Valuing farm data for agronomic research and operational purposes : the example of yield forecasting in viticulture

Cécile Laurent

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Sciences Agronomiques

École doctorale [GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau]
Portée par

Les Unités de recherche
ITAP – Technologies et méthodes pour les agricultures de demain
ABSYS – Accompagner la transition agroécologique dans les systèmes de cultures pérennes

Valuing farm data for agronomic research and operational purposes: the example of yield forecasting in viticulture

Présentée par Cécile LAURENT
Le 3 décembre 2021

Sous la direction d'Aurélie METAY et Bruno TISSEYRE

Devant le jury composé de

Benjamin SULTAN, Directeur de recherche, IRD	Président
Justine VANDEN HEUVEL, Associate Professor, Cornell University	Rapportrice
Julien JACQUES, Professeur, Université Lyon II	Rapporteur
Nicolas GUILPART, Maître de conférence, AgroParisTech	Examinateur
Anne PELLEGRINO Maître de conférence, Institut Agro	Examinatrice
Aurélie METAY, Professeure, Institut Agro	Co-directrice de thèse
Bruno TISSEYRE, Professeur, Institut Agro	Invité
Thibaut SCHOLASCH, Docteur, Fruition Sciences	Invité

UNIVERSITÉ DE MONTPELLIER

Montpellier SupAgro

FRUITION SCIENCES

Valuing farm data for agronomic research and operational purposes : the example of yield forecasting in viticulture

Valorisation des données agricoles pour la recherche agronomique et à des fins opérationnelles : Exemple de la prévision du rendement en viticulture.

Cécile LAURENT

Under the supervision of Aurélie METAY and Bruno TISSEYRE

Agrosystèmes Biodiversifiés ABSys Biodiversified Agrosystems

FRUITION SCIENCES CULTIVATE KNOWLEDGE

l'institut Agro agriculture • alimentation • environnement | Montpellier SupAgro

Unité Mixte de Recherche ITAP Technologies & méthodes pour les agricultures de demain INRAE - Montpellier SupAgro

#DigitAg

Loin d'être une fin en soi, la soutenance de ma thèse est quand même l'occasion de me retourner sur le chemin parcouru et de remercier les personnes qui, parfois sans vraiment le savoir, m'ont donné une direction.

Merci aux personnes qui m'ont fait découvrir le monde viticole…

A Benoît Laurent et sa compagne Claire, qui avaient bien compris que mon stage de première année était un test, qui ont tout fait pour que ça me plaise et qui ont réussi.
A Laurence Jougla pour m'avoir montré qu'on peut toujours travailler sans forcer et qu'être une femme dans le milieu du vin n'est pas un problème.
A l'équipe du Château de Meursault,
pour m'avoir transmis un bout de votre passion du vin.
A Nikki Lohitnavy et son équipe, pour m'avoir donné l'occasion de remettre en cause tout ce que j'avais appris, à peine sortie de l'école !

Merci aussi aux personnes des domaines qui ont accepté de partager leurs données et d'échanger autour de la prévision de rendement. Ces échanges m'ont permis d'élargir et de préciser ma vision de la problématique terrain.

Merci aux personnes qui m'ont fait découvrir la recherche…

A Nilo Mejia dont la rigueur et le sens critique resteront un exemple pour moi.
Nilo, c'est toi qui m'as donné envie de faire de la recherche.
A Edith Lecadre pour m'avoir accompagnée
dans le début confus de ma réflexion autour de la thèse.
À toutes les personnes qui ont pris le temps de discuter avec moi
quand j'avais besoin de faire mûrir mon projet.
C'est bien en pensant à vous que je rendrai toujours la pareille !

Et puis, bien-sûr merci à ceux qui m'ont fait prendre goût à la recherche

A Aurélie et Bruno. Vous avez formé un super duo d'encadrement et je suis très fière d'apprendre avec vous. Merci d'avoir trouvé le juste équilibre entre autonomie et assistance, merci pour votre bienveillance, votre ouverture d'esprit et votre réactivité, toujours dans la bonne humeur !
Je garderai en tête le rire de Bruno qui traverse les murs
et les sms d'Aurélie qui s'assure que tout va bien.

A Thibaut et Sébastien qui ont mis au point un microcosme très particulier entre recherche et réalité de terrain. Merci aussi d'oser la différence et l'humilité. Je suis aussi très fière de faire partie d'une entreprise qui met autant d'importance dans le respect de ses valeurs.

A Gilles, qui m'avait dit être pénible question statistiques mais qui m'a surtout appris à bien poser mon problème. J'ai sacrément progressé avec toi.

A James pour ton expertise avisée de chercheur international.
A Meïli aussi, pour avoir continué à suivre ma thèse, même de loin.

Aux doctorants ou jeunes docteurs qui m'ont inspirée : Corentin, Julien,
avec qui j'ai pris plaisir à travailler : Baptiste, Léo, Anice, Eva, Daniel
et ceux qui commencent : Maxime et Yulin.
Prenez bien soin des points doctorants !

Merci aussi aux deux super équipes dont j'ai pu partager l'ambiance appliquée et bienveillante, même en télétravail : j'ai nommé les équipes Agrotic et Fruition !

Enfin, quelques mots pour mon entourage…
Merci à mes parents, qui m'ont transmis leur amour de l'agriculture et cette force d'entreprendre qui me pousse toujours plus loin. Ce n'est pas facile d'être parent et les enfants ne disent pas souvent merci. Alors pour une fois, merci. Merci d'avoir toujours fait du mieux que vous pouviez pour nous et merci de nous avoir emmenées jusqu'aux portes de l'âge adulte.
Merci à ma famille de m'avoir donné un ancrage et des principes.

Merci à mes amis et colocs
dont la diversité des personnalités et des idées ne cessera jamais de me faire avancer.

Et merci à toi Max, parce que cette thèse aura aussi été ton quotidien pendant ces trois ans.
J'ai beaucoup de chance de te compter dans mon équipe.

Contents

List of Figures **9**

List of Tables **11**

1 Introduction **21**

2 Bibliography and presentation of the research strategy for the PhD project **23**
2.1 Intention note . . . 23
2.2 A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture . . . 24
2.2.1 Details about the paper . . . 24
2.2.2 Introduction . . . 25
2.2.3 What are the main issues in assessing grapevine yield ? . . . 28
2.2.4 How do existing methods address yield assessment issues ? . . . 36
2.2.5 Conclusion : what are the perspectives to improve yield assessment ? . . . 58
2.2.6 Acknowledgments . . . 60
2.3 Presentation of the research questions addressed in the PhD project . . . 61
2.3.1 The PhD project focuses on yield modelling approaches based on farm data . . . 61
2.3.2 The PhD project aims at developing an operational, temporal and site-specific approach of yield development . . . 61
2.3.3 Presentation of the research approach of the PhD project . . . 63

3 Designing a yield forecasting method from the characterization of farm data sets **65**
3.1 Intention note . . . 65
3.2 Presentation of the material for the PhD project . . . 65
3.2.1 Characterization of the studied vineyards . . . 65
3.2.2 Characterization of farm data on the basis of the collected data sets . . . 66
3.2.3 Selection of three data sets for further study . . . 67
3.2.4 Presentation of the conceptual model of yield development . . . 69
3.2.5 Presentation of the yield forecasting method . . . 70

4 Computing a site-specific timeline based on a thermal index to perform data analysis consistently with the vineyard phenology **73**
4.1 Intention note . . . 73
4.2 Building new temperature indices for a local understanding of grapevine physiology . . . 74
4.2.1 Details about the paper . . . 74
4.2.2 Introduction . . . 75
4.2.3 Material and Method . . . 76
4.2.4 Results and discussion . . . 77
4.2.5 Conclusion . . . 81
4.3 eGDD : an extension of the Growing Degree Days approach to compute a site-specific thermal index in viticulture using a constrained optimization method. Part I : Method development. . 81
4.3.1 Details about the paper . . . 81
4.3.2 Introduction . . . 82
4.3.3 Materiel and Method . . . 84

4.3.4 Results . . . 89
4.3.5 Discussion . . . 94
4.3.6 Conclusion and perspectives . . . 98
4.3.7 Acknowledgements . . . 98
4.4 eGDD : an extension of the Growing Degree Days approach to compute a site-specific thermal index in viticulture using a constrained optimization method. Part II : Method validation . . . 98
4.4.1 Details about the paper . . . 98
4.4.2 Introduction . . . 99
4.4.3 Materiel and Method . . . 100
4.4.4 Results . . . 101
4.4.5 Discussion . . . 106
4.4.6 Conclusion . . . 108
4.5 General conclusion of the chapter . . . 108

5 Handling the temporality of yield development through the analysis of time series of weather data 111
5.1 Intention note . . . 111
5.2 Evaluation of a functional Bayesian method to analyse time series data in precision viticulture 113
5.2.1 Details about the paper . . . 113
5.2.2 Introduction . . . 113
5.2.3 Material and Method . . . 115
5.2.4 Results . . . 116
5.2.5 Discussion . . . 118
5.2.6 Conclusion . . . 118
5.2.7 Acknowledgements . . . 119
5.3 Identifying operational weather indicators of grapevine yield development by exploring time series of farm data synchronized according to an eGDD thermal index with a linear functional Bayesian method . . . 119
5.3.1 Details about the paper . . . 119
5.3.2 Introduction . . . 120
5.3.3 Material and Method . . . 120
5.3.4 Results . . . 125
5.3.5 Discussion . . . 129
5.3.6 Conclusion . . . 132
5.4 General conclusion of the chapter . . . 133

6 Discussion and perspectives 134
6.1 Looking back at the results of the PhD project . . . 134
6.2 Discussing the response of the PhD project to the issues and challenges identified for grape yield forecasting in an operational context . . . 135
6.2.1 Aiming at operational relevance . . . 136
6.2.2 Taking into account the temporality of yield development . . . 137
6.2.3 Proposing a site-specific approach of yield development . . . 137
6.3 Discussing the response of the PhD project to valuing farm data to perform agronomic research with operational purposes . . . 138
6.3.1 Handling the characteristics of farm data . . . 138
6.3.2 Interpreting the results of analyses from farm data . . . 139
6.4 How will the results of the PhD project be integrated into a grape yield forecasting model ? . 139
6.5 How could the PhD proposal be further assessed ? . . . 140
6.6 How far can the PhD case study be generalized to the agronomic approach for operational purposes using farm data? . . . 141
6.6.1 How far is generalizable to other variables than grape yield ? . . . 141
6.6.2 How far is generalizable to other crops than grapevine ? . . . 141

7 Résumé étendu de la thèse **143**
7.1 Chapitre I : Introduction, problématique et structure du document 143
7.2 Chapitre II : Bibliographie et présentation de la démarche de recherche 144
7.2.1 Etat de l'art des questions, méthodes et perspectives pour l'estimation, prédiction et prévision de rendement en viticulture (Laurent et al., 2021) 144
7.2.2 Présentation des questions de recherche abordées dans la thèse 145
7.3 Chapitre III : Conception d'une méthode de prévision du rendement à partir de la caractérisation de jeux de données agricoles 147
7.3.1 Caractérisation des jeux de données agricoles et sélection de trois cas d'étude 147
7.3.2 Développement d'un modèle conceptuel pour l'élaboration du rendement adapté au contexte opérationnel 147
7.3.3 Développement d'une méthode de prévision du rendement viticole en contexte opérationnel 148
7.3.4 Présentation des verrous scientifiques explorés dans le projet de doctorat 149
7.4 Chapitre IV : Etalonnage d'un temps local exprimé en fonction d'un indice thermique pour effectuer des analyses des données en cohérence avec la phénologie du vignoble 152
7.4.1 Objectif du chapitre 152
7.4.2 Présentation de la méthode Extended Growing Degree Days (eGDD) 152
7.4.3 Validation de la méthode eGDD 153
7.4.4 Définition d'une échelle spatiale d'importance primaire pour la mise en oeuvre de la méthode eGDD 155
7.5 Chapitre V : Prise en compte de la temporalité de l'élaboration du rendement par l'analyse de séries temporelles de données météorologiques 156
7.5.1 Objectif du chapitre 156
7.5.2 Présentation de la démarche analytique incluant les méthodes eGDD et BLiSS 156
7.5.3 La démarche analytique permet d'identifier des indicateurs météorologiques locaux de l'élaboration du rendement 157
7.6 Chapitre VI : Discussion et perspectives 158

List of Figures

2.1 Outline of a longitudinal section of a grapevine latent bud . . . 29
2.2 Overlapping successive reproductive and vegetative cycles that compose grapevine physiology 31
2.3 Influences of temperature, light, water status and nutrition storage from season n-1 to n on yield components observed in season n that have been reviewed in literature. . . . 33
2.4 A framework for constitutive steps of any yield assessment method. . . . 37
2.5 Illustrative real-world examples of different sampling methods . . . 48
2.6 Sources of uncertainty in the process of yield assessment. . . . 52
2.7 Challenges and scientific questions for grape yield assessment in an operational context addressed in the PhD project . . . 62

3.1 Conceptual model of annual grape yield elaboration . . . 70
3.2 Target design of the yield forecasting method implemented for a given vineyard . . . 71
3.3 Positioning of the two research axes aiming at valuing the time series of meteorological data in the framework of yield the forecasting method. . . . 72

4.1 Positioning of chapter 4 in the framework of the yield forecasting method . . . 73
4.2 Annual profiles of local thermal indices from budbreak to veraison for one block . . . 79
4.3 PAST functions and related thermal indices corresponding to different time expressions . . . 85
4.4 Various shapes of the eGDD PAST function . . . 86
4.5 Values taken by the Synchronization Optimization Criterion . . . 88
4.6 Example of the outputs of the eGDD model . . . 90
4.7 Distribution over the 79 blocks of the optimized temperature thresholds values . . . 91
4.8 Distribution over the 79 blocks of the cross-validated Prediction errors . . . 93
4.9 Distribution over the 79 blocks of the cross-validated maximum number of periods that are discretized in time series for each block . . . 94
4.10 Cross-validated Prediction errors for vineyards A,B and C . . . 102
4.11 Cross-validated maximum number of periods that are discretized in time series for vineyards A, B and C . . . 103
4.12 Distribution of the daily mean temperature across the whole history of vineyards A, B and C 104
4.13 Profiles of the PAST function obtained with with a Prediction or Synchronisation use for vineyards A, B and C . . . 105
4.14 Results from the Principal Components Analysis of optimized temperature thresholds by block 106

5.1 Positioning of chapter 5 in the framework of the yield forecasting method . . . 112
5.2 Key phenological stages and the temporal pattern of the grapevine yield determining process over two seasons . . . 114
5.3 Example of BLiSS outputs . . . 117
5.4 Example of a PAST function in the eGDD method framework . . . 122
5.5 Specifications of an analytical framework for time series data . . . 124
5.6 Daily mean temperature data in vineyard A, B and C . . . 126
5.7 PAST functions obtained with the eGDD method for vineyards A, B and C . . . 127
5.8 BLiSS estimation for the synchronized time series of vineyards A, B and C . . . 128

6.1 Procedure for the construction of a yield forecasting method in viticulture for operational purposes and using farm data . . . 134

6.2 Positioning of the scientific contributions of the PhD project to the issues and challenges identified for grape yield forecasting in an operational context. 136
6.3 Preliminary representation of a hierarchical Bayesian model for grape yield forecasting 140

7.1 Positionnement des axes de recherche en fonction de la méthode de prévision du rendement . 146
7.2 Modèle conceptuel de l'élaboration du rendement viticole 148
7.3 Cahier des charges de la méthode de prévision de rendement attendue 149
7.4 Positionnement des axes de recherche traités dans le cadre la thèse 150
7.5 Exemple des sorties de la méthode eGDD pour une parcelle 153
7.6 Erreurs de prédiciton obtenues avec la méthode eGDD pour différents vignobles 154
7.7 Résultats de synchronisation obtenus avec la méthode eGDD 155
7.8 Différents profils de PAST functions 156
7.9 Estimation BLiSS pour un série temporelle synchronisée 157
7.10 Estimation BLiSS pour la série temporelle synchronisée des données de température moyenne quotidienne pour les vignobles A, B et C (respectivement a, b et c) 159

List of Tables

2.1 Most common vineyard yield indicators described in literature . . . 30
2.2 Summary of the main uses cases for grape yield assessment (based on technical conversations) 35
2.3 Operational issues for grape yield components measurement . . . 40
2.4 Methods for yield components measurement reviewed in literature . . . 42
2.5 Main methods for plant ancillary indicators measurement in commercial vineyards . . . 44
2.6 Methods for yield estimation and prediction reviewed in literature . . . 53
2.7 Comparison of operational advantages of yield estimation and prediction methods reviewed in literature . . . 58

3.1 Non-exhaustive listing of the variables and indicators which may compose any farm data set . 67
3.2 Characteristics of the 9 vineyards which shared their data and of their respective data sets. . 67
3.3 Characteristics of the 3 vineyards selected as case studies in terms of environment, organization and data set. . . . 68

4.1 Characteristics of the 3 vineyards selected as case studies in terms of environnement, organization and data set. . . . 76
4.2 Equations and parameters for the computation of new thermal indices . . . 77
4.3 Minimum phenological deviation and corresponding thermal indices for vineyards A, B and C. 77
4.4 Maximum deviation between the dates of flowering and veraison for vineyards A, B and C. . 80
4.5 Optimization criterion values obtained with the eGDD method for the Prediction and Synchronization uses . . . 91
4.6 Characteristics of the three blocks of vineyards A, B and C. . . . 100

5.1 Characteristics of the three vineyards A, B and C. . . . 121
5.2 Results of the time series discretization for vineyard A, B and C. . . . 127
5.3 Site-specific periods of weather influence on yield development . . . 129

7.1 Liste non-exhaustive des variables et indicateurs présents dans les jeux de données agricoles . 147
7.2 Périodes locales d'influence du climat sur le rendement . . . 158

Publications and international conferences

Laurent, C., Baragatti, M., Taylor, J., Scholasch, T., Metay, A., Tisseyre, B., 2019. Evaluation of a functional Bayesian method to analyse time series data in precision viticulture, in: Precision Agriculture '19. Presented at the 12th European Conference on Precision Agriculture, Wageningen Academic Publishers, Montpellier, France, pp. 67–73. https://doi.org/10.3920/978-90-8686-888-9$_7$

Laurent, C., Scholasch, T., Tisseyre, B., Metay, A., 2020. Building new temperature indices for a local understanding of grapevine physiology. Presented at the XIIIth International Terroir Congress, Virtual Congress, Adelaïde, Asutralia, pp. 1–8.

Laurent, C., Oger, B., Taylor, J.A., Scholasch, T., Metay, A., Tisseyre, B., 2021. A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture. European Journal of Agronomy 130, 126339. https://doi.org/10.1016/j.eja.2021.126339

Laurent, C., Rançon, F., Lopez Fornieles, E. , Scholasch, T., Metay, A., Taylor, J., Tisseyre, B., 2021. 18. Is it relevant to account for grapevine phenology in time series of satellite images?, in: Precision Agriculture?21. Wageningen Academic Publishers, pp. 163–169. https://doi.org/10.3920/978-90-8686-916-9$_1$8

Oger, B., **Laurent, C.**,,Vismara, P., Tisseyre, B., 2021. Is the optimal strategy to decide on sampling route always the same from field to field using the same sampling method to estimate yield? OENO One 55, 133–144. https://doi.org/10.20870/oeno-one.2021.55.1.3334

Short abstract

Farm data are data commonly collected in commercial farms in the course of their everyday operation. They offer the advantage of accurately representing the production context of a given farm by providing information on the crop system and management, environment, etc. but they are characterized by lower data quality and a parsimonious number of variables. **This study is based on the hypothesis that valuing farm data with adapted statistical methods can help in performing agronomical research that complies with the operational needs and constraints of a commercial farm and result in the delivery of actionable site-specific information.** This PhD thesis was founded by a precision viticulture company, Fruition Sciences, and therefore aimed at investigating this hypothesis in the viticultural industry with the case study of yield forecasting. The main focus was placed on leveraging time series of weather data because they are commonly encountered in farm data sets and because climate influence on grapevine yield development is prevalent. **All the methods proposed in this PhD project constitute an original framework for valuing farm data with various possible applications in agriculture.**

Keywords

Grapevine, Wine, Estimation, Prediction, Time Series of Weather Data, Yield Development, Bayesian

Résumé court

Les données agricoles sont des données couramment collectées dans les exploitations agricoles au cours de leur fonctionnement quotidien. Elles présentent l'avantage de représenter fidèlement le contexte de production d'une exploitation donnée en fournissant des informations sur le système et la conduite des cultures, l'environnement, etc. mais elles sont caractérisées par une moindre qualité et un nombre parcimonieux de variables. **Cette thèse est basée sur l'hypothèse que la valorisation des données agricoles avec des méthodes statistiques adaptées peut aider à réaliser des recherches agronomiques tenant compte des besoins et contraintes opérationnelles d'une exploitation et aboutir à la livraison d'informations locales et pertinentes pour l'aide à la décision sur le terrain.** Ce projet a été initié par une entreprise de viticulture de précision, Fruition Sciences, et visait donc à étudier cette hypothèse au sein de la filière viticole à travers l'étude de cas de la prévision du rendement. L'accent a été mis sur l'exploitation des séries temporelles de données météorologiques parce qu'elles sont couramment rencontrées dans les jeux de données agricoles et parce que l'influence du climat sur le développement du rendement de la vigne est prédominante. **Toutes les méthodes proposées dans ce projet de thèse constituent un cadre original pour la valorisation des données agricoles avec diverses applications possibles en agriculture.**

Mots clés

Vigne, Vin, Estimation, Prédiction, Séries Temporelles de Données Météorologiques,
Elaboration du Rendement, Bayésien

Abstract

Farm data are data commonly collected in commercial farms in the course of their everyday operation for traceability and management purposes. For example, time series of weather or yield data are commonly collected. The block characteristics and cultural practices are also recorded with varying degrees of precision. These farm data offer the advantage of accurately representing the production context of a given farm by providing information on the cropping system, environment, crop management, logistics etc. They also offer larger data sets than those resulting from research experiments, especially in terms of data histories. However, they are characterized by lower data quality and a parsimonious number of variables. **This study is based on the hypothesis that valuing farm data with adapted statistical methods can help in performing agronomical research that complies with the operational needs and constraints of a commercial farm and result in the delivery of actionable site specific knowledge and/or information.** This PhD project was founded by a precision viticulture company, Fruition Sciences, and therefore aimed at investigating this hypothesis in the viticultural industry. The case study of yield forecasting was chosen because it corresponds to a performance indicator used to decide most operations in the vineyard and winery as well as in the commercial and economic management of various viticultural, winemaking and wine trading organizations. It is also a cornerstone of agronomic and ecophysiological research in viticulture. Nevertheless, the relative contribution of the numerous research proposals to this topic remains unclear and the industry expectations unsatisfied. Thus, a literature review was first undertaken to propose a comprehensive framework for yield estimation, prediction and forecasting methods. It was decomposed into the three processes of measurement, sampling and modelling. Positioning each literature proposal in this framework while evaluating their consideration of operational needs and constraints revealed yet unexplored opportunities for improving grape yield forecasting. These conclusions were translated into the following scientific questions : **how to define a site-specific model of grape yield development which could comply with the operational constraints and expectations and which could be informed by farm data ? how to adapt the required statistical analyses accounting for characteristics of farm data ? then how to develop a grape yield forecasting approach based on these adapted conceptual model and statistical analyses ?** The Phd project aimed at answering these questions from the study of 9 farm data sets among which 3 were selected as case studies.

Climate influence on grapevine yield development is prevalent and weather time series data are commonly encountered in the data sets of commercial vineyards, even when data from other influencing variables might be missing. Therefore, the main focus was placed on leveraging time series of weather data. Their analysis first required the computation of a timeline that was consistent with the vine phenology in order to study the weather influence over several years at comparable stages of the on-going yield development. This work led to **the development of a constrained optimization method called extended Growing Degree days (eGDD method).** It allows the computation of site-specific timelines based on thermal indices. The thermal indices computed with the eGDD method showed better results than the Gregorian calendar and the classical Growing Degree Days approach in predicting the achievement dates of phenological stages and in synchronizing time series in accordance with grapevine phenology. The so-synchronized time series of weather data were then analyzed with a Bayesian functional Linear regression with Sparse Steps functions (Grollemund et al. 2019) and resulted in **the detection of vineyard-specific periods of weather influence.** These meteorological indicators are intended to be integrated as covariates into a yield forecasting model.

Based on the case study of yield forecasting in viticulture, this work provided an example of leveraging farm data to conduct agronomic research which complies with field needs and constraints in order to provide relevant information for decision support in the field. All the methods proposed in this PhD project constitute an original framework for valuing farm data with various possible applications in agriculture.

Résumé

Les données agricoles sont des données couramment collectées dans les exploitations agricoles au cours de leurs activités quotidiennes et à des fins de traçabilité et/ou de gestion. Par exemple, des séries temporelles de données météorologiques ou de rendement sont couramment collectées. Les caractéristiques des parcelles et les pratiques culturales sont également renseignées avec différents degrés de précision. Ces données d'exploitation présentent l'avantage de représenter avec précision le contexte de production d'une exploitation donnée en fournissant des informations sur le système de culture et les pratiques culturales, l'environnement, la logistique, etc. Elles offrent également des jeux de données plus importants que ceux résultant des expériences de recherche, notamment en termes d'historique des données. Cependant, les données agricoles sont aussi caractérisées par une moindre qualité et un nombre de variables réduit. **Cette thèse est basée sur l'hypothèse que la valorisation des données agricoles par des méthodes statistiques adaptées peut aider à conduire des recherches agronomiques appropriées aux besoins et contraintes opérationnels d'une exploitation et générer des connaissances et/ou informations locales pertinentes pour l'aide à la décision sur le terrain.** Ce projet a été initié par une entreprise de viticulture de précision, Fruition Sciences, et vise donc à étudier cette hypothèse au sein de la filière viti-vinicole. L'étude de cas de la prévision du rendement a été privilégiée car elle correspond à un indicateur agronomique utilisé pour décider de la plupart des opérations au vignoble et au chai ainsi que dans la gestion commerciale et économique de diverses organisations viticoles, vinicoles et de négoce. Le rendement constitue également une pierre angulaire de la recherche agronomique et écophysiologique en viticulture. Néanmoins, la contribution relative des nombreuses propositions de recherche sur ce sujet reste floue et les attentes de l'industrie insatisfaites. Ainsi, une revue de la littérature a d'abord été entreprise afin de proposer un cadre global pour les méthodes d'estimation, de prédiction et de prévision du rendement. Ce cadre a été décomposé en trois processus : mesure, échantillonnage et modélisation. Le positionnement de chaque proposition de la littérature dans ce cadre et l'évaluation de leur prise en compte des besoins et contraintes opérationnels a révélé des possibilités encore inexplorées d'amélioration de la prévision du rendement viticole. Ces conclusions ont été traduites dans les questions scientifiques suivantes : **comment définir un modèle de d'élaboration du rendement viticole local et adapté aux contraintes et attentes opérationnelles sur la base de données agricoles ? comment adapter les analyses statistiques requises en tenant compte des caractéristiques des données agricoles ? puis comment développer une approche de prévision du rendement viticole basée sur ce modèle conceptuel et ces analyses statistiques ?** Le projet de doctorat vise à répondre à ces questions à partir de l'étude de 9 jeux de données viticoles, dont 3 ont été sélectionnés comme cas d'étude.

