2, la valeur de  $X_{s_{max}}$  ne sera donc pas déterminée par la valeur de *Infl* ou de *Sl*.

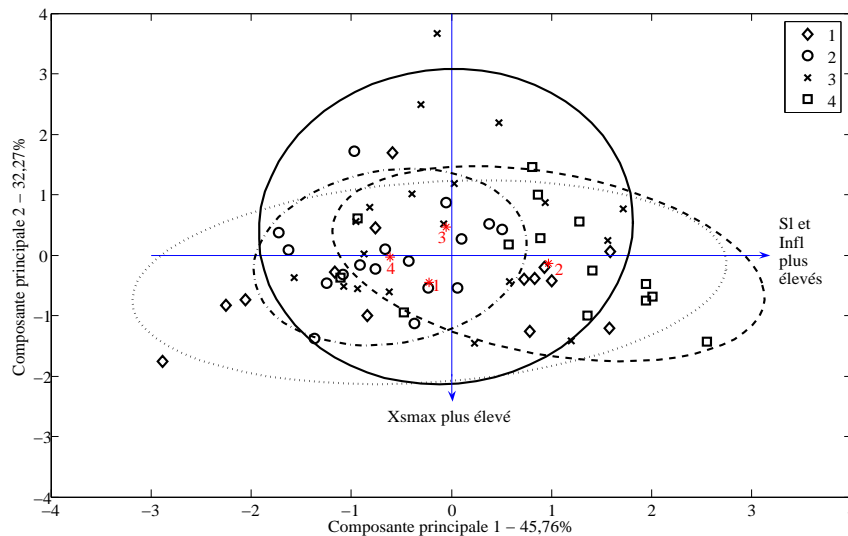


FIGURE 6.15 – Carte factorielle du plan 1-2 de l'ACP réalisée sur les 65 individus de la base de données 2

La figure 6.15 présente à nouveau le plan factoriel 1-2. Sur cette figure, la signification des deux axes est rappelée.

Les individus appartenant à chaque zone sont très étalés sur tout le plan factoriel. L'effet du climat pour l'étude de cette base de données est donc moins net que dans le cas de l'étude de la base de données 1.

Toutes les parcelles de cette étude sont irriguées. Le niveau de contrainte hydrique de chaque individu est donc à la discrétion de chaque viticulteur. La formule d'indice de sécheresse ne prend pas en compte les quantités d'eau apportées par irrigation. La valeur d'indice de sécheresse de chaque classe n'est donc pas représentative de la contrainte hydrique de chaque individu. En insérant les quantités d'eau apportées à chaque parcelle par irrigation, un indice de sécheresse corrigé pourrait être calculé pour chaque individu et cet indice pourrait alors être utilisé pour l'analyse des résultats.

Pour la base de données 2, des trois indices utilisés par Tonietto et Carbonneau pour leur classification agroclimatique, les indices calculés à partir des températures de l'air, IH et IF, seront donc plus représentatifs du climat subi par les individus que l'indice de sécheresse IS.

Pour l'étude de la position des différentes classes sur le plan factoriel 1-2, plus d'attention sera donc portée aux indices IH et IF qu'à l'indice IS. L'indice IF est utile pour connaître le type d'arômes qui seront présents dans les baies [92]. L'indice héliothermique, IH, est donc celui qui a le plus d'influence sur la rapidité de l'accumulation des sucres dans les baies et le potentiel de concentration maximale en sucres dans les baies de raisin.

Le tableau 6.6 donne la moyenne de l'indice IH et des moyennes de températures moyenne, maximale et minimale sur la période allant du 01/04 au 30/09 pour chacune des 4 classes agroclimatiques.

Les classes 3 et 4 présentent une valeur de IH plus faible que les classes 1 et 2. Les classes 3 et 4 présentent aussi des températures maximales plus faibles que les classes 1

### Application 3 : Modélisation de l'accumulation de sucres dans les baies de raisin pendant la maturation

---

TABLE 6.6 – Moyenne d'IH et des températures moyenne, maximale et minimale sur la période 01/04 au 30/09 par zone agroclimatique

Classe	IH (°C)	TM (°C)	TX(°C)	TN(°C)
1	2435	18	28,6	10,1
2	2378	17,7	28	9,3
3	2158	17,1	26,5	9,7
4	2128	17,4	25,9	10,1

et 2. Ce sont probablement ces températures maximales qui impliquent une valeur de IH plus faible. Les valeurs moyennes de températures sont quant à elles très proches entre les 4 classes.

Entre la classe la plus précoce, la classe 2, et la classe la plus tardive, la classe 4, il y a donc une différence au niveau du cumul des températures et des températures maximales de l'air. La précocité des individus de la classe 2 peut donc être éventuellement induite par ces températures plus chaudes, favorisant une activité photosynthétique plus importante. A l'inverse les individus de la classe 4 sont plus tardifs à cause de températures plus faibles.

Les températures de l'air subies par les individus ont donc eu un effet sur la précocité de maturation. Ces conclusions recourent des connaissances sur l'effet des températures sur la rapidité de maturation.

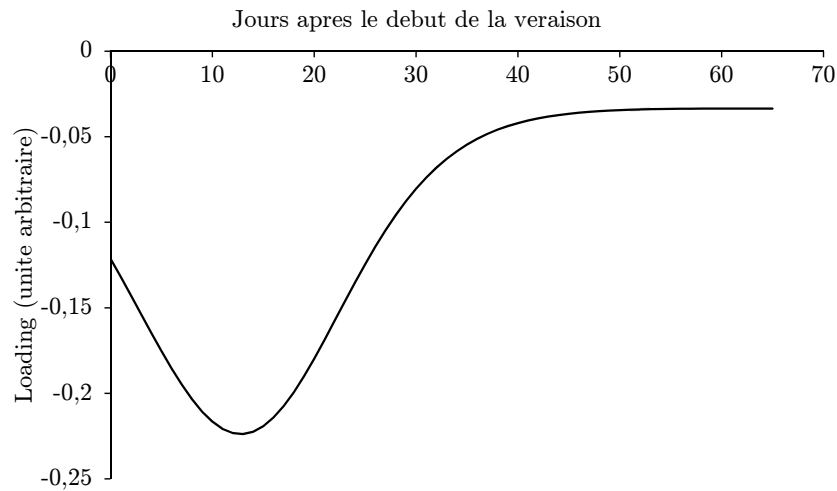
### Résultats de l'approche extensive

Pour cet exemple, les deux premières composantes principales déterminées par l'ACP ont été conservées. Elles permettent d'expliquer 73,57 puis 22,41 % de la variance totale. Le plan défini par les deux premières composantes principales explique donc 95,98 % de la variance totale.

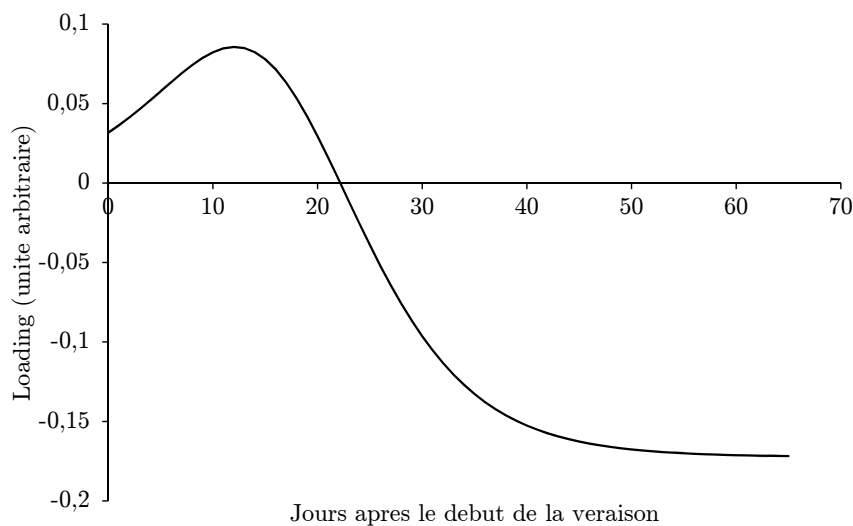
La figure 6.16 donne le loading de chacune des deux premières composantes principales.

Le loading de la première composante principale est toujours négatif. Comme lors de l'étude de la base de données 1, il suit une courbe de forme approximativement gaussienne. Les jours après le 30ème jour après le début de la véraison ont un poids très faible dans la détermination de cette composante. Pour faciliter l'analyse de ce loading, il est donc possible de simplifier cette composante principale en un créneau entre le jour où la véraison commence et le 30ème jour après le début de la véraison. Dans ce créneau, les individus vont se différencier en fonction de l'intégrale de la concentration en sucres dans les baies entre les deux bornes du créneau.

En première approximation, la première composante principale met donc en avant que presque 74% de la variabilité entre les 65 courbes étudiées se situe dans un créneau compris entre le début de la véraison et 30 jours après celui-ci, soit sur moins de la première moitié de la cinétique.



(a) % de variance expliquée : 73.57%



(b) % de variance expliquée : 22.41%

FIGURE 6.16 – Loadings des deux premières composantes principales de l'ACP réalisée sur l'utilisation extensive des modèles ajustés à partir de la base de données 2 : (a) Loading de la première composante principale, (b) Loading de la seconde composante principale

Comme il avait été noté lors de l'analyse de la figure 6.5, la cinétique d'accumulation des sucres peut avoir commencé sa phase d'accumulation linéaire avant la date du début de la véraison observée. C'est pourquoi, contrairement au cas de la base de données 1, les jours juste après le début de la véraison ont un poids important dans la détermination de cette composante principale.

Les individus avec les concentrations les plus élevées entre la véraison et le 30ème jour après le début de la véraison seront projetés dans le sens négatif sur cet axe et inversement.

Comme il a été noté pendant l'étude de la base de données 1, les individus présentant

les intégrales de concentration les plus élevées avant le 30ème jour après le début de la véraison sont les individus précoces et avec une vitesse importante d'accumulation des sucres dans les baies.

Cette composante principale peut donc être décrite comme un axe permettant de différencier les individus à accumulation précoce des individus à accumulation tardive.

Le loading de la seconde composante principale est positif puis négatif. Il est positif jusqu'au 20ème jour après le début de la véraison mais le poids de chaque jour dans la détermination de la composante principale est très faible. Après le 30ème jour après le début de la véraison, chaque jour a un poids négatif important dans la détermination de cette composante principale.

En première approximation, cette composante principale peut donc être décrite comme faisant l'intégrale de la concentration en sucres dans les baies à partir du 30ème jour après le début de la véraison et sur toute la fin de la maturation. Cette partie du loading est négative. Les individus les plus concentrés seront projetés négativement et inversement.

Les individus les plus concentrés après le 30ème jour après le début de la véraison sont ceux qui ont des  $X_{s_{max}}$  élevés et atteignent le plateau de  $X_{s_{max}}$  tôt. Cette composante principale discrimine les individus avec une forte valeur de  $X_{s_{max}}$ , assez précoces et de vitesse assez forte pour atteindre le plateau de  $X_{s_{max}}$  tôt, des individus avec une valeur de  $X_{s_{max}}$  faible ou très tardifs ou lents qui n'atteignent le plateau que très tardivement.

Comme dans le cas de l'étude de la base de données 1, dans cette base de données, les deux loadings présentés en figure 6.16 indiquent qu'il y a nettement plus de variabilité entre les courbes dans la première partie de la cinétique que dans la seconde.

D'une manière générale, les formes des loadings sont très similaires pour les deux bases de données. Les courbes représentant l'accumulation des sucres dans les baies de raisin présentent donc des fenêtres de date de variabilité maximum semblables. Ces résultats sont obtenus sur deux bases de données de cépage, provenance et millésimes différents. Les formes similaires des loadings montrent que, malgré toutes les différences entre les deux bases de données, les fenêtres de jours, après le début de la véraison, où les courbes présentent le plus de variabilité sont communes. L'hypothèse que de tels loadings puissent être observés quelle que soit la base de données étudiée peut donc être émise. Si cette hypothèse est vraie, il y aurait donc systématiquement plus de variabilité entre les courbes sur la période allant de la véraison au point d'inflexion qu'au niveau des plateaux représentant les concentrations maximales dans les baies.

La figure 6.17 présente la position des individus sur le plan constitué par les deux



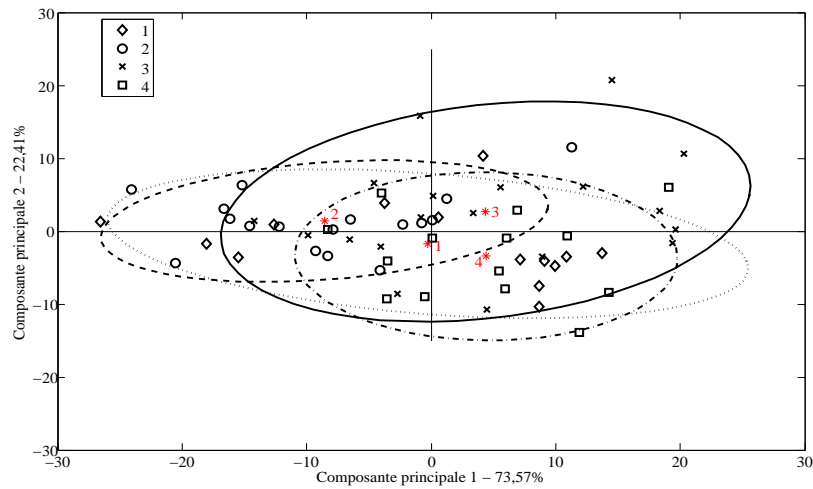


FIGURE 6.17 – Carte factorielle du plan 1-2 de l'ACP en extensif réalisée sur les 65 individus de la base de données 2. La classe 1 est représentée par l'ellipse en pointillés, la classe 2 par l'ellipse en tirets, la classe 3 par l'ellipse en trait continu puis la classe 4 par l'ellipse en pointillés mixtes.

premières composantes principales. Les individus sont représentés en fonction de la classe agro-climatique à laquelle ils appartiennent.

Les ellipses représentatives de chaque classe ont été définies de façon semblable à celles utilisées pour l'approche intensive. Le centre de l'ellipse de chaque classe agro-climatique est le centre de gravité du nuage de points représentant la classe. Il est mis en avant par un point rouge. L'ellipse de chaque classe représente la surface dans laquelle 95% des individus de cette classe devraient être présents si la distribution de ceux-ci était gaussienne.