L'influence du climat sur le développement du rendement de la vigne est prédominante et les données de séries temporelles de données météorologiques sont couramment rencontrées dans les jeux de données des domaines en production, même lorsque les données correspondant à d'autres variables d'influence peuvent manquer. L'accent a donc été mis sur l'exploitation des séries temporelles de données météorologiques. Leur analyse a d'abord nécessité une expression du temps cohérente avec la phénologie de la vigne afin d'étudier l'influence de la météo sur plusieurs années à des stades comparables de l'élaboration continue du rendement. Ce travail a conduit au **développement d'une méthode d'optimisation sous contraintes appelée méthode extended Growing Degree Days ou Degrés Jours étendus (méthode eGDD).** Elle permet d'établir un calendrier spécifique à chaque site sur la base d'un indice thermique. Les indices thermiques calculés avec la méthode eGDD ont montré de meilleurs résultats que le calendrier grégorien ou l'approche classique des Degrés Jours pour prédire les dates d'atteinte des stades phénologiques et pour synchroniser des séries temporelles en fonction de la phénologie de la vigne. Les séries temporelles de données météorologiques ainsi synchronisées ont ensuite été analysées à l'aide d'une régression linéaire fonctionnelle bayésienne avec des fonctions en escalier (Grollemund et al. 2019) et ont permis de **détecter des périodes d'influence météorologique spécifiques au vignoble.** Ces indicateurs météorologiques ont vocation à être intégrées à un modèle de prévision du rendement.

Sur la base de l'étude de cas de la prévision du rendement en viticulture, ce travail a fourni un exemple de valorisation des données agricoles pour mener des recherches agronomiques en tenant compte des besoins et contraintes du terrain et en fournissant des informations pertinentes pour l'aide à la décision sur le terrain. Toutes les méthodes proposées dans ce projet de thèse constituent un cadre original pour la valorisation des données agricoles avec diverses applications possibles en agriculture.

Chapter 1

Introduction

Agronomy aims at improving the understanding of agrosystems mechanisms and of supporting the decision making process in the field, by providing relevant information and decision rules at different spatial and temporal scales (Damour, Navas, and Garnier 2018). In the purpose of generating knowledge about the agrosystems, numerous models have been built by integrating the results of research experiments. They can be categorized according to the state variables they involve and their precision in describing the mechanisms they study (Lamanda et al. 2012; Doré et al. 2006). **However, few of these models have been designed to be functional in an operational context i.e. these models have not been designed neither to account for the needs and constraints of commercial farms nor to work from farm data. Therefore, there is a big stake in addressing this issue to support the decision making process in the field. Firstly, the diversity of agrosystems embraced by the operational context exceeds the one studied by the research, which may challenge the validity of the proposed models.** However, this diversity is likely to generate knowledge on yield development in a wide variety of situations (Allan et al. 2013). **Secondly, the data available in commercial farms (farm data) is not adapted to the parameters commonly used in agronomic models. Farm data is characterized by lower data quality than data from research experiments.** This prevents a precise understanding of agronomic phenomena but corresponds to a degree of uncertainty the operational decision making must deal with. **In addition, farm data is often characterized by a parsimonious number of variables.** Indeed, the long cycles (one or two years) and the high workload that characterize agricultural work create a relatively constrained and slow potential to change practices and collect more data. Besides, the current development of many measurement technologies often correspond to ancillary data or are not yet operational in the field (Weiss, Jacob, and Duveiller 2020; Laurent, Oger, et al. 2021; Tardaguila et al. 2021). Therefore, the number of variables likely to be used as parameters of agronomic models is little changing. However, farm data sets often contain long data histories that are assumed not leveraged to their full informational potential by research works that had little access to such histories. **Thirdly, the operational constraints and expectations are often not taken into account in the way the agronomic information is built and delivered.** Therefore, the operational reality calls for a rethinking of the conceptual models and of the statistical methods when agronomic research aims at supporting decision-making more than generating knowledge.

To address this issue, an agronomic research area more oriented towards an operational application has been developed, often in collaboration with private companies. Fruition Sciences is one of these companies. It is based in California, USA and in France. It helps vineyards to understand and react to vintage conditions according to their terroir, by analyzing their own data. The expertise of Fruition Sciences mainly consists in a reading of the season according to the plant water status, which is confronted with other agronomic variables such as vegetative expression or yield to provide viticultural recommendations. Its client vineyards/wineries are located all over the world and correspond to very different operational situations in terms of environment, cropping system, socio-economic background and production objectives. **The present PhD project was hosted by Fruition Sciences with a CIFRE funding and logically developed on the international example of these vineyards. It was led under the hypothesis that doing agronomical research using farm data and complying with operational constraints and expectations could result in the delivery of actionable site-specific information for decision making in the field.** The viticultural industry is an interesting sector to be considered under this hypothesis because it is a relatively conservative sector, for which many decisions are still taken in an

expert way. In contrast, other crops such as wheat, maize or soybean benefit from a large number of models to support their management decisions (Dury et al. 2012; Silva and Giller 2020; Young, Ros, and de Vries 2021). In addition, the perennial character of grapevine has encouraged the constitution of data sets across years for constant blocks. Therefore, it is assumed that the year and site effects in agronomic processes are more easily identifiable in viticultural data sets than in a data set including blocks in rotation.

The PhD project focused on the perspective of yield forecasting in viticulture as a case study. Yield forecast is indeed one of the performance indicators used to decide most operations in the vineyard, in the winery as well as in the commercial and economic management of various viticultural, wine-making and wine trading organizations. It is also a cornerstone of agronomic and ecophysiological research in viticulture. There are numerous recent publications on yield estimation, prediction and forecasting (Nuske, Wilshusen, et al. 2014; Nogueira Júnior et al. 2018; Sirsat et al. 2019; Zhu, Fraysse, et al. 2020; Arab et al. 2021) but few really address the operational context i.e. the needs, constraints and data of commercial vineyards. Hence, the research questions of the PhD project focused on using farm data to provide relevant yield information for operational decision-making. It consisted in identifying and developing a conceptual model of yield elaboration, designing a yield forecasting method, proposing appropriate statistical methods and performing an implementation that could be adapted to the diversity of operational situations in viticulture.

For this purpose, a review of the scientific literature and field knowledge was first performed in order to identify the issues, methods and challenges of grape yield forecasting in an operational context (chapter 2). This review determines the research questions and approach of the PhD project which are also presented in chapter 2. Chapter 3 presents the first results of the PhD project : i) the collected material and its characterization, ii) a conceptual model of yield development in an operational context, iii) the design of the corresponding yield forecasting method and iv) the presentation of the analytical steps to concretely implement the yield forecasting method from farm data sets. These analytical steps are the subject of chapters 4 and 5. Finally, chapter 6 discusses the contribution and perspectives of the PhD work to yield forecasting in viticulture but also to valuing farm data for agronomic research and operational purposes in a wider perspective.

Chapter 2

Bibliography and presentation of the research strategy for the PhD project

2.1 Intention note

The case study of the PhD project addresses grape yield forecasting at the block level in an operational context. Many proposals have been made in the scientific literature to improve yield estimation, prediction and forecasting. In particular, each technological development such as remote sensing and derived vegetation indices (Sun et al. 2017; Arab et al. 2021), image capture and analysis (Nuske, Wilshusen, et al. 2014; Di Gennaro and Matese 2020), or artificial intelligence methods (Sirsat et al. 2019) has been followed by a chorus of proposals for use in yield assessment. The relative contribution of each of these proposals remains unclear considering the whole yield development process. It is even more unclear because these proposals don't necessarily address the same definition of grape yield and because numerous terms such as yield estimation, prediction or forecast are used without being precisely defined. In this manuscript, the term yield assessment is used to refer to all methods that aim at providing some kind of yield information. However, despite all these proposals, the research results in yield assessment are still below the industry's expectations. Thus, **a first working hypothesis is that reviewing the process of yield development in an operational context could help in determining if the industry expectations are reasonable and which challenges need to be addressed and considered to better capture the drivers of yield development.** Such a review should address the agronomic process of yield development but also the cultural practices which come with it, whose motivations may be agronomic but also linked to habits or logistic constraints. Then, **a second working hypothesis is that developing an understanding framework for the entire yield assessment process could help in appreciating the relative contributions of the literature and hence identifying as yet unexplored research questions which show potential for improving yield assessment.**

Moreover, only a few literature studies have considered the adaptation of yield assessment approaches to the operational context. As a result, yield assessment methods are often adopted by the industry on the basis of scientific work although these methods have not necessarily been designed to be effective or even valid in such operational conditions. Thus, a third working hypothesis was that reviewing operational constraints and expectations regarding yield assessment could help in identifying scientific or technical limitations and challenges. The design of yield assessing methods could then be consequently improved for increased relevance in commercial vineyards and wineries.

Under these three hypotheses, a literature review was performed in the section 2.2 of this chapter. Hence, the conclusions of the literature review were translated into scientific questions in section 2.3. Finally, the approach used along the PhD project to address the scientific questions is presented in section 2.4.

2.2 A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture

2.2.1 Details about the paper

2.2.1.1 Title and publication informations

This section was published as a literature review in the European Journal of Agronomy in June 2021. The corresponding article was entitled "A review of the issues, methods and perspectives for yield estimation, prediction and forecasting in viticulture".

2.2.1.2 Authors

C. Laurent[1,2,3,*], B. Oger[4], J. Taylor[3], T. Scholasch[1], A. Metay[2], B. Tisseyre[3]

[1] Fruition Sciences, 34000 Montpellier, France
[2] ABSYS, Univ. Montpellier, CIRAD, INRAE, Institut Agro, 34060 Montpellier, France
[3] ITAP, Univ Montpellier, INRAE, Institut Agro, 34060 Montpellier, France
[4] MISTEA, Univ Montpellier, Institut Agro, INRAE, 2 Place Pierre Viala, 34060 Montpellier, France
[*] corresponding author : cecile@fruitionsciences.com

2.2.1.3 Abstract

Grapevine yield is defined as the quantity of harvest, expressed as either grape mass or wine volume units, which has been collected per surface unit are and per crop cycle. The information about current and future yield, termed a yield assessment in this paper, is an essential decision-making element for the grape and wine industry. Crop management, wine-making, commercial, accounting and strategic operations are all adapted to and all impact on the expected yield of the current vintage. Numerous yield assessment approaches have been proposed in the scientific literature. However, only a few of them have considered their adaptation to the operational context under the constraints, needs and strategies of commercial vineyards and wineries. The few studies that have worked on the operational implementation of yield assessment methods have only partially addressed this issue, concentrating their improvement efforts on a single step in the yield assessment process. This paper first proposes to review the characteristics of yield development in an operational context that must be taken into account by yield assessment methods. These characteristics are consolidated into three main challenges for yield assessment methods: (i) addressing the complex temporality of yield development, (ii) ensuring a local monitoring of yield development and (iii) fitting to the operational needs and constraints to allow for relevant decision support systems in the field. The approaches of yield estimation, prediction and forecast are discussed in the context of these challenges. In a second step, the paper proposes a generic framework for the yield assessment process, including a review of the variables that are used to explain grapevine yield. Issues and proposals from the literature associated respectively with measurement, sampling and yield modelling are reviewed and the need for improved modelling of relationships between explanatory variables and the desired, reported yield variable is discussed. The yield assessment methods found in the literature are categorized and compared according to measurement, estimation and modelling approaches, and then according to the three challenges identified for yield assessment in operational conditions, such that the yield assessment method is adapted to commercial needs and not to research objectives. In conclusion, concrete proposals for new grape yield assessment methods are discussed in order to investigate the as yet unexplored opportunities for the improvement in yield assessment in operational contexts that have been identified in the paper. These considerations could easily be transposed to other perennial crops.

2.2.1.4 Highlights

- **Operational grapevine yield development includes vineyard/winery specific operations.**
- **Operational needs drive the challenges that yield reporting methods have to meet.**
- **Yield reporting methods are related to measurement, sampling and modelling issues.**
- **Yield estimation, prediction and forecast address different levels of uncertainty.**
- **Yield reporting methods should ensure a temporal, local, operational yield monitoring.**

2.2.1.5 Keywords

Measurement, Sampling, Yield Models, Operational, Wine, Climate, Uncertainty

2.2.2 Introduction

For the wine industry, yield is agronomically defined as the quantity of harvest, either expressed in grape mass or wine volume units, that has been collected per surface unit area and per crop cycle. Since the introduction of wine regulations at the beginning of the 20th century, grape and wine production has been seen as a trade-off between harvest quantity, i.e. yield, and quality (Ravaz and Sicard 1911; Champagnol 1984; Guilpart, Metay, and Gary 2014). However, this trade-off is not bijective i.e. a given harvest quality does not imply a unique yield but can exist across a range of possible yields (Tardáguila et al. 2008; Intrigliolo and Castel 2009; McClymont et al. 2012; Martínez et al. 2016). Therefore, grape yield can be optimized for a given harvest quality by applying appropriate technical operations throughout the production chain. To this end, decisions on operations, both in the vineyard and in the winery, are based on an expected final yield and expected growing conditions from the start of the season. Numerous approaches to report the expected final yield have been proposed in the scientific literature (Clingeleffer et al. 2001; Diago, Correa, et al. 2012; Cunha, Ribeiro, and Abreu 2016; Nogueira Júnior et al. 2018; Sirsat et al. 2019; Zhu, Fraysse, et al. 2020). In this paper, they are referred to as yield assessment methods when considered as a whole. Most of these studies are conducted in a context of research experiments aimed at statistically linking total yield to a yield component (Diago, Correa, et al. 2012; Lopes et al. 2016; Cunha, Ribeiro, and Abreu 2016), another plant-related variable (Cunha, Marçal, and Silva 2010; González-Flor et al. 2014; Sun et al. 2017) or environmental variables (Nogueira Júnior et al. 2018; Sirsat et al. 2019; Zhu, Fraysse, et al. 2020).

However, there are only a few studies that consider the adaptation of yield assessment approaches to the operational context. As a result, yield assessment methods are often adopted by the wine sector on the basis of scientific work although they have not necessarily been defined to be effective or even valid in such operational conditions. The operational context includes additional needs and constraints to be met to ensure the smooth running of the production chain of any commercial wine-growing structure. The operational context also generates data that differs from data collected for a research experiment. Operational data, e.g. weather data or field observations, is collected throughout the season for immediate decision-making purposes but it is usually not intended to support any statistical analysis. Therefore, there is an issue of whether or not these scientific studies have properly accounted for operational needs, constraints and capabilities in terms of data acquisition in order to enable wine sector professionals to rigorously apply the methods proposed in the scientific literature in a production context.

Such a question is also of real interest from a scientific research perspective since adapting to production conditions requires reporting on a wide variety of situations for grape yield development, which is likely to generate knowledge on this subject. In that respect, operational datasets often offer larger amounts of data, particularly in terms of time series, which can be used to improve yield modelling by supporting novel statistical approaches. However, the development of operational methods presents scientific challenges related to the quality, heterogeneity and low number of site specific data to be taken into account. It is also dependent on the definition of indicators to be used, which in turn depends on the working habits of each vineyard/winery. Finally, yield assessment methods constitute both a technical and social issue for the wine industry and it is the role of scientific research to address both. In particular, the development of more relevant reporting methods should encourage their adoption by the wine industry, promoting a virtuous approach to developing agri-services that collect data for their own improvement and to support further research studies.

The few studies that have adapted yield assessment to operational conditions have often focused on improving only one step of the yield assessment process. For example, some studies have attempted to improve measurement issues by working on the automation of yield components counting (Aquino, Millan, Diago, et al. 2018; Aquino, Millan, Gutiérrez, et al. 2015; Liu, Zeng, and Whitty 2020) or total yield weighing

(Tarara et al. 2014; Lopes et al. 2016). Others studies have sought to optimize sampling strategies (Araya-Alman, Leroux, et al. 2019; Oger, Vismara, and Tisseyre 2021). Although relevant, these studies remain limited in the way that they respond to operational issues since they aim to improve only one constitutive step of the yield assessment process. As a consequence, the yield assessment methods that are currently available to the industry have significant limitations, such as a high degree of imprecision and an inability to characterize the uncertainty associated with the yield assessment. Moreover, the extent of these limitations is difficult to quantify in regard to what can be reasonably expected under operational conditions as no single yield assessment approach has dealt with the problem in its entirety.

Any proposal for an operational yield assessment method can not be based solely on the optimization of any single step of the yield assessment process. Instead, it must collectively assess issues associated with measurement and sampling approaches for both the estimation of explanatory variables and of the yield response to be explained, as well as modelling issues for the development of a yield assessment model. New methods will also require an analytical approach that considers the entire yield development process in relation to the operational needs and constraints resulting from the production context. However, there is no holistic synthesis of the existing literature on yield assessment methods to help to identify the key research and industry questions that remain to be addressed in order to achieve a robust, accurate method of yield assessment in commercial vineyards. To address this deficit, this paper first provides an overview of the yield development process under operational conditions and a summary of the subsequent operational needs and constraints related to yield assessment to identify the challenges to be accordingly addressed. Secondly, a knowledge framework of yield assessment methods is proposed. It is framed in terms of measurement, sampling and modelling approaches for a yield assessment purpose. For each of these three topics, issues and literature propositions are presented. Finally, the yield assessment methods proposed in the literature are reviewed with regard to their characteristics in terms of measurement, sampling and modelling and to the challenges identified in the first part. In conclusion, concrete proposals for a new grape yield assessment method are discussed. These considerations are primarily aimed at the production of wine grapes, which constitutes the vast majority of published literature, but could easily be transposed to the production of table grapes, juice grapes or potentially to other perennial crops. As the terminology and nuances around yield assessment are often complex and diverse, a series of definitions for the terms used in the paper are also provided.

Definitions

GRAPE (ACTUAL) YIELD: quantity of harvest that is effectively reached, expressed in mass (kg or t) or volume units (L or hL) per plant or surface unit area (ha or a)

INPUT INDICATOR: a variable that influences yield development without being reciprocally influenced by it

MEASUREMENT: observation in the vineyard or in the winery, may be performed with or without the help of instrumentation

MEASURAND: real value of a particular quantity to be estimated

MODELLING: establishing a statistical relationship between explanatory variables and a response variable to be explained (here, grapevine yield)

OPERATIONAL CONDITIONS/DATA: referring to the conditions/data of a commercial vineyard or winery, as opposed to research ones. This often implies that not all conditions are equal because different choices can be made from one year to the next or from one block to the next for agronomic, logistical and human reasons.

OUTPUT INDICATOR: a variable that stands as an outcome of yield development, mainly referring to yield components

SAMPLING: choice of measurement sites (spatial sampling) and dates (temporal sampling)

SITE-SPECIFIC: including effects of the environment (soil, climate, topography etc.), cultural practices, operational constraints, needs and strategies in particular the qualitative orientation of the production

SYSTEMIC INDICATOR: a variable that influences yield development and that is reciprocally influenced by it

VINEYARD: refers to both grapevine blocks and the company that cultivates it, as understood in vineyard estate

WINERY: refers to both the cellar in which the operations of wine-making take place and the company that actually produces wine and commercializes it. N.B.: sometimes the terms vineyard and winery refer to the same enterprise

YIELD ASSESSMENT: any kind of yield information, including yield estimation, prediction and forecast without distinction

YIELD ASSESSMENT UNCERTAINTY: a distribution of values attributable to the real yield performance once the measurement, sampling and modelling steps have been performed

YIELD COMPONENTS: grapevine reproductive anatomical structures that are successively settled during the vineyard part of a yield development cycle

YIELD DEVELOPMENT CYCLE: overall process of grape or wine production, which includes different stages depending on the enterprise and target markets. N.B. this is not limited to the vineyard and may be extended into the winery

YIELD ESTIMATE: yield assessment made in the same unit, time and space than the measurement and sampling processes it results from

YIELD FORECAST: yield assessment made in different units, time or space than the measurement and sampling processes it results from. It is most commonly associated with a yield performance that is expected to be reached in the same space but at a future date and consequently expressed in different units. A yield forecast corresponds to a statistical distribution computed on the basis of the training dataset and a priori knowledge.

YIELD PREDICTION: yield assessment made in different units, time or space than the measurement and sampling processes it results from. It is most commonly associated with a yield performance that is expected to be reached in the same space but at a future date and consequently expressed in different units. A yield prediction corresponds to a statistical expectation (single value) computed on the basis of the training dataset.

2.2.3 What are the main issues in assessing grapevine yield ?

2.2.3.1 Contextual information about yield development

Grape actual yield (grape yield) corresponds to the grape quantity that is effectively reached in the field. It is the dynamic output of a yield development cycle, which includes effects in season n-1 and ends with the harvest in season n (Howell 2001; Clingeleffer et al. 2001; Carmona et al. 2008; Vasconcelos et al. 2009; Guilpart, Metay, and Gary 2014) . Grape yield may be considered during the season while it is still developing but in most cases, it is considered close to harvest. Grape yield is often expressed in mass units per stock (Guilpart, Metay, and Gary 2014; Nogueira Júnior et al. 2018) or per area units (De la Fuente et al. 2015; Araya-Alman, Leroux, et al. 2019).

However, in a commercial context, each production organization (vineyard or winery) has its own yield definition, depending on when it considers yield performance as being final and the operations it includes in the definition of the yield development cycle. In some cases, final yield is considered beyond the harvest operation and up until grapes or wine are marketed, as suggested by the definition of marketable yield for vegetable crops (Vigiuer et al. 2018). In such a commercial case, the reproductive cycle not only refers to the production of grapes but also of wine. In such cases, yield may be expressed in volume units per block, groups of blocks when wine from different blocks is blended, vineyard or a whole production area. For example, one vineyard may directly weigh harvest baskets in the vineyard while another winery estimates the volume of must after pressing or wine after bottling. Similarly, for table grapes or raisins (dried fruits), the definition of a yield development cycle may include transformation and marketing operations, such as eventual drying, storage or transport.

2.2.3.1.1 Yield is gradually established at the plant scale over two seasons

The grapevine yield development process is considered to start in the latent buds established in the year n-1. The general structure of a grapevine latent bud is described in Fig. 2.1. These latent buds generally include three structures, or buds, that follow the same organizational pattern, but are in different states of progress (Pratt 1971). Two of the structures, termed secondary buds, will only develop in season n if the primary bud is damaged (Vasconcelos et al. 2009). Within each of the three buds, a shoot apical meristem (SAM, Fig. 2.1) initiates foliar primordia and lateral meristems. Lateral meristems are also called anlage or uncommitted primordium and may later result either in a tendril or an inflorescence. The physiological process of inflorescence induction corresponds to the modification of gene expression that alters the balance of endogenous hormones in response to environmental stimuli (Boss et al. 2003; Vasconcelos et al. 2009; Li-Mallet, Rabot, and Geny 2016). The subsequent floral initiation is characterized by repeated branching in the lateral meristems, which promotes immature inflorescences. Depending on the experimental conditions, the floral initiation seems to occur approximately four to seven weeks after budbreak in season n-1 (Vasconcelos et al. 2009). Once there are one to four inflorescences established, the latent bud may enter into different levels of dormancy (Lavee and May 1997; May 2004; Jones, Menary, and Wilson 2009). Tourmeau 1976 further suggested that some of the inflorescence primordia may also be implemented prior to the floral differentiation in season n.

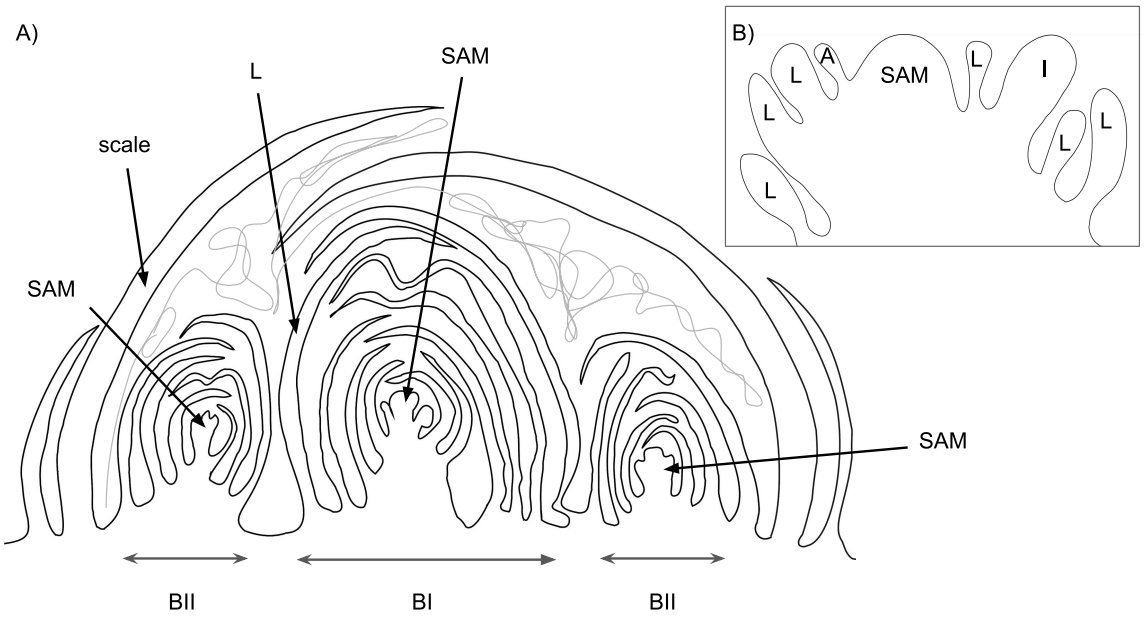

Figure 2.1 – Outline of a longitudinal section of A) a grapevine latent bud (x10) and B) the upper zone of a grapevine primary bud (x100)

with BI : primary bud, BII : secondary bud, SAM : shoot apical meristem, L : leaf primordium, I: inflorescence primordium, A : anlage

At the beginning of season n, only a portion of the latent buds are left after pruning. Shortly before and during the budbreak of season n, inflorescence primordia differentiate into flowers (Srinivasan and Mullins 1981; May 2000; Vasconcelos et al. 2009) . Inflorescence architecture has been described by (May 2004) and Meneghetti, Gardiman, and Calo 2006. During flowering, cross-pollination by insects or wind has mainly been reported (Pratt 1971). A few days are required for the completion of the pollination and fertilization stages that lead to the fruit set (Vasconcelos et al. 2009). Berries develop for approximately a hundred days after the flowering period in a double sigmoid pattern with two phases of active growth separated by a latency phase (Ollat et al. 2002; Bigard et al. 2019). After fertilization, the herbaceous phase corresponds to an accumulation of water (mainly via the xylem) and various assimilates, including malic and tartaric acids (via xylem and phloem), as well as to seed formation(Ollat et al. 2002). During this stage, operations of bunch thinning, foliar fertilization and irrigation may be undertaken using different strategies to promote berry development. Veraison is commonly detected by a change in berry colour but it seems to be more accurately approximated by the observation of berry softening (Bigard et al. 2019). This stage marks the turn in berry metabolism. From then on, phloem unloading of sugars and polyphenols increases while xylem water supply is progressively stopped. During the ripening phase, sugar accumulation results in a second berry growth (Keller, Zhang, et al. 2015). Part of the malic acid is also metabolized (Ollat et al. 2002). The phloem flow then progressively stops and a decrease in berry volume may be observed due to dehydration in relation to microclimatic conditions rather than plant water status(Keller, Zhang, et al. 2015; Gambetta et al. 2020).This last stage is called the berry concentration phase and may last until the harvest is triggered.