Les 4 classes sont relativement confondues sur ce plan factoriel. Les individus de chacune des classes n'ont donc pas une position induite par la classe d'où ils proviennent. Cependant, certaines tendances peuvent être observées.

La classe 2 présente des individus projetés principalement sur la gauche du plan factoriel. Les individus de cette classe agroclimatique présentent donc une intégrale de concentration en sucres dans les baies de raisin élevée entre le jour où commence la véraison et le 30ème jour après le début de la véraison. Les individus de la classe 2 sont donc globalement plus concentrés en début de maturation.

La classe 3 est très étendue. Certains individus de cette classe sont situés dans le quart en haut à droite du plan. Ces individus présentent donc (i) une intégrale de la concentration en sucres dans les baies de raisin entre le jour où commence la véraison et le 30ème jour après le début de la véraison faible puis (ii) une intégrale de la concentration en sucres dans les baies de raisin après le 30ème jour après le début de la véraison faible. La classe 3 est donc la classe où les individus peu concentrés en sucres dans les baies sur toute la période de maturation sont présents.

La classe 4 présente des individus projetés principalement dans le quart en bas à gauche de ce plan. L'intégrale de la concentration en sucres dans les baies de raisin

des individus de cette classe est (i) plutôt faible avant le 30ème jour après le début de la véraison puis (ii) plutôt élevée après le 30ème jour après le début de la véraison. Cette classe est donc globalement caractérisée par des individus avec une concentration en sucres dans les baies de raisin faible en début de maturation, mais qui augmente rapidement et est élevée en fin de maturation.

Les caractéristiques des différentes classes recourent les observations réalisées lors de l'analyse des résultats de l'approche intensive. Les individus se regroupent tout de même moins selon leur classe agro-climatique que lors de l'analyse des résultats de l'approche intensive.

Il y a donc une plus grande variabilité des individus au sein d'une même classe agro-climatique au niveau de la forme des cinétiques. L'effet du climat sur la forme réelle suivie par la cinétique d'accumulation des sucres dans les baies de raisin est donc moins important que celui visible sur la valeur des paramètres du modèle défini en équation 6.1.

Grâce à ces deux analyses, un grand nombre de parcelles d'un même millésime a pu être comparé alors qu'aucune comparaison n'aurait pu être réalisée sur la base des séquences temporelles. Il est maintenant possible de rapprocher les parcelles ayant eu un comportement proche. Les comparaisons peuvent être effectuées à partir de séquences temporelles provenant de différents domaines. Une parcelle pourra donc être comparée à un grand nombre de parcelle et celle(s) dont elle aura le comportement le plus proche sera / seront mise(s) en avant. Ces analyses confirment donc le potentiel de la méthodologie proposée pour comparer des individus avec des séquences temporelles hétérogènes.

## 6.4 Conclusion : Une méthodologie générique qui permet de comparer des individus hétérogènes

Le test de la méthode proposée dans cette thèse, effectué sur chacune des deux bases de données, a prouvé que cette méthode permet de comparer des individus initialement incomparables.

Dans le cas de l'étude de la base de données 1, 7 millésimes d'une même parcelle ont été comparés. Pour chaque millésime, des notions de précocité et de concentration en sucres dans les baies en comparaison aux autres millésimes ont pu être mises en avant. De plus, les résultats des modèles ont aussi pu être comparés à une classification agro-climatique des millésimes. Des connaissances expertes de l'effet du climat sur la précocité et la vitesse de maturation ainsi que sur la quantité maximale de sucres accumulée dans les baies de raisin ont été retrouvées. Pour la base de données 1, la méthodologie proposée permet donc de créer de la connaissance locale, entre les millésimes et en fonction du climat, du comportement de la cinétique d'accumulation des sucres dans les baies de

raisin. Cette connaissance locale ne contredit pas la connaissance experte de l'effet du climat sur l'accumulation des sucres dans les baies de raisin.

Dans le cas de l'étude de la base de données 2, 65 parcelles d'un même millésime ont été comparées. Pour chaque parcelle, la ou les parcelles de comportement le plus proche peuvent être définies. Ces proximités peuvent être définies à partir de l'étude des paramètres du modèle (équation 6.1) ajustés aux données de chaque séquence temporelle. Elles peuvent aussi être déduites de l'étude de la cinétique complète de l'accumulation de sucres dans les baies de chaque parcelle. La différenciation des parcelles en classes agro-climatiques a permis de retrouver de la connaissance experte sur l'effet du climat sur la précocité de maturation par exemple. Pour cette base de données, de nouvelles connaissances du contexte local ont été mises en avant. Cette nouvelle connaissance ne contredit pas la connaissance experte de l'effet du climat sur la précocité de la maturation par exemple.

Pour chacune des bases de données, la méthode proposée a donc permis de rendre les individus comparables. Les comparaisons se sont de plus avérées utiles pour la compréhension du phénomène spécifique, pour chaque millésime ou parcelle, d'accumulation des sucres dans les baies de raisin. La méthode proposée est donc générique pour représenter l'accumulation des sucres dans les baies de raisin. Elle permet de plus de créer une nouvelle connaissance à propos de chaque individu, complémentaire des connaissances expertes déjà disponibles.

Que ce soit pour l'approche intensive ou pour l'approche extensive, les composantes principales étaient à peu près déterminées par soit (i) les mêmes paramètres, soit (ii) les mêmes créneaux de dates dans lesquelles la variabilité des courbes était importante. De plus, les pourcentages de variabilité expliqués par chaque composante principale étaient du même ordre de grandeur pour les deux analyses.

L'utilisation de la méthode objet de cette thèse semble donc apporter une nouvelle connaissance spécifique à l'étude de l'accumulation de sucres dans les baies :

- Le paramètre  $X_{s_{max}}$  est assez indépendant des paramètres  $Infl$  et  $Sl$  qui sont quant à eux bien corrélés.
- Il y a moins de variabilité entre individus au niveau du paramètre  $X_{s_{max}}$  qu'au niveau du paramètre  $Infl$
- Les courbes présentent environ 70% de variabilité en début de maturation et sont donc peu variables sur la fin de celle-ci.

Dans le cadre d'une utilisation accrue de cette méthode, à partir des cartes factorielles obtenues lors des analyses descriptives, l'effet de certaines pratiques culturales ou d'un type de sol par exemple sur la cinétique d'accumulation de sucres dans les baies de raisin pourrait être étudiée. Des regroupements comme celui présenté avec les classes agroclimatiques pourraient être réalisés. L'effet d'une pratique culturale sur la concentration maximale en sucres dans les baies de raisin pourrait ainsi en être compris. Ce travail

est un travail de valorisation de la méthode mise en place dans cette thèse qui n'a pas été réalisé. Il devra être entrepris pour la mise en place d'un outil d'aide à la décision utilisant la méthode développée.

Dans cette partie, l'échelle de temps choisie était le calendrier julien. Cependant, l'utilisation du calendrier thermique aurait permis de comparer les individus en prenant en compte la quantité de chaleur que chacun a pu accumuler et non en fonction du nombre jours passés depuis la véraison. Ainsi, les individus pourraient être comparés sans artefacts dus à des différences du climat entre eux.

Le calendrier thermique n'est pas encore répandu chez les professionnels de la filière viti-vinicole. C'est pourquoi, les cinétiques ont été analysées en temps julien mais le travail sur un temps thermique reste une perspective de développement qui sera étudiée prochainement. Les cinétiques pourront ainsi être comparées en travaillant sur la vitesse réelle d'accumulation des sucres dans les baies de raisin.

# Chapitre 7

## Conclusion et perspectives

---

### Sommaire

<b>7.1</b>	<b>Bilan des travaux réalisés . . . . .</b>	<b>118</b>
7.1.1	Une méthode de modélisation paramétrique de séquences temporelles pour permettre les comparaisons . . . . .	118
7.1.2	Trois applications différentes d’une même méthode . . . . .	119
7.1.3	Atteinte de l’objectif de la thèse et généralité de la méthode . . . . .	121
<b>7.2</b>	<b>Perspectives de développement du travail proposé . . . . .</b>	<b>122</b>
7.2.1	Elargissement du nombre de grandeurs modélisées, du nombre de paramètres calculés . . . . .	122
7.2.2	Analyse multivariée des individus . . . . .	124
7.2.3	Modification des modèles pour représenter au mieux les cinétiques réelles des différentes grandeurs . . . . .	125
7.2.4	Utilisation d’autres méthodes d’analyses pour comparer les individus . . . . .	126
7.2.5	Calcul de distance à partir des positions des individus sur les plans factoriels des analyses de données . . . . .	127
7.2.6	Mise en place d’un système à base de cas . . . . .	127
7.2.7	Évaluation de la qualité d’ajustement des modèles en fonction du type de séquence temporelle . . . . .	127
7.2.8	Simulation de la fin d’une cinétique . . . . .	129
<b>7.3</b>	<b>Utilisation du travail de recherche réalisé pour aider au pilotage de la vigne . . . . .</b>	<b>130</b>

---

## 7.1 Bilan des travaux réalisés

### 7.1.1 Une méthode de modélisation paramétrique de séquences temporelles pour permettre les comparaisons

L'objectif de cette thèse est de valoriser les bases de données de traçabilité dans une optique d'aide à la décision. Il a été choisi d'utiliser les données de bases de données de traçabilité pour proposer un formalisme permettant de réaliser des analogies entre parcelles et millésimes.

La première étape d'un raisonnement par analogies est la comparaison de couples parcelle  $\times$  millésime. De ce fait, l'objectif de ce travail de thèse est d'utiliser les données de traçabilité pour comparer ces couples.

Une première constatation a été que les données brutes des bases de données de traçabilité sont asynchrones. Une seconde constatation a été que le nombre d'évaluations d'une grandeur varie d'une parcelle à l'autre et d'un millésime à l'autre. Les ensembles de données caractérisant chaque couple parcelle  $\times$  millésime sont donc dans des espaces de dimension différente.

L'objectif opérationnel du travail de thèse est de proposer un formalisme permettant de retraiter les données de traçabilité pour que tous les couples parcelle  $\times$  millésime soient représentés dans un espace commun afin de les rendre comparables.

Pour représenter les couples parcelle  $\times$  millésime dans un espace commun, une méthode de modélisation paramétrique de l'évolution de chaque grandeur pour chaque couple parcelle  $\times$  millésime a été proposée.

Une nomenclature des différents éléments présents dans les bases de données de traçabilité a d'abord été proposée.

La connaissance experte en viticulture permet de connaître (i) l'évolution de chaque grandeur ou (ii) l'effet de chaque grandeur sur la plante. A partir de cette connaissance, des modèles paramétriques ont été définis pour représenter ces évolutions ou évaluer ces effets.

Pour chaque individu, la séquence temporelle relative à la grandeur modélisée a permis d'ajuster le ou les paramètres du modèle. A partir de ces paramètres ajustés, une estimation de la valeur de la grandeur dans toutes les fenêtres temporelles de la discrétisation temporelle choisie pour l'étude de la grandeur a aussi pu être calculée.

Pour chaque individu, deux représentations de l'évolution de chaque grandeur sont donc disponibles : (i) un vecteur constitué des paramètres du modèle ajusté ou (ii) un vecteur constitué de toutes les estimations de la grandeur dans toutes les fenêtres de la discrétisation de la base temporelle. Ces représentations constituent les espaces communs entre individus permettant de les rendre comparables.

Trois applications ont été proposées pour évaluer les potentialités de la méthode mise en place : (i) la première traite de la comparaison du comportement climatique de

millésimes pour différents cépages, (ii) la seconde présente l'étude des potentialités d'un modèle pour comparer les cinétiques d'augmentation du pH et (iii) la troisième permet d'évaluer les potentialités d'un modèle pour comparer les cinétiques d'accumulation des sucres dans les baies et présente les premiers résultats de la comparaison d'individus.

### 7.1.2 Trois applications différentes d'une même méthode

Pour les trois applications choisies, une ou plusieurs grandeurs de mesure ont été modélisées.

Dans tous les cas, l'approche méthodologique proposée a permis de comparer des individus auparavant non comparables car les séquences temporelles étaient de dimensions différentes et / ou asynchrones.

#### Comparaison du comportement climatique de couples cépage $\times$ millésime

La première application étudiée était la comparaison du comportement climatique de différents millésimes et pour différents cépages entre la floraison et la véraison. L'étude a porté sur 7 millésimes et 11 cépages. Partant de mesures effectuées par une station météorologique automatique de l'INRA, située dans le vignoble de la station expérimentale INRA Pech Rouge, un ensemble de grandeurs de mesure climatiques a été retenu pour leur importance sur la vigne. Ces grandeurs ont été modélisées grâce à des modèles très simples : moyenne, somme, écart-type ou compteurs d'un nombre de jours ayant dépassé un seuil sur toute la base temporelle choisie. Un grand nombre de paramètres a donc été évalué pour chaque individu.

La floraison et la véraison ont observées à des dates différentes pour chaque individu. Les dates de la base temporelle sur laquelle les modèles ont été ajustés sont donc spécifiques à chaque individu. Les valeurs des paramètres représentent le climat subi par chaque individu et sont différentes d'un individu à l'autre même au sein d'un millésime.

Les paramètres ajustés pour chaque individu ont été analysés grâce à une analyse en composantes principales. Cette analyse a permis la comparaison des individus.

Le climat global de chaque millésime a pu être comparé à celui des autres millésimes, entre la floraison et la véraison.

Pour un cépage particulier, l'effet réel du climat du millésime pouvait être comparé à celui des autres millésimes.

Au sein d'un millésime, les cépages ont pu être comparés entre eux et l'effet du climat sur les cépages précoces ou tardifs a pu être différencié. Cette étude a donc été plus approfondie qu'une analyse qui aurait juste pris en compte le climat global du millésime.

Grâce à cette analyse en composantes principales, avec 4 composantes principales, 83% de la variabilité entre les 7 millésimes et les 11 cépages a été représentée. L'extraction de la position de chaque individu sur ces 4 composantes principales permet de représenter, dans un espace de dimension 4, 83% de la variabilité entre les individus.

Une analyse de l'intervalle entre deux stades phénologiques peut être envisagée quels que soient les stades. Ainsi, sur différentes périodes des analogies entre millésimes et cépages peuvent être proposées. Le climat subi par la vigne transforme son potentiel et modifie les interventions des viticulteurs. Pour un individu en cours d'étude, ces analogies sur le climat ont un fort potentiel d'aide à la décision. A partir de la connaissance des individus passés les plus proches, de leur itinéraire technique, de la composition du raisin vendangé et du vin obtenu, des décisions stratégiques peuvent être prises pour soit obtenir le même type de vin / raisin, soit obtenir un type de vin / raisin différent.

### **Modélisation de l'augmentation de pH et comparaison d'individus à partir des courbes modélisées**

L'augmentation du pH dans les baies de raisin a été étudiée pour 53 parcelles de Cabernet-Sauvignon de la Napa Valley en Californie, uniquement en 2011.

L'augmentation du pH dans les baies de raisin a été modélisée grâce à une sigmoïde adaptée de la fonction logistique. Trois paramètres ont été ajustés pour chaque individu. Ces trois paramètres permettent de reconstituer toute la cinétique d'augmentation du pH depuis les premières mesures jusqu'à la maturité du raisin.

Une approche de comparaison de cinétiques sur une base graphique a ensuite été proposée. La comparaison des courbes a permis de comparer les cinétiques d'augmentation du pH dans les baies de raisin de différents individus dans leur ensemble.

Auparavant, le pH dans les baies de différents individus pouvait être comparé à la vendange et / ou lors de mesures synchrones. L'approche méthodologique proposée permet donc de comparer les cinétiques à tout moment entre les premières mesures et la vendange.

Les différences de comportement des cinétiques restent encore mal expliquées. Des informations complémentaires sur le climat ou le régime hydrique subis par les vignes, par exemple, semblent nécessaires. Ces informations, associées à la comparaison de nombreux individus, pourraient permettre de proposer des règles de décision utiles pour le pilotage de la plante. Ces questions pourraient faire l'objet de futures recherches.

### **Modélisation de l'accumulation des sucres dans les baies de raisin et étude du potentiel des deux modes d'utilisation des modèles pour comparer les individus**

Deux bases de données ont servi pour cette application. La première était constituée de 7 individus : suivi de l'accumulation de sucres dans les baies de raisin pour une parcelle de Mourvèdre de la station expérimentale INRA Pech Rouge entre 2003 et 2009. La seconde était constituée de 65 individus : suivi de l'accumulation de sucres dans les baies de raisin pour 65 parcelles de Cabernet-Sauvignon de la Napa Valley sur le millésime 2011.

L'accumulation de sucres dans les baies de raisin a été modélisée grâce à une sigmoïde adaptée de la fonction logistique. Trois paramètres ont été ajustés pour chaque individu.



Deux types de représentations de chaque individu, issues des deux types d'utilisation possibles pour un modèle, ont ensuite été analysés : (i) le vecteur constitué des trois paramètres ajustés pour chaque individu et (ii) le vecteur constitué des estimations journalières de la concentration en sucres sur toute la base temporelle d'étude.

Une analyse en composante principale a été réalisée sur les deux types de vecteurs cités ci-dessus, pour chacune des deux bases de données d'étude.

L'approche méthodologique proposée a permis de comparer les individus de chacune des bases de données et de définir l'individu avec l'accumulation de sucres la plus précoce ou avec l'accumulation la plus rapide dans la phase linéaire d'accumulation. L'analyse des individus à partir de l'estimation journalière de la concentration en sucres dans les baies de raisins sur toute la base temporelle étudiée a permis de comparer les individus en étudiant l'ensemble de leur cinétique d'accumulation.

Les résultats de l'ACP sur l'utilisation en extension des modèles ont mis en avant des fenêtres de dates, à l'intérieur de la base temporelle, dans lesquelles les cinétiques étaient les plus différentes. Ces fenêtres étaient semblables pour les deux bases de données étudiées. Les cinétiques des différents individus ont été comparées en évaluant l'intégrale de la concentration en sucres dans les baies de raisin sur ces fenêtres.

Sur les plans factoriels des ACP, les individus les plus proches ont pu être identifiés. Ces plans peuvent donc servir de base à un raisonnement par analogies.

Afin de pouvoir appuyer les comparaisons et analogies entre individus, une classification agro-climatique des individus a été utilisée. Trois indices ont été calculés pour chacun des individus.

Des connaissances expertes de l'effet du climat sur la précocité, la rapidité et l'intensité de l'accumulation des sucres dans les baies de raisin ont été retrouvées pour les deux bases de données.

Cette connaissance du type de climat intervenu sur chaque individu a donc permis de comprendre une partie des différences entre les cinétiques d'accumulation de sucres dans les baies de raisin des différents individus. Cette relation pourrait être, à terme, utilisée pour un meilleur pilotage de la culture.

### **7.1.3 Atteinte de l'objectif de la thèse et généralité de la méthode**

La méthode proposée dans le chapitre 3 a été validée pour comparer des individus en travaillant avec différents types de grandeur de mesure. Les modèles ont permis de représenter la cinétique de toutes sortes de grandeurs pour des individus de provenance, cépage, pratiques culturales très différents, avec une bonne qualité d'ajustement des modèles. Les modèles choisis permettent donc de représenter la cinétique de chaque grandeur quel que soit le type de séquence temporelle initiale.

La méthode proposée pourrait s'appliquer à tout type de données asynchrones issues du suivi de l'évolution d'un système agricole ou industriel à condition qu'un modèle

puisse être déterminé pour la grandeur analysée. Pour ce faire, il faudrait se référer à l'expertise de ces systèmes pour définir les modèles.

Les comparaisons ont permis d'acquérir une nouvelle connaissance du contexte local. Ces comparaisons se sont avérées utiles quelle que soit la provenance des données.

L'approche proposée permet donc de comparer des individus pour toute grandeur de mesure quelle que soit la provenance de la base de données de traçabilité.

L'objectif de ce travail de thèse, de proposer une méthode de traitement des informations des bases de données de traçabilité dans un but de comparaison d'individus a donc été atteint. La méthode nécessite néanmoins qu'un modèle de forme fixe puisse être défini pour la grandeur analysée.

## 7.2 Perspectives de développement du travail proposé

### 7.2.1 Elargissement du nombre de grandeurs modélisées, du nombre de paramètres calculés

Dans l'application 1, un grand nombre de grandeurs climatiques ont été modélisées. Les modèles choisis pour comparer le comportement climatique de différents individus entre floraison et véraison peuvent être appliqués à tout type de fenêtre de temps. L'étude du climat à toute période du cycle biologique de la vigne est donc possible. Dans ce cas-ci, il n'est pas nécessaire de produire d'autres modèles, mais l'étude des paramètres à d'autres périodes pourrait faire correspondre d'autres individus entre eux. De telles analogies à tout moment du cycle biologique peuvent ensuite être utilisées pour un meilleur pilotage de la plante.

Dans les applications 2 et 3, la fonction logistique a été adaptée à l'étude de l'augmentation du pH et de l'accumulation de sucres dans les baies de raisin. Ces deux grandeurs de mesure ne représentent qu'un échantillon de toutes les grandeurs de mesures qui peuvent être évaluées sur la plante ou dans les baies.

Des modèles représentant la cinétique d'autres grandeurs doivent donc maintenant être proposés. Ces modèles permettront de comparer les individus en travaillant sur le comportement de la plante et de la baie dans son ensemble.

Deux types de modèles peuvent être envisagés :

1. Les modèles qui seront adaptés de la fonction logistique comme pour les deux grandeurs présentées
2. Les modèles pour lesquels la fonction à choisir pour représenter la cinétique de la grandeur est différente de la fonction logistique

Dans le premier cas, des cinétiques telles que l'augmentation du poids des baies ou celle de la teneur en sucres dans les baies ( $\text{mg.baie}^{-1}$ ), l'accumulation des anthocyanes

ou du cation potassium dans les baies de raisin, de la véraison à la récolte, pourront être modélisées en adaptant la fonction logistique selon la même logique que celle présentée dans les applications 2 et 3.

Certains auteurs ont utilisé la fonction logistique pour modéliser la cinétique de  $\frac{1}{AT}$  [75]. Ce postulat devra être testé pour permettre une représentation de la cinétique de dégradation des acides dans les baies de raisin.

Les courbes de croissance de chaque rameau et de chaque feuille ont la forme d'une sigmoïde [19]. La surface foliaire de la vigne est une combinaison du nombre et de la taille des rameaux, qui conditionne le nombre de feuilles, et de la taille de chaque feuille. Il est donc envisageable que la cinétique de la surface foliaire de chaque plante, à condition qu'aucune action visant à réduire cette surface n'ait été effectuée, suive une sigmoïde sur tout le cycle végétatif. Elle pourrait alors être modélisée grâce à la fonction logistique. La mise en place d'un tel modèle permettrait une comparaison de la surface foliaire de différents individus à tout moment du cycle végétatif. Plus la surface foliaire d'une plante est importante, plus les produits de la photosynthèse seront synthétisés en grand nombre. La comparaison des individus au regard de la surface foliaire pourrait donc permettre de faire rapidement des prédictions quand au potentiel qualitatif de chaque vigne.

La fonction logistique devrait donc permettre de représenter la cinétique d'un grand nombre de grandeurs évaluées sur la plante ou dans les baies. Les individus pourront donc être comparés, comme présenté dans ce travail de thèse, sur la base d'un ensemble de vecteurs de paramètres ou d'estimations représentatifs de l'activité de la vigne et du raisin.

Dans le second cas, la fonction logistique ne pourra pas être utilisée pour modéliser la cinétique de ces grandeurs. Des grandeurs comme l'état hydrique de la plante ou l'évolution de la concentration en acide tartrique ou en tanins dans les baies sont concernées.

Pour modéliser la cinétique de l'acide tartrique ou des tanins dans les baies de raisin, des connaissances expertes devront à nouveau être mobilisées pour définir les fonctions les plus à même de représenter leur évolution au cours de la maturation.

L'état hydrique de la vigne, comme cela a été rappelé dans l'application 1, est évalué par le bilan entre l'eau reçue par la plante par précipitations ou irrigation et l'eau perdue par transpiration de la plante, drainage ou ruissellement au niveau des sols. Cette grandeur a donc une cinétique aussi aléatoire que celle des grandeurs climatiques entrant dans le bilan permettant de calculer sa valeur. Il n'est donc pas possible d'utiliser une fonction pour représenter cette cinétique. Des connaissances expertes existent sur la gravité de la contrainte hydrique sur la vigne à tout moment du cycle végétatif [70]. Ces connaissances pourront permettre de comparer les individus en fonction de cette gravité et ainsi faciliter le pilotage de la culture.

Dans un même ordre d'idées, dans le cas, par exemple, d'une action visant à diminuer la quantité de feuillage de la vigne comme un effeuillage ou un rognage, la cinétique de la surface foliaire de la plante ne suivrait alors plus une fonction logistique. Plus

particulièrement, de telles actions rendraient l'utilisation d'une fonction impossible pour appliquer la méthode objet de cette thèse. Il faudrait alors envisager de travailler avec soit (i) des modèles non paramétriques, soit (ii) sur des bases temporelles représentant l'intervalle entre deux de ces actions pour pouvoir revenir à l'utilisation d'une fonction dérivable telle que la fonction logistique, soit (iii) de se référer à des modèles de type moyenne et écart-type comme ceux mis en place pour l'étude du climat présentée dans le chapitre 4. Une telle problématique n'a par contre pas été envisagée ni étudiée durant cette thèse et mérite un développement futur.

La modélisation d'autres grandeurs est une perspective de développement de la méthode proposée de longue haleine. A terme, l'objectif est de pouvoir comparer des individus sur la base de tout type de grandeur mesurée sur la plante, la baie ou l'environnement.

Les comparaisons se feront donc sur le comportement global de la baie. Elles permettront de faire des raisonnements par analogie prenant en compte l'évolution de la plante et de la baie en entier à tout moment.

### 7.2.2 Analyse multivariée des individus

Dans les applications 2 et 3, les paramètres ajustés pour chaque individu ont été utilisés pour comparer les individus sur la base de la cinétique d'une seule grandeur.

L'augmentation du pH et l'accumulation des sucres dans les baies de raisin sont modélisées à l'aide de la même fonction. Les paramètres ajustés sont donc en nombre identique et représentent des points clefs de la courbe semblables : valeur maximale, abscisse du point d'inflexion et pente de la courbe au point d'inflexion.

Pour ces deux grandeurs, le paramètre *Infl* représente la moitié de la maturation. Pour chaque individu, les valeurs d'abscisses du point d'inflexion pourraient être utilisées pour définir une vitesse globale comparée de la maturation de chaque grandeur. Une telle connaissance pourrait être utilisée pour estimer l'ordre d'atteinte de la valeur maximale pour chaque individu à l'issue de la maturation.

La date de maturité du raisin est définie par une combinaison, la plus optimale possible, de la maturité de toutes les grandeurs mesurées dans le raisin. La comparaison des paramètres *Infl* pourrait donc, à terme, permettre de définir une date de maturité estimée.

Plus que cette comparaison entre les abscisses des points d'inflexion, la comparaison entre individus pourrait bénéficier d'une analyse de données en combinant les deux grandeurs modélisées. Pour chaque individu, le paramètre le plus limitant entre les deux grandeurs serait identifié. Les individus se regrouperaient alors en fonction de leur ressemblance sur la base de l'accumulation des sucres dans les baies et de l'augmentation du pH.

Après la mise en place de modèles pour représenter la cinétique d'autres grandeurs

représentant des composés évalués dans les baies de raisin, une comparaison multivariée pourrait alors être réalisée. Une telle comparaison pourrait permettre de faire des analogies entre individus sur la base de la composition chimique de la baie et des cinétiques modélisées.

### 7.2.3 Modification des modèles pour représenter au mieux les cinétiques réelles des différentes grandeurs

Les modèles permettant de représenter l'augmentation du pH ou l'accumulation de sucres dans les baies de raisin sont définis à partir de la même fonction. Les équations proposées pour modéliser la cinétique de ces grandeurs sont semblables. De tels modèles pourraient être définis pour d'autres grandeurs mesurées sur la plante ou dans les baies.

Les caractéristiques de ces équations sont les suivantes :

- 3 paramètres :  $X_{max}$  la valeur maximale de la grandeur,  $Infl$  l'abscisse du point d'inflexion de la courbe et  $\tau$  un paramètre permettant d'ajuster la valeur de la pente  $Sl$  de la courbe au point d'inflexion
- une valeur d'initialisation correspondant à la concentration ou au niveau de la grandeur mesurée au moment où la véraison commence

La valeur d'initialisation permet de forcer une valeur basse du modèle si la séquence temporelle à laquelle le modèle est ajusté ne possède pas de descripteurs juste avant ou au moment de la véraison.