2.2.3.1.2 Yield development is monitored at the vineyard scale over the second season through the observation of already established yield components

Earlier stages of the yield development process occur in the bud at a cellular level and are not observable without specialized laboratory equipment. It is not until after budbreak of season n that producers can observe most of the successive reproductive structures, e.g. flowers, bunches and berries. These so-called yield components are used as indicators of on-going reproductive development (Pagay and Collins, 2017). Table 2.1 shows a list of the main successive reproductive development stages post-budbreak and counts of reproductive organs that could be performed at each stage. These counts are often used to establish

a percentage of the successful or deficient completion of the reproductive stages. Therefore, they are also considered as indicators of the on-going yield development. For example, the counting of flowers that have turned into berries compared to the total number of flowers is used to establish the fruit-set rate. In some stages, counterpart phenomena, associated with a loss of yield, can also be observed and these are also mentioned in Table 2.1.

Few indicators are reported in the literature to monitor the post-harvest yield development operations in a commercial context. However, any estimate of the total yield may be considered as an indicator of the final total yield as expressed according to the winery practices. For example, the number of harvest baskets or containers may be an indicator of a final yield that would be captured after bottling.

Table 2.1 – Most common vineyard yield indicators described in literature

Main reproductive development stage	**Counter-part phenomenon**	**Indicators**	**References**
Inflorescence induction		Bud fertility (number of primary inflorescences per latent bud)	May 2000
Inflorescence evocation	Bunch necrosis	Fertility index (number of inflorescences per cane)	Collins, Coles, et al. 2006
Inflorescence differentiation		Number of inflorescences per vine stock	Guilpart, Metay, and Gary 2014
Floral differentiation	-	Number of flowers per inflorescence	May 2000; Gourieroux et al. 2016
Bloom Pollination Fertilization Fruit set	Millerandage Coulure	Pollen concentration Number of berries per bunch	May 2000 Collins and Dry 2009 Guilpart, Metay, and Gary 2014 Baby et al. 2015 Cunha, Ribeiro, and Abreu 2016
Berry growth	Water and sugar accumulation disorder Berry shrivel Berry shrinkage	Berry mass	Dokoozlian and Kliewer 1996 May 2000 Tilbrook and Tyerman 2009 Guilpart, Metay, and Gary 2014 Crespo-Martínez et al. 2019

The yield components in Table 2.1 are commonly associated within a formula of the same type as Eq. 2.1 to provide an assessment of grape yield in units of mass at the plant or block scale (Dry 2000; Clingeleffer et al. 2001) by decomposing actual yield into its constituent components. Please note that some of the components of Eq. 2.1 are fixed (vines/field) while others are sequentially set during the reproductive cycle and will be dependent on effects in previous stages. As it relies on berry mass at harvest, Eq. 2.1 is a theoretical equation that can only be fully derived retrospectively (post-harvest). Of course, growers and winemakers need information on yield potential in-season, not post-harvest, therefore historical averages or surrogate observations can be substituted into Eq. 2.1 for early or mid-season assessments of yield. For example, the number of inflorescences at bloom can be used instead of final bunch counts. The implications of the different timing of each yield component implementation are not considered in this section but are discussed later in the paper.

$$\text{Actual grape yield (mass unit)} = \frac{\text{N°of vines}}{\text{field}} * \frac{\text{N° of buds}}{\text{vine}} * \frac{\text{N°of bunches}}{\text{bud}} * \frac{\text{N° of berries}}{\text{bunch}} * \text{berry mass (mass unit)} \tag{2.1}$$

2.2.3.1.3 Yield development depends on internal factors

Internal factors mainly refer to genetic and molecular factors associated with the physiological processes from induction to growth and maturation of all yield components (Boss et al. 2003; Carmona et al. 2008; Houel et al. 2015). Their effects are common to all species in the Vitis genus but are sometimes nuanced across varieties (Boursiquot, Dessup, and Rennes 1995; Dry et al. 2010; Martínez-Zapater et al. 2019; Ibáñez et al. 2019). For instance, such differences may apply to the anatomy of yield components (e.g. bunch and berry size, bunch compactness) or to the plant response to environmental factors (e.g. heat and water requirements, sensitivity to some diseases). These characteristics are settled by the choice of planted grape variety and rootstock.

2.2.3.1.4 Yield development is interacting with other grapevine physiological developments

Since the grape yield (reproductive) development cycle occurs over two-seasons, but vegetative development only occurs within a season, there are periods when one vegetative cycle and two reproductive cycles are occurring simultaneously within the grapevine. This leads to inter-cycle dependencies as a result of nutrient and water partitioning over the season (Petrie, Trought, and Howell 2000; Bates, Dunst, and Joy 2002; Zapata, Deléens, et al. 2004; Zhu, Génard, et al. 2018). Moreover, within each development cycle, a competition for nutrients and water is also observed between organs of the same plant, which are implemented either simultaneously or successively (Keller, Tarara, and Mills 2010; Carrillo et al. 2015; Poni and Gatti 2017). Therefore, the yield development cycle process is constantly integrating inter- and intra-cycle dependencies. By showing the temporal overlapping of different cycles in grapevine physiology, Fig. 2.2 highlights the cycles that may be interdependent.

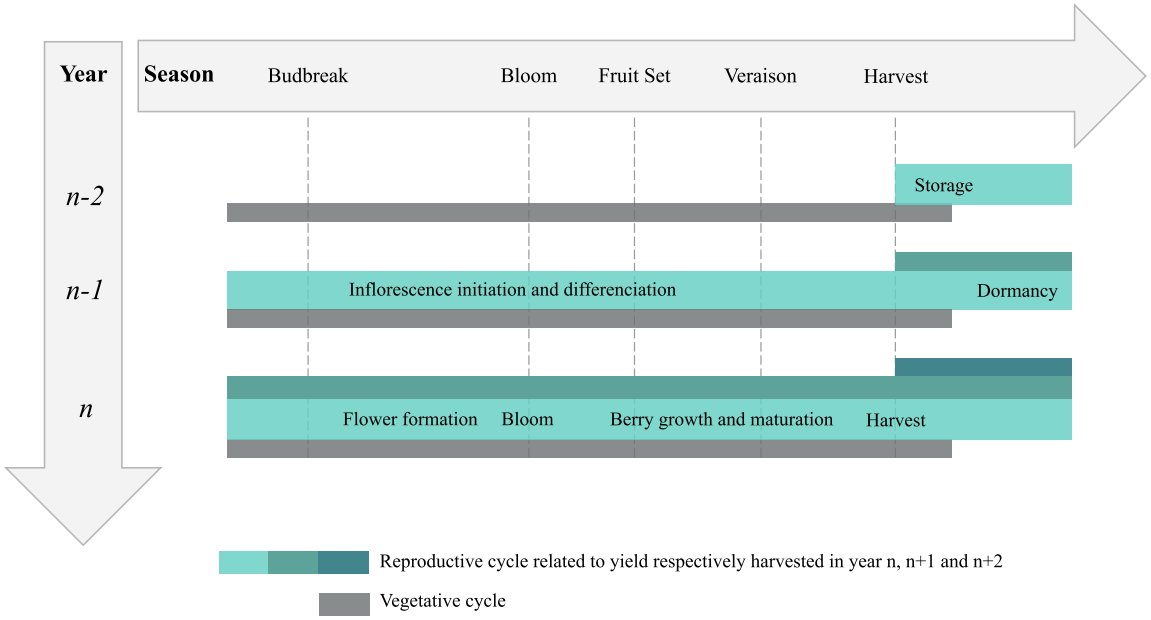

Figure 2.2 – Overlapping successive reproductive and vegetative cycles that compose grapevine physiology throughout the on-going season

2.2.3.1.5 Yield development is interacting with various external factors

External factors include environmental influences, interactions with neighboring plants and cultural practices. Environmental influences may be biotic or abiotic. Abiotic environmental influences mainly refer to resource availability in relation to climatic conditions (e.g. temperature, light, rain, relative humidity and wind) and soil properties (Coipel et al. 2006; Van Leeuwen, Roby, and De Rességuier 2018). Biotic environmental influences correspond to disease or pest development (Valdés-Gómez et al. 2009; Leroy et al. 2013; Guilpart, Roux, et al. 2017; Ouadi et al. 2019). Inter-plant interactions may refer to intra-species interactions with neighboring vines or inter-species interactions with cover and inter-row crops or under-vine weeds. They are mainly described in terms of competition for light, water and nutrients (Champagnol 1984; Garcia et al.

2018; Van Leeuwen, Pieri, et al. 2019). Cultural practices are relative to the vineyard establishment i.e. vine density and spacings (Champagnol 1984; Van Leeuwen, Pieri, et al. 2019)., training system and canopy manipulation (Duchêne, Jaegli, Salber, and Gaudillère 2003; Duchêne, Jaegli, and Salber 2003; Reynolds and Vanden Heuvel 2009; Poni and Gatti 2017), pruning and fruit thinning (Naor, Gal, and Bravdo 2002; Keller, Mills, et al. 2005; Reynolds and Vanden Heuvel 2009), soil preparation (Ripoche et al. 2011; Guerra and Steenwerth 2012), cover cropping (Celette and Gary 2013; Garcia et al. 2018), fertilization (Metay et al. 2015) and irrigation (Intrigliolo and Castel, 2009; Scholasch and Rienth, 2019). They may affect yield by modulating abiotic and biotic environmental factors as well as inter-vine competition.

2.2.3.2 Subsequent issues for yield assessment

2.2.3.2.1 The yield development system is often simplified to allow the identification of indicators

Grapevine yield development follows a fixed internal determinism, due to the selected plant material, that is modulated by interacting external factors and physiological intra- and inter-cycle dependencies. This is why grape actual yield is defined as the quantity of harvest that is reached in the field when limiting and reducing external influences have interfered with potential yield (Van Ittersum et al. 2013; Savary et al. 2018). Thus, the number of possible combinations of all these factors of variation is infinite and grapevine yield development may be considered as being specific to each site and year. This statement is further amplified if extended to the commercial context, for which the definition of a yield development cycle varies depending on the vineyard/winery and sometimes includes post-harvest factors.

To study such a complex system, many scientific studies simplified it to the effects of factors initiating external complex influences. These input factors primarily influence yield development without any reciprocal influence. They may also influence other external factors of variation but again, without reciprocal influence (Lamanda et al. 2012). For example, weather variables are considered as input factors because they are not influenced back by yield development. In contrast, biotic aggressions are both influenced by climatic conditions and yield development. Therefore, they are not considered as input factors, but as part of the studied yield development system. Under this definition, the yield development system includes plant internal and physiological factors as well as external factors that are in interaction with yield development or other external factors that interact with each other to influence yield development. The yield development system is considered as a black box by numerous literature works that study the effects of input factors on output variables, which are yield or yield components (Dunn and Martin 2000; Ojeda, Deloire, and Carbonneau 2001; Petrie and Clingeleffer 2005; Sánchez and Dokoozlian 2005; Keller, Tarara, and Mills 2010; Greer and Weston 2010; Pagay and Collins 2017).

2.2.3.2.2 The timing for the observation of indicators should be driven by grapevine phenology

The influence of an external factor on yield development depends on i) the factor considered (e.g temperature, mildew pressure, pruning etc.) ii) the date of occurrence of the influence in relation to the grapevine development stage (e.g. temperature effects around budbreak or bloom), iii) the intensity of the influence (e.g low or high temperatures) and iv) the duration of the influence (e.g. short or long exposure). Therefore, many scientific studies have decomposed the problem by focusing on the effects on yield components yield of a specific input factor (e.g. high temperatures for a constant duration) applied at different dates (Buttrose 1974; Dokoozlian 2000; Dokoozlian and Kliewer 1996; Dunn and Martin 2000; Ebadi, May, and Coombe 1996; Gouot et al. 2019a; Gouot et al. 2019b; Greer and Weston 2010; Guilpart, Metay, and Gary 2014; Jones, Menary, and Wilson 2009; Keller, Tarara, and Mills 2010; Matthews and Anderson 1989; Ojeda, Deloire, and Carbonneau 2001; Pagay and Collins 2017; Petrie and Clingeleffer 2005; Pouget 1981; Sánchez and Dokoozlian 2005; Triolo et al. 2019). Brought together, these works allow for the identification of a broad timeline for yield development which is presented in Fig. 2.3.

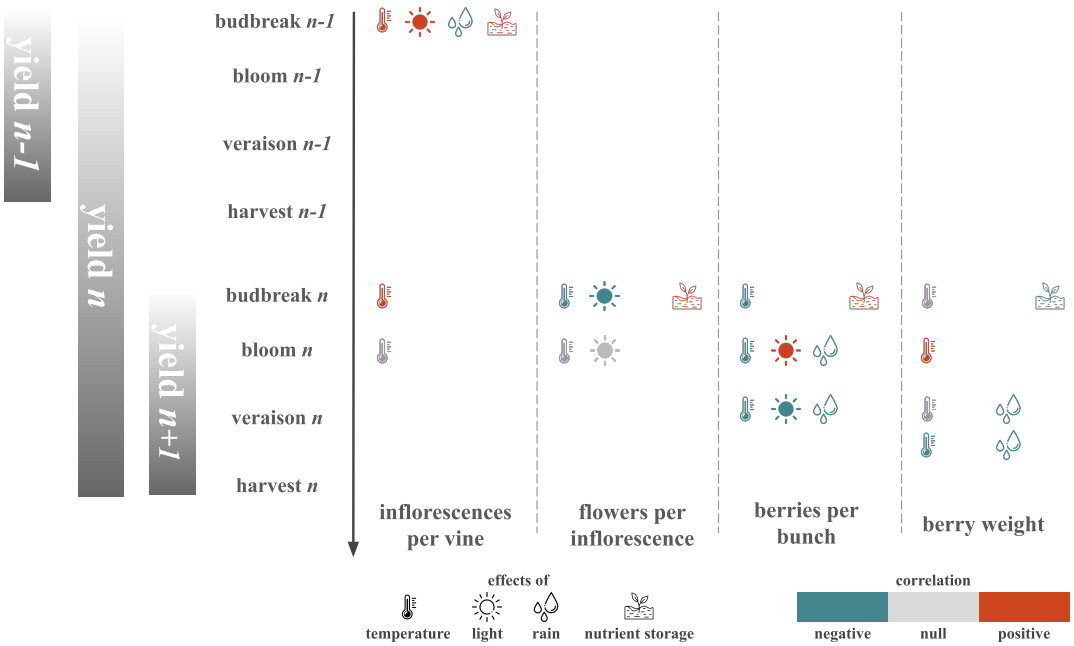

Figure 2.3 – Influences of temperature, light, water status and nutrition storage from season n-1 to n on yield components observed in season n that have been reviewed in literature.

A null correlation means that the relationship has been shown to be absent by at least one study whereas an empty cell mean that no study has been conducted on the considered influence.

In this timeline (Fig. 2.3), key steps may be identified as the moment when a yield component is settled and becomes fixed for the remainder of the season. Hence, there appears to be optimal periods during the growing season for capturing stable information on yield variability. For example, the number of bunches per vine does not evolve after bloom in year n and can be definitively counted. Therefore, to optimize field observations and to consolidate yield assessment reliability, it seems relevant to perform observations and yield modelling on a few key dates in the season. These dates will be driven by the grapevine phenology in a given season and will not be inter-annually fixed dates.

2.2.3.2.3 Trajectory effects of external influences should be taken into account by yield indicators

Due to physiological interdependencies, yield development is a dynamic process whose progress depends on the external influence it is currently experiencing but also on the previous influences it has integrated. The past influences may modulate the effect of the current influence through both their individual realization and succession. Such trajectory effects are complex to investigate but have been suggested from multi-year studies (Duchêne, Jaegli, Salber, and Gaudillère 2003; Duchêne, Jaegli, and Salber 2003; Sadras, Moran, and Petrie 2017; Vaillant-Gaveau et al. 2014; Guilpart, Metay, and Gary 2014; Guilpart, Roux, et al. 2017; Netzer et al. 2019). This means that the information contained in time series of data could be explicative of yield development by recording trajectory effects. However, most literature works have focused on punctual indicators based on a few phenological stages or time steps that are often considered independent when analysed with classical methods, such as linear regression (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017).

2.2.3.2.4 Yield development should be considered site-specifically and over time

Because of trajectory effects of external influences on yield development and particularly on the implementation of each yield component, the decomposition of yield variability (cf. Eq. 2.1) will differ across sites and years. Moreover, for the same block, the proportion of yield variability explained by the implementation of each component will also vary depending on whether the decomposition of yield is considered spatially or

temporally. For instance, using a regularly distributed sampling design on an intra-field scale (nine vineyard blocks in one season), Carrillo et al. 2015 showed that the number of bunches per vine spatially explained 60% of average field yield variability, while the number of berries per bunch, the berry mass and the interaction between the number of bunches per plant and the number of berries per bunch respectively explained 11%, 4% and 20% of the spatial yield variability. Clingeleffer et al. 2001 also reported that the number of bunches per vine explained 61% of the spatial yield variability along a 130 m long vineyard row in Australia. However, when separately studying several fields for 7 seasons, Clingeleffer et al. 2001 found that the role of bunches per vine was more variable when explaining temporal yield variability, explaining from 39% to 99% of the temporal yield variance. Thus, the classical rule of thumb, according to which the number of bunches per vine, the number of berries per bunch and the mass of a berry respectively explain 60%, 30% and 10% of yield variability, should not be taken as granted in any situation. Instead, it should be site-specifically checked with a temporal, not spatial, analysis of yield variability if done for the purpose of a classical yield assessment i.e. from year to year.

2.2.3.2.5 Site-specific study of yield development implies analyzing operational data and managing uncertainty

Site-specific considerations of yield development often implies analyzing operational data. Operational data corresponds to data that has been collected on the considered commercial vineyard according to its own needs, constraints and habits. It is characterized by a strong heterogeneity and parsimony in terms of the indicators available for analysis (choice of input, systemic or output variables) and of the data collection protocol. Moreover, these operational data inevitably contain noise associated with data collection as well as errors on data traceability and management in the daily operational functioning of the vineyard/winery. Differential or variable management of operations that influence yield development may vary, generating mismatches or overlaps that make data traceability a real issue for a comprehensive analysis. For example, grapes from two blocks that underwent different bunch thinning may be picked together and then assembled in different proportions into two tanks. A vineyard operation may also be carried out imprecisely in time and/or space, which then causes a subsequent heterogeneous influence on yield development that is not captured, or is unable to be captured, in the yield assessment.

2.2.3.3 Operational expectations about yield assessment

2.2.3.3.1 Operational use cases require a yield assessment at different dates, space and units

At the vineyard and winery scale, for each yield development cycle, a yield assessment is needed to support decision making associated with cultural practices, harvest, wine-making logistics, commercialization and managing inputs and outputs for accounting. On a supply area or territorial scale, yield assessment is also an important decision-making support for trade purposes. In the longer-term, yield assessment may be used for strategic purposes, either at the vineyard or winery scale or even larger (label area, supply area, etc.). Table 2.2 summarizes the main expectations related to these different use cases according to three criteria: i) what definition is used for a yield development cycle in the considered vineyard/winery i.e. at which stage of the production is yield considered as final and how is it measured? ii) the date at which the yield assessment information is required to support decision making and iii) the spatial scale at which the yield assessment information is needed to support decision making. Information presented in Table 2.2 is based on the consolidated interpretation of numerous conversations that the authors held in the field with producers and experts, both in France and abroad.

Table 2.2 – Summary of the main uses cases for grape yield assessment (based on technical conversations)

The expected date, spatial scale and unit respectively refer to the date, the spatial scale and unit to assess grape yield that are wished by the industry. The associated operational decisions refer to the operations of the vineyard/winery whose decision is based on a yield assessment. The expected benefits refer to the reason for using a yield assessment to decide on such operations.

Expected date for a yield assessment	**Expected spatial scale for a yield assessment**	**Expected yield units for a yield assessment**	**Associated operational decisions**	**Expected benefits**
before budbreak of season n	field or within-field zones	mass per unit area	vineyards operations : pruning intensity and soil fertilization	optimized management of marketable yield
	wine blends	volume of wine blends after press	winery logistics : purchase of the barrels	costs saving, possibility to order any required material
during season n	block or within blocks zones	mass per unit area	vineyard operations : bunch thinning intensity, eventual fertilisation and irrigation level	optimized management of marketable yield
	one or several blocks	final volume of wine blends	accounting : managing stocks, planning revenue	good accounting, investment reasoning
	production area	final volume of wine blends	territorial agency : planning of marketing and commercialisation operations wine traders : purchase contracting	profitable sales and purchases
just before harvest n	field or within-field zones in regards to all the fields to be harvested	mass per field	vineyards operations: organizing harvest, planning work force and allocating transport equipment	optimized harvest decision
	one or several blocks	mass per wine blend	winery logistics: making and allocating space in tanks, purchasing wine-making consumables, planning tasks and work force, scheduling harvest intakes and treatment	optimal harvest blending and gain in wine quality
	production area	wine final volume	territorial agency : announcement of an eventual regulation for harvest volumes	optimized commercialization
longer-term	production area	wine final volume	whole industry : anticipation of the effects of future contexts on wine production, market price etc. for research orientation and strategic development	business sustainability

2.2.3.3.2 Adoption of the yield forecasting method also depends on complying operational constraints

If a yield assessment method respects the requirements of a given use case in terms of yield definition, date and spatial scale, then its implementation in the vineyard depends on additional criteria related to its ease of use. Concomitant workload and measurement time (including automated approaches) are important choice criteria: the higher the workload, the shorter the measurement time required then the more beneficial automation becomes. Easy to use equipment, as well as an easy protocol, will favor adoption. Cost also undeniably influences adoption but is rarely evaluated in scientific literature as equipment is seldom tested and designed under commercial conditions. Requirements for labor are also crucial, and access and cost of labor may considerably vary from one vineyard/winery to another.

First and foremost, yield assessment methods need to be quick and easy to perform, as well as robust. In that sense, using a low number of indicators whose influences on final yield are well understood and can be modelled should be preferred. Some indicators, such as weather data, are commonly encountered in operational datasets and correspond to input factors of the yield development system (Lamanda et al. 2012). These data are usually easily accessible in the field in terms of time, materials, expertise and cost. Their use

should then be prioritized in yield forecasting methods in order to allow a large number of vineyards/wineries to benefit from the easier to use and more robust yield forecast.

2.2.3.4 Challenges for yield assessment

By examining the issues and operational expectations about yield development, several challenges have emerged for yield assessment. Firstly, it is necessary to select indicators whose effects can be clearly analyzed (input or output indicators) and are accessible in a commercial context. Their measurement must be as precise as possible and judiciously positioned in time and space in order to represent the on-going yield development process. These data should be coherent with operational expectations in terms of the temporal and spatial scale of the yield assessment. Moreover, as yield development should be site-specifically described, these indicators should be collected where the yield assessment is made. This implies analyzing operational data, whose characteristics may be different from the data commonly analyzed in research experiments.

In order to extract the maximum information contained in these data, interactions need to be considered, e.g. inter- and intra-cycle dependencies and environmental influences in interaction that include potential future effects etc. This requires an acknowledgment of the temporality of yield development and consistently processing the data, for example by leveraging time series. Table 2.2 shows that the operational use cases of yield assessment differ by the dates at which yield assessment is required. However, the complexity of grape yield can't be assessed with the same accuracy throughout the season since the development of reproductive organs is successive and some key components that determine the final yield (Eq. 2.1) are not determined until very late in the season (e.g. final berry mass). Therefore, uncertainty regarding the temporal evolution of yield has to be handled by yield assessment methods in order to provide the user with sensible information for a sound, operational decision support system. Finally, yield assessment methods should comply with operational constraints to be adopted in the field.

2.2.4 How do existing methods address yield assessment issues ?

The previous section has outlined the complexity of grapevine yield development as well as the subsequent diversity of issues and challenges associated with yield assessment. This diversity has resulted in a multiplication of methods that have addressed improvements in yield assessment. The focus of the following section is to review the scientific literature associated with yield assessment and to propose a framework that allows these diverse methods to be integrated in a unique, comprehensive and cohesive way. The framework is composed of three steps: measurement, sampling and modelling. It is designed to support the development of future yield assessment methods and to help identify addressed and unaddressed challenges to orient new research efforts. For this purpose, the issues, including strengths and limitations, of existing methods are specifically discussed within the context of this framework after a general description of it.

2.2.4.1 A general framework to review existing scientific literature answers on yield assessment challenges

2.2.4.1.1 Any yield assessment method comprises three steps : measurement, sampling and modelling

The process of yield assessment anticipates a future yield performance based on the analysis of the mathematical relationship that links an explanatory variable(s) to a final yield response. To achieve this there must be (i) a measurement process of a yield attribute(s) and/or indicator(s), (ii) a sampling process i.e. an estimation of the attribute/indicator using some sampling method, and (iii) a final process of modelling the yield response to provide an actual final yield assessment. The whole yield assessment process is illustrated in Fig. 2.4.