Pour les séquences temporelles couvrant toute la maturation, avec des mesures ayant commencé avant la véraison, cette valeur d'initialisation n'est pas nécessaire.

Dans l'application 3, la figure 6.3 montre que les modèles ajustés surestiment les premières valeurs de °Br. La valeur d'initialisation choisie pour cette grandeur était de 5,5°Br. Cette valeur était supérieure aux mesures contenues dans les deux séquences temporelles modélisées (<5°Br). C'est donc à cause de cette valeur d'initialisation que les premières mesures de la séquence temporelle étaient surestimées. Cette surestimation de la valeur d'initialisation fait augmenter la RMSE d'ajustement du modèle. Elle implique donc, artificiellement, une augmentation du pourcentage d'erreur d'ajustement du modèle à la séquence temporelle.

Lors d'un ajustement du modèle à une séquence temporelle dont les mesures ont commencé avant la véraison, la valeur de la grandeur à véraison pourrait être considérée comme un quatrième paramètre. Ce paramètre serait alors ajusté au même titre que les 3 autres paramètres. La valeur réelle de la grandeur pour l'individu considéré au moment où la véraison commence serait ainsi calculée. Ce paramètre pourrait ensuite servir pour les analyses de données.

La composition des baies à véraison ne serait donc plus une connaissance issue de l'expertise. Les individus pourraient alors être comparés par rapport à leur composition à véraison tout comme à tout autre date de la maturation.

L'ajustement d'un quatrième paramètre lorsque cela est possible va être développé

très prochainement. Deux types de modèles seront donc disponibles.

Le modèle qui sera ajusté pour chaque individu sera choisi en fonction du type de séquence temporelle à laquelle le modèle sera ajusté, et notamment de la présence de descripteurs avant la véraison ou non.

En fonction des séquences temporelles, 3 ou 4 paramètres seront ajustés. Il faudra donc comparer les individus séparément en fonction du type de séquence temporelle permettant leur ajustement.

Ce choix du nombre de paramètres ajustés par modèle en fonction du type de séquence temporelle disponible constitue une perspective d'amélioration de la méthode mise en place, notamment au niveau de la définition des modèles paramétriques.

## 7.2.4 Utilisation d'autres méthodes d'analyses pour comparer les individus

Dans les applications présentées, les données issues de la modélisation ont été analysées par analyse en composante principale. Une première évolution pourrait constituer dans le fait de réaliser ces ACP en prenant en compte différentes grandeurs. Ainsi, les analogies entre individus pourraient être réalisées en prenant en compte la globalité de chaque plante ou baie.

D'autres analyses pourraient être proposées. Parmi celles-ci, des analyses en correspondances multiples (ACM) ou des analyses factorielles multiples (AFM) peuvent être envisagées [34].

L'analyse en correspondances multiples permet de travailler sur des classes de valeurs. Chaque paramètre peut être divisé en classes. Pour chaque individu, la classe à laquelle appartient le paramètre est alors définie. Les grandeurs se regroupent ensuite en fonction du nombre de fois où deux classes apparaissent en même temps.

Une analyse en correspondances multiples pourrait être envisagée pour intégrer aux analyses la notion de gravité de la contrainte hydrique de chaque individu. De plus, les corrélations entre grandeurs ne sont pas forcément linéaires. Des effets non linéaires d'une grandeur sur une autre pourraient être alors mis en évidence. Une analyse conjointe des grandeurs climatiques et des grandeurs agronomiques pourrait être envisagée dans ce cadre. Grâce à de telles analyses, des groupes d'individus de comportement semblable pourraient être identifiés. Ces analogies pourraient ensuite être utilisées par les viticulteurs comme une information complémentaire pour leurs décisions.

L'analyse factorielle multiple est une analyse adaptée au traitement de matrices dans lesquelles un ensemble d'individus est décrit par plusieurs groupes de variables. Ce type d'analyse est une alternative à l'ACP classique. Une sous-matrice constituée avec chaque groupe de variable est analysée par ACP séparément puis les résultats sont combinés.

Cette analyse pourrait être utilisée pour analyser conjointement les grandeurs climatiques et les grandeurs agronomiques. Une ACP serait réalisée pour chacun des deux styles de grandeurs puis les résultats seraient combinés. Des liens entre grandeurs climatiques et grandeurs agronomiques pourraient alors être mis en évidence.

### **7.2.5 Calcul de distance à partir des positions des individus sur les plans factoriels des analyses de données**

Après chaque analyse de données, la position de chaque individu sur les différentes composantes principales peut être extraite.

A partir de ces positions, une distance entre deux individus peut être calculée. Les individus analogues peuvent ensuite être définis comme ayant une distance faible.

Deux individus proches sur un plan factoriel peuvent être éloignés sur un autre plan. Un calcul de distance pourrait permettre de définir les individus analogues, toutes composantes principales confondues.

De tels calculs de distances permettraient de faire des analogies de façon objective. Dans le cas d'une analyse avec un très grand nombre d'individus, qui se superposeraient sur les plans factoriels, un calcul de distance rendrait les analogies plus faciles.

### **7.2.6 Mise en place d'un système à base de cas**

La réalisation d'analogies entre individus pourrait être automatisable. Une technique d'analyse, le raisonnement à base de cas, permet de réaliser un raisonnement par analogies [62].

Pour un individu que l'on veut rapprocher d'un autre, un système à base de cas effectue une recherche dans la base de connaissance et renvoie un cas similaire à l'individu en cours d'analyse. Puis, à partir du cas similaire, un conseil de gestion de l'individu en cours d'analyse peut être fourni pour, soit suivre la même trajectoire que le cas similaire, soit suivre une trajectoire différente.

Mettre en place un tel outil permettra de répondre pleinement à la problématique pratique initiale de cette thèse. Mettre en place un système à base de cas en complément du travail déjà réalisé constituera un travail de recherche futur.

Un outil d'aide à la décision pourrait être extrait de ce système à base de cas. Il fournirait des préconisations d'actions sur la vigne pour obtenir au mieux possible un raisin de composition voulue.

### **7.2.7 Evaluation de la qualité d'ajustement des modèles en fonction du type de séquence temporelle**

Dans le cas de la modélisation paramétrique de grandeurs à partir de fonctions telles que la fonction logistique, la qualité de l'ajustement du modèle en fonction du type de séquence temporelle peut-être questionnée. Ce cas correspond aux suivis réalisés sur le

terrain. Lors de ces suivis, comme les applications 2 et 3 le montrent, en général moins de 20 mesures permettent de suivre la cinétique de la grandeur sur toute la base temporelle d'étude.

Mathématiquement, le nombre de descripteurs nécessaires par séquence temporelle pour permettre l'ajustement d'un modèle correspond au nombre de paramètres du modèle. Dans le cadre de l'utilisation de la fonction logistique, comme dans dans les applications 2 et 3, cela indique que 3 mesures au cours de la maturation pourraient être suffisantes pour ajuster le modèle. Si elles sont suffisantes, quelle est la qualité et la représentativité du modèle ainsi ajusté et à partir de combien de données un modèle ajusté pourrait être considéré comme présentant peu d'erreur? Quelle est la meilleure répartition des mesures pour un bon ajustement des modèles? Afin de répondre à ces questions, l'évaluation de la qualité d'ajustement des modèles en fonction du type de séquence temporelle va être étudiée.

Dans le cas d'un modèle adapté de la fonction logistique, si, par exemple, les mesures servant à ajuster la sigmoïde ne couvrent pas la fin de la base temporelle, aucun descripteur ne permettra l'ajustement de la valeur de  $X_{max}$  du modèle. Le modèle sera donc ajusté avec un plateau de  $X_{max}$  qui risque d'être sous-estimé.

Ce cas n'est qu'un cas particulier des possibles erreurs d'ajustement du modèle en fonction de chaque type de séquence temporelle.

Rappelons aussi qu'afin d'éviter une erreur d'ajustement du modèle dans les cas où la séquence temporelle ne couvrirait pas la période proche de la véraison, un point d'initialisation expert a systématiquement été inséré aux modèles ajustés.

L'effet de tout type de séquence temporelle sur la qualité d'ajustement du modèle, quels que soient ses dimensions et intervalles de temps entre deux mesures doit donc être évalué.

Afin de tester l'erreur d'ajustement du modèle en fonction du type de séquence temporelle, un test de sensibilité devra être envisagé. Ce test visera à ajuster un modèle à partir de différentes séquences temporelles représentant la même cinétique. A partir d'une cinétique d'équation connue, différentes séquences temporelles devront être générées. Le modèle sera ajusté avec ces séquences temporelles. L'écart entre chaque paramètre ajusté et la valeur réelle du paramètre de la cinétique initiale sera alors estimé. A partir de ces écarts, la sensibilité de l'ajustement de chaque paramètre au type de séquence temporelle permettant l'ajustement pourrait être estimée.

Le type de séquence temporelle permettant l'ajustement le plus précis possible de chacun des paramètres pourrait ainsi être défini. Un indice de qualité globale de l'ajustement de chaque paramètre et du modèle serait ainsi donné pour chaque séquence temporelle. Cet indice pourrait être utilisé pour donner une qualité, *a priori*, d'ajustement de chaque paramètre en fonction des caractéristiques de la séquence temporelle. Lors de l'analyse des résultats, cet indice permettrait de prendre en compte une imprécision de chaque paramètre sur la position des individus en fonction de la qualité d'ajustement de chacun.



### **Définition de dates de mesures optimales à partir du test de sensibilité d'ajustement du modèle en fonction de la séquence temporelle utilisée**

Dans un objectif de pilotage, une telle approche pourrait être utilisée pour définir les dates auxquelles les mesures d'une grandeur doivent être réalisées.

La séquence temporelle minimale à partir de laquelle l'ajustement du modèle est toujours de bonne qualité pourrait être définie. De plus, la position approximative des descripteurs au sein de la séquence temporelle pour minimiser les erreurs d'ajustement des paramètres et la RMSE pourrait être fournie.

A partir de ces connaissances, les dates optimales de mesures, dans un but d'utilisation des modèles proposés, pourraient être proposées. Ces dates seraient proposées sous la forme d'une fenêtre de dates possibles pour la mesure. La position de la mesure en fonction des stades de développement clefs pourrait aussi être définie. A titre d'exemple, il pourrait entre autres être conseillé de réaliser systématiquement une mesure à proximité de la date de véraison.

Un tel conseil sur le positionnement des dates d'échantillonnage permettrait de réduire le nombre de mesures effectuées et de les répartir aux moments cruciaux de l'évolution de chaque grandeur.

### **7.2.8 Simulation de la fin d'une cinétique**

Dans le cas notamment de l'étude de la maturation ou de l'évolution de la surface foliaire, il faut simuler la fin de la cinétique pour comparer un individu dès les premières mesures avec les individus passés. L'ajustement des modèles sur une séquence temporelle incomplète conduirait à des valeurs des paramètres ajustés incertaines.

Une approche pour simuler la fin d'une courbe est d'utiliser les données passées contenues dans la base de données de traçabilité. Ces données permettent de définir des intervalles dans lesquels chaque paramètre peut se situer.

Un des paramètres serait alors choisi et serait fixé pour la simulation. Sa valeur serait définie dans l'équation de la courbe et ne serait donc plus ajustée.

Plusieurs valeurs de ce paramètre seraient alors testées entre les bornes de l'intervalle défini grâce aux données passées. Une valeur des deux autres paramètres serait alors ajustée. Si ces valeurs sont entre les deux bornes de l'intervalle défini pour chacun des paramètres, alors le triplet de paramètres serait défini comme possible pour représenter la cinétique en cours de simulation. Un ensemble de triplets possibles serait ainsi défini.

Ces triplets de paramètres pourraient alors être comparés aux individus de la base de données de traçabilité. Des analogies pourraient ainsi être réalisées dès les premières mesures de chaque individu.

### 7.3 Utilisation du travail de recherche réalisé pour aider au pilotage de la vigne

Comme cela a été présenté dans les perspectives d'utilisation des trois applications de cette thèse, une étude de l'effet de différents paramètres agronomiques sur la vigne pourrait être réalisée à partir des cartes factorielles réalisées après analyse descriptive. Pour l'ensemble d'individus représenté sur la carte factorielle, si un paramètre agronomique a un effet sur un des paramètres des modèles, cela serait mis en avant. De nouvelles connaissances propres au contexte local seraient ainsi définies. Ces connaissances pourraient ensuite amener à la mise en place de règles de décision propres à chaque base de données en cours d'étude. Un tel travail sera à effectuer au cas par cas et est une perspective d'utilisation de la méthode objet de cette dans un cadre de production.

La modélisation paramétrique des grandeurs de mesure évaluées sur la vigne, le raisin, l'environnement de la plante et les comparaisons entre individus permettent de caractériser de façon objective chaque individu en fonction des autres individus renseignés dans la base de données de traçabilité. Une meilleure connaissance objective de chaque individu en découle.

Des analogies entre individus peuvent être réalisées. Les raisonnements par analogies sont habituellement utilisés en agriculture et *a fortiori* en viticulture pour prendre des décisions. Pour un individu en cours de production, en fonction de la connaissance (i) du produit fini souhaité et (ii) de l'itinéraire technique et du produit fini obtenu pour les cas analogues déjà résolus, des décisions stratégiques de gestion de la vigne, du raisin ou de l'environnement de la plante peuvent être prises par le viticulteur. Permettant des analogies entre parcelles et millésimes, le travail de recherche réalisé facilite cette prise de décision. Il constitue donc une aide potentielle à la décision objective et utile.

Le travail réalisé pourrait donc être utilisé pour piloter la culture et faciliter l'obtention d'un produit fini fixé. Une production de qualité supérieure à la production actuelle pourrait ainsi résulter de l'utilisation de ce travail.

Enfin, les analogies sont faites en se basant sur des connaissances acquises par l'analyse de données représentant le contexte local. De nouvelles connaissances pour chaque base de données pourraient donc être produites. Ces connaissances seraient spécifiques au contexte local. Elles pourraient à terme permettre de mettre en place des règles de décision propres à chaque domaine ou ensemble de domaines constituant une base de données de traçabilité.