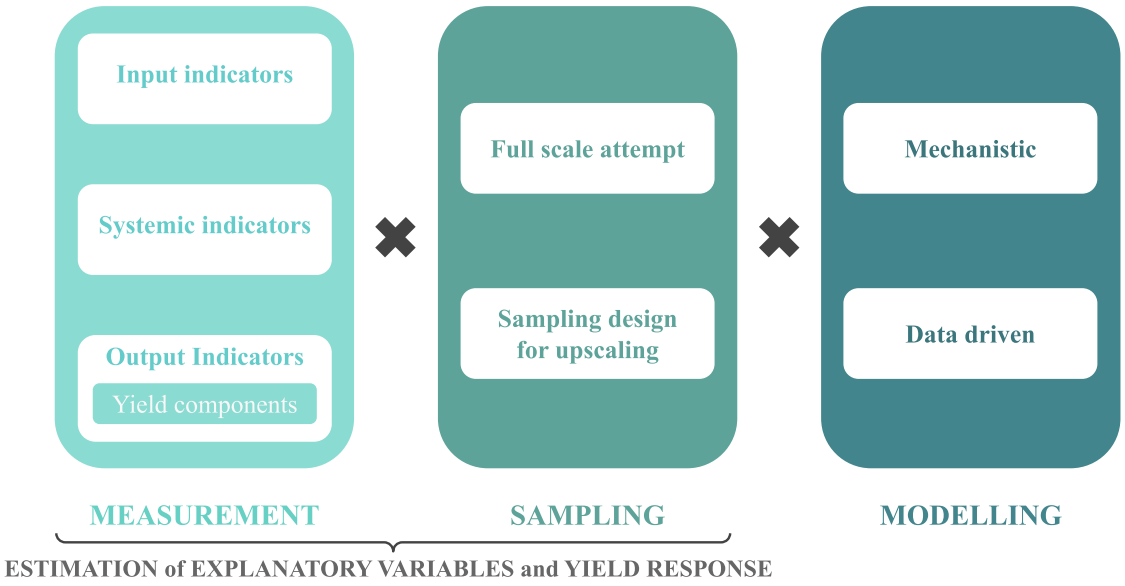

Figure 2.4 – A framework for constitutive steps of any yield assessment method.

An input indicator refers to a variable that influences yield development without being reciprocally influenced by it. A systemic indicator refers to a variable that influences yield development and that is reciprocally influenced by it. An output indicator refers to a variable that stands as an outcome of yield development, mainly referring to yield components.

Measurements may be of internal or external indicators that are not directly associated with yield components but either influence or interact with yield development. These are termed input and systemic indicators respectively depending on whether the effect is an influence or an interaction. For instance, grapevine variety is an internal factor considered to only influence yield development (input indicator), whereas grapevine water status is an internal factor that interacts with yield development during the season (systemic indicator). Similarly, precipitation is an external factor that only influences yield development (input indicator) whereas soil water content is an external factor that interacts with yield development (systemic indicator). Indicators that are directly associated with yield components at any stage of development are termed output indicators. These are direct measures of some aspect of yield potential, such as the number of bunches per vine, number of berries per bunch, berry mass, etc. Identifying the nature and type of the indicator (input, systemic or output) to be measured defines the method(s) by which it should be measured, the sampling approach needed and the mathematical relationships that could be considered in the modelling step.

Whatever the considered indicator to be measured, observations in the field are highly unlikely to ever be exhaustive due to technical and economic limitations. For example, it is either impractical or impossible, depending on the method of measurement (manual or via image analysis), to count all the berries in a vineyard. If the entire population of an attribute cannot be observed, then a sampling approach must be adopted to estimate the indicator value. Two broad sampling approaches have been identified from the viticulture literature. In the first instance, field observations are made as exhaustive as possible and the statistical inference is meant to only correct errors associated with the measurement process (Sun et al. 2017; Millan et al. 2018; Ballesteros et al. 2020). This approach is termed a 'full scale attempt'. Alternatively, observations are performed on representative sites of the distribution of the considered indicator at a given scale. A statistical inference is then used to upscale the limited number of observations. This approach is termed 'sampling design for upscaling' and is the main approach used in the field with various types of sampling designs having been proposed to improve the resulting estimate (Wulfsohn et al. 2012; Carrillo et al. 2015; Araya-Alman, Acevedo-Opazo, et al. 2017; Meyers et al. 2020).

Combined, the processes of measurement and sampling generate an estimation of any explanatory or yield response that will be used for yield assessment (Fig. 2.4). The estimation of an explanatory variable or of the yield response is not by itself a yield assessment. For instance, observations of the number of bunches per vine may be performed on several vines in a vineyard block using a sampling design to estimate the mean number of bunches per vine. The mean number of bunches per vine is an explanatory variable that requires some form of mathematical relationship to be linked to the yield response to achieve a yield assessment. The establishment of such a mathematical relationship corresponds to the modelling step of the yield assessment framework (Fig. 2.4). From the literature there are again two general ways that this is achieved. Firstly, and most commonly in operational contexts, a 'data-driven' approach can be used to establish a mathematical relationship between the explanatory variables and the yield response using statistical tests on experimental datasets. Alternatively, the mathematical relationship between the explanatory variables and the yield response can be established using a priori knowledge of yield development, and this is termed a 'mechanistic' approach. Depending on the chosen model, the modelling process will generate either a yield prediction or a yield forecast.

2.2.4.1.2 A yield estimation, prediction and forecast refer to different yield assessment approaches

Yield estimation, prediction and forecast define different approaches to yield assessment. Yield estimation is the simplest approach and refers to a yield assessment made in the same unit, time and space as the measurement and sampling processes it results from, while a prediction or forecast will alter the time, space or units between measurement and the reporting of a yield assessment. Therefore, a yield estimation can be obtained by simply multiplying an output indicator, such as the number of bunches or total grape mass measured per vine, by the number of vines in the block using a sampling design for upscaling approach. In this example, the yield estimation is reported as total bunches or total mass in the block for the same date as when the measurement and sampling steps were performed, and using the same unit.

In contrast, yield prediction or forecast refers to a yield assessment made in different unit, time or space than the measurement and sampling processes it results from and it additionally implies a modelling step in the yield assessment. It is most commonly associated with a yield performance that is expected to be reached in the same space but at a future date and is consequently often expressed in different units. Following the same example, mean bunches per vine can be multiplied by an expected or historical mean bunch mass to anticipate final bunch mass and by the number of vines in the block to achieve a yield prediction/forecast as a mass per block (i.e. different unit at a different time). The total grape mass per vine could also be multiplied by the number of vines in the block and weighted by a reduction coefficient to represent the effect of the numerous influences that yield development still has to undergo before the harvest. In this second case, the yield prediction/forecast is expressed in the same space and unit but doesn't correspond to the same date the yield performance is considered at. These two results are no longer estimates but yield predictions/forecasts. In these examples, quite simple models were involved as a modelling step for pedagogical purposes but much more complex ones can be considered (c.f. Table 2.6).

Usually, a yield prediction and forecast both refer to a date ulterior to the measurement and sampling date. However, they differ in the mathematical expression of the yield response. A yield prediction corresponds to a single value as a modelled yield response (Wonnacott and Wonnacott 1990; Saporta 2011) whereas a yield forecast corresponds to a distribution as a modelled yield response (Robert 1992). This difference mostly implies different uncertainty management, the uncertainty handled by forecast approaches being wider. The following paragraph aims to provide a more detailed description of the statistical approaches that support a yield prediction or forecast.

Both a yield prediction and forecast involve a yield model whose parameters represent the effect of each explanatory variable on the yield response. The residuals of this model represent the portion of yield variability that is not covered by the explanatory variables for each statistical individual e.g yield performance per year and block. A yield prediction only refers to frequentist statistics (classical statistics). In this case,

the model parameters are estimated as a statistical expectation on the basis of the dataset (Wonnacott and Wonnacott 1990; Saporta 2011). The effect of each explanatory variable on yield response is represented by a single value that depends on the analyzed data. As a consequence, the modelled yield response also corresponds to a single value. In contrast, a yield forecast corresponds to the distribution of the response variable (Robert 1992). It can be indirectly computed using frequentist statistics or directly calculated using Bayesian statistics. The former implies the computation of different intervals to encompass the single-valued yield prediction and the integration of some expert knowledge to generate a distribution of the yield responses. It takes into account that the model parameters are not accurately known and that the model is not exhaustive in its representation of the yield development phenomenon. The Bayesian approach of yield forecast achieves a similar objective, a distribution of potential yield responses, but by directly estimating the model parameters with distributions.

2.2.4.2 Issues and methods for measurement of yield components

2.2.4.2.1 Indicators measurement involves various technologies and methods

Issues for measurement are component specific

For output indicators, different yield components have different measurement issues. In chronological order, the first measurement of yield potential is performed via bud counts and bud dissection to determine the proportion of fertile buds and the number of potential bunches per bud. The number of fertile buds should preferably be assessed during a short period at the very end of winter when the final necrosis rate has been reached. Destruction is compulsory to observe the primordia dimensions, which are small, and some of the undifferentiated inflorescences may not be detected or incorrectly counted.

Following budbreak, inflorescences are visible to the naked-eye in the developing shoots and can be more easily differentiated from the canopy if counted early in the season (Wolpert and Vilas 1992). As the canopy develops, identifying the inflorescence becomes more difficult. Bloom is a short phenological stage. Flowers are small, of high number and may be hidden in the inflorescence architecture or whole inflorescences may be hidden in the rapidly developing canopy. The destruction of whole inflorescences for flower counting is often a sensitive issue for grape growers as the risk of losing inflorescences to an external event is still high, and manually removing inflorescences reduces yield potential.

Inflorescences become larger bunches but these may still be hidden in the canopy, which is continuing to develop and is becoming denser, depending on the trellis design and vine management (Nuske, Wilshusen, et al. 2014; Rose et al. 2016). While bunch number per vine and berries per bunch is fixed after fruit set, berry development is highly variable and the differentiation of verjuice, shoulder or deteriorated berries is not trivial in the field. Berries are numerous, relatively small and the more compact the bunch, the more hidden some berries may be. Moreover, still green or mature white berries are not easily distinguished from the canopy. Unripe berries are sometimes not counted, according to the protocol adopted by a particular vineyard/winery, while mature berries may be damaged during handling, thus creating a difference between harvested yield and delivered yield. Additionally, berry dimensions change continuously during their growth and maturity, including the possibility for berry mass to decrease before the harvest due to dehydration. Table 2.3 summarizes from the authors knowledge the measurement issues to be considered for each yield component. The levels of difficulty indicated are benchmarks that have been qualitatively determined from experience and a consolidated interpretation of numerous conversations that the authors have held with producers and experts in the field, in France and abroad.

Table 2.3 – Operational issues for grape yield components measurement

with +++ : very high difficulty, ++ : high difficulty, + : medium difficulty, empty cell : low difficulty

Measured yield component	Short implementation	Small dimensions and mass	High number	Risk of visual obstruction	Risk of misinterpretation	Destruction requirement	Concomitant workload	Global difficulty
Inflorescences in the bud	+	++		+	++	+++		+++
Inflorescences post-budbreak			+	+				
Flowers	+++	++	+++	+++	+	++	+	+++
Bunches			+	++	+++		++	+
Berries	++	+	+++	+++	+	++	++	+++

Review of currently proposed approaches for yield components estimation

There are three methods of measuring yield components: counts, sizing (area and volume measures) and weights (Table 2.4).The counting of inflorescences, bunches and berries per vine(s) are the most common in-vineyard measurements. Sizing pre-harvest usually focuses on bunch or berry dimensions, while sizing post-harvest mainly relates to volume measurements during pressing or wine-making processes. Weighing is generally used to estimate final yield mass, either by use of on-harvester yield monitors, weighting baskets/bins/gondolas in the field or truck weights on delivery to a grape crushing facility. Measurements can be considered as either destructive or non-destructive and performed manually or using sensor technology. Destructive measurement often requires the observation to be performed quickly after removal of the yield components from the vine or the yield component to be stored, e.g. frozen, for later analysis. Destructive sampling does allow further analysis to be performed indoors in more controlled conditions and to overcome field difficulties, including changing environmental conditions. At-harvest or post-harvest, measurements are by definition destructive, but in-season vineyard measurements may be performed destructively or non-destructively. Manual measurement is still commonly performed in-vineyard, as it requires a limited investment in equipment and allows a better observation of occluded yield components. However, manual measurements are prone to errors of concentration, perception and protocol interpretation as well as to the different capabilities and decisions of different operators (Carrillo et al. 2015). Moreover, manually measuring yield components often involves time and labor costs, which constrains the number of measurements that can be performed.

Given these limitations, there has been more research into the development of non-destructive yield component sensors in the past decade that has been enabled by advances in computer science. The main sensing method is image analysis coupled to modern artificial intelligence approaches (Aquino, Millan, Gutiérrez, et al. 2015; Aquino, Millan, Diago, et al. 2018; Aquino, Barrio, et al. 2018; Liu, Li, et al. 2018; Coviello et al. 2020). These are typically deployed as on-the-go automated sensing systems on terrestrial vehicles (Lopes et al. 2016; Millan et al. 2018) and increasingly on unmanned aerial platforms (UAVs)(Di Gennaro, Toscano, et al. 2019) to generate high-resolution information on yield components. Image sensors can also be deployed as low-cost mobile phone applications to automate the counting of yield components (Aquino, Barrio, et al. 2018; Aquino, Millan, Gaston, et al. 2015) either in a laboratory or in the field. Manual picture taking is still subject to operator error but the prevalence of smartphone technology makes it very accessible. Different artificial intelligence algorithms allow yield component detection based on differences in colour, shape or texture (Nuske, Wilshusen, et al. 2014; Liu and Whitty 2015; Pothen and Nuske 2016; Abdelghafour, Keresztes, et al. 2017; Abdelghafour, Rosu, et al. 2019). Image analysis generally implies an increased repeatability and a reduced measurement time compared to traditional manual measurement, especially for components that are numerous, of small dimensions or of very short duration in the field. The main limitation to image analysis is a requirement for the yield component(s) to be completely visible and recognizable by the artificial intelligence. For example, image analysis cannot be used to count berries on the reverse side of a bunch, nor can it see through leaves to count occluded bunches or berries. To circumvent this limitation, calibration methods to correct images to total yield have been proposed. Results have been positive but these have only been tested on limited datasets to date and their reproducibility is poorly evaluated (Nuske, Wilshusen, et al. 2014). When performed in the vineyard, methods based on image analysis are also challenged by the

varied shapes, dimensions and color of grape yield components, as well as a changing background and variable light conditions within the picture frame (Nuske, Wilshusen, et al. 2014; Grimm et al. 2019). Finally, the feasibility of in-vineyard imaging sensor systems is also dependent on the need to correctly deploy and to maintain the related equipment.

It should also be noted that some direct mass measurements in-season within vineyards are possible, although not always commercially relevant. Yield monitoring systems on grape harvesters have been demonstrated as a possible method of destructive yield estimation mid-season in juice grapes (Taylor, Sanchez, et al. 2016; Bates, Dresser, et al. 2018) as well as for yield-mapping at harvest. Alternatively, measuring the change in wire tension in single high-wire trellis systems (Blom and Tarara 2009; Tarara et al. 2014) has been proposed as the only dynamic method to follow crop development via changes in mass, but it requires permanent infrastructure that is likely cost-prohibitive outside of research activities. Table 2.4 presents the main methods established for yield component measurement. Most of them have been mainly used in research situations so far.

Table 2.4 – Methods for yield components measurement reviewed in literature

Type of measure		Type of technology	Yield component measured	Date of measurement	Experimental dataset	Announced results	References
count	destructive	image analysis in laboratory	flowers	bloom	3 inflorescence development stages 4 varieties 533 images	percentage error of 0.16 on validation dataset	Liu, Li, et al. 2018
count	destructive	image analysis in laboratory	berries	just prior to harvest	10 varieties 100 bunches	r^2 of 0.71 on calibration dataset	Ivorra et al. 2015
count	destructive	image analysis in laboratory	berries	just prior to harvest	7 varieties 10 bunches per variety	r^2 of 0.62 on calibration dataset	Diago, Tardaguila, et al. 2015
count	destructive	manual	bud fertility	before budbreak	3 blocks 1 variety	no reference data	Rawnsley and Collins 2005
count	destructive	manual	flowers	bloom	no published data	no published data	commonly used on field
count	destructive	manual	berries	after veraison up to just prior to harvest	no published data	no published data	commonly used on field
count	non destructive	analysis of manually taken images	flowers	bloom	11 varieties 2 devices 140 images	recall from .80 to 0.91	Aquino, Millan, Gaston, et al. 2015
count	non destructive	analysis of manually taken images	berries	from bloom to veraison	1 device 150 images	r^2 of 0.92 between observed and simulated data	Grossetete et al. 2012
count	non destructive	analysis of manually taken images	berries	from bloom to veraison	8 varieties multiple devices 145 images	mean absolute error from 0.01% to 0.12% on calibration dataset	Coviello et al. 2020
count	non destructive	analysis of manually taken images	berries	from bloom to veraison	2 varieties 529 images	r^2 from 0.88 to 0.95 on calibration dataset	Liu, Zeng, and Whitty 2020
count	non destructive	analysis of manually taken images	berries	from pea-size stage to bunch closing	12 varieties 2 devices	recall from 0.96 to 0.98	Aquino, Barrio, et al. 2018
count	non destructive	on-board image analysis	flowers	bloom	6 varieties 16 vines per variety 1 device	recall from 0.84 to 0.89 on validation dataset	Palacios et al. 2020
count	non destructive	on-board image analysis in field	bunches	unknown	1 variety 5 meters of wire 2 treillis systems	recall of 0.77 and 0.8	Rose et al. 2016
count	non destructive	on-board image analysis	bunches	harvest	respectively 190 and 35 images of white and red wines	respectively 0.91 and 0.97 of images with correct count	Reis et al. 2012
count	non destructive	on-board image analysis	berries	around veraison	3 varieties 229 plants 1 devices	undetected berries from 0.54 to 0.74 of the total count	Nuske, Achar, et al. 2011 Grocholsky 2011
count	non destructive	on-board image analysis	berries	from fruitset to harvest	6 varieties 1 device 1212 images	recall from 0.12 to 0.96	Nuske, Wilshusen, et al. 2014
count	non destructive	on-board image analysis	berries	before and after thinning, just prior to harvest	3 dates 3 varieties 2 training systems 60 or 38 images (2019 or 2020)	from 0.85 to 0.94 correctly detected berries (2019)or r^2 of 0.98 on calibration dataset (2020)	Zabawa et al. 2020
count	non destructive	on-board image analysis	berries	unknown	1 variety 10 bunches	recall of 0.77	Rose et al. 2016
count	non destructive	manual	inflo-rescences	around bloom	no published data	no published data	commonly used on field
count	non destructive	manual	bunches	after bloom and before veraison	no published data	no published data	commonly used on field
size	destructive	image analysis in laboratory	bunches berries	just prior to harvest	10 varieties 100 bunches	r^2 of 0.82on calibration data set	Ivorra et al. 2015
size	destructive	image analysis in laboratory	berries	just prior to harvest	no published data	no published data	Dyostem®, Vivelys
size	destructive	manual	berries	just prior to harvest	no published data	no published data	commonly used in field
size	non destructive	analysis of manually taken images	berries	pea-size stage, veraison, harvest	3 varieties 3 phenological stages 750 berries per stage	r^2 of 0.88 between observed and simulated data	Roscher et al. 2014
size	non destructive	on-board (but manually moved) image analysis	bunches	at harvest	49 bunches 1 device	average absolute error of 0.67 on calibration dataset	Kurtser et al. 2020
weight	destructive	manual	bunches	just prior to harvest	no published data	no published data	commonly used in field
weight	destructive	manual	berries	just prior to harvest	no published data	no published data	commonly used in field
weight	destructive	image analysis in laboratory	bunches	just prior to harvest	7 varieties 10 bunches per variety	r^2 from 0.65 to 0.97 on calibration dataset	Diago, Tardaguila, et al. 2015
weight	destructive	image analysis in laboratory	berries	just prior to harvest	7 varieties 10 bunches per variety	r^2 of 0.84 between observed and simulated data	Diago, Tardaguila, et al. 2015
weight	non destructive	analysis of manually taken images	bunches	fruitset, bunch closing and veraison	6 seasons 4 varieties 50-200 bunches per variety	prediction error from 0.06 to 0.15on validation dataset	Serrano, Roussel, et al. 2005
weight	non destructive	on-board image analysis	bunches	after veraison	1 season 1 block 30 contiguous plants	r^2 of 0.80 between observed and simulated data	Lopes et al. 2016 VINBOT®
weight	non destructive	wire tension	total yield	dynamic	3 seasons 2 blocks 2*3 consecutive rows per block	r^2 from 0.84 to 0.98 on calibration dataset	Blom and Tarara 2009 Tarara et al. 2014

2.2.4.2.2 Issues and methods for measurement of input and systemic indicators are numerous and diverse

Input or systematic indicators are mainly related to vine water status, nutrient status or canopy conditions. All three refer to complex variables that can be estimated in many different ways. Therefore, the choice of the indicator to be measured is very influential on the yield assessment process. The methods available are also very different in terms of cost, ease of implementation, and the temporal and spatial support of implementation (punctual or continuous). There is often a trade-off between the information desired and the methods implemented to measure it. In this section, only an overview of the numerous and diverse measurement methods is provided for each type of indicator, with an emphasis on methods that are used in operational vineyard situations (not just in research). Further details and specific references are given in Table 2.5.

In the field, vine water status may be estimated by direct observation. Methods such as the Shoot Tip Index (Rodriguez Lovelle, Trambouze, and Jacquet 2009) have been developed to guide visual assessment (Brunel et al. 2019). Vine water status may also be indirectly estimated through in-field soil moisture measurement, especially with tensiometric measurements (Dobriyal et al. 2012; Rienth and Scholasch 2019). However, the reference method for vine water status remains leaf water potential measurement with a pressure chamber. Pressure chambers are more expensive than tensiometers and their use requires a demanding protocol. This permits only a few measurements to be performed per day. In addition, the stability of the balance between the potential measured on the leaf petioles and the water potential of the rest of the plant is debated (Rienth and Scholasch 2019). Continuous measurements of vine water status are possible using sap flow technologies, but these require expensive semi-permanent installations (Rienth and Scholasch 2019). Other methods do exist to directly measure plant water status but are mainly experimental and have only been used for research purposes to date (Santesteban et al. 2015; Lavoie-Lamoureux et al. 2017).

Vine nutritional status is estimated during winter by measuring wood biomass or in-season using either petiole laboratory analysis or an optical measurement of leaf tissue. Wood biomass gives an indication about the amount of carbohydrate reserves (Demestihas, Debuisson, and Descotes 2018) and vine size and is measured manually or using dedicated sensors (Physiocap®, Ereca). Laboratory analysis of petioles provides detailed information on the concentration of nitrogen and other minerals, especially potassium (Cozzolino et al. 2020), but it is a destructive method that requires a demanding protocol to be performed quickly. Leaf nitrogen content can be manually and non-destructively in the field estimated using chlorophyll fluorescence sensing (Cerovic et al. 2015).

Canopy dimensions are still mostly evaluated by manual measurement but there is a shift toward using proximal and remote sensing for biophysical vine parameters. Canopy sensing can be performed manually (pedestrian transport) or on-board a terrestrial vehicle, air-borne vehicle (UAV, plane) or satellite. The characteristics of different remote and proximal canopy sensing methods are very different depending on the signal characteristics (active or passive, wavelengths etc.), the need for correction of the raw signal as well as the spatial and temporal resolution of the captured images or spectra (Weiss, Jacob, and Duveiller 2020; Gautam and Pagay 2020). Based on these characteristics, numerous vegetation indices have been proposed to serve different operational applications.

Table 2.5 – Main methods for plant ancillary indicators measurement in commercial vineyards

Plant ancillary data		Type of measure		Type of technology	Temporal support	Spatial support	References
water status	soil	non destructive	mass water content	gravimetric method	punctual	punctual	
water status	soil	non destructive	volumetric soil moisture content	neutron probe time domain reflectometry capacitance gamma ray attenuation ground penetrating radar	punctual or continuous	punctual	Rienth and Scholasch 2019 Dobriyal et al. 2012
water status	soil	non destructive	water potential	tensiometer pressure plate method	punctual or continuous	punctual	
water status	atmosphere	non destructive	evapotranspiration	weather sensors	continuous	punctual or continuous	
water status	water balance	non destructive	total or fraction of transpirable soil water computation	weather sensors soil granulometric analysis root depth measurement	continuous	punctual or continuous	
water status	plant	destructive	carbon isotope discrimination	mass spectrometry	punctual	punctual	Pichon et al. 2021 Rienth and Scholasch 2019
water status	plant	destructive	water potential	pressure chamber	punctual	punctual	Lavoie-Lamoureux et al. 2017 Santesteban et al. 2015
water status	plant	non destructive	visual observation	field observation apex method	punctual	punctual	Herrero-Langreo et al. 2013
water status	plant	non destructive	stomatal conductance leaf gas exchange	porometer infrared gas analyzer	punctual	punctual	
water status	plant	non destructive	sap flow	stem heat balance	continuous	punctual	
nutrition and storage	wood	non destructive	shoot number and diameter	manual	punctual	punctual	Champagnol 1984
nutrition and storage	leave and petiole	destructive	nitrogen and minerals content	laboratory analysis including near infra-red spectroscopy	punctual	punctual	Cozzolino et al. 2020
nutrition and storage	leave and petiole	non destructive	chlorophyll content	transmittance sensing	punctual	punctual	Cerovic et al. 2015
nutrition and storage	fruit and must	destructive	nitrogen and minerals content	laboratory analysis including near infra-red spectroscopy	punctual	punctual	
canopy	height, leaf area, porosity	non destructive	sizing	manual	punctual	punctual	
canopy	height, leaf area, porosity	non destructive	passive reflectance	visible or RGB image analysis	punctual	continuous	
canopy	height, leaf area, porosity	non destructive	active reflectance	visible, laser or multispectral image analysis	punctual	continuous	Weiss, Jacob, and Duveiller 2020
canopy	dimensions (height, width, density)	non destructive	3D modelling	stereo vision or structure from motion algorithms from multispectral imagery LiDAR	punctual	continuous	Gautam and Pagay 2020 Di Gennaro and Matese 2020
canopy	vegetation indices	non destructive	active reflectance	RGB, multispectral and thermal image analysis	punctual	punctual or continuous	Cheraïet et al. 2020
canopy	vegetation indices	non destructive	passive reflectance	RGB, multispectral and thermal image analysis	punctual or continuous	punctual	

Weather data is the main external systematic indicator measured in vineyards. This is often done using fixed weather stations (either on-farm or from a nearby reference point) together with some form of extrapolation. Weather stations for agriculture are well developed and temperature, rainfall, relative humidity, wind and global radiations information can be routinely obtained at very high temporal resolutions. The authors have chosen not to venture into a description of weather sensors here, although it is important to note that the appearance of virtual weather stations, through weather modelling and forecasting, has opened up a new and easier access to local weather data for producers for use in crop model applications (Launspach, Taylor, and Wilson 2017) beyond having fixed weather stations.

2.2.4.3 Issues and methods associated with sampling for yield assessment

Grapevine yield has been shown to be highly variable, both temporally at the block-scale (Chloupek, Hrstkova, and Schweigert 2004; Clingeleffer et al. 2001) and spatially at a within-block scale in a variety of different production systems (Taylor, Tisseyre, et al. 2005). To achieve a representative estimate of any explanatory or response variable from punctual measurements, the number of measurements needs to reflect the expected variance of the variable at the desired scale. However, measurements represent a significant effort in terms of labour and time costs, logistical organization and increasingly in equipment and technology costs. Manually measuring indicators is often an arduous task that is usually required (and performed) at periods during the season when concomitant workload is high. Thus, there is an optimization equation to be solved between the precision of the related estimate, the cost of the effort required to obtain it and the offset cost of not performing other concomitant vineyard activities. While this optimization equation has not been formally defined or solved to our knowledge, it is inherent in the development of sampling designs in viticulture. A sampling design corresponds to a reasoned number, timing and location of measurements aiming at estimating an explanatory variable (yield component or other variables) or the response yield variable that is operationally acceptable in terms of precision and required efforts. It is important to note that each grower is likely to have a differing idea of the level of precision required and the affordable effort available, although to date, proposed sampling designs for grape yield assessment have not considered this constraint. Issues and constraints in sampling for yield assessing that have been addressed in the literature are reviewed in the following section.

2.2.4.3.1 Sampling issues are related to the selection of representative sites

Data collection is more relevant during key phenological periods when the yield components of the final yield potential (Eq. 2.1) become fixed (Wolpert and Vilas 1992). The optimal timing of a measurement can be determined by considering the date when a yield component is no longer evolving. However, it may not be easy to identify these key dates in the field, especially when they occur asynchronously in space and time (Verdugo-Vásquez, Acevedo-Opazo, Valdés-Gómez, Ingram, et al. 2020). This is further complicated by a lack of consolidation in the literature on reported timings. Some studies have reported the timing of their observations in terms of Gregorian days from the completion of a commonly observed phenological stage e.g. budbreak, bloom, fruitset or veraison (Petrie and Clingeleffer 2005; Molitor and Keller 2017). The start and end of these stages are open to different interpretations in different years and in different regions or locales. The use of a fixed-day time step also ignores local environmental effects, particularly thermal time effects, on vine and berry development. This limits any global comparison or the derivation of general conclusions on the timing of reproductive development in vineyards outside the studied areas. This shows the necessity to work with a time expression that captures the reproductive development conditions being experienced by the vine.

Within the field, environmental influences may generate random or structured spatial patterns that may or may not be temporally stable (Clingeleffer et al. 2001; Kaspar et al. 2003; Tisseyre, Mazzoni, and Fonta 2008). Therefore, every berry, bunch, vine and zone within a block or a vineyard experiences a different combination of environmental conditions, which are rarely measured. Furthermore, as grapevine physiology and development is subject to fixed acrotony rules under spatially heterogeneous environmental influences, phenological asynchronicity is expected at all scales, from berries within a bunch (Bigard et al. 2019) to zones within a field (Verdugo Vásquez et al. 2015; Verdugo-Vásquez, Acevedo-Opazo, Valdés-Gómez, García de Cortázar-Atauri, et al. 2017; Verdugo-Vásquez, Acevedo-Opazo, Valdés-Gómez, Ingram, et al. 2020). This phenological asynchronicity also implies that every berry, vine or every block will not respond in the same way to these external influences, which in themselves will vary spatially. Correctly sampling under these conditions implies the correct selection and location of samples that are able to represent population distributions across the area to be studied i.e. in the geographic space and/or across the known variability of the indicator i.e. in the attribute space. If this is to be achieved then it is preferable to avoid measurements associated with rare events or abnormal values e.g. dead vines, diseased vines, vines suffering from a localised stress or vines located on the edge of a row or block etc. Nevertheless, the number of missing plants must be accurately estimated in order to upscale yield that is assessed at the individual vine scale. This has been investigated in particular through remote sensing approaches (Robbez Masson and Foltete 2005; Primicerio

et al. 2017). When upscaling measurements, the weight given to each sample site may depend on the number of individuals in the population that it represents. This weight may vary when the variance is not uniform across the block (or sampled area).

Another point to consider is the ease of travelling from one sample site to another in the field. Most grape blocks are trellised, thus restricting movement across rows within a block. Travel time between sampling sites may be considerably increased if the sampling design is poorly organized (Oger, Vismara, and Tisseyre 2021). Slopes or difficulties when walking around, such as the absence of grass cover, may also increase the effort required. For these reasons, sampling has to meet a trade-off between the effort or time invested and the desired estimate precision. The number of sampling sites needed to achieve sufficient estimation precision depends on the local stochastic variance (Wolpert and Vilas 1992) and the size of the area sampled. On average, the number of recommended measurements required is in the range of 20 to 30 (Clingeleffer et al. 2001). However, operational constraints do not always permit an operator to achieve a sufficient number of measurements. This recommendation is also based on generating an average yield estimation for a block/vineyard and does not consider the effects of intra-block spatial variance in the yield components.

2.2.4.3.2 Sampling designs for yield assessment depend on the data available, the accuracy of the aimed estimate and the allocated operational means

The complete measurement of yield components during the season (number of bunches, number of berries, berry mass, etc.) is not yet possible. However, imaging and sensing technologies are being developed at a rapid pace with the aim to observe the entirety of a yield component in the field, in line with the full-scale attempt approach (Sun et al. 2017; Millan et al. 2018; Ballesteros et al. 2020). Until this is routinely and operationally possible in commercial vineyards, the industry will need to rely on sampling designs and upscaling. To achieve this, there have been different designs presented in the scientific literature (Oliver, Bishop, and Marchant 2013). These approaches can be adapted to all explanatory variables, particularly to yield components, as well as to the yield attribute (response). The required sampling effort always needs to be considered in relation to the effort required to measure the chosen indicator(s). In most cases, the final yield variance and its spatial structure is unknown prior to sampling. Consequently, ensuring a representative criterion for yield or yield components when sampling is challenging.

Random sampling

In a situation where no a priori information is available, a random sampling method is generally recommended (Clingeleffer et al. 2001). This sampling strategy gives each site of the population an equal chance to be selected in the final sample. The lack of a priori information does not allow for the preferential selection of one individual over another, so the choice of sampling sites is completely random. This approach is difficult to correctly implement in the field, as randomness is often biased by practical constraints, such as the distances to be covered or the point of entry into the block. Therefore, from field expertise, samples often tend to be located disproportionally close to block edges and there is an acknowledged operational bias as the actual sampling sites are driven to some extent by operator expertise and visual observations.

Grid/Systematic sampling

Alternatives to random sampling under conditions with no a priori information are based on carrying out measurements on a regular basis (Clingeleffer et al. 2001).This can be achieved by locating measurement sites on the nodes of a regular grid overlying the block (grid sampling) or by visiting all the rows of the block and carrying out a measurement each time a certain number of vine stocks have been covered (systematic sampling) (Wulfsohn et al. 2012). Regular grid sampling or systematic sampling is open to bias from periodicity in the data, e.g. sampling an individual row that was pruned differently or perhaps missed a spray application and which may not be representative of the block. If grid or systematic sampling is used, sampling theory dictates that some degree randomness should be incorporated to minimize this risk.

Stratified/Target sampling

When available, integrating ancillary data into the sampling design can significantly improve the quality of the estimate. Stratified (or target) sampling proposes to select sites to be measured from a classification of all the candidates sites based on their ancillary data value. It relies on the hypothesis that there is a relationship between the ancillary data and the yield components or that the ancillary data spatial structure reflects the

spatial yield variability. Ancillary data can correspond to other yield components, vine variables or other variables, such as input indicators. These data often have the advantage of already being collected or being accessible at a low cost before sampling. In the case of grape yield components or yield sampling, ancillary data may correspond to vegetation indices measured by canopy response imagery such as NDVI (Carrillo et al. 2015; Meyers et al. 2020) or historical yield data (Araya-Alman, Acevedo-Opazo, et al. 2017).The statistical process for selecting target sites can vary (e.g. using quantiles or k-means classification) but with the same objective of defining a set of sampling sites that s representative of the ancillary data distribution. These approaches are widely used, especially in soil studies (Adamchuk et al. 2011) and have been proposed for yield assessment in viticulture (Bramley and Williams 2001; Carrillo et al. 2015; Meyers et al. 2020) but have not yet been widely adopted in commercial vineyards. Some variants, such as the Ranked Set sampling, have also been deployed into other perennial horticultural fruit crops (Uribeetxebarria et al. 2019) or for other vineyard parameters. Since the correlation between yield and ancillary data can vary greatly depending on location and time, the use of such data must be carefully considered and integrated (Carrillo et al. 2015). The resolution of the ancillary data, as well as the transformations carried out to upscale the measurement information (aggregations, interpolations, changes in resolution), must also be tailored to the objective pursued.

Model sampling

Model sampling follows the principles of target sampling but goes further in exploiting the available ancillary data. This sampling strategy uses the observations made at the measurement sites to calibrate the parameters of a model linking the sampled variable to the ancillary data. In a second step, the constructed (or newly calibrated) model is used to predict values for the sampled variable from the set of available ancillary data. The final estimation of the sampled variable is then performed using the mean of all the predicted values. This approach has already been presented for grape yield estimation (Carrillo et al. 2015).

Sampling more complex populations

Other methods have been developed to sample complex populations that can be divided into subpopulations. The criteria used to form these subpopulations should be selected according to the sampling objectives. For instance, this type of method can be used at the territory scale to select blocks belonging to different areas or at the vine scale to select bunches on different shoots. There are different ways of sampling these populations but common approaches are cluster sampling or multi-stage sampling (Etikan and Bala 2017). These sampling methods propose to choose measurement sites by randomly selecting a subpopulation and then an individual from the subpopulation, thus allowing the freedom to assign different probabilities to sub-population and individuals. The weight assigned to each observation in the final mean may also vary according to the original subpopulations. Some variants of these stratified sampling approaches have already been applied in agronomy (Wulfsohn et al. 2012).

Sampling to build a yield map

In certain situations, such as that of selective harvesting, it can be beneficial to build an accurate yield (or yield component) map instead of deriving a mean yield assessment. If this is done from punctual observations, then the sampling design (number and location of samples) needs to respect limitations with the interpolation method, e.g. kriging or nearest neighbour or inverse distance, and the desired resolution of the map. The same sampling design is not appropriate for estimating mean block statistics and for mapping intra-block spatial patterns. Interpolation and map production tends to require a larger number of samples to generate useful information, e.g. kriging typically requires more than 100 samples (Webster and Oliver 1992) and manually measured yield component maps have only been reported in research studies to date. It is cost prohibitive in commercial situations to map yield estimates from manual measurements, despite the desire to have this information. The need for affordable, timely, higher resolution data to map yield components is a principal reason for the recent activity in the development of on-the-go sensors for yield components (Nuske, Wilshusen, et al. 2014).

Fig. 2.5 illustrates how the data from some of these different sampling designs for upscaling (random vs stratified) can be modelled to derive a yield assessment. From the same punctual observations, it shows that different inference models lead to different estimations of the sampled variable and yield assessment compared to the final, exhaustively measured harvest yield.

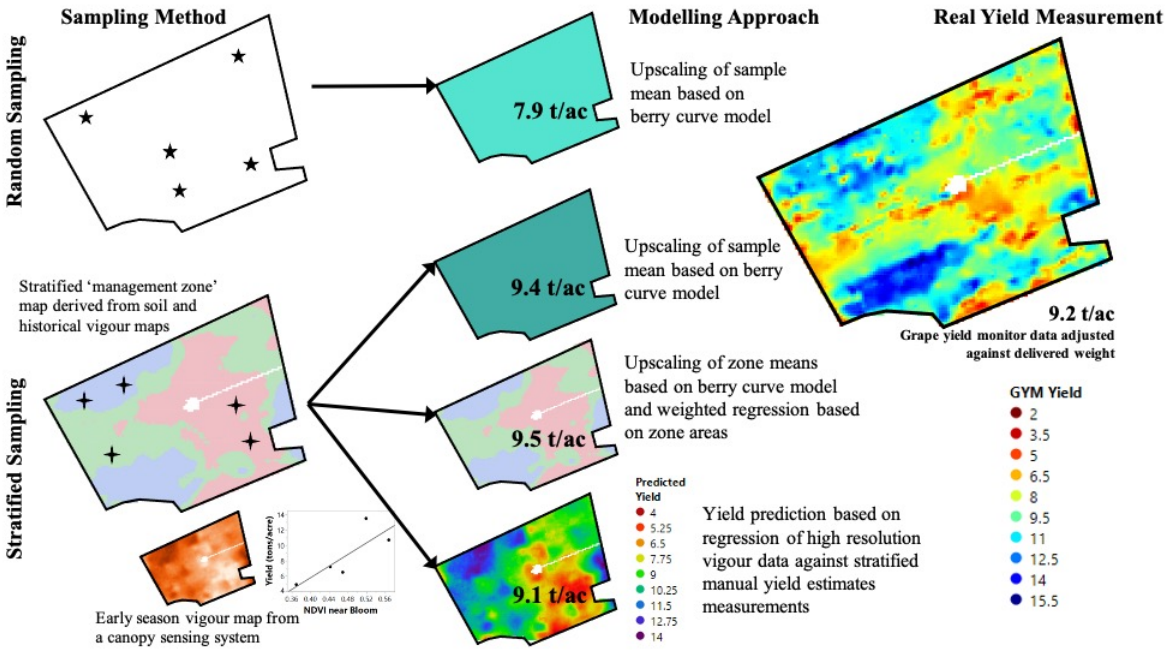

Figure 2.5 – Illustrative real-world examples of different sampling methods applied to different modelling approaches to derive various mid-season yield assessments for the same juice grape block in the Lake Erie AVA.

A destructive 'full scale attempt' using data from an on-harvester grape yield monitor (GYM) is also presented as the validation data for the mid-season yield assessments. Data shown is courtesy of Dr T. Bates, Cornell Lake Erie Research and Extension Laboratory, and Mr T. Betts, Betts' Vineyards LLC, Westfield, NY.

2.2.4.4 Issues and methods for yield modelling

The modelling step consists in establishing a mathematical relationship between explanatory variables and the yield response variable. Two main approaches are identified. In the first instance, an a priori mechanistic model can be established using available scientific knowledge on yield development which is calibrated and validated over experimental datasets (Valdés-Gómez et al. 2009; Cola et al. 2014; Nogueira Júnior et al. 2018). This approach is mainly used to capture and understand the dominant mechanisms of yield development but it restricts the modelling to knowledge already discovered by previous work and may rely on generic, rather than site-specific, interactions. This kind of approach also requires numerous parameters to be specified, which restricts the field use of these mechanistic models as these parameters are generally not easily accessible in commercial vineyard conditions. Empirical approaches in the field may often be considered as degraded mechanistic models since they consider a theoretical yield equation fed by historical data, such as the average number of bunches per vine over several years (Nogueira Júnior et al. 2018). Alternatively, a data-driven modelling approach can be performed to find mathematical relationships using empirical data with reduced or no a priori knowledge. It assumes that some influences are not explicitly integrated into the mathematical relationship but rather included in a black box where only inputs and outputs are known (Cunha, Ribeiro, and Abreu 2016; Di Gennaro, Toscano, et al. 2019; González-Fernández et al. 2020). This approach has the advantage of requiring fewer parameters to be specified and allows the detection of site-specific relationships. For this reason it is more commonly used in the field in commercial conditions. However, it strongly relies on the quantity and quality of the studied data. It may lead to the detection of non-significant or erroneous relationships depending on the dataset and may not always be easily interpreted from an agronomic point of view. This approach will be enhanced by new methods that use artificial intelligence (Sirsat et al. 2019; Ballesteros et al. 2020).

Additionally, regardless of the approach considered, other technical modelling elements need to be addressed. Firstly, data collection, whether from experimental research or commercial vineyards, induces interdependencies in the dataset. It comes from the repetition of observations on the same entities. For example, field observations are often collected on the same block over several vintages or on multiple blocks in the same vintage. These interdependencies are rarely taken into account in the modelling process of reviewed yield assessment methods even though such interdependencies may affect model establishment as part of the observed variability is due to the data collection design. Using mixed effects models is a possible response to the problem of data interdependencies (Zhu, Fraysse, et al. 2020) but it requires specialized modelling expertise and sometimes limits the predictive capabilities of a model by the choice of explanatory variables to be put into either fixed or random effects.

Secondly, explanatory variables may also present intercorrelation. It generally comes from a common determinism that is not necessarily known. For example, radiation and temperature may be intercorrelated because high temperatures are often observed when radiation is high. Data intercorrelation may be spatial or temporal and can be addressed by specific statistical analyses, such as geostatistics or times series analyses, when data resolution allows it. However, most reviewed works will typically only study a few punctual indicators that focus on summary data (e.g. means or variances) for a given spatial area (Cunha, Marçal, and Silva 2010; De la Fuente et al. 2015; Liu, Cossell, et al. 2017) for a few phenological stages or time steps (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017) in order to assume data independence. Such a data simplification may imply a loss in the information that is potentially contained in high resolution datasets (Araya-Alman, Leroux, et al. 2019; Laurent, Baragatti, et al. 2019)

Regarding the type of statistical model considered, most reported studies have used a linear model (Dunn and Martin 2004; Cunha, Ribeiro, and Abreu 2016; Zhu, Fraysse, et al. 2020). However, a more complete consideration of environmental influences and of the vine-management-environmental dynamics on temporal yield development (accumulation, threshold effect, succession or trajectory effect etc.) is likely to require other, more complex types of models. These new complex models need to permit data to be fitted non-linearly for an explicative aim and be robust to the introduction of new data to extrapolate or expand applications (Parker, García de Cortázar-Atauri, et al. 2020). Models based on artificial intelligence methods may be more suitable for this (Sirsat et al. 2019).

Model selection can be done according to various criteria whose implications for model performances are different. In the case of yield assessment, criteria allowing the selection of robust models should be favoured even if the data is less fitted. For example, criteria such as the Akaike or Bayesian information criteria (respectively AIC and BIC) may be preferable to the coefficient of determination (r^2) for prediction or forecasting models. Finally, the size of the studied datasets may not allow for independent training and validation dataset for model evaluation (Sirsat et al. 2019; Diago, Correa, et al. 2012). In these cases, options such as cross-validation are recommended (Cunha, Ribeiro, and Abreu 2016; Molitor and Keller 2017) and the size of the training dataset and the number of different modalities it contains will provide an indication of the adaptability and transferability of the modelling approach.

2.2.4.5 Uncertainty in yield assessment

In general terms, the notion of uncertainty refers to the inability to exactly know the outcome of a given phenomenon. This is a rather broad concept that remains unclear in many discussions as few formalized definitions are proposed. The fact that any yield assessment method is composed of three different steps means that the uncertainty associated with a yield assessment incorporates notions of uncertainty associated with metrology, sampling and modelling. Thus, this section reviews what the scientific literature calls uncertainty and sources of uncertainty for each step in order to consolidate a first understanding of uncertainty in the whole yield assessment process. Given the variety of combinations of measurement, sampling and modelling methods used to generate the yield assessment, this consolidated understanding is neither exhaustive nor quantified. However, it should help further investigation in the particular case of each yield assessment method.

2.2.4.5.1 *Uncertainty in the measurement step refers to some dispersion around a real measurand value*

In metrology, measurement uncertainty refers to both the doubt about the validity of the result of a measurement and quantitative measures of this doubt. It is formalized as a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand e.g. standard deviation (JCGM 100 : 2008). The measurand refers to the value of a particular quantity (variable) to be measured (JCGM 100:2008). An empirical approach of uncertainty in measurement distinguishes random or systematic contributions, that respectively refer to the concepts of precision or fidelity and bias (Ramsey, Ellison, and Rostron 2019). In this sense, the concept of uncertainty is to be distinguished from the concept of error. Uncertainty refers to a range of attributable values whereas error is defined as a single-valued difference between the measurement and the true (or reference) value (Ramsey, Ellison, and Rostron 2019). Sources of uncertainty in measurement include variations in the operator activity (inattention, finite perception, misinterpretation or misreading etc.), in the equipment analytical capacity (finite resolution, non exhaustive detection, low metrological fidelity and accuracy etc.) and in the measurement protocol (incomplete definition of the measurand, assumptions, inadequate knowledge of environmental effects on the measurement etc.) in relation to the measurement issues cited in the dedicated paragraph.

2.2.4.5.2 *Uncertainty in the sampling step refers to the measurement definition of uncertainty but integrates more uncertainty sources*

Uncertainty in sampling is mostly considered with regards to the variable estimate. Sources of uncertainty arising from sampling include the heterogeneity of the mesurand over the target sample i.e. over the portion of measurand that is intended to be estimated (Ramsey, Ellison, and Rostron 2019). This portion may be defined in terms of space, time, composition etc.
A second source of uncertainty in the sampling step is linked to the sample representativeness of the measurand distribution in the target sample. This depends on the number of sites or dates and the criteria with which they are selected, in accordance with the chosen sampling method (Kruskal and Mosteller 1979; Ramsey, Ellison, and Rostron 2019). In the case of target or model sampling, this representativeness issue also implies the magnitude and robustness of the link between the auxiliary variable and the variable to be sampled.
A third source of uncertainty refers to the model that is used to infer an estimation of the considered variable from punctual measurement performed at the a superiori scale. For instance, it refers to the model that is used to estimate a number of bunches per vine at the block scale from the number of bunches measured on a few vines in the block. The sampling model can never fully capture the reality and induces some additional variability in the estimate.

There are two approaches to quantify uncertainty in an estimation process. The empirical, or experimental, approach refers to the notion of repeatability and corresponds to the replication of the estimation process in order to give a direct assessment of the estimation uncertainty. The theoretical, or modelling, approach aims at individually identifying all the uncertainty sources and to combine them into a model (Ramsey, Ellison, and Rostron 2019). Following sampling, the uncertainty associated with an estimation is also defined as the dispersion of the values that could reasonably be attributed to the measurand, but it is a combination of uncertainty resulting from both measurement and sampling and so should be wider than the measurement uncertainty.

2.2.4.5.3 *Uncertainty in the modelling step refers to the variability the model can not account for*

Uncertainty in modelling has various definitions and is rarely formalized. However, uncertainty due to the inherent variability of the considered phenomenon (inherent uncertainty) is to be distinguished from epistemic uncertainty. Inherent uncertainty refers to the fact that yield development integrates complex mechanisms which are determined in some parts and random in other parts. It follows that the output of a yield elaboration cycle should always be considered as variable and that it can never be known in advance. Inherent variability refers to the range of values attributable to this output, it can not be exhaustively quantified. Epistemic uncertainty is due to the fact that only part of the yield development phenomenon is

understood and that, even if understood, it is not always possible to theoretically and numerically draw a perfect representation of it (Walker et al. 2003). Thus, epistemic uncertainty corresponds to the portion of inherent variability which remains uncovered by the modelling effort. It may also integrate an extra-portion of variability because imperfect modelling may lead to expecting yield responses that might not even be integrated in the inherent variability (that are not possible but not recognized as such).

Epistemic uncertainty can be characterized and minimized by handling three main uncertainty sources in modelling. The first source corresponds to the choice of a model structure and of a functional form for the model (Walker et al. 2003). This choice is often based on model-fitting criteria with the dataset and sometimes integrates a priori knowledge on the phenomenon to be modelled. However, it is always an approximation of the real phenomenon and the approximation is even more disrupted when extrapolated to conditions not contained in the dataset. The second source comes from the estimation of the model parameters of the model. This second uncertainty source causes a distinction between predictive and forecasting modelling approaches. Based on a distribution instead of single-valued expectation for each parameter, forecasting approaches handle a bigger part of uncertainty than predictive approaches. In other words, it can be considered that forecasting approaches allow to model a larger panel in yield development possibilities than the panel captured in the dataset i.e. than what is modelled by predictive approaches. In this way, the model uncertainty (i.e. the remaining uncertainty that is not handled by the model) should be lower in forecasting approaches than in predictive approaches. Finally, it should be remembered that the variables handled in the modelling step are estimates. Therefore, a third source of uncertainty is linked to way that the model manages the uncertainty associated with the explanatory and response variables. Some models, such as mixed effects or weighted regressions, are designed to better deal with this source of uncertainty.

2.2.4.5.4 Uncertainty in yield assessment requires an integration of uncertainty sources from the measurement, sampling and modelling steps

Each year, in an empirical situation, i.e. when no effort of yield assessment is done, uncertainty regarding the upcoming yield performance (i.e. the yield measurand at any spatial scale) primarily corresponds to inherent uncertainty. It is associated with the natural variability of yield development.
In order to recognize part of this natural variability, field observations are performed and the corresponding variables are estimated. The estimation is aimed at both the yield response variable and at the explanatory variables that could illustrate the mechanisms by which yield is determined. It is carried out through the implementation of measurement and sampling methods. However, these methods are imperfect in their ability to quantify the real quantity of each variable (i.e. the measurand of each variable). Therefore, they induce additional uncertainty associated with some additional variability to the inherent variability of yield development, as illustrated in Fig. 2.6.
The objective of a model is to best capture the uncertainty in yield development and in the associated variable estimations in order to result in the narrowest possible model uncertainty. Among the proposals of the scientific literature to achieve this objective, forecasting approaches are to be distinguished from predictive approaches because they enable a greater level of uncertainty to be managed. The portion of uncertainty that is not managed by the modelling step is considered as the yield assessment uncertainty. It could be formalized in the metrological way i.e. as a distribution of values attributable to the yield measurand once the measurement, sampling and modelling steps have been performed (Fig. 2.6).

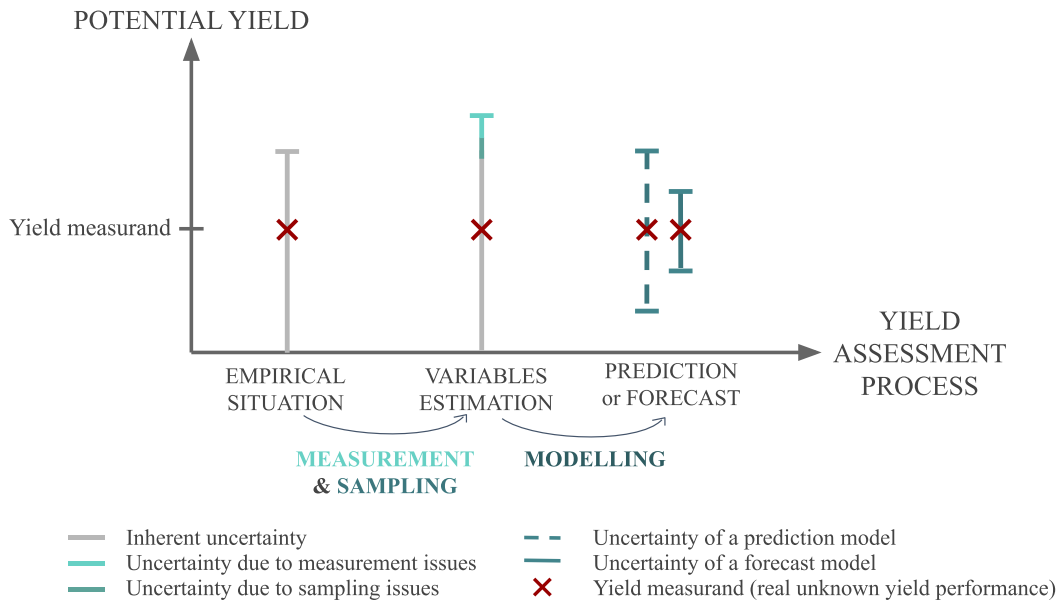

Figure 2.6 – Sources of uncertainty in the process of yield assessment.