# Table des figures

1.1	Prise de décision par analogies en agriculture . . . . .	9
2.1	Cycle biologique annuel de la vigne [79] . . . . .	14
2.2	Dates de débourrement et de floraison et nombre de mesures de la température journalière pour une même parcelle et 2 millésimes différents . . .	16
2.3	Mesures ponctuelles de la concentration en sucres dans les baies de raisin pour une même parcelle et 3 millésimes différents . . . . .	18
2.4	Cinétique du volume des baies de raisin de la nouaison à la maturité et évolution globale des autres grandeurs mesurées dans les baies . . . . .	20
2.5	Collaboration entre données de traçabilité et expertise pour comparer les parcelles et / ou millésimes . . . . .	22
3.1	Discrétisation de l'échelle temporelle pour l'étude d'un phénomène temporel	25
3.2	Schéma récapitulatif des étapes de retraitement proposées . . . . .	30
4.1	Liste des paramètres utilisés dans l'application et symbole associé . . . . .	47
4.2	Etude de deux individus de même dimension : TS asynchrones . . . . .	48
4.3	Position des individus sur le plan constitué des deux premières composantes principales . . . . .	49
4.4	Cercle de corrélation des variables - plan 1-2 - ACP réalisée sur les paramètres climatiques entre floraison et véraison . . . . .	52
4.5	Carte factorielle des individus analysés dans l'ACP avec les directions mises en avant . . . . .	54
4.6	Carte factorielle des individus analysés dans l'ACP sans paramètres prenant en compte la précocité des cépages avec les directions déterminées par les paramètres mises en avant . . . . .	56
4.7	Carte factorielle des individus analysés dans l'ACP sans paramètres prenant en compte la précocité des cépages avec les directions déterminées par les paramètres mises en avant pour la troisième et la quatrième composante principale . . . . .	58
5.1	Dates de mesures du pH des trois parcelles choisies . . . . .	67
5.2	GDD de mesures du pH des trois parcelles choisies . . . . .	68
5.3	Modèles de la cinétique du pH pendant la maturation de trois parcelles de Cabernet-Sauvignon . . . . .	73

TABLE DES FIGURES

---

6.1	Date des relevés effectués pour chacun des 7 individus de la base de données 1	85
6.2	Deux exemples de modèle paramétrique ajusté et sa séquence temporelle pour la base de données 1 : (a) cinétique du millésime 2004, (b) cinétique du millésime 2009 . . . . .	87
6.3	Deux exemples de modèle paramétrique ajusté et sa séquence temporelle pour la base de données 2 : (a) cinétique de la parcelle 32, (b) cinétique de la parcelle 36 . . . . .	88
6.4	Cinétiques d'accumulation des sucres dans les baies de raisin modélisées pour les 7 individus de la base de données 1 . . . . .	93
6.5	Quelques cinétiques d'accumulation des sucres dans les baies de raisin pour la base de données 2 . . . . .	94
6.6	Carte factorielle du plan 1-2 de l'ACP réalisée sur les 7 individus de la base de données 1 en prenant comme variables les 3 paramètres ajustés pour chaque modèle . . . . .	95
6.7	Analyse en composante principale des 7 millésimes de la base de données 1 par les trois paramètres de chaque modèle : (a) cercle des corrélations, (b) carte factorielle du plan 1-2 . . . . .	97
6.8	Loadings des deux premières composantes principales de l'ACP réalisée sur l'utilisation extensive des modèles ajustés à partir de la base de données 1 : (a) loading de la première composante principale, (b) loading de la seconde composante principale . . . . .	100
6.9	Variabilité des 7 cinétiques étudiées dans le créneau déterminé par la première composante principale . . . . .	101
6.10	Mise en avant de la notion d'amplitude de courbe pour la différenciation des individus sur la seconde composante principale . . . . .	103
6.11	Variabilité des 7 cinétiques étudiées dans les fenêtres déterminées par la seconde composante principale . . . . .	104
6.12	Carte factorielle du plan 1-2 de l'ACP réalisée sur l'utilisation extensive des modèles ajustés à partir de la base de données 1 . . . . .	105
6.13	Carte factorielle du plan 1-2 de l'ACP réalisée sur les 65 individus de la base de données 2. La classe 1 est représentée par l'ellipse en pointillés, la classe 2 par l'ellipse en pointillés mixtes, la classe 3 par l'ellipse en trait continu puis la classe 4 par l'ellipse en tirets. . . . .	106
6.14	Cercle des corrélations des variables de l'ACP réalisée sur les 65 parcelles de la base de données 2 en prenant en variables les trois paramètres ajustés pour chaque modèle . . . . .	108
6.15	Carte factorielle du plan 1-2 de l'ACP réalisée sur les 65 individus de la base de données 2 . . . . .	109
6.16	Loadings des deux premières composantes principales de l'ACP réalisée sur l'utilisation extensive des modèles ajustés à partir de la base de données 2 : (a) Loading de la première composante principale, (b) Loading de la seconde composante principale . . . . .	111

6.17	Carte factorielle du plan 1-2 de l'ACP en extensif réalisée sur les 65 individus de la base de données 2. La classe 1 est représentée par l'ellipse en pointillés, la classe 2 par l'ellipse en tirets, la classe 3 par l'ellipse en trait continu puis la classe 4 par l'ellipse en pointillés mixtes. . . . .	113
C.1	Carte factorielle des composantes principales 3 et 4 de l'ACP réalisée sur les paramètres climatiques entre floraison et véraison . . . . .	153
C.2	Cercle de corrélation des variables - plan 3-4 - ACP réalisée sur les paramètres climatiques entre floraison et véraison . . . . .	155
C.3	Carte factorielle des composantes principales 3 et 4 de l'ACP réalisée sur les paramètres climatiques entre floraison et véraison avec les directions mises en avant . . . . .	157



# Liste des tableaux

4.1	Dates en jours julien de la floraison et de la véraison et nombre de jours entre ces deux stades pour les 11 cépages et les 7 millésimes étudiés . . . .	40
4.2	Grandeurs utilisées pour la comparaison du comportement climatique sur différents individus . . . . .	42
4.3	Type de calcul réalisé par grandeur et nom du paramètre extrait . . . . .	43
4.4	Compteurs du nombre de jours ayant dépassé un seuil et description des seuils . . . . .	44
4.5	Moyenne ou somme des variables climatiques principales entre le 01 juin (jour 152) et le 15 juillet (jour 196) . . . . .	50
4.6	Temps entre floraison et véraison et position sur l'axe 2 par millésime . .	51
5.1	Description des individus de la base de données - pH . . . . .	64
5.2	Répartition des paramètres ajustés lors de la modélisation du pH - Base de données pH - 53 individus . . . . .	69
5.3	Valeur des paramètres ajustés lors de la modélisation de l'augmentation de pH de trois parcelles de la base de données étudiées . . . . .	72
6.1	Description des individus de la base de données 1 - INRA Pech Rouge, France . . . . .	80
6.2	Description des individus de la base de données 2 - Napa Valley, Californie	82
6.3	Valeurs ajustées des paramètres pour chaque individu de la base de données 1 . . . . .	89
6.4	Statistiques de répartition des paramètres pour les individus de la base de données 2 . . . . .	89
6.5	Moyenne et écart-type de chaque paramètre par zone agroclimatique . . .	107
6.6	Moyenne d'IH et des températures moyenne, maximale et minimale sur la période 01/04 au 30/09 par zone agroclimatique . . . . .	110
C.1	Correspondance entre le temps entre floraison et véraison et la position sur l'axe 3 par millésime . . . . .	154





# Bibliographie

- [1] C. Acevedo-Opazo, B. Tisseyre, H. Ojeda, S. Ortega-Farias, and S. Guillaume, *Is it possible to assess the spatial variability of vine water status ?*, Journal International des Sciences de la Vigne et du Vin **42** (2008), no. 4, 203–219.
- [2] V.I. Adamchuk, J.W. , Hummel, M.T. , Morgan, and S.K. , Upadhyaya, *On-the-go soil sensors for precision agriculture*, Computers and Electronics in Agriculture **44** (2004), no. 1, 71–91.
- [3] G. Anastasi, O. Farruggia, G. L. Re, and M. Ortolani, *Monitoring high-quality wine production using wireless sensor networks*, Proceedings of the 42nd Annual Hawaii International Conference on System Sciences, HICSS (Waikoloa, HI), 42nd Annual Hawaii International Conference on System Sciences, HICSS, 2009.
- [4] J. Arno, J. A. Martinez-Casasnovas, M. Ribes-Dasi, and J. R. Rosell, *Review. precision viticulture. research topics, challenges and opportunities in site-specific vineyard management*, Spanish Journal of Agricultural Research **7** (2009), no. 4, 779–790.
- [5] J. M. Attonaty, M. H. Chatelin, and F. Garcia, *Interactive simulation modeling in farm decision-making*, Computers and Electronics in Agriculture **22** (1999), no. 2-3, 157–170.
- [6] C. Aubry, F. Papy, and A. Capillon, *Modelling decision-making processes for annual crop management*, Agricultural Systems **56** (1998), no. 1, 45–65.
- [7] T. Becker and M. Knoche, *Deposition, strain and microcracking of the cuticle in developing 'riesling' grape berries*, Vitis **51** (2012), no. 1, 1–6.
- [8] R. Beckwith, D. Teibel, and P. Bowen, *Unwired wine : Sensor networks in vineyards*, Proceedings of IEEE Sensors (Vienna) (D. Rocha, P. M. Sarro, and M. J. Vellekoop, eds.), IEEE Sensors 2004, vol. 2, 2004, pp. 561–564.
- [9] J. Bergqvist, N. Dokoozlian, and N. Ebisuda, *Sunlight exposure and temperature effects on berry growth and composition of cabernet sauvignon and grenache in the central san joaquin valley of california*, American Journal of Enology and Viticulture **52** (2001), no. 1, 1–7.
- [10] S. Best, L. Leon, and M. Claret, *Use of precision viticulture tools to optimize the harvest of high quality grapes*, Frutic (Montpellier, France), 2005.
- [11] J. Blouin and G. Guimberteau, *Maturation et maturité des raisins*, Feret, Bordeaux, 2000.

- [12] J. Blouin and E. Peynaud, *Connaissance et travail du vin, 4ème édition*, la vigne ed., Dunod, Bordeaux, 2005.
- [13] R. G. V. Bramley, *Progress in the development of precision viticulture - variation in yield, quality and soil properties in contrasting australian vineyards*, Precision Tools for Improving Land Management (2001), 25–43.
- [14] J. Branas, *Viticulture*, Imprimerie Dehan, Montpellier, 1974.
- [15] A. Carbonneau, A. Deloire, and B. Jaillard, *La vigne : Physiologie, terroir, culture*, Dunod, Paris, 2007.
- [16] A. Carbonneau, C. Riou, D. Guyon, J. Riom, and C. Schneider, *Agrométéorologie de la vigne en france*, Office des publications officielles des Communautés européennes, Luxembourg, 1992.
- [17] M. Cerf and M. Sébillotte, *Approche cognitive des décisions de production dans l'exploitation agricole*, *Économie Rurale* **239** (1997), 11–18.
- [18] Z. G. Cerovic, N. Moise, G. Agati, G. Latouche, N. Ben Ghazlen, and S. Meyer, *New portable optical sensors for the assessment of winegrape phenolic maturity based on berry fluorescence*, *Journal of Food Composition and Analysis* **21** (2008), no. 8, 650–654.
- [19] F. Champagnol, *Éléments de physiologie de la vigne et de viticulture generale*, F. Champagnol, Montpellier, 1984.
- [20] C. Combi, F. Pinciroli, and G. Pozzi, *Managing different time granularities of clinical information by an interval-based temporal data model*, *Methods of Information in Medicine* **34** (1995), no. 5, 458–474.
- [21] B. G. Coombe and M. G. McCarthy, *Dynamics of grape berry growth and physiology of ripening*, *Australian Journal of Grape and Wine Research* **6** (2000), no. 2, 131–135.
- [22] A. K. Covington, R. G. Bates, and R. A. Durst, *Definition of ph scales, standard reference values, measurements of ph and related terminology*, *Pure Appl. Chem.* **57** (1985), no. 3, 531–542.
- [23] D. Cozzolino, M. B. Esler, R. G. Damberg, W. U. Cynkar, D. R. Boehm, I. L. Francis, and M. Gishen, *Prediction of colour and ph in grapes using a diode array spectrophotometer (400-1100 nm)*, *Journal of Near Infrared Spectroscopy* **12** (2004), no. 2, 105–111.
- [24] F. Cus, *Influence of crop load on yield and grape quality of cv. chardonnay*, *Acta agriculturae slovenica* (2004), 73 – 83.
- [25] Z. W. Dai, P. Vivin, F. Barrieu, N. Ollat, and S. Delrot, *Physiological and modelling approaches to understand water and carbon fluxes during grape berry growth and quality development : A review*, *Australian Journal of Grape and Wine Research* **16** (2010), no. SUPPL. 1, 70–85.
- [26] Z. W. Dai, P. Vivin, T. Robert, S. Milin, S. H. Li, and M. Genard, *Model-based analysis of sugar accumulation in response to source-sink ratio and water supply in grape (vitis vinifera) berries*, *Functional Plant Biology* **36** (2009), no. 6, 527–540.