The displayed figures are only examples. The distribution of the bars at each stage show the potential distribution of yield outcomes at each step. This does not necessarily reflect the actual accuracy of the yield assessment at a given stage, which should improve from left to right.

Table 2.2 highlights that most of the operational needs are oriented towards an assessment of the final (harvest/production) yield rather than current actual yield. For this reason the grape and wine sector should be more interested in yield prediction and forecast than in yield estimation. In the field, a yield estimate is sometimes interpreted to anticipate the final yield performance. To do so, some human expertise is used to introduce a notion of uncertainty relating to the future outcome of yield development. This represents a poorly traceable and reproducible approach. The user expertise is engaged to expand from the technical uncertainty to be handled through the yield estimation process to the uncertainty that is required to be addressed when assessing yield future evolution. The intent of yield assessment is to inform the decision-making processes. Therefore the quality (certainty) of a yield assessment, and not just the yield value of the assessment, should be considered when making decisions. It is clear that future approaches to yield assessment should handle uncertainty in a more objective and reproducible way (Walker et al. 2003). This is an area of yield assessment that requires further development and corresponds to the implementation of at least yield prediction and, preferably, yield forecasting approaches.

2.2.4.6 Review of yield assessment methods

2.2.4.6.1 Yield assessment methods address different use cases through different choices in the measurement, sampling and modelling steps

Table 2.6 summaries the different yield assessment methods that have been reviewed. It mainly contains references from the scientific literature, as the methods used by the industry are poorly documented. This is evidence of the fact that commercial wineries still have their own practices and often empirical habits when it comes to yield assessment methods. Yield assessment methods are presented according to the use case they address. These use cases are described according to three characteristics: i) the date at which the yield assessment is provided, ii) the spatial scale at which the yield assessment is realized and iii) the definition and units that are used to express the yield response. Afterwards, the propositions are split in terms of the measurement and sampling strategies associated with the estimation of the explanatory variables and in terms of the models used. Finally, the experimental datasets that have been used to establish the method are

indicated as an appraisal of the reproducibility and robustness of the method in order to properly interpret the results.

Table 2.6 – Methods for yield estimation and prediction reviewed in literature

Use Case			Variables estimation							
Date	Spatial Scale	Yield unit	Indicators type	Indicators nature	Measurement method	Sampling method	Model	Experimental dataset	Announced results	References
season n-1	plant	kg	systemic	leaf predawn water potential, leaf nitrogen content	pressure chamber, destructive measure in laboratory	random sampling	mechanistic	3 seasons 2 blocks 5 treatments 8 to 30 plants per treatment	r^2 from 0.65 to 0.7 between observed and simulated data	Guilpart, Metay, and Gary 2014
season n-1	territorial	hL/ha	systemic	NDVI imagery	satellite 1 km resolution	none	data driven: linear regression with cross-validation	10 seasons 36 images per season 4 regions of 3x3 km	r^2 from 0.73 to 0.88	Guilpart, Metay, and Gary 2014
after budbreak, bloom, after fruitset, close to harvest of season n	block	t	output	historical yield data, yield components	manual measurements	random sampling	empirical	1 to 4 seasons 40 blocks	r^2 from 0.40 to 0.95 on calibration dataset	Clingeleffer et al. 2001
bloom, veraison, close to harvest of season n	block	kg/ha	input	weather data phenological dates	unknown	none	artificial intelligence: random forest, lasso, elasticnet, spikeslab	3 seasons, 128 blocks	relative root mean squared error from 0.24 to 0.29	Sirsat et al. 2019
anytime in season n	block	t/ha	input	calibration parameters: weather, soil, cultural practices, genetic data (see Brisson et al. 2003)	manual measuring	random sampling	mechanistic: STICs	3 seasons, 3 blocks, 2 varieties, 5 plants per block	r^2 from 0.88 to 0.91 between observed and simulated data	Valdés-Gómez et al. 2009
anytime in season n	block	t/ha	input		unknow	unknown	mechanistic: STICs	10 to 21 seasons for calibration, 3 seasons for validation, 3 varieties, 6 vineyards, 2 regions	r^2 of 0.86, RMSE of 2.07	Fraga, García de Cortázar Atauri, et al. 2016
fragaModellingClimateChange2016 anytime in season n	intra-block	kg/m	input	weather data, calibration parameters: density, coefficient of light extinction, base temperature, thermal requirements for phenology	manual measuring	random sampling	mechanistic	calibration : 2 seasons, 1 block, 2 treatments, 4 plants per treatment; validation : 5 seasons, unspecified blocks	r^2 of 0.96 on validation data set	Cola et al. 2014
anytime in season n	plant	kg	input		manual measuring, literature review	none	mechanistic	2 seasons, 1 block, 1 variety, 60 plants	r^2 of 0.97 between observed and simulated data	Nogueira Júnior et al. 2018
bloom of season n	territorial	kg/ha	output	airborne pollen	suction trap with optical microscopic analysis at 10m above ground level	none	data driven: linear regression	5 seasons	r^2 of 0.92 on calibration data set	Cristofolini and Gottardini 2000
bloom of season n	territorial	hL	output	airborne pollen, weather data	filter trap with optical microscopic analysis at 15m above ground level	none	data driven: logistic regression with cross-validation	15 seasons 1 site	r^2 of 0.79, average relative error of 0.056 on validation data set	Cunha, Ribeiro, and Abreu 2016

Use Case			Variables estimation							
Date	Spatial Scale	Yield unit	Indicators type	Indicators nature	Measurement method	Sampling method	Model	Experimental dataset	Announced results	References
between bloom and veraison of season n	block	t	input and output	number of shoots per plant, historical number of bunches per shoot, historical bunch mass	on-board video-analysis	mechanistic	2 seasons, 4 blocks, 1 video per row	undefined error from 0.012 to 0.36 on calibration data set	r^2 of 0.79, average relative error of 0.056 on validation data set	Liu, Cossell, et al. 2017
between bloom and veraison of season n	block	t/ha, yield map 3m resolution	systemic	NDVI and LAI imagery	satellite imagery 30m resolution	target sampling	data driven: linear regression	2 seasons, 2 blocks	NDVI : r^2 from 0.63 to 0.77 ,LAI : r^2 from 0.48 to 0.77 on calibration data set	Sun et al. 2017
between bloom and veraison of season n	territorial	kg/ha	input output	weather data, number of inflorescences per vine	manual counting	none	data driven: multivariate linear regression with cross validation	10 seasons, 10 plants, 1 region	r^2 of 0.99 between observed and simulated on validation data set	González-Fernández et al. 2020
between fruitset and veraison of season n	block	kg	input systemic	downy mildew spore concentration, cumulative rainfall	suction trap with optical microscopic analysis at 2m above ground	none	data driven: multivariate linear regression	6 seasons, 1 block, 1 variety, 20 plants per block	r^2 of 0.98 on calibration dataset	Fernández-González et al. 2011
between fruitset and veraison of season n	intra-block	kg/m	output	historical mean bunch mass, mean berry mass	manual counting and weighing	none	empirical	4 seasons, 14 blocks, 4 rows per block	r^2 from 0.6 to 0.75 between observed and simulated data on validation data set	De la Fuente et al. 2015
between fruitset and veraison of season n	intra-block	kg /vine	output	bunch count and dimensions	image analysis	target sampling	data driven: linear regression	2 seasons, 1 block, 2 zones	r^2 of 0.82 between observed and simulated data	Di Gennaro, Toscano, et al. 2019
between fruitset and veraison of season n	bunch	g	output	bunch dimensions	manually taken pictures on field	none	data driven: linear regression with cross -validation	6 seasons, 4 varieties, 50 to 200 bunches per variety	prediction error from 0.06 to 0.15 on validation data set	Serrano, Roussel, et al. 2005
veraison of season n	block	t/ha	output	historical yield maps, post-harvest pruning weight, NDVI imagery, number of buds per plant, number of shoots per plant, number of bunches per plant	yield monitor installed on harvester, satellite imagery, 1-2m resolution manual counting	none	data driven : multi-variate linear regression	1 season, 9 blocks	r^2 from 0.2 to 0.72 on calibration data set	Martínez-Casasnovas, Martín-Montero, and Auxiliadora Casterad 2005
veraison of season n	plant	kg	input	weather data	weather station	none	data driven: mixed effects linear regression	15 seasons, 4 blocks, 8 replicates, 4 plants by replicate	r^2 of 0.8 between observed and simulated data	Zhu, Fraysse, et al. 2020
veraison of season n	plant	kg	systemic	NDVI imagery,WI imagery	NDVI : ground-based passive sensing	none	data driven : linear regression	2 seasons,7 blocks, 3 plants per block	NDVI : r^2up to 0.63, WI: r^2 up to 0.56 on calibration data set	Serrano, González-Flor, and Gorchs 2012, González-Flor et al. 2014
close to harvest of season n	territorial	hL/ha	input	weather data	weather : provided by a national service	none	data driven : window pane analysis and linear regression with cross validation	21 seasons, regional production records	r^2 of 0.82 and 0.85 on calibration data set	Molitor and Keller 2017

Use Case			Variables estimation							
Date	Spatial Scale	Yield unit	Indicators type	Indicators nature	Measurement method	Sampling method	Model	Experimental dataset	Announced results	References
close to harvest of season n	block	kg	output	proportion of fruit pixels	image analysis of manually taken pictures	none	data driven : linear regression	1 season, 10 plants with 2 stages of defruitation	0.73 of observed data variability explained by predicted yield on validation data set	Diago, Correa, et al. 2012
close to harvest of season n	block	kg /plant	output	actual yield per plant, historical geolocalized yield per plant	manual weighing	target sampling	data driven : principal component analysis and linear regression	1season, 3 blocks, 2 to 10 plots of 5 plants per block	estimation error lower than 0.1 for more than 5 sampling sites	Araya-Alman, Leroux, et al. 2019
close to harvest of season n	block	kg	systemic	NDVI imagery,yield per plant	manual weighing, airborne passive , sensing, 0.25m resolution	target sampling	data driven : linear regression with bootstrap	2 seasons, 1 block, 54 sites, 1m per site	r^2 of 0.33 and 0.46 on calibration data set	Hall et al. 2011
close to harvest of season n	block	kg /plant	systemic	NDVI imagery	airborne (UAV)	none	artificial intelligence : artificial neuronal network	1 season, 1 block	relative error of 0.05	Ballesteros et al. 2020
close to harvest of season n	block	t/ha	systemic	vegetation indices	satellite imagery (30m resolution)	random sampling	artificial intelligence : artificial neuronal network	2 seasons,4 varieties,31 blocks	$r^2$from 0.76 to 0.97 and RMSE from 0.11 to 1.45 on validation data set	Arab et al. 2021
close to harvest of season n	intra-block	kg/m	output	proportion of fruit pixels	on-board image analysis	none	data driven: linear regression	1 season, 16 plants with 4 stages of defruitation, 1 picture per meter	r^2 of 0.85 on calibration data set	Dunn and Martin 2004
close to harvest of season n	plant	kg	output	proportion of fruit pixels	on-board image analysis	none	data driven: linear regression	1 season, 1 block, 1 row of 30 contiguous plants	mean absolute error of 0.28	Lopes et al. 2016
close to harvest of season n	plant	kg	output	fruit pixel	on-board image analysis	none	data driven: linear regression	1 to 3 seasons, 6 blocks	yield prediction error from 0.025 to 0.29 on calibration data set	Nuske, Wilshusen, et al. 2014
close to harvest of season n	plant	t	output	number of nodes per plant ,number of bunches per plant, number of berries per bunch, number of seeds per berry, bunch mass, minimum december temperature	manual counting and weighing	blocks typology according to historical yield level	data driven: linear and non-linear regression	13 seasons, 8 blocks, 2 rows per block, 4 plants per row	mean absolute percent error from 0.07 to 0.12 on calibration data set	Folwell et al. 1994
close to harvest of season n	bunch	kg	output	proportion of fruit pixels	on board image analysis	none	data driven : linear regression	25 non-occluded bunches	error of 0.17	Font et al. 2015

The diversity of the experimental conditions under which the methods were established and validated as well as the criteria used prevents any comprehensive comparison of methods or results of yield assessment. However, it can be seen that a wide variety of use cases has been studied in the literature, ranging in time-scales from the season n-1 to periods close to harvest, and at spatial scales from individual vines to entire regions. Yield was expressed in either mass or volume units at the territorial scale, while it was mostly expressed in mass units at higher scales.

All three types of indicators (Fig. 2.4) were reported in the scientific literature. Input indicators mostly corresponded to weather data or block characteristics, such as density or topography, and were mainly used by mechanistic or artificial intelligence models, perhaps because of the complexity in the relationship between the input indicator(s) and the yield response. Systemic indicators mainly corresponded to vegetation indices and related studies were predominantly applied to the vineyard block, whole vineyard or regional scale. Indicators of water and nitrogen stress or biotic aggression were also reported. Output indicators mostly corresponded to yield components that were measured manually or measured using on-the-go image-based sensors. It appears that the diversity of measurement methods reviewed earlier in the paper has not yet been transferred to whole yield assessment processes.

It should be noted that few studies investigated sampling issues, even though the estimation of yield components at the vineyard block scale (or larger) using vine observations at sub-block scale is often proposed. Moreover, studies employing random sampling often seemed to be biased by practical constraints or by some form of expert knowledge that induced a subjective selection of measurement points.

Modelling is predominantly done using data-driven models that have focused on the use of the linear, uni- or multivariate models. Mechanistic and artificial intelligence models propose other functional forms. The bias due to data dependency or over-fitting was rarely taken into account since mixed effects or cross validation methods were rarely used. In addition, results were generally announced in terms of model fit to the data (e.g Coefficient of Determination), even though other selection criteria seem more interpretable from an operational point of view (e.g. Root Mean Square Error).

2.2.4.6.2 Some issues in yield assessment are still poorly addressed by existing methods

Challenges to be addressed by yield assessment methods in an operational context were identified in the first section of this paper. In addition to the challenges related to measurement, sampling and modelling processes, any given yield assessment also needs to : i) address temporal yield development, ii) address yield development as being a site-specific phenomenon and iii) account for operational constraints. The main practical implications of these challenges for the establishment of yield assessment methods are detailed in Table 2.7. They are used to comprehensively compare the methods that have been presented in Table 2.6.

Few methods have addressed the temporal nature of yield development. There has only been one recent reported use of time-series analysis for yield assessment (Arab et al., 2021), regardless of the type of explanatory variables used. Intra-seasonal variables, such as weather variables, are the most likely to support time series analysis because of their recording modalities (continuous records in 15 minutes to daily time steps). However, most data-driven methods that use weather variables were based on the computation of punctual or aggregated indicators, e.g. the mean temperature around the bloom period (Zhu, Fraysse, et al. 2020) or the amount of rain accumulated during a hundred growing degree days (Guilpart, Metay, and Gary 2014), which are considered independent for statistical analysis. Additionally, most mechanistic models can be applied at any time of the season as long as the data is available (Valdés-Gómez et al. 2009; Cola et al. 2014; Nogueira Júnior et al. 2018). They are not dynamic models as such, but they can be used to provide yield assessments at regular time periods. Finally, there have only been scientific and industrial proposals for yield estimation or prediction methods to date. No yield forecasting method for yield assessment has been proposed yet. Moreover, no information on the yield assessment uncertainty has ever been given. This implies that the actual methods in use in operational situations do not maximize the portion of yield observable variability that they address. They also do not meet the operational need to have a yield assessment whose uncertainty is characterized and announced.

The challenge related to the consideration of a site-specific system for yield development is only partly addressed by methods that have worked on improving sampling strategies (spatial or temporal) that are adaptable to various local indicators (De la Fuente et al. 2015; Araya-Alman, Leroux, et al. 2019). Methods that have worked to improve the measurement of explanatory variables, particularly those that offer automated measurement of yield components, assume the measurement of a specific indicator, which may be

different from the indicator historically used by the local vineyard/winery (Lopes et al. 2016; Liu, Cossell, et al. 2017). In the reported literature, there are few methods that try to deal with operational data as small and heterogeneous datasets of commonly performed field observations. This gap clearly shows that scientific viticulture research has yet to address the issue of using operational data to support a yield assessment method. Most methods have grasped the operational importance of proposing a yield assessment based on non-destructive (pre-harvest) observations that can be automated. Data-driven methods do assume a small number of parameters to be operationally accessible. This is not true for mechanistic methods, whose possibilities of operational implementation are much more limited. Operational implementation is also limited by the fact that approximately half of the methods reported did not provide a yield assessment at the spatial scale required by the operational needs, thus forcing the user to upscale the results with limited means and with additional sources of uncertainty. Finally, the fact that most of the required equipment was not accessible to the industry, and that most of the methods were not yet commercially implemented, showed that the transfer from scientific research to the industry on the subject of yield assessment is still very restricted.

Table 2.7 – Comparison of operational advantages of yield estimation and prediction methods reviewed in literature

Temporality			Site-specific		Operational								
Time series	Intra-seasonal indicators	Dynamic method	Site-specific indicators	Operational data sets	No required up scaling	Low measuring time	Automatisable	Non destructive	Few parameters	Accessible equipment	Uncertainty management	Already implemented on field	References
								x	x		prediction		Guilpart, Metay, and Gary 2014
	x				x	x	x	x	x		prediction		Cunha, Marçal, and Silva 2010
			x	x	x				x	x	prediction	x	Clingeleffer et al. 2001
	x		x		x	x	x		x		prediction		Sirsat et al. 2019
	x				x		x	x			estimation		Valdés-Gómez et al. 2009
	x				x		x	x			prediction		Fraga, Santos, et al. 2016
	x	x			x		x	x			prediction		Cola et al. 2014
	x	x					x	x			estimation		Nogueira Júnior et al. 2018
					x		x	x	x		prediction		Cristofolini and Gottardini 2000
	x				x		x	x	x		prediction	x	Cunha, Ribeiro, and Abreu 2016
					x		x	x			prediction		Liu, Cossell, et al. 2017
					x	x	x	x		x	prediction		Sun et al. 2017
													González-Fernández et al. 2020
	x			x		x	x	x	x		prediction		Fernández-González et al. 2011
			x							x	prediction	x	De la Fuente et al. 2015
							x	x			estimation		Di Gennaro, Toscano, et al. 2019
								x	x	x	estimation		Serrano, Roussel, et al. 2005
					x						prediction		Martínez-Casasnovas, Martín-Montero, and Auxiliadora Casterad 2005
	x					x	x	x	x	x	prediction		Zhu, Fraysse, et al. 2020
						x	x	x	x		prediction		Serrano, González-Flor, and Gorchs 2012,González-Flor et al. 2014
x	x			x	x	x	x	x	x	x	prediction		Molitor and Keller 2017
						x	x	x	x		estimation		Diago, Correa, et al. 2012
			x	x	x				x	x	estimation		Araya-Alman, Leroux, et al. 2019
						x	x	x	x		prediction		Hall et al. 2011
	x				x	x	x	x	x		prediction		Ballesteros et al. 2020
x					x	x	x	x	x		prediction		Arab et al. 2021
						x	x	x	x		estimation		Dunn and Martin 2004
						x	x	x	x		estimation		Lopes et al. 2016
						x	x	x	x		prediction		Nuske, Wilshusen, et al. 2014
					x					x	prediction		Folwell et al. 1994
						x	x	x	x		estimation		Font et al. 2015

2.2.5 Conclusion : what are the perspectives to improve yield assessment ?

The adaptation of yield assessment approaches to the production conditions of any commercial viticulture enterprise raises interesting scientific questions linked to the management of operational needs, constraints and data that the scientific literature only partially addresses. Based on the assumption that an entire survey of the yield assessment process is required to improve yield assessment in a commercial context, this paper reviewed issues and answers in the literature that have already been developed for measurement, sampling, and modelling (cf. Fig. 2.5). Comparing literature contributions to the operational needs and constraints of yield assessment has highlighted the need for new yield assessment thinking and methods that are readily

transferable to and between commercial, operational systems. Three still unaddressed scientific topics have been identified for their potential of yield assessment improvement based on operational data.

2.2.5.1 Aiming at operational relevance

First, there is a need to comply with the practical nature of operational constraints, which are common to all vineyards/wineries. This mainly refers to the ease of implementation of the yield assessment methods in the field and interpretability of the provided yield assessment information. Ease of implementation requires the yield development conceptualization to be as simple and robust as possible and a parsimonious number of accessible indicators to be taken into account. According to the methodology presented by Lamanda et al. 2012, the effects of input and output factors would be easier to model and would require fewer parameters to be taken into account. Yield components, which can be considered as outputs of the yield development system, have been foremost and most commonly studied in the literature. Therefore, numerous studies have sought to automate their measurement (Diago, Correa, et al. 2012; Nuske, Wilshusen, et al. 2014; Aquino, Millan, Gutiérrez, et al. 2015; Aquino, Millan, Gaston, et al. 2015; Aquino, Millan, Diago, et al. 2018; Liu, Cossell, et al. 2017). At this time, these efforts do not allow for an exhaustive measurement of any component, and the observation of yield components remains subject to spatial sampling issues (Carrillo et al. 2015; Araya-Alman, Leroux, et al. 2019; Oger, Vismara, and Tisseyre 2021). Similarly, with yield components being time-defined and representative of evolving indicators of the dynamic process of yield development, their observation also involves temporal sampling issues (Oger, Vismara, and Tisseyre 2021), which combines biological and operational difficulties (cf. Table 2.3). Given this, the potential of considering input factors, such as weather data, that can be continuously and automatically recorded, has been poorly explored (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017; Zhu, Fraysse, et al. 2020).

Secondly, good interpretability of a yield assessment involves providing unequivocal information as accurately as possible. The objective is to remove the need for the user to exercise their judgment, and therefore their subjectivity, in their understanding of the provided yield assessment. This can be achieved by ensuring that the model can manage as much of the observable yield variability as possible within the model by using indirect or direct forecasting approaches. Good interpretability also implies an ability to express the expected yield with an operationally used definition and units, as well as at the correct spatial and temporal scales for the user. These last two points commit to improving knowledge of yield development from both a local and temporal perspective.

2.2.5.2 Accounting for yield development temporality

Operational datasets that may support a local modelling of yield development often contain time series. So far in the literature, these times series have been used to compute indicators, based on a few phenological stages or time steps, and are often considered as independent when analysed with linear regression analysis (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017). This approach significantly restricts the potential of time series data analysis in yield assessment by considering only part of the contained information. However, yield development has been recognised in this paper as a dynamic process that includes trajectory or memory effects due to temporal inter-dependencies in grapevine physiology. Thus, the use of novel methods, such as specific or functional time series analysis, could help in further extracting information from time series data. Nonlinear relationships should also be investigated in order to improve the modelling of some biological yield-determining phenomena. Such methods could advance the detection of external influences to be preferentially considered in a site-specific model of yield development (Laurent, Baragatti, et al. 2019). However, the multiplication of extreme and unusual climatic events in the context of climate change raises the question of the data history on which to train any yield assessment model.

Furthermore, leveraging time series requires improved noise reduction methods in the analysis that may be induced by phenological shifts between blocks or years of the same dataset. Computing Growing Degree Days to express time in a more grapevine phenology consistent metric has been a first answer to this issue (Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017). However, recent work has challenged the way thermal indicators are computed (Parker, García de Cortázar-Atauri, et al. 2020; Camargo-Alvarez

et al. 2020) or the time step at which temperature is summarized (Gaiotti et al. 2018; Gouot et al. 2019a; Gouot et al. 2019b) to model grapevine physiology. This should inspire future work aimed at expressing the timing of grapevine development in a more precise and accurate metric. For the expression of time to be locally relevant, it also seems important to adapt the calculation of these metrics to the site-specific conditions e.g. plant material, pedo-climate, topography, orientation and cultural practices. The date when the reproductive cycle is considered to start may also be reconsidered in favor of seasons prior to the season n-1 (Duchêne, Jaegli, and Salber 2003; Vaillant-Gaveau et al. 2014; Pagay and Collins 2017).

Finally, the literature has identified some key steps in the dynamics of yield development when a portion of the final yield variability is fixed by the stable implementation of successive yield components (c.f. Fig. 2.3). These steps permit a consideration of what can be expected of yield assessment methods in terms of accuracy. Conducting a yield decomposition analysis (cf. Eq. 2.1) from a temporal perspective on a local dataset should provide an indication on the accuracy that can be expected at each yield component implementation date through the portion of yield variability explained by this component. For example, if the number of bunches per vine is found to temporally and locally explain 40% of final yield variability, then around 40% of accuracy could be expected from yield assessment methods at the date when inflorescences are observed in the field. Such a number should nevertheless be only considered as an approximation for reasons of physiological dependencies in the yield development cycle that may affect other, subsequent components (Duchêne, Jaegli, and Salber 2003; Duchêne, Jaegli, Salber, and Gaudillère 2003; Pagay and Collins 2017). In order to accommodate operational needs and to dynamically monitor yield development, it seems relevant to develop a yield forecast at several dates in the season, with accordingly increasing accuracy and certainty. Depending on the case, these dates could correspond to the progression of certain yield components, to the availability of key ancillary data or to periods that are identified as highly influential at the end of which it would be justified to update the yield forecast.

2.2.5.3 Proposing a site-specific approach to yield development

Operational yield definition, units and scales vary from one commercial vineyard/winery to another. Similarly, there will be enterprise-specific differences in other agronomic indicators and ancillary data that may be collected during the season. As a result, the use of operational data in yield assessment models is likely to generate new knowledge on yield development by allowing a wide variety of situations to be studied with large datasets. Data heterogeneity among vineyards/wineries should not be a reason to avoid using operational data in yield assessment. However, a sufficiently flexible yield assessment method is needed to allow for different site-specific schemes of yield development to be considered. This site-specific scheme should depend on the main site-specifically identified external influences and on the available data. To identify the site-specific variables that should be used as main explanatory variables, data-driven approaches may be used to support expert conceptualization. Therefore, there is a real issue for vineyards/wineries to consolidate data, whose modalities of data collection remain constant over time and space. Such a consolidation should allow for the dataset to be analyzed in its entirety as a spatio-temporal dataset and to rigorously support data-driven approaches. Faced with the growth of artificial intelligence, the challenge for research is also to develop more complete and objective yield assessment methods when supported by sufficiently large datasets, without losing the possibility of interpreting the results at a local level so that they can be fully exploited from an agronomic point of view.

2.2.6 Acknowledgments

This work was funded by Fruition Sciences and supported by the French National Research Agency under the Investments for the Future Program, named # DigitAg and referred as ANR-16-CONV-0004. The authors would also like to thank Gilles LeMoguédec for his revision of the statistical elements presented in this paper. Fig. 2.5 was generated from real-world data provided by Dr Terence Bates, Cornell Agri-Tech.

2.3 Presentation of the research questions addressed in the PhD project

2.3.1 The PhD project focuses on yield modelling approaches based on farm data

Through the case study of yield assessment in viticulture, the previous article formalized the implementation of agronomic research in three steps: measurement and estimation of agronomic indicators and modeling of a response variable based on these indicators. It revealed that most of the research works aiming at an operational application are positioned on the measurement and sampling issues and little on the modelling ones. Therefore, there is a significant opportunity to work on modelling approaches to improve yield forecasting in an operational context. Moreover, the literature review highlighted the potential of leveraging farm data to address operational purposes. There are two types of farm data: those that have already been collected, whose measurement, sampling and even traceability are imposed and can not be revisited, and those that will be collected in the future, for which it is possible to decide on the measurement and sampling processes. **The PhD project focuses on the already collected farm data. As a result, a wide variety of previously collected data sets can be accessed.** This decision also implies the opportunity to immediately start working on improving the research answer to the field demands in terms of yield assessment. It also addresses the issue of valuing already collected data that is currently lying idle in some paper or computer files. Therefore, this PhD project aims at providing an answer to farmers who often question the benefit of collecting data when they are not analyzed and when most of their decisions are made on the basis of expertise. Finally, it is a way to prepare for the future by anticipating the processing of data that will be collected and eventually providing thoughts for improving the measurement and estimation of indicators in order to facilitate the modeling step later on.

2.3.2 The PhD project aims at developing an operational, temporal and site-specific approach of yield development

Challenges regarding the modelling process based on farm data and for operational purposes were spotted in section 2. 2. These challenges were classified into three categories : i) aiming for operational relevance, ii) accounting for yield development temporality and iii) developing a site-specific approach of yield development. Therefore, the scientific questions of the PhD project relate to the implementation of an agronomic research addressing these three requirements with the example of yield forecasting in viticulture: **i) which conceptual model of grape yield development could comply with such an operational, temporal and site-specific approach of yield development ? ii) which yield assessment method could be implemented to account for such a model ? iii) which statistical modelling approaches could be used to deploy such a model on the basis of farm data ? iv) how to mobilize such statistical approaches into the yield forecasting method ?**

A mapping of the challenges identified in chapter 2 to the scientific questions is proposed in Fig. 2. 7.

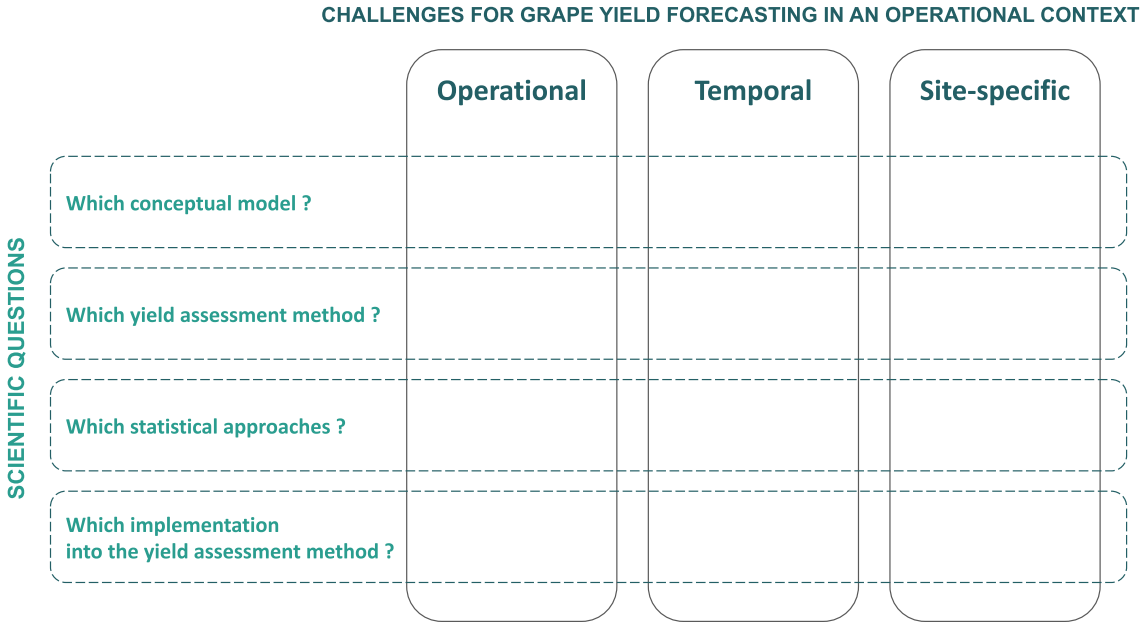

Figure 2.7 – Challenges and scientific questions for grape yield assessment in an operational context addressed in the PhD project

2.3.2.1 Which conceptual model of yield development ?

Addressing the operational context of commercial vineyards and wineries implies addressing a wide variety of yield development conditions. Taking the definition of marketable yield (Vigiuer et al. 2018), these conditions correspond to the interaction of numerous factors including plant internal factors, environmental factors, cultural practices and possibly transport and wine making practices. Considering that each of these factors has several modalities, which will interact in different ways with all the modalities of the other factors, the number of possible combinations to describe the conditions of yield development in an operational context becomes unmanageable. Moreover, there is no capacity to collect the agronomic indicators that would inform all the corresponding variables in commercial farms. In addition, grape yield development is a dynamic process which takes place over two years. It may involve trajectory and memory effects in relation to the adaptive physiology of grapevine. This dynamic character is often transcribed by mechanistics models (Valdés-Gómez et al. 2009; Cola et al. 2014; Nogueira Júnior et al. 2018). However, their parameterization also requires a high number of variables to be informed (e.g. leaf area, radiation use efficiency, coefficient of light extinction, soil water capacity etc.) that are not available in farm data sets.
These constraints related to the operational context prevent the conceptual models commonly used in agronomy from being applied. **Instead, they call for a new conceptual model which would i) be site-specific so as to address only a small portion of the possible yield development conditions and ii) take the yield temporality into account while being deployed from farm data.** Therefore, a first scientific question addressed the conception of such a yield development model.

2.3.2.2 Which yield forecasting method ?

Yield development being dynamic, the operational decisions that are made on the basis of the expected yield performance also require a dynamic approach of yield assessment. However, the already set portion of the final yield performance is evolving during the season, which implies that yield assessments made at different dates may vary in certainty. Moreover, the data quality of farm data sets also vary from one vineyard to another in relation with measurement and sampling issues. For the yield assessment to be relevant in supporting operational decision-making, it must then involve a characterization of the uncertainty. Thus, according to the definition set in section 2, a yield forecasting method is required. Still, **this forecast should be achieved while mobilizing a site-specific and temporal model of yield development. Therefore, a second scientific question addresses the design of such a grape yield forecasting method.**

2.3.2.3 Which statistical modelling approaches could be used to deploy such a model ?

The implementation of a site-specific yield forecasting method can only be done using farm data. These data are characterized by larger volumes, corresponding to more blocks and more years, than most experimental data sets. These data also contain time series. However, they are also characterized by a lower quality in relation to their measurement, sampling and traceability. Although larger than experimental data sets, these data sets still correspond to small data sets from a statistical point of view and contain numerous spatial and temporal interdependencies. In this PhD project, it is assumed that the volume of data and the use of adapted statistical methods will allow farm data to be leveraged despite these limitations. Therefore, a third scientific question addresses the selection of such statistical methods.

2.3.2.4 Which statistical modelling approaches could be used to deploy such a model ?

Once the statistical approaches adapted to farm data have been identified, a fourth scientific question relates to their implementation into analyses which will constitute the analytical steps of the yield forecasting method : on which data should they focus and in which order ?

2.3.3 Presentation of the research approach of the PhD project

To address these four scientific questions, the research approach was organized into 5 stages.

2.3.3.1 Collection of farm data sets and consolidation of field feedbacks about grape yield forecasting

A call for volunteers was sent by email to Fruition Sciences network of customers and prospects. An oral prospection was also made during the 2018 Vinitech-Sifel exhibition. A technical manager from each volunteer vineyard/winery was then interviewed. The aim of the discussions was to define the ideal yield forecast from an operational point of view. The questionnaire used to guide the interviews is given in Appendix 1. It purposely contained few questions in order to conduct a relatively free interview and not subjectively restrict the vision of yield forecasting that the vineyard/winery could have. A first evaluation of the available data and of the vineyard commitment to share them was also conducted.
20 vineyards/wineries were interviewed. 9 of them had available data and accepted to share them. A data agreement was established for each vineyard wishing to share its data. It was verified that the shared data did not fall within the scope of personal data as understood by the European General Data Protection Regulation (2016/679).

During the first year of the PhD project, an important work of data cleaning and characterization was then performed on the shared data sets. The data characterization was conducted in two stages. First, for each vineyard, it was necessary to fully understand what each indicator collected corresponded to i.e. to understand i) the reason why the indicator was chosen, ii) the protocol according to which it was measured and estimated and iii) the reason why such a protocol was designed. This work was achieved through iterative interviews and email exchanges.
Secondly, a comparison of the data sets was carried out. It aimed at identifying i) the variables and indicators that were always present in farm data sets, they are called essential variables and indicators and ii) the variables and indicators that were available only in few certain cases, they are called optional variables and indicators. This comparison also identified the diversity of indicators collected to inform both essential and optional variables. The consequence of their collection protocols on the data set quality was also investigated, in particular in terms of consistency in time and space.

2.3.3.2 Selection of case study data sets

3 of the 9 collected data sets were selected to represent the diversity of farm data sets on which the yield forecasting method should be able to work. First, the data sets were selected for their different variants of the conceptual yield development model (Fig. 2.8). They represented different environmental and vineyard management settings. The indicators that were available to inform essential variables also varied, which was likely to lead to different model training. Secondly, the data sets were selected in order to represent different

production contexts as well as different data sets volume and quality. The data set volume was defined according to the number of blocks and years in the data sets. The data set quality was defined according to the data consistency in time and space, the precision of the measurement and sampling protocols (qualitative assessment from the interviews) and of the data traceability as well as the presence of evident outliers in the data.

2.3.3.3 Definition of a conceptual model for yield development

Based on the study of the collected data sets, a conceptual model of yield development was proposed to gather knowledge about yield development and contribute to create a generic method of yield forecasting able to i) address as many vineyards/wineries as possible and ii) fulfill the criteria of operational relevance, site-specificity and temporality from farm data sets. It is presented in chapter 3.

2.3.3.4 Design of a grape yield forecasting method

On the basis of farm data characterization and of the conceptual model of yield development, a design for a grape yield forecasting method was proposed (Chapter 3). The analytical steps that compose its implementation were also specified.

2.3.3.5 Research implementation for each step of the yield forecasting method

The development of the yield forecasting method required to address two scientific issues prior to the forecasting model as such. The two corresponding research axes are presented in specific chapters (chapters 4 and 5). For each chapter, working hypotheses are proposed in a first section. These hypotheses were first empirically validated in a preliminary work which is presented in the second section of the chapter. It corresponds to an oral communication in a conference to gather peer feedback. A method or analytical framework addressing the scientific issue was then developed and tested on the three selected data sets. This work was built to be submitted as a research article in a scientific journal and corresponds to the third part of the chapter. Finally, a conclusion provides a perspective on the scientific contributions of each research axis, in relation to the hypotheses announced at the beginning of the chapter.

Chapter 3

Designing a yield forecasting method from the characterization of farm data sets

3.1 Intention note

Chapter 3 presents the first results of the PhD project. In section 3.2, the discussions that were held with commercial wineries and the data collected thereby are summarized. **A first working hypothesis was that the collected data sets allowed a good representation of the different data sets which can be found in commercial vineyards/wineries.** Following this hypothesis, criteria for characterizing any farm data sets were established from the study of the collected data sets. This characterization included a listing of essential variables and corresponding indicators and a non-exhaustive listing of optional variables and corresponding indicators. **A second working hypothesis was that the data present in all data sets could be sufficient to support an operationally relevant yield forecasting method. This implied that any vineyard possessing essential data could have access to a yield forecast, whose precision would then depend on the volume and quality of its data set.** Thus, a conceptual model of yield development was proposed on the basis of the listing of essential variables set in section 3.3 of this chapter. Subsequently, the integration of this yield development model into an actionable yield forecasting method is presented in section 3.4. This model was designed in order to meet the criteria of operational relevance, site-specificity and temporality on the basis of farm data, as recommended in chapter 2. Finally, the statistical implementation steps of such a method were defined in section 5. They are specifically presented in the following chapters.

3.2 Presentation of the material for the PhD project

3.2.1 Characterization of the studied vineyards

3.2.1.1 The vineyards/wineries showed two main motivations for responding the call for volunteers

About twenty vineyards participated in qualitative interviews. They were mainly situated in France (Bordeaux, Languedoc and Provence regions), in California (Sonoma, Napa and Sacramento regions) and in Israel. They were mainly oriented towards the production of qualitative wines and managed from 20 to 100ha approximately. They came forward for two main reasons: i) they had already started collecting data, sometimes a long time ago, and had already established their own empirical method to estimate or predict yield but they considered it unsatisfactory with regard to their support decision-making needs or ii) they perceived a need to assess future yield in another way than by expertise but had not started collecting data yet and were looking for guidelines to start with.

3.2.1.2 The vineyards/wineries expected different benefits from an improved yield assessment

For some vineyards, the yield assessment error and therefore the need for improvement was financially quantified, for example through the negotiation of short-term loans in advance. This was particularly true for wineries that buy grapes at the harvest time. Indeed, every deviation from the yield forecast meant as much money to be rapidly found, with less ability to negotiate interest rates than with anticipation. Other wineries quantified the benefit of a yield assessment improvement through the proportion of qualitative grapes/juice that had to be vatted with lower quality grapes/juice in order to fill the tanks at the end of each harvest day or the day after.
About half of the interviewed vineyards/wineries had developed an empirical yield assessment method based on a yield decomposition into a multiplication of yield components as described by Clingeleffer et al. 2001 and De la Fuente et al. 2015.They had heard that 60% of the yield performance was established in year n-1 in relation with the inflorescences development and 30% very close to the harvest in relation with the berries development. Therefore, they were interested in a validation of this statement in order to judiciously position their data collection effort. They also wanted to know if they could count on such a reliability of yield assessment at the time of deciding the cultural practices. These practices mainly referred to pruning and bunch thinning for the first yield assessment and fertilization, irrigation and date of harvesting in order to not let berries dehydrate for the second yield assessment. Finally, most wineries had never made the effort to quantify the benefit of a yield assessment improvement. However, they knew that such a benefit was spread over a wide range of operations, from the vineyard to the marketing and even the strategic management of the vineyard/winery. The results of these discussions were transcribed in the literature review about yield forecasting in an operational context presented in chapter 2.

3.2.1.3 The vineyards/wineries had different needs and constraints to perform the specifications of an ideal yield assessment

The results related to this topic were transcribed in the literature review about yield forecasting in an operational context presented in chapter 2. Several use cases of a yield assessment were identified according to the date, spatial scale and unit the yield information should comply with. It highlighted that the yield assessment has to be performed during periods of high workload and had to be quick and easy to perform. Moreover, a good adoption of the yield assessment method would be encouraged by accessible equipment and protocol, in terms of cost and competences.

3.2.2 Characterization of farm data on the basis of the collected data sets

Among the interviewed vineyards, 9 of them had data to share and accepted to share them. On the basis of their explorative study, variables and their corresponding indicators were classified into two categories : i) the variables and indicators which were observed in almost all the data sets composed a so-called essential data set and iii) the variables and indicators which were found in some data sets but not in others were considered as composing the optional part of data sets. They are non-exhaustively listed in Table 3.1.

Table 3.1 – Non-exhaustive listing of the variables and indicators which may compose any farm data set

Variable type	Essential indicators	Optional indicators	Temporal scale	Spatial scale
Block characteristics	Topographic context	Exposition	sub-annual	block
	Area	Soil characteristics	sub-annual	block
	Variety, rootstock		sub-annual	block
	Plantation density	Number of missing vines	sub-annual	block
	Plantation date		sub-annual	block
	Pruning system		sub-annual	block
Yield	Final yield performance	Number of berries per bunch	annual	block
	Bunch number per vine	Bunch mass	annual	block
		Berry mass	annual	block
Weather	Temperature	Radiation	daily	vineyard
	Rain	Relative humidity	daily	vineyard
Water status		Soil tension	daily or infra-annual	vineyard
		Water potential	daily or infra-annual	vineyard
		Sap flow	daily or infra-annual	vineyard
Nutrition		Soil analyses	annual	block
		Petiolar analyses	annual	block
Phenology	Budbreak, bloom	Any other phenological stage	annual	block
	and veraison dates	Shoot growth	annual	block
Cultural practices		Pruning (date and intensity)	annual	block
		Bunch thinning (date and intensity)	annual	block
		Cover cropping (area, date of destruction)	annual	block
		Fertilization (dates and doses)	infra-annual	block
		Irrigation(dates and volumes)	infra-annual	block

3.2.3 Selection of three data sets for further study

3.2.3.1 Presentation of the 9 collected data sets

The characteristics of the 9 vineyards which shared their data and their corresponding data sets are summarized in Table 3.2.

Table 3.2 – Characteristics of the 9 vineyards which shared their data and of their respective data sets.

		1	2	3	4	5	6	7	8	9
Vineyard	Localization	France	France	France	France	France	France	USA	USA	Israel
	Climate	oceanic	oceanic	oceanic	oceanic	semi-arid	semi-arid	semi-arid	semi-arid	semi-arid
	Irrigation	no	no	no	no	no	no	yes	yes	yes
	Vineyard consistent management	yes	yes	yes	yes	yes	yes	yes	yes	doubtful
Data set	Presence of all essential varibales	yes	yes	no	no private weather station	yes	no private weather station	yes	yes	yes
	Presence if optional variables	yes	yes	yes	yes	no	yes	yes	no	yes
Data volume and quality	Number of blocks	79	86	75	51	36	84	41	17	132
	Number of years	26	6	23	19	11	3	15	13	18
	Missing data	no	no	yes	yes	yes	yes	no	yes	yes
	Data traceability	good	good	good	doubtful	good	doubtful	good	doubtful	good
	Presence of evident outliers	no	no	no	yes	no	no	no	yes	yes

6 of the 9 vineyards were localized in France, 2 in California, USA and 1 in Israel. They were divided into two main types of climate : oceanic and semi-arid. Their localization, sometimes inducing specifications of geographical indications to be respected, and their climate mostly determined whether they irrigated or not. This information is important to appreciate the behaviour of the vineyard-specific system in the conceptual model of yield development. The non-limitation in water indeed modulates the influence of most of the other variables, whether categorized as input or systemic. 8 out of 9 vineyards presented all the essential variables although 2 of them had weather data from a station localized outside the vineyard. 7 vineyards also presented optional variables, which varied from one vineyard to another.

The vineyards presented 17 to 132 blocks and 3 to 26 years in their data sets. Please note that the announced number of years corresponded to the duration of the longest time series, which may not be true for the times series of all the variables of the same data set. In particular, most data sets counted more years of yield data than weather data. 6 out of 9 data sets were consistent according to years i.e. the quantity and quality of data was consistent for a given year. Only 2 data sets were consistent according to the blocks i.e. the data quantity and quality was consistent between all years of a given block. This was explained during the interviews by a variety of reasons going from a frost that cancelled the whole production of the year to a logistic failure (omission or too high workload). These reasons were year-dependent and not block-dependent. Therefore, the data sets were more consistent in terms of years than blocks i.e. data corresponding to almost all the blocks were present for one year and not for another year while for a given block, data were present or absent depending on the years. The hypothesis of a consistent vineyard management at the vineyard scale seemed valid for 8 data sets. It could not be verified in the last case because the management of a part of the vineyard was performed by an exterior company for a few years included in the data set. It was very difficult to match the data scattered in various files for 3 vineyards. Therefore, the quality of the data and of their correspondence in the data set was considered as doubtful because many assumptions had to be made to match the data. Finally, three vineyards presented a warning proportion of evident outliers, in particular in the weather data. For example, several periods greater than a week of evident outliers in a one year of temperature data invited to question the quality of the rest of the data set.

3.2.3.2 Presentation of the 3 selected data sets

3 data sets were selected in order to study different production contexts as well as data sets of different volume and quality. 2 of them were composed of different ranches i.e. different groups of blocks spaced a few kilometers apart.
Vineyard 7 will be named vineyard A henceforth. It represented a vineyard in a semi-arid region and using irrigation. It counted a relatively small number of blocks which were split into 4 ranches with different environments due to different topography (hillside or valley). It presented a relatively good consistency between blocks and years.
Vineyard 9 will be named vineyard B henceforth. It represented a vineyard in a semi-arid region and using irrigation but in a different country from vineyard A and localized at a higher altitude. It counted a relatively high number of blocks which were split into 4 ranches with similar environments. It presented a relatively low consistency between blocks and years with a lot of missing data.
Vineyard 1 will be named vineyard C henceforth. It represented a vineyard in an oceanic region and with no irrigation. It counted a relatively high number of blocks which were all situated in the same ranch. It presented the highest consistency between blocks and years of all the collected data sets.
The three vineyards presented common varieties i.e Cabernet-Sauvignon, Merlot, Petit Verdot and Syrah but counted a wide range of rootstocks.
The characteristics of the 3 vineyards are summarized in Table 3.3

Table 3.3 – Characteristics of the 3 vineyards selected as case studies in terms of environment, organization and data set.

	Vineyard A	Vineyard B	Vineyard C
Localization	California, USA	Israel	Bordeaux, France
Latitude	38	32	45
Altitude	210	770	20
Topographical context	Valley and Hillside	Plateau	River bank
Type of climate	semi-arid	semi-arid	oceanic
Type of soil	loam or clay, more or less rocky	volcanic	loam or clay gravels
Number of ranches	4	3	1
Number of blocks per ranch	3,25,8,5 (41 in total)	58,32,42 (132 in total)	79
Plantation date	from 1994 to 2012	from 1984 to 2012	from 1952 to 2012
Average plantation density (stocks/hectare)	2220	2500	8700
Average aera	0.6	1.7	1.1
Varieties	Cabernet-Sauvignon Merlot Petit Verdot	Cabernet-Sauvignon Merlot Syrah	Cabernet-Sauvignon Merlot Petit Verdot
Years of concomitant phenological, yield and weather data	2007 to 2018	2000 to 2018	2001-11 and 2014-15

With each data set, a data selection was also performed. Data corresponding to juvenile blocks i.e. which were less than 3 years old, were discarded. In addition, blocks presenting data corresponding to less than three years (not including the juvenile years) were also discarded. Some blocks had also to be discarded for intricate tractability. For example, a block could be divided into 2 for a given year, each part being managed as a block in itself, and into 3 the year after. The corresponding areas were not reported and there was no information on the difference in the vineyard management of the subdivided blocks. In addition, some varieties corresponded to a very low number of blocks in the vineyards, especially for white varieties. The related blocks were also discarded. In contrast, blocks with a large amount of data but recently destroyed were maintained in the data sets.
Moreover, years of yield or phenological data that were not concomitant with some weather data were discarded. However, while respecting this requirement, yield and phenological data could correspond to different years.
The figures reported in Table 3.3 do not include discarded data. As an indication, approximately 40, 80 and 60% of the data sets of vineyards A, B and C respectively had to be discarded. It was mainly due to missing years of weather data compared to yield and phenological observations for vineyard A and C. It was due to low block numbers for many grape varieties and above all to a lot of missing data among the blocks and years for no apparent reason in the case of vineyard B.

3.2.4 Presentation of the conceptual model of yield development

The block scale is the spatial scale for most grape cropping systems and a common denominator for most wine-making assemblages. Therefore, it is used to establish the conceptual model of grape yield development. Following the protocol for the conceptualization of agro-ecosystem proposed by Lamanda et al. 2012, the conceptual model focuses on one output variable that was the yield performance at the block scale (Fig. 3.1).
The studied system was defined as including the vines and all the influencing variables that are retro-influenced by the vines. It non-exhaustively included i) physiological variables in relation with the vegetative and reproductive developments, ii) variables in relation with water, nutrients and minerals availability, iii) variables referring to the plant sanitary status, iv) cultural practices that were decided as an adaptation to the state of the whole system as it has just been described and v) eventually wine-making operations. For a given year and across years, a certain consistency in the way this system works is assumed at the vineyard level (or possibly at the winery level with certain precautions related to its spatial organization). This consistency could be summarized into a vineyard effect which was considered as not decomposable due to lack of data and knowledge regarding the multiplicity of possibilities to be studied. This vineyard effect should then be estimated as a whole at the vineyard scale.
Input variables influence the system behavior without being retro-influenced by it. They non-exhaustively corresponded to the characteristics of each block : plant material i.e. variety and rootstock, plantation date and density, type of soil, exposition, topography etc. Input variables also included climate factors such as temperature, rain, relative humidity, radiation etc. As an integration of block characteristics and climate influence that determined the system response to numerous influences, the vine phenology was also understood as an input variable. Finally, the plant memory and past trajectory in terms of yield development were also considered as input variables because they couldn't be changed any more. They were indicated by variables such as the previous yield performances or already settled yield components. These input variables were estimated at the block scale and therefore corresponded to a block effect which mitigates the vineyard effect defined by the system.
The resulting conceptual model is illustrated in Fig. 3.1.