- 
- [27] C. Davies and S. P. Robinson, *Sugar accumulation in grape berries : Cloning of two putative vacuolar invertase cdnas and their expression in grapevine tissues*, *Plant Physiology* **111** (1996), no. 1, 275–283.
- [28] P. Debaeke, N. Munier-Jolain, M. Bertrand, L. Guichard, J. M. Nolot, V. Faloya, and P. Saulas, *Iterative design and evaluation of rule-based cropping systems : Methodology and case studies. a review*, *Agronomy for Sustainable Development* **29** (2009), no. 1, 73–86.
- [29] G. Delgado, V. Aranda, J. Calero, M. Sanchez-Maranon, J. M. Serrano, D. Sanchez, and M. A. Vila, *Using fuzzy data mining to evaluate survey data from olive grove cultivation*, *Computers and Electronics in Agriculture* **65** (2009), no. 1, 99–113.
- [30] A. Deloire, E. Vaudour, V. Carey, V. Bonnardot, and C. Van Leeuwen, *Grapevine responses to terroir : A global approach*, *Journal International des Sciences de la Vigne et du Vin* **39** (2005), no. 4, 149–162.
- [31] C. Deola, B. Leger, O. Naud, and A. Ugaglia, *A bioeconomic model to train interdisciplinary research on reducing phytosanitary operations in the vineyards*, *Oenométrie XIV* (Trèves, DE), 2007.
- [32] S. Dimitriadis and C. Goumopoulos, *Applying machine learning to extract new knowledge in precision agriculture applications*, *Proceedings - 12th Pan-Hellenic Conference on Informatics, PCI 2008 (Samos Island), 12th Pan-Hellenic Conference on Informatics, PCI 2008, 2008*, pp. 100–104.
- [33] D. L. Elliott-Fisk, *Viticultural soils of california, with special reference to the napa valley*, *Journal of Wine Research* **4** (1993), no. 2, 67–77.
- [34] B. Escofier and J. Pages, *Analyses factorielles simples et multiples, deuxième édition*, Dunod, Paris, 1993.
- [35] J. Fernandez-Novales, M. I. Lopez, M. T. Sanchez, J. Morales, and V. Gonzalez-Caballero, *Shortwave-near infrared spectroscopy for determination of reducing sugar content during grape ripening, winemaking, and aging of white and red wines*, *Food Research International* **42** (2009), no. 2, 285–291.
- [36] H. Friesland, *Review of the scientific literature on the effect of climate and weather, especially during the ripening period, on the quality and storage capacity of grapes, spring barley and potatoes*, *Report of the RA VI working group on agricultural meteorology (Geneva, Switzerland)*, vol. 1, 2002, pp. 81 – 118.
- [37] P. Galet, *Précis de viticulture - 7 ème edition*, JF Impression, Saint Jean de Vedas, 2000.
- [38] S.R. Garner, G. Holmes, R.J. McQueen, and I.H. Witten, *Machine learning from agricultural databases : Practice and experience*, *J. Computing* **6** (1997), no. 1 A, 69–73.
- [39] P. Geladi, *Chemometrics in spectroscopy. part 1. classical chemometrics*, *Spectrochimica Acta - Part B Atomic Spectroscopy* **58** (2003), no. 5, 767–782.
-

- [40] V. Geraudie, *Système d'aide a la conduite des procédés viticoles et oenologiques, basé sur l'utilisation d'un capteur piéton de suivi de maturité*, Ph.D. thesis, Montpellier SupAgro, 2009.
- [41] V. Geraudie, J.M. Roger, and H. Ojeda, *Development of a tool allowing the prediction of grape maturity using near-infrared spectroscopy*, The Australian and New Zeland Grapegrower and Winemaker (2010), no. Annual technical issue, 10–16.
- [42] J. Gladstones, *Viticulture and environment*, Winetitles, Adelaide, 1992.
- [43] I. A. Goralwalla, Y. Leontiev, M. T. Ozsu, D. Szafron, and C. Combi, *Temporal granularity : Completing the puzzle*, Journal of Intelligent Information Systems **16** (2001), no. 1, 41–63.
- [44] J-P. Goutouly and Z. G. Cerovic, *Capteurs optiques portables pour une viticulture de précision*, Revue des œnologues **127** (2008), 17–18.
- [45] O. M. Grant, L. Tronina, H. G. Jones, and M. M. Chaves, *Exploring thermal imaging variables for the detection of stress responses in grapevine under different irrigation regimes*, Journal of Experimental Botany **58** (2007), no. 4, 815–825.
- [46] R. Guidetti, R. Beghi, and L. Bodria, *Evaluation of grape quality parameters by a simple vis/nir system*, Transactions of the ASABE **53** (2010), no. 2, 477–484.
- [47] H. E. Holt, I. L. Francis, J. Field, M. J. Herderich, and P. G. Iland, *Relationships between berry size, berry phenolic composition and wine quality scores for cabernet sauvignon (vitis vinifera l.) from different pruning treatments and different vintages*, Australian Journal of Grape and Wine Research **14** (2008), no. 3.
- [48] G. S. Howell, *Sustainable grape productivity and the growth-yield relationship : A review*, American Journal of Enology and Viticulture **52** (2001), no. 3, 165–174.
- [49] P. Huglin, *Nouveau mode d'évaluation des possibilités héliothermiques d'un milieu viticole*, C.R. Acad. Agric. (1978), 1117–1126.
- [50] C. Interieri and I. Filippetti, *Planting density and physiological balance : Comparing approaches to european viticulture in the 21st century*, ASEV 50th Anniversary Annual Meeting (Seattle, Washington), 2000.
- [51] L. Irimia, C. Tardea, M. Niculaua, and L. Rotaru, *Influence of canopy on must and wine quality in the zweigelt variety*, (2006).
- [52] D. I. Jackson and P. B. Lombard, *Environmental and management practices affecting grape composition and wine quality-a review*, American Journal of Enology and Viticulture **44** (1993), no. 4, 409.
- [53] C. Jensen, C. Dyreson, M. Bohlen, J. Clifford, R. Elmasri, S. Gadia, F. Grandi, P. Hayes, S. Jajodia, W. Kafer, N. Kline, N. Lorentzos, Y. Mitsopoulos, A. Montanari, D. Nonen, E. Peressi, B. Pernici, J.F. Roddick, N.L. Sarda, M.R. Scalas, A. Segev, R.T. Snodgrass, M.D. Soo, A. Tansel, Tiberio P., and G. Wiederhold, *The consensus glossary of temporal database concepts - february 1998 version*, Temporal Databases : Research and Practice (1998), 367–405.
- [54] L. F. Johnson, *Temporal stability of an ndvi-lai relationship in a napa valley vineyard*, Australian Journal of Grape and Wine Research **9** (2003), no. 2, 96–101.

- 
- [55] L.F. Johnson, D.E. Roczen, S.K. Youkhana, R.R. Nemani, and D.F. Bosch, *Mapping vineyard leaf area with multispectral satellite imagery*, Computers and Electronics in Agriculture **38** (2003), no. 1, 33–44.
- [56] G. V. Jones and R. E. Davis, *Climate influences on grapevine phenology, grape composition, and wine production and quality for bordeaux, france*, American Journal of Enology and Viticulture **51** (2000), no. 3, 249–261.
- [57] H. G. Jones, M. Stoll, T. Santos, C. D. Sousa, M. M. Chaves, and O. M. Grant, *Use of infrared thermography for monitoring stomatal closure in the field : Application to grapevine*, Journal of Experimental Botany **53** (2002), no. 378, 2249–2260.
- [58] M. Keller, *Managing grapevines to optimise fruit development in a challenging environment : a climate change primer for viticulturists*, Australian Journal of Grape and Wine Research **16** (2010), 56–69.
- [59] W. M. Kliewer, *Influence of environment on metabolism of organic acids and carbohydrates in vitis vinifera. i. temperature*, Plant Physiol. **39** (1964), 869–880.
- [60] W. M. Kliewer and N. K. Dokoozlian, *Leaf area/crop weight ratios of grapevines : Influence on fruit composition and wine quality*, American Journal of Enology and Viticulture **56** (2005), no. 2, 170.
- [61] W. Koblet, M. C. Candolfi-Vasconcelos, W. Zweifel, and G. S. Howell, *Influence of leaf removal, rootstock, and training system on yield and fruit composition of pinot noir grapevines*, American Journal of Enology and Viticulture **45** (1994), no. 2, 181–187.
- [62] David B. Leake, *Case-based reasoning*, Knowledge Engineering Review **9** (1994), no. 1, 61–64.
- [63] K. Mace, P. Morlon, N. Munier-Jolain, and L. Quere, *Time scales as a factor in decision-making by french farmers on weed management in annual crops*, Agricultural Systems **93** (2007), no. 1-3, 115–142.
- [64] H. Martens and T. Naes, *Principal component analysis*, Multivariate Calibration, John Wiley and Sons, Chichester, New York, 1989, pp. 97–100.
- [65] A. Matese, S. F. Di Gennaro, A. Zaldei, L. Genesio, and F. P. Vaccari, *A wireless sensor network for precision viticulture : The nav system*, Computers and Electronics in Agriculture **69** (2009), no. 1, 51–58.
- [66] S. Matsui, K. Ryugo, and W. M. Kliewer, *Growth inhibition of thompson seedless and napa gamay berries by heat stress and its partial reversibility by applications of growth regulators*, American Journal of Enology and Viticulture **37** (1986), no. 1, 67.
- [67] M. Moller, V. Alchanatis, Y. Cohen, M. Meron, J. Tsipris, A. Naor, V. Ostrovsky, M. Sprintsin, and S. Cohen, *Use of thermal and visible imagery for estimating crop water status of irrigated grapevine*, Journal of Experimental Botany **58** (2007), no. 4, 827–838.
-

- [68] D. Moncomble and L. Panigai, *Caractérisation et gestion de la maturation par terroir en champagne*, Colloque International sur les Terroirs Viticoles (Angers, France), 1996, pp. 1–7.
- [69] F.J. Montero, J. Melia, A. Brasa, D. Segarra, A. Cuesta, and S. Lanjeri, *Assessment of vine development according to available water resources by using remote sensing in la mancha, spain*, *Agricultural Water Management* **40** (1999), 363–375.
- [70] H. Ojeda, *Irrigation qualitative de précision de la vigne*, *Le Progrès agricole et viticole*(Montpellier) **124** (2007), no. 7, 133–141.
- [71] H. Ojeda, A. Deloire, A. Carbonneau, A. Ageorges, and C. Romieu, *Berry development of grapevines : Relations between the growth of berries and their dna content indicate cell multiplication and enlargement*, *Vitis* **38** (1999), no. 4, 145–150.
- [72] N. Ollat, P. Diakou-Verdin, J. P. Carde, F. Barrieu, J. P. Gaudillere, and A. Moing, *Grape berry development : A review*, *Journal International des Sciences de la Vigne et du Vin* **36** (2002), no. 3, 109–131.
- [73] N. Ollat and J. P. Gaudillere, *The effect of limiting leaf area during stage i of berry growth on development and composition of berries of vitis vinifera l. cv. cabernet sauvignon*, *American Journal of Enology and Viticulture* **49** (1998), no. 3, 251–258.
- [74] D. Olle, J. L. Guiraud, J. M. Souquet, N. Terrier, A. Ageorges, V. Cheynier, and C. Verries, *Effect of pre- and post-veraison water deficit on proanthocyanidin and anthocyanin accumulation during shiraz berry development*, *Australian Journal of Grape and Wine Research* **17** (2011), no. 1, 90–100.
- [75] L. Panigai and D. Moncomble, *L'étude de la maturation en champagne - contribution de la fonction logistique*, *Le Vigneron Champenois* (1988), 507–514.
- [76] A. K. Parker, I. G. De Cortazar-Atauri, C. Van Leeuwen, and I. Chuine, *General phenological model to characterise the timing of flowering and veraison of vitis vinifera l*, *Australian Journal of Grape and Wine Research* **17** (2011), no. 2, 206–216.
- [77] P. R. Petrie and V. O. Sadras, *Advancement of grapevine maturity in australia between 1993 and 2006 : Putative causes, magnitude of trends and viticultural consequences*, *Australian Journal of Grape and Wine Research* **14** (2008), no. 1, 33–45.
- [78] D. R. E. Possner and W. M. Kliewer, *The localisation of acids, sugars, potassium and calcium in developing grape berries*, *Vitis* **24** (1985), no. 4, 229–240.
- [79] A. Reynier, *Manuel de viticulture, 10ème edition*, tec et doc ed., Lavoisier, Paris, 2007.
- [80] P. Ribereau-Gayon, D. Dubourdieu, B. Doneche, and A. Lonvaud, *Traité d'œnologie 1. microbiologie du vin, vinifications*, la vigne ed., Dunod, Paris, 1998.
- [81] S. P. Robinson and C. Davies, *Molecular biology of grape berry ripening*, *Australian Journal of Grape and Wine Research* **6** (2000), no. 2, 175–188.
- [82] L. G. Santesteban and J. B. Royo, *Water status, leaf area and fruit load influence on berry weight and sugar accumulation of cv. 'tempranillo' under semiarid conditions*, *Scientia Horticulturae* **109** (2006), no. 1, 60–65.

- 
- [83] G. Saporta, *Probabilité, analyse des données et statistique - deuxième édition*, Technip, Paris, France, 2006.
- [84] X. Sean Wang, S. Jajodia, and V. S. Subrahmanian, *Temporal modules : An approach toward federated temporal databases*, *Information Sciences* **82** (1995), no. 1-2, 103–128.
- [85] L. Serrano, C. Gonzalez-Flor, and G. Gorchs, *Assessing vineyard water status using the reflectance based water index*, *Agriculture, Ecosystems and Environment* **139** (2010), no. 4, 490–499.
- [86] Arie Shoshani and Kyoji Kawagoe, *Temporal data management*, Twelfth International Conference on Very Large Data Bases, Proceedings (VLDB '86), Publ by Morgan Kaufmann Publ Inc, 1986, pp. 79–88.
- [87] R. Snodgrass, *Temporal databases*, *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space* (1992), 22–64.
- [88] M. Stoll and H. G. Jones, *Thermal imaging as a viable tool for monitoring plant stress*, *Journal International des Sciences de la Vigne et du Vin* **41** (2007), no. 2, 77–84.
- [89] J. A. Taylor, *Digital terroirs and precision viticulture : investigation into the application of information technology in australian vineyards*, Ph.D. thesis, University of Sydney, Australia, 2004.
- [90] B. Tisseyre, C. Mazzoni, N. Ardoin, and C. Clipet, *Yield and harvest quality measurement in precision viticulture - application for a selective vintage*, Proceedings III (Montpellier, France), ECPA - European Conference on Precision Agriculture, June 18-21, 2001, pp. 133–138.
- [91] B. Tisseyre, H. Ojeda, and J. Taylor, *New technologies and methodologies for site-specific viticulture*, *International Journal of Wine and Vine Research* **41** (2007), no. 2, 63–76.
- [92] J. Tonietto and A. Carbonneau, *A multicriteria climatic classification system for grape-growing regions worldwide*, *Agricultural and Forest Meteorology* **124** (2004), no. 1-2, 81–97.
- [93] C. Van Leeuwen, *Incidences de l'alimentation en eau de la vigne, appréciée par l'état hydrique du feuillage, sur le développement de l'appareil végétatif et la maturation du raisin (vitis vinifera variété cabernet franc, saint-emilion 1990)*, *Journal International des Sciences de la Vigne et du Vin* **28** (1994), no. 2, 81–110.
- [94] C. Van Leeuwen, P. Friant, M. E. Jaeck, S. Kuhn, and O. Laviaille, *Hierarchy of the role of climate, soil and cultivar in terroir effect can largely be explained by vine water status*, Joint International Conference on Viticultural Zoning (Cap Town, South Africa), 2004, pp. 433–439.
- [95] C. Van Leeuwen, C. Garnier, C. Agut, B. Baculat, G. Barbeau, E. Besnard, I. Cortazar, E. Marguerit, C. Monamy, and S. Koundouras, *Heat requirements for grapevine varieties is essential information to adapt plant material in a changing climate*, VII International Terroir Congress (Changins - Switzerland), 2008.
-

- [96] P. F. Verhulst, *Notice sur la loi que la population suit dans son accroissement*, *Corr. Math. et Phys* **10** (1838), 113–121.
- [97] G. Wiederhold, S. Jajodia, and W. Litwin, *Dealing with granularity of time in temporal databases*, *Advanced Information Systems Engineering* (1991), 124–140.
- [98] A. J. Winkler, J. A. Cook, W. M. Kliever, and L. A. Lider, *General viticulture - 4th edition*, University of Calif. Press, Berkeley and Los Angeles, 1974.
- [99] S. Wold, K. Esbensen, and P. Geladi, *Principal component analysis*, *Chemometrics and Intelligent Laboratory Systems* **2** (1987), no. 1-3, 37–52.
- [100] N. Zhang, M. Wang, and N. Wang, *Precision agriculture - a worldwide overview*, *Computers and Electronics in Agriculture* **36** (2002), no. 2-3, 113–132.



# Appendices



# Annexe A

## Détail du choix des valeurs seuils pour les compteurs

La connaissance experte présentée dans la partie 4.1 a été utilisée pour définir les seuils des compteurs utilisés pour l'étude comparative du climat entre les 7 millésimes et les 11 cépages sur la période allant de la floraison à la véraison. Les choix effectués sont détaillés ci-dessous. Certains seuils ont été adaptés au contexte particulier de la station expérimentale INRA Pech Rouge, de climat méditerranéen.

Les sens des inéquations utilisées pour le calcul des différents compteurs sont donnés dans le tableau 4.4. La liste ci-dessous ne détaille que les raisons ayant mené au choix des valeurs présentées dans ce tableau.

1. L'activité photosynthétique de la plante est maximale lorsque les températures sont proches de 25°C. La vitesse de maturation est maximale lorsque les températures sont comprises entre 20 et 25°C. Le climat du domaine étudié est méditerranéen. Les températures sont donc chaudes.
  - Températures fraîches, ne permettant pas une activité métabolique maximale : TM20 et TX25
  - Températures élevées, activité métabolique maximale puis pouvant être limitée par des températures extrêmes : TM25, TX30 et TX35
  - Températures comprises entre deux bornes qui représentent une vitesse de maturation maximale : TMMat
  - Températures comprises entre deux bornes qui représentent une activité photosynthétique maximale : TMCR et TXOPT
2. Les précipitations permettent une incorporation de l'eau dans le sol pour qu'elle soit utilisée par la plante à partir d'une intensité de 30 mm
  - Jours avec des précipitations, mais ces précipitations ne présentent pas un cumul assez élevé pour que l'eau ait été utilisée par la plante : RR15
  - Jours avec des précipitations suffisantes pour que l'eau ait été utilisée par la plante : RR30 et RRX30

3. L'activité photosynthétique est maximale à partir de  $400\text{W.m}^{-2}$  de rayonnement global. Elle est en augmentation avant ce seuil.  $400\text{W.m}^{-2}$  correspondent à  $2200\text{J.cm}^{-2}$ .
  - Jours avec un rayonnement global ayant permis une activité photosynthétique maximale : RG22
  - Jours avec un rayonnement global ayant impliqué une limitation importante de l'activité photosynthétique : RG11
4. L'activité photosynthétique de la vigne est maximale lorsque l'humidité de l'air est comprise entre 60 et 70%.
  - Jours avec un temps d'humidité relative très faible de longue durée : U4F
  - Jours avec un temps d'humidité relative très faible de courte durée : U46
  - Jours avec un temps d'humidité relative très forte de longue durée : U8F et U9F
  - Jours avec un temps d'humidité relative très forte de courte durée : U86 et U96

# Annexe B

## Calcul détaillé des indices bioclimatiques entre floraison et véraison

Les cumuls de températures seront détaillés dans un premier temps puis le calcul de l'indice de sécheresse sera détaillé dans un second temps.

### B.1 Les indices basés sur des cumuls de températures

#### B.1.1 Cumul de températures base 10 : Les degrés jour de Winkler (Win) [98]

Pour chaque jour, une valeur de degrés jour dite base 10 a été déterminée comme présenté en équation B.1 :

$$\forall i \in [F; V], \text{ } ^\circ j_{Win}(i) = \begin{cases} 0, & \text{si } TM \leq 10 \\ TM - 10, & \text{si } TM > 10 \end{cases} \quad (\text{B.1})$$

Puis, l'indice (équation B.2) a été calculé en additionnant toutes les valeurs de degrés jours base 10 de la floraison à la véraison.

$$Win = \sum_{i=F}^{i=V} \text{ } ^\circ j_{Win}(i) \quad (\text{B.2})$$

#### B.1.2 Cumul de températures base 10 et sans prise en compte des températures supérieures à 19 °C : Indice de Gladstones (Glad) [42]

Pour chaque jour, une valeur de degrés jour de base 10 et ne prenant pas en compte les températures supérieures à un seuil fixé à 19 °C a été déterminée selon la formule

donnée en équation B.3 :

$$\forall i \in [F; V], \text{ } \circ j_{Glad}(i) = \begin{cases} 0, & \text{si } TM \leq 10 \\ TM - 10, & \text{si } 10 < TM < 19 \\ 9, & \text{si } TM \geq 19 \end{cases} \quad (\text{B.3})$$

Puis, l'indice (équation B.4) a été calculé en additionnant toutes les valeurs de degrés jours base 10 et seuillés à 19 °C de la floraison à la véraison.

$$Glad = \sum_{i=F}^{i=V} \circ j_{Glad}(i) \quad (\text{B.4})$$

### B.1.3 Cumul de températures moyennes et maximales : indice héliothermique de Huglin (IH) [49]

Pour chaque jour, une valeur de degrés jour entre la température moyenne et la température maximale auxquelles 10 °C sont enlevés a été déterminée. Le calcul est présenté en équation B.5.

$$\forall i \in [F; V], \text{ } \circ j_{IH}(i) = (TM(i) - 10) + (TX(i) - 10) \quad (\text{B.5})$$

Puis, l'indice a été calculé en additionnant toutes les valeurs de degrés jours en les pondérant d'un paramètre représentant la longueur du jour déterminé en fonction de la latitude du domaine (équation B.6).

$$IH = \sum_{i=F}^{i=V} k \times \frac{\circ j_{IH}(i)}{2} \quad (\text{B.6})$$

k est un coefficient de longueur du jour. Sa valeur est de 1.03 pour les latitudes de la station expérimentale INRA Pech Rouge.

### B.1.4 Cumul de précipitations soustrait du cumul de températures : indice thermo-pluviométrique (ITP) de Ribéreau-Gayon et Peynaud, in [11]

Cet indice était calculé en réalisant la différence entre la somme des températures entre floraison et véraison et la somme des précipitations sur la même période (équation B.7).

$$ITP = \sum_{i=F}^{i=V} TM(i) - \sum_{i=F}^{i=V} RR(i) \quad (\text{B.7})$$

## B.2 Les indices d'évaluation de l'état hydrique

Un seul indice a été utilisé pour évaluer l'état hydrique de la plante.

## B.2.1 L'indice de sécheresse (IS) [92]

Cet indice a été calculé à partir du 1er avril et jusqu'à la véraison. Sa formule globale est donnée en équation B.8.

$$IS = W_0 + RR_{SV} - Tv_V - Es_V \quad (B.8)$$

$W_0$  correspondait à la réserve hydrique initiale du sol.  $W_0=200$  mm.

$RR_{SV}$  était la somme des précipitations du 1er avril à la véraison. Elle était calculée selon la formule de l'équation B.9

$$RR_{SV} = \sum_{i=01/04}^{i=V} RR(i) \quad (B.9)$$

$Tv_V$  était la transpiration potentielle de la plante et  $Es_V$  était l'évaporation directe du sol. Ils étaient tous deux calculés sur une base mensuelle grâce aux formules détaillées ci-dessous : la transpiration potentielle mensuelle de la plante ( $Tv_{mois}$ ) est détaillée en équation B.10 tandis que l'évaporation directe du sol mensuelle ( $Es_{mois}$ ) est détaillée en équation B.11.

$$Tv_{mois} = k_{mois} \times ETP_{mois} \quad (B.10)$$

$$Es_{mois} = \frac{ETP_{mois}}{N_{mois}} \times (1 - k_{mois}) \times JPM_{mois} \quad (B.11)$$

$k_{mois}$  était le coefficient d'absorption radiative de la vigne. Il était évalué à 0,1 en avril, 0,3 en mai et 0,5 de juin à septembre.

$ETP_{mois}$  était la somme mensuelle d'évapotranspiration potentielle.

$N_{mois}$  était le nombre de jours par mois.

$JPM_{mois}$  représentait le nombre de jours d'évaporation efficace depuis le sol pour chaque mois. Il était calculé grâce à la formule donnée en équation B.12.

$$JPM_{mois} = \frac{RR_{mois}}{5} \quad (B.12)$$

$RR_{mois}$  était la somme mensuelle de précipitations.

Le calcul de l'indice de sécheresse pouvait présenter deux cas différents : la date de véraison pouvait être en juillet ou en août.

Si la date de véraison était en juillet,  $Tv_{mois}$  et  $Es_{mois}$  étaient calculés normalement pour les mois d'avril, mai et juin. Puis pour le mois de juillet,  $Tv_{JV}$  correspondait à la transpiration potentielle de la plante jusqu'à la véraison (équation B.13) et  $Es_{JV}$  correspondait à l'évaporation directe du sol en juillet jusqu'à la véraison (équation B.14).

$$Tv_{JV} = 0,5 \times ETP_{JV} \quad (B.13)$$

$ETP_{JV}$  était la somme d'évapotranspiration potentielle sur le mois de juillet jusqu'à la véraison.

$$Es_{mois} = \frac{ETP_{JV}}{N_{JV}} \times 0,5 \times \frac{RR_{JV}}{5} \quad (\text{B.14})$$

$N_{JV}$  était le nombre de jours en juillet avant la véraison

$RR_{JV}$  était la hauteur totale des précipitations sur le mois de juillet jusqu'à la véraison

Les valeurs de  $Tv_V$  et  $Es_V$  étaient alors la somme de ces calculs comme présentés en équation B.15 pour la transpiration potentielle et en équation B.16 pour l'évaporation directe du sol.

$$Tv_V = Tv_{avr} + Tv_{mai} + Tv_{juin} + Tv_{JV} \quad (\text{B.15})$$

$$Es_V = Es_{avr} + Es_{mai} + Es_{juin} + Es_{JV} \quad (\text{B.16})$$

Si la date de véraison était en août,  $Tv_{mois}$  et  $Es_{mois}$  étaient calculés normalement pour les mois d'avril, mai, juin et juillet. Puis pour le mois d'août,  $Tv_{AV}$  correspondait à la transpiration potentielle de la plante jusqu'à la véraison (équation B.17) et  $Es_{AV}$  correspondait à l'évaporation directe du sol en août jusqu'à la véraison (équation B.18).

$$Tv_{AV} = 0,5 \times ETP_{AV} \quad (\text{B.17})$$

$ETP_{AV}$  était la somme d'évapotranspiration potentielle sur le mois d'août jusqu'à la véraison.

$$Es_{mois} = \frac{ETP_{AV}}{N_{AV}} \times 0,5 \times \frac{RR_{AV}}{5} \quad (\text{B.18})$$

$N_{AV}$  était le nombre de jours en août avant la véraison

$RR_{AV}$  était la hauteur totale des précipitations sur le mois d'août jusqu'à la véraison

Les valeurs de  $Tv_V$  et  $Es_V$  étaient alors la somme de ces calculs comme présentés en équation B.19 pour la transpiration potentielle et en équation B.20 pour l'évaporation directe du sol.

$$Tv_V = Tv_{avr} + Tv_{mai} + Tv_{juin} + Tv_{Juil} + Tv_{AV} \quad (\text{B.19})$$

$$Es_V = Es_{avr} + Es_{mai} + Es_{juin} + Es_{Juil} + Es_{AV} \quad (\text{B.20})$$



## Annexe C

# Analyse détaillée du plan factoriel 3-4 de l'ACP réalisée à partir des paramètres climatiques sur les 11 cépages et les 7 millésimes

La figure C.1 donne la position des individus analysés par l'ACP en cours d'étude dans la carte factorielle constituée de la troisième et de la quatrième composantes principales.

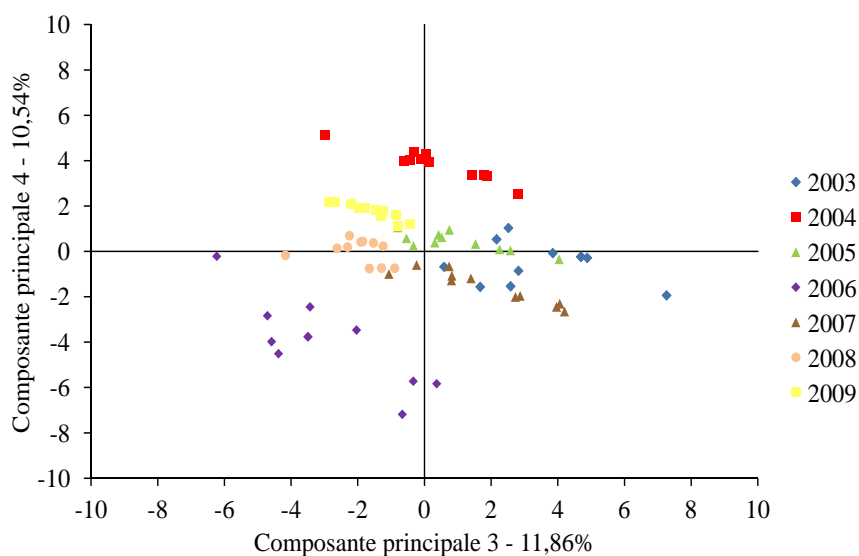


FIGURE C.1 – Carte factorielle des composantes principales 3 et 4 de l'ACP réalisée sur les paramètres climatiques entre floraison et véraison

Contrairement à la carte factorielle des deux premières composantes principales, les millésimes sont assez confondus sur ce graphique. Deux millésimes, 2004 et 2006, se différencient nettement des autres millésimes.

Sur la troisième composante principale, les millésimes sont tous très étendus et les positions des individus présentent des recouvrements d'un millésime à l'autre. Un millésime

TABLE C.1 – Correspondance entre le temps entre floraison et véraison et la position sur l’axe 3 par millésime

2003		2004		2005		2006		2007		2008		2009	
FV	CP3	FV	CP3	FV	CP3	FV	CP3	FV	CP3	FV	CP3	FV	CP3
47	2,52	45	-2,97	47	-0,33	40	-6,23	50	-1,07	55	-4,17	49	-2,69
47	2,16	45	-0,61	47	-0,54	42	-4,59	51	-0,24	56	-2,25	50	-2,86
53	4,69	46	-0,43	48	-0,81	42	-4,38	51	0,74	56	-2,62	51	-2,20
53	3,85	47	-0,10	50	0,42	45	-4,71	54	0,82	57	-2,31	52	-1,78
53	4,69	49	-0,30	51	0,51	46	-3,50	55	1,39	58	-1,85	52	-1,97
53	0,59	49	0,04	52	0,74	46	-3,50	55	0,81	58	-1,91	53	-1,44
54	4,88	51	0,13	52	0,31	48	-3,43	59	2,73	59	-1,53	54	-1,24
55	2,82	52	1,43	52	1,53	50	-0,67	61	2,87	60	-1,24	54	-1,31
57	1,67	53	1,79	53	2,25	51	-2,04	63	4,06	62	-1,65	55	-0,87
58	2,58	53	1,89	54	2,58	54	-0,34	63	3,96	63	-0,89	56	-0,81
63	7,26	56	2,81	58	4,04	56	0,37	65	4,20	63	-1,29	56	-0,44

FV : nombre de jours entre floraison et véraison, CP3 : Position de l’individu sur la troisième composante principale

sime se différencie d’un autre par le climat global de floraison à la véraison. Puis, au sein du millésime, les individus ou cépages se différencient (i) par la date exacte d’atteinte de la floraison et de la véraison, qui permettent d’ajuster les valeurs des paramètres pour chacun et (ii) par la longueur de la période entre floraison et véraison qui, comme cela a été noté dans l’analyse du plan composé des deux premières composantes principales, a un effet sur la valeur des paramètres correspondant à des sommes de descripteurs.

Le tableau C.1 regroupe les positions de chacun des individus sur la troisième composante principale, séparées par millésimes et ordonnées en fonction du nombre de jours qu’il s’est passé entre la floraison et véraison. Comme dans le cas du tableau 4.6 pour l’axe 2, il y a une correspondance globale entre le temps entre floraison et véraison et la position sur l’axe 3. Au sein de chaque millésime, plus le nombre de jours entre floraison et véraison est élevé et plus la coordonnée de l’individu sur l’axe 3 est orientée vers le positif. Les cépages avec un nombre de jours important entre floraison et véraison seront donc projetés à droite du nuage de points représentant chaque millésime tandis que ceux avec une véraison d’apparition précoce après la floraison seront positionnés à gauche du nuage de points. Entre les individus, des inversions peuvent néanmoins être observées. Le nombre de jours entre floraison et véraison n’est donc pas le seul paramètre influençant la disposition des individus au sein de chaque nuage de points.

La troisième composante principale pourrait donc permettre de différencier les individus selon le nombre de jours entre floraison et véraison auquel seraient associés des paramètres qui pourraient être mis en avant dans le cercle des corrélations.

Les millésimes sont très peu étendus sur la quatrième composante principale. Cette

composante principale est donc une composante qui permet de travailler sur le climat global du millésime.

Sur la quatrième composante principale, 2003, 2005, 2007, 2008 et 2009 sont confondus, au centre de l'axe. Cette composante ne permet donc pas de les différencier entre eux. 2004 ressort des autres millésimes dans le sens positif de cette composante principale, tandis que 2006 se sépare dans le sens négatif de l'axe. Cette composante principale permet donc d'explicitier les différences climatiques entre 2004 et 2006.

Dans les moyennes fournies dans le tableau 4.5, aucune valeur ne permet de différencier 2004 et 2006 entre eux et des autres millésimes. Dans le cas présent, il n'est donc pas possible de tirer une conclusion *a priori* sur la signification de cette composante principale. L'analyse du cercle des corrélations va permettre d'essayer de comprendre quelles grandeurs permettent de différencier ces millésimes des autres millésimes.

Les différents millésimes semblent suivre une direction oblique toujours semblable quel que soit le millésime. Le cercle des corrélations pourrait donner une signification à cette direction.

La figure C.2 donne le cercle des corrélations des paramètres dans le plan constitué de la troisième et de la quatrième composantes principales.

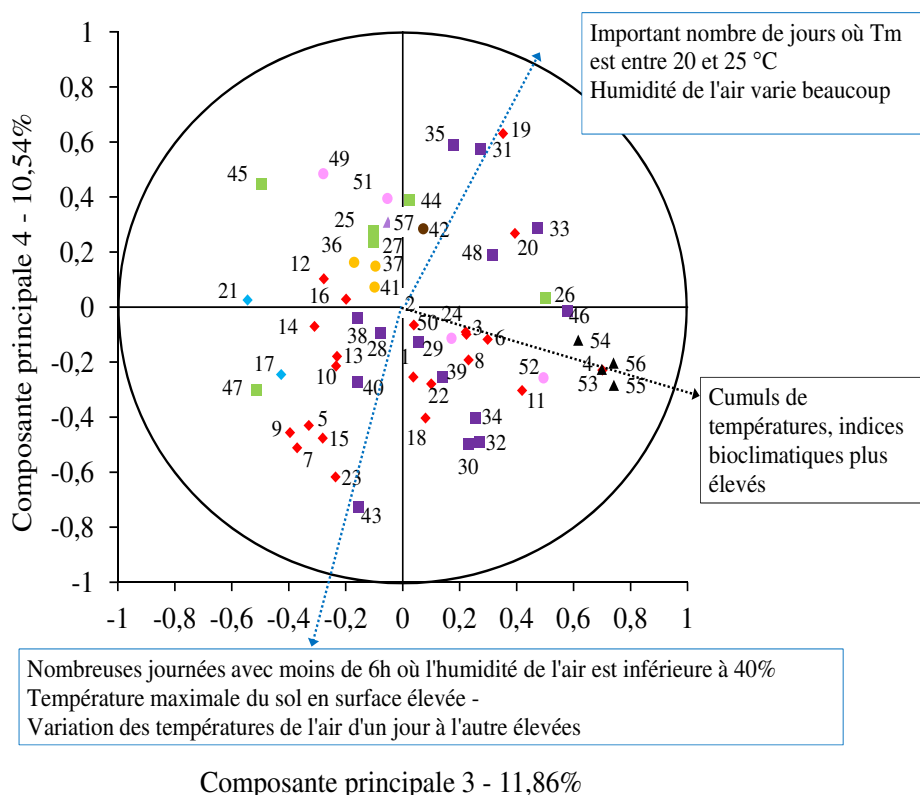


FIGURE C.2 – Cercle de corrélation des variables - plan 3-4 - ACP réalisée sur les paramètres climatiques entre floraison et véraison

Contrairement au cercle des corrélations du plan constitué des deux premières composantes principales, les paramètres sont plutôt mal représentés sur ce cercle.

La troisième composante principale est assez bien corrélée positivement à un ensemble de paramètres représentant des cumuls de températures et le calcul des différents indices bioclimatiques entre floraison et véraison hormis l'indice de sécheresse. Cet axe est donc principalement déterminé par ces sommes. La valeur d'un cumul de températures entre floraison et véraison va être augmentée d'abord par un nombre de jours sur cette période plus long puis éventuellement par des températures plus élevées sur celle-ci.

La troisième composante principale est donc, comme cela avait été supposé à l'analyse de la position des individus, un axe qui pourra permettre de différencier au sein d'un millésime, les cépages avec le temps entre floraison et véraison le plus long, de ceux avec le temps le plus court. Il pourra aussi permettre de séparer les millésimes avec des cumuls de température globalement plus élevés et ceux avec des indices plus faibles.

La quatrième composante principale est assez bien corrélée positivement avec le compteur de jours où la température moyenne de l'air a été comprise entre 20 et 25°C (optimum pour la maturation) et d'importantes variations de l'humidité de l'air d'un jour à l'autre. En négatif, elle est déterminée par le compteur du nombre de jours où le temps pendant lesquels l'humidité de l'air a été inférieure à 40% est inférieur à 6 heures, le compteur du nombre de jours où la température maximale a été supérieure à 35°C, la température maximale du sol en surface et d'importantes variations de température d'un jour à l'autre.

Cet axe est donc déterminé par l'opposition entre :

- des températures maximales du sol et de l'air élevées associées à de fortes variations de la température d'un jour à l'autre et un grand nombre de jours où la température moyenne a été raisonnable.
- un nombre de jours où l'air a rarement été extrêmement sec et d'importantes variations d'humidité de l'air d'un jour à l'autre.

De tels antagonismes n'auraient jamais pu être mis en avant sans la modélisation des paramètres climatiques et la réalisation de cette ACP.

La figure C.3 présente la position des individus dans la carte factorielle de la troisième et de la quatrième composante principale sur laquelle les directions des groupes de paramètres présentées auparavant ont été tracées.

Il avait été noté que chaque millésime était étendu selon une même direction oblique. L'ensemble de paramètres correspondant aux cumuls de températures entre floraison et véraison détermine cette direction. Sur cette direction, les millésimes se confondent les uns aux autres.

Cet axe permet donc de différencier les cépages au sein d'un millésime sans réellement prendre en compte les spécificités climatiques de chaque millésime. Au sein d'un millésime, une plus forte valeur d'indice bioclimatique reflète un nombre de jours entre floraison et véraison plus long. Pour un millésime, les cépages avec une floraison précoce et / ou une véraison tardive seront donc vers la droite du nuage de points tandis que ceux avec une floraison tardive et / ou une véraison précoce seront vers la gauche du nuage de points.

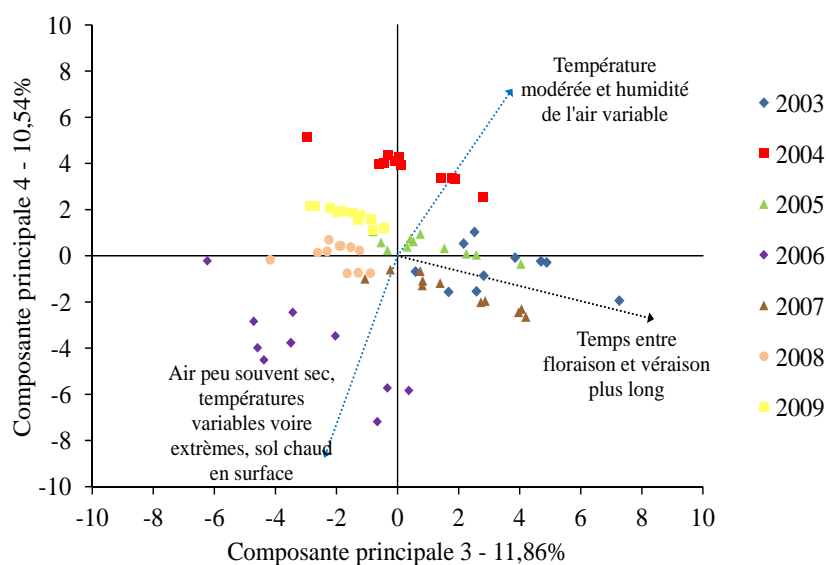


FIGURE C.3 – Carte factorielle des composantes principales 3 et 4 de l'ACP réalisée sur les paramètres climatiques entre floraison et véraison avec les directions mises en avant

Le tableau 4.1 montre qu'il n'y a ni un ordre de longueur de la période entre floraison et véraison entre les cépages ni un temps entre floraison et véraison pour un même cépage prédéfini. La position de chaque cépage à l'intérieur du nuage de points caractérisant le millésime sera donc spécifique à chaque millésime. Le tableau 4.1 rappelle aussi que pour un même nombre de jours entre floraison et véraison, les dates de floraison et véraison ne sont pas fixes. Les sommes de températures ne seront donc pas toutes identiques. Il y aura donc une variabilité de la valeur des indices bioclimatiques entre les cépages due aux températures à proximité de la floraison et à proximité de la véraison. Les températures autour de ces deux stades phénologiques ont donc un effet déterminant sur la position d'un cépage au sein du nuage de points représentatif d'un millésime.

Perpendiculairement à cette direction correspondant aux indices bioclimatiques à l'exception de l'indice de sécheresse, un axe est défini par l'opposition entre (i) des températures de l'air modérées, optimales pour la maturation, et des températures maximales extrêmes associées à un sol chaud en surface et (ii) une humidité de l'air très variable d'un jour à l'autre et un air qui n'est que peu souvent extrêmement sec.

C'est cet antagonisme qui permet de différencier 2004 et 2006 des autres millésimes.

Il est donc possible de dire que la période entre floraison et véraison du millésime 2006 a présenté un grand nombre de jours où la température a dépassé 35°C, un sol chaud en surface et des températures variables d'un jour à l'autre. De l'autre côté, 2004 a été caractérisé par une période de floraison à véraison avec des températures optimales pour la croissance de la baie et beaucoup de variabilité de l'humidité de l'air.

Ce plan permet donc principalement, au sein d'un millésime, de différencier les cépages en fonction du temps passé entre floraison et véraison et des dates de floraison et véraison. La position relative d'un cépage au sein d'un millésime peut-être comparée à celle du même cépage au sein des autres millésimes. Des analogies entre millésimes pour un même

cépage peuvent donc être réalisées. Ces analogies sont complémentaires à celles réalisées sur le plan factoriel 1-2 et pourront servir à une caractérisation du climat de chaque individu en fonction des autres individus.