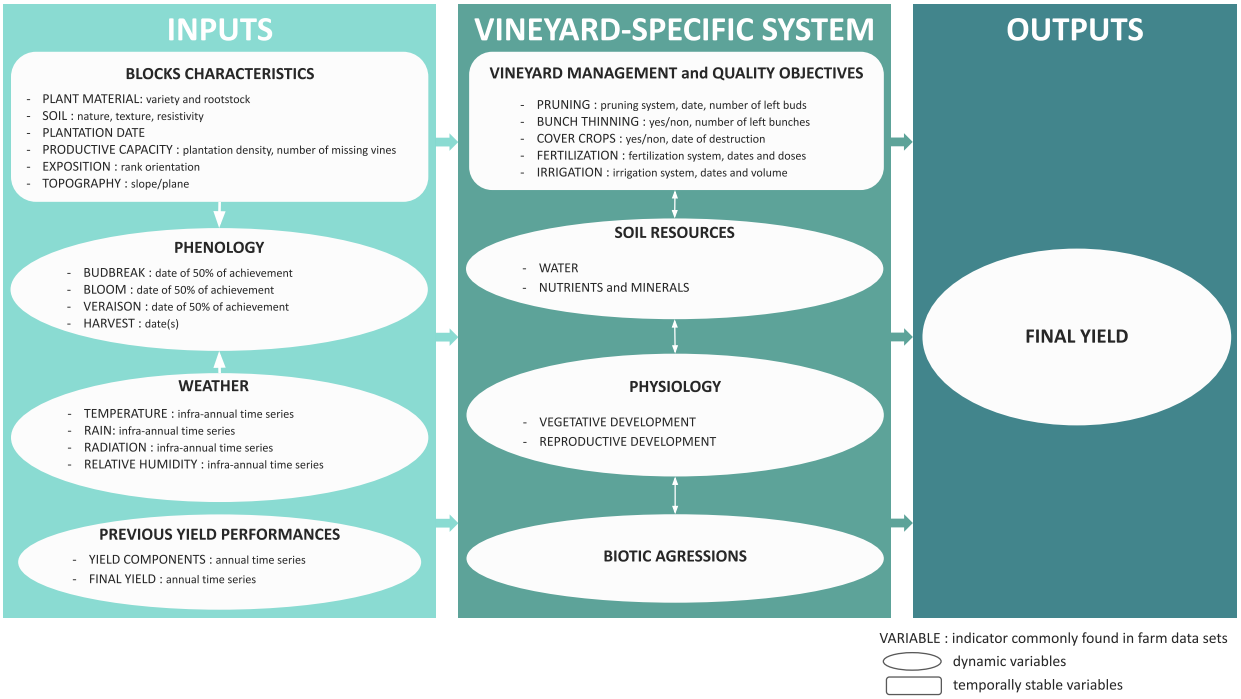

Figure 3.1 – Conceptual model of annual grape yield elaboration from literature in an operational context, only the input and output variables are studied and the vineyard-specific system is considered as a black box.

Most of the farm data encountered in the field belong to the class of input variables. Therefore, the system behavior can be estimated like a black box with a data-driven approach at the block scale relying on a vineyard-specific effect. The variables involved in the model can vary from one vineyard to another depending on their available data. The indicators which inform each variable may vary in their measurement and sampling protocol. Therefore, they may vary in their definition and unit depending on the vineyard. For example, the yield performance can be measured in tons per acre before or after transport of the harvest or it can be measured in hectoliters per hectare after pressing or after racking. The weather data can also be measured with several stations localized within the vineyard or with only one station localized outside the vineyard.

3.2.5 Presentation of the yield forecasting method

3.2.5.1 General specifications of the the yield forecasting method

The variables and indicators that can be available to implement the previous conceptual model can vary from one vineyard to another. Most operational constraints are defined at the vineyard or ranch scale but the yield performance is generally managed at the block scale. **Therefore, the yield performance was chosen to be modelled at the block scale by mobilizing a vineyard-specific model that will take block variables as covariates. The model was considered as vineyard-specific because it was trained on the own farm data set of each vineyard in line with its unique operational context. Therefore, the intended method was a method that could be adapted to different vineyards in order to build a yield assessment model only valid for each of them respectively.**

Complying with the temporality of grape yield development and of operational needs, the yield assessment should be dynamic. However, it was assumed that a yield forecast was not equally valuable depending on the dates on which it was provided. Indeed, based on the literature highlighting periods of increased sensitivity in the yield development process (cf. Fig. 2.3), it could be hypothesized that a yield forecast following a highly

influential period would provide more information on the evolution of the expected yield than following a period with little influence. In addition, a yield forecast produced when a decision has to be made regarding cultural practices in the vineyard would be more operationally relevant than if the forecast was established according to a random timing. **Thus, the intended method should be able to produce several yield assessments judiciously positioned in the season. Each one would involve a different but still vineyard-specific model, based on the data available i.e. which has already been collected at that time of the season.**

For this yield assessment to be a relevant aid in operational decisions in the field, it must be accompanied by a characterization of its reliability. In other words, this yield assessment should be a yield forecast. This will be expressed by a distribution of probable yields instead of a unique predicted value.

An illustration of the expected deliverable for a given block is given in Fig. 3.2.

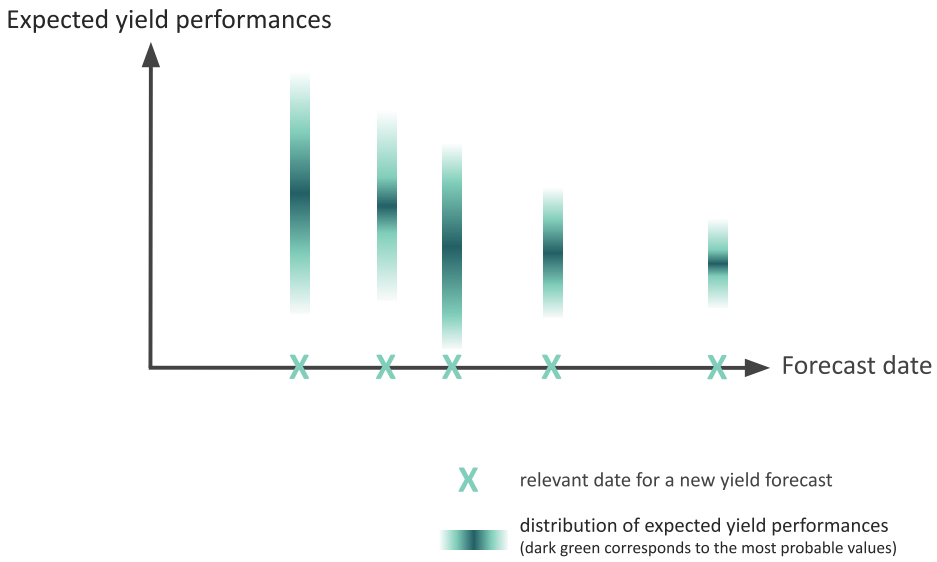

Figure 3.2 – Target design of the yield forecasting method implemented for a given vineyard over several, here 5, forecast dates

The most probable yield performance would decrease along the season in relation with the limiting and reducing influences of the environment and cultural practices. The variance of the expected yield performances would also decrease with the exception of extreme events such as for the third prediction date. After such events, caused damages may be difficult to quantify and the vines must be given time to recover. This may lead to greater uncertainty about the yield performances to be expected. The variance of the expected yield performances close to harvest should be the smallest of the season.

3.2.5.2 Scientific issues explored in the PhD project

When a new farm data set is received, it is characterized according to the criteria set out in this chapter (see Table 3.1). Essential data have then to be processed in order to be further integrated as covariates in a yield forecasting model.

In particular, a common feature of the studied farm data sets was the inclusion of time series. These time series can be classified into two categories: i) series with an annual time step that have one observation per year, for example time series of yield or yield components observation and ii) series with an

infra-annual time step which have numerous observations throughout the season, for example time series of weather data (cf. Table 3.1). Data series with an infra-annual time step are interesting because they provide continuous information and therefore allow to run more yield forecasting models than other indicators, which are positioned only once in the season. However, they are not trivial to analyze for several reasons. The first reason is that the infra-annual data series are collected simultaneously with the ongoing yield development. Therefore, their comparison within any analysis implies that the rate of yield development is the same from one series to another i.e. from one block to another or from one year to another. It is known from numerous literature works that this is not the case when time is expressed according to the Gregorian calendar Wang, 1960; Cross and Zuber, 1972; Grigorieva et al., 2010). Therefore, the time series of infra-annual data must be synchronized according to the rhythm of yield development during or after their collection. Furthermore, it is assumed that yield development follows a site-specific pattern, at least at the vineyard scale. Therefore, the synchronization method should take this site effect into account. **Therefore, a first research axis focused on the site-specific synchronization of data series at an infra-annual time step. It is addressed in chapter 4.** The focus is set on weather data series because they are essential data, found in most farm data sets, and because they correspond to input variables whose influence on the yield elaboration is recognized as prevalent.

The second reason why time series data at an infra-annual time step is not easy to analyze is that they contain a lot of information, more or less relevant with regard to yield development and also a lot of noise. However, the information extracted from the time series is intended to be incorporated into a yield forecasting model with other variables. Therefore, it should be selected with parsimony so as not to systematically over-fit the yield forecasting model. Finally, a third reason relates to the fact that time series can not be analyzed with classical statistical analysis methods because of their autocorrelation. **Thus, a second research axis addresses the valorization of infra-annual time series data synchronized according to the rate of yield development. It aims at extracting relevant and parsimonious information regarding yield development. This research axis is addressed in chapter 5 and results in the building of weather indicators.**

The two research axis developed during the PHD project are presented in line with the implementation of the yield forecasting method in Fig. 3.3. The corresponding manuscript organization is also mentioned.

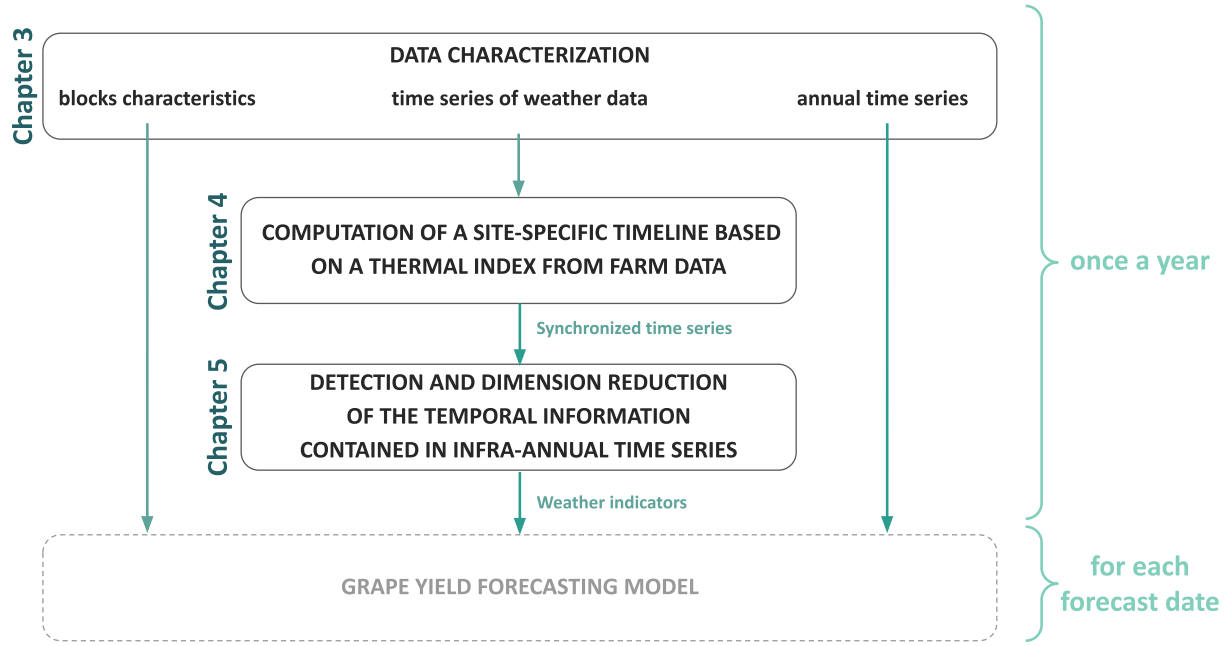

Figure 3.3 – Positioning of the two research axes aiming at valuing the time series of meteorological data in the framework of yield the forecasting method. The yield model, in light grey, constitutes a a further work perspective to achieve the yield forecasting method.

Chapter 4

Computing a site-specific timeline based on a thermal index to perform data analysis consistently with the vineyard phenology

4.1 Intention note

The research work presented in this chapter refers to the first analytical step of the grape yield forecasting method, as shown in Fig. 4.1

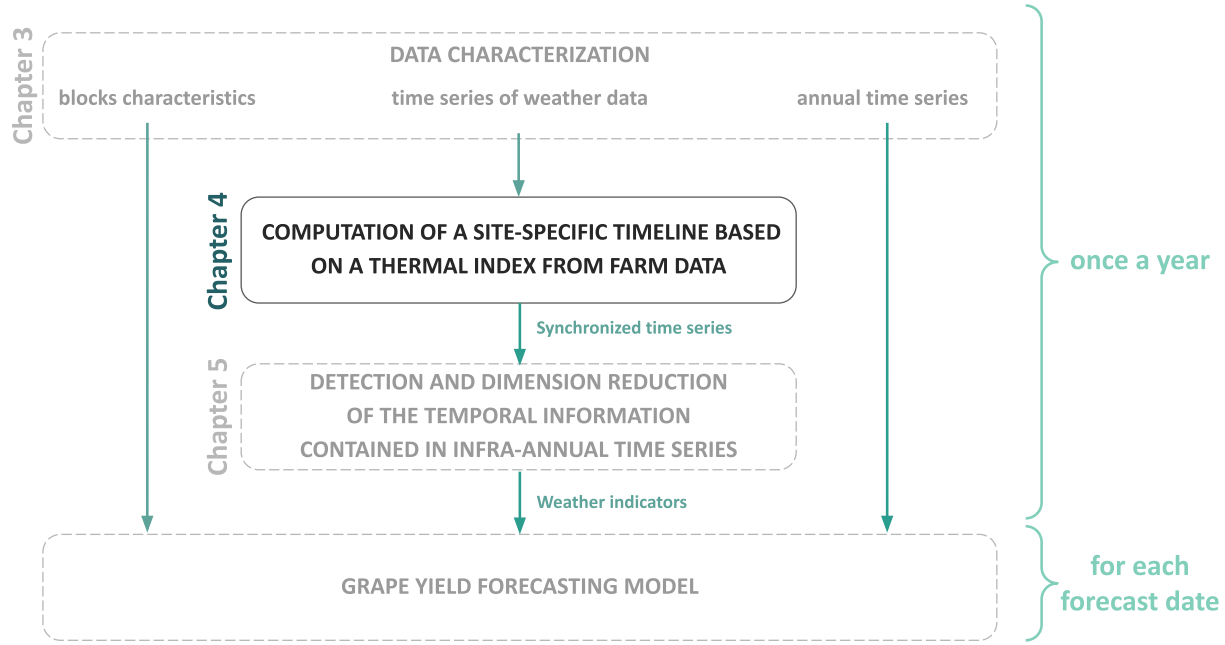

Figure 4.1 – Positioning of chapter 4 in the framework of the yield forecasting method

To compare the effect of an environmental influence on grapevine, here expressed in terms of yield, the environmental data is required to be expressed according to equal development stages of the vines across different years for the same site or different sites for the same year. Consequently, time series data are required to be synchronized consistently with grapevine phenology.

Temperature corresponds to one of the main terroir factors influencing grapevine physiology and a fortiori its phenology. Therefore, thermal indices have been developed to predict the achievement date of key phenological stages (Parker, De Cortázar-Atauri, et al. 2011; Fila et al. 2014; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017). They have subsequently been used to synchronize time series data to study the effect of an environmental variable on a grapevine physiological response (Guilpart, Metay, and Gary 2014) . In addition, several studies have suggested the interest of site-specifically calibrating these thermal indices in order to account for the site-specific interaction of numerous variables that determine grapevine response to temperature in terms of phenology (Nendel 2010; Caffarra and Eccel 2010; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017; de Rességuier et al. 2020; Laurent, Scholasch, et al. 2020).
Therefore, a first working hypothesis was that using a site-specific thermal index as a timeline would improve the synchronization of the time series of data and thus reduce the noise due to their phenological shift for further analyses. In the context of this PhD, a second working hypothesis corresponded to the feasibility of implementing such a thermal index from operational data. Finally, in the related literature, a given thermal index can implicitly be used for both prediction or synchronization purposes. **Thus, a third working hypothesis stated that the site-specific calibration should be driven by the intended use of the thermal index for better performance.** This hypothesis led to clarify and formalize the difference between a prediction or synchronization use of a given thermal index.
Therefore, the research work presented in this chapter was performed in two stages. First, the three working hypotheses were explored for proof of concept through a rapid study involving the site-specific calibration of empirically constructed thermal indices for vineyards A, B and C. Second, a method called eGDD for Extended Growing Degree Days was designed and developed to generalize the construction of a site-specific thermal index. Such a method should work from farm data and account for the intended use of the thermal index. It was developed based on the vineyard C data set and then tested on the data sets of vineyards A, B and C.

4.2 Building new temperature indices for a local understanding of grapevine physiology

4.2.1 Details about the paper

4.2.1.1 Title and publication informations

This section was presented as an oral communication entitled "Building new temperature indices for a local understanding of grapevine physiology" at the XIIIth International Terroir Congress (2020, Adelaide, Australia). It was awarded by the prize of high quality presenter.

4.2.1.2 Authors

C. Laurent[1,2,3,*], T. Scholasch[1],B. Tisseyre[3], A. Metay[2]

[1] Fruition Sciences, 34000 Montpellier, France
[2] ABSYS, Univ. Montpellier, CIRAD, INRAE, Institut Agro, 34060 Montpellier, France
[3] ITAP, Univ Montpellier, INRAE, Institut Agro, 34060 Montpellier, France
[*] corresponding author : cecile@fruitionsciences.com

4.2.1.3 Abstract

4.2.1.3.1 Aim

Temperature corresponds to one of the main terroir factors influencing grapevine physiology, primarily evidenced by its impact on phenology. Thus, numerous studies have aimed at expressing time with thermal indices such as growing degree days (GDD) and have enabled a better modelling of grapevine responses to temperature. However, some works have highlighted the need to adapt GDD to the considered pedo-climatic context and grape variety or to refine the time step at which temperature variables are computed. The present study aims to investigate the hypothesis that grapevine response to temperature depends on the

production context, i.e. plant material, pedo-climate, topography, orientation and cultural practices, and that thermal indices should then be locally adapted.

4.2.1.3.2 Methods and results

GDD with different base temperatures but also other indices based on other algebraic equations on daily average temperature were calculated starting from the budbreak date and using data from weather stations located in the Bordeaux region (France), California (USA) and Israel. The dates of flowering and veraison were expressed according to each of these indices for three commercial blocks located near each weather station. For each block, the relative differences in the flowering and veraison dates were calculated for any couple of years and summed squared. The number of studied years considered ranged from fifteen to five depending on the blocks. The relative difference between two dates was computed as their difference in index-related degrees divided by the average index-related amount of degrees to reach veraison. The thermal index which minimizes the sum of the relative differences of flowering and veraison dates for all the years of the same block is considered to best illustrate the temperature local effect. As such, this local effect includes both grapevine physiological response to temperature and the difference between the weather station data and the conditions actually experienced by the vines. Dates of flowering and veraison of all years coincide when expressed in a given thermal index for most of the blocks. Thus, the hypothesis whereby temperature is a predominant factor in grapevine phenology may be confirmed. Moreover, the thermal indices allowing such an adjustment are different between blocks of different locations, thus demonstrating that temperature effects on grapevine phenology are better captured when considered according to locally calibrated indices.

4.2.1.3.3 Conclusion

Temperature effects may be better captured by different thermal indices depending on the local context. In a precision viticulture context, a growing access to local and higher resolution weather data and grapevine observations enables models to be used locally. Therefore, the present study corresponds to a first attempt to highlight the importance of calibrating a local thermal index to improve the performance and operational relevance of any temperature-based model.

4.2.1.4 Keywords

Local thermal index, precision viticulture, terroir factors

4.2.2 Introduction

Temperature corresponds to one of the main terroir factors influencing grapevine physiology, primarily evidenced by its impact on phenology (Tonietto and Carbonneau 2004; Pagay and Collins 2017; Prats-Llinàs et al. 2020). Thus, many models have shown the importance of considering a temperature-based timeline, called thermal indice in this paper, to better represent grapevine phenology (Sadras and Moran 2013; Cola et al. 2014; Leolini et al. 2020; Prats-Llinàs et al. 2020). This thermal index is then used to express the timing of any variable aimed at explaining grapevine development (Suter et al. 2019; Parker, García de Cortázar-Atauri, et al. 2020). However, some studies have shown the interest of parameterizing the computation of thermal indices according to the variety (Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017) or to the site characteristics such as topography, elevation or orientation (Neethling et al. 2019; Verdugo-Vásquez, Acevedo-Opazo, Valdés-Gómez, Ingram, et al. 2020; de Rességuier et al. 2020). Thus, when the objective is not to compare varieties or management methods in terms of heat requirement (Molitor, Fraga, and Junk 2020) but to use the most relevant timeline to study grapevine development, it seems interesting to locally calibrate a thermal index. The calibration of such a local index is of real operational interest for commercial vineyards, since it would allow them to better understand the site-specific performance of their vines when research work is often based on experiments in very different environmental conditions that cannot be applied for them. In order to do this, calibration of the thermal index must be possible from operational data, which are commonly available in commercial vineyards. In this paper, two types of operational data were considered for the calibration of a local thermal index: weather data and the dates of achievement of the three most observed phenological stages, i.e. budbreak, bloom and veraison.

The working hypothesis of this study is that there exists a thermal index that allows the best expression of the thermal determinism of grapevine phenology. Therefore, the dates of achievement of the different phenological stages expressed in this index would be consistent over the years since resulting from the same

temperature driven process. The objective of the study is to identify the thermal index that allows the best matching of the dates of achievement of the phenological stages for a given vineyard and to test whether this index is the same from one vineyard to another.
Therefore, data from three different commercial vineyards respectively situated in the Napa Valley (California, USA), Israel and the Bordeaux region (France) have been analyzed to support the work presented in this paper. As these vineyards are located at different latitudes, the study also tested the impact of introducing photoperiod in the thermal index as a duration of exposure to effective growing temperatures.

4.2.3 Material and Method

4.2.3.1 Data description

Data was collected from three commercial vineyards situated in the Napa Valley (California, USA), Israel and the Bordeaux region (France). They are respectively noted vineyard A, B and C in this paper. For each vineyard, three blocks planted with Cabernet Sauvignon have been studied. The three vineyards differ according to their pedo climatic, topographic and altimetric environment, exposure as well as management strategies in terms of harvest yield and quality. As a result, the choice of plant material and cultural practices differ between vineyards, as an illustration of adaptation to local i.e. site-specific environmental and operational contexts. Some of these differences are summarized in Table 4.1.

Table 4.1 – Characteristics of the 3 vineyards selected as case studies in terms of environnement, organization and data set.

	Vineyard A	Vineyard B	Vineyard C
Localization	California, USA	Israel	Bordeaux, France
Area per block (hectare)	0.5 to 2	1.2	1.1 to 1.4
Plantation density (stocks/hectare)	2220	2500	8700
Plantation date	from 1994 to 2012	from 1984 to 2012	from 1952 to 2012
Rootstocks	1103P or 110R	101-14 MGt or none	101-14MGt or 3309C

Data of vineyards A, B and C were available respectively from 2008 to 2017 (ten seasons), from to 2015 to 2019 (five seasons) and from 2001 to 2015 (fifteen seasons). For each vineyard, daily average temperatures have been recorded using a commercial weather station located within the vineyard. For each block, the dates of budbreak, bloom and veraison have been collected according to the Gregorian calendar at 50% of achievement.

4.2.3.2 Computation of the thermal indices

For each block and each year, three types of thermal indices have been calculated from the date of budbreak. They are respectively noted 1, 2 and 3. Index 1 corresponds to cumulation of daily average temperature above a given base temperature and weighted by photoperiod (indice 1b) or not (indice 1a). Base temperatures from 0 to 25°C have been tested. Index 2 corresponds to the cumulation of daily average temperature between a minimum and a maximum threshold temperature and weighted by photoperiod (index 2b) or not (2a). Minimum and maximum thresholds have respectively been tested from -5°C to 10°C and from 10°C to 30°C. Index 3 corresponds to the cumulation of daily average temperature if situated between a minimum and a maximum threshold temperature and to a subtraction of the difference of daily average temperature with one of the thresholds if it is exceeded. The corresponding growing degrees were weighted by photoperiod (index 3b) or not (3a). Minimum and maximum thresholds have respectively been tested from -5°C to 10°C and from 10°C to 30°C. The six indices tested in the study are summarized in Table 4.2.

Table 4.2 – Equations and parameters for the computation of indices 1a, 1b, 2a, 2b, 3a and 3c. GDDi: growing degree days of DOY i, i going from the budbreak date to the veraison date, T_a : daily average temperature, T_b : base temperature, T_1 : minimum temperature threshold, T_2 : maximum temperature threshold, P : photoperiod.

	Index 1		Index 2		Index 3	
	1a	1b	2a	2b	3a	3b
Equation	$if\ T_b \leq T \leq T_a$ $GDD_i = GGD_{i-1}$ $+Ta - T_b$ $else$ $GDD_i = GGD_{i-1}$	$if\ T_b \leq T \leq T_a$ $GDD_i = GGD_{i-1}$ $+(Ta - T_b) * P$ $else$ $GDD_i = GGD_{i-1}$	$if\ T_1 \leq T_a \leq T_2$ $GDD_i = GGD_{i-1}$ $+Ta$ $else$ $GDD_i = GGD_{i-1}$	$if\ T_1 \leq T_a \leq T_2$ $GDD_i = GGD_{i-1}$ $+Ta * P$ $else$ $GDD_i = GGD_{i-1}$	$if\ T_1 \leq T_a \leq T_2$ $GDD_i = GGD_{i-1}$ $+Ta$ $if\ Ta \leq T_1$ $GDD_i = GGD_{i-1}$ $-\lvert T_1 - T_a \rvert$ $if\ T_2 \leq T_a$ $GDD_i = GGD_{i-1}$ $-\lvert T_2 - T_a \rvert$	$if\ T_1 \leq T_a \leq T_2$ $GDD_i = GGD_{i-1}$ $+Ta * P$ $if\ Ta \leq T_1$ $GDD_i = GGD_{i-1}$ $-\lvert T_1 - T_a \rvert$ $if\ T_2 \leq T_a$ $GDD_i = GGD_{i-1}$ $-\lvert T_2 - T_a \rvert$
Parameters	T_b from 0 to 25°C	T_b from 0 to 25°C	T_1 from -5 to 10°C T_2 from 10 to 30°C	T_1 from -5 to 10°C T_2 from 10 to 30°C	T_1 from -5 to 10°C T_2 from 10 to 30°C	T_1 from -5 to 10°C T_2 from 10 to 30°C

4.2.3.3 Comparison of the thermal indices

For each block and each year, the dates of bloom and veraison have been expressed in each thermal index and normalized by the average date of veraison expressed in the given index and calculated for all the years of the given block. For each block and each thermal index, the Euclidean distances between the dates of respectively bloom and veraison have been calculated two by two and summed squared. The results per vineyard were obtained by summing the corresponding distances of the three blocks. They are noted phenological deviation of the corresponding vineyard. For each vineyard, the thermal index resulting in the minimum phenological deviation has been selected, apart from calculation artefacts due to too elevated base temperature (index 1) or too close temperature thresholds (indices 2 and 3).
In order to better understand the meaning of a degree in the selected thermal indices, the maximum observed deviation of flowering and veraison dates was calculated in civil days i.e. day of the year (DOY) and in DOY equivalents of the thermal indices. These equivalents were approximated by dividing the date of flowering or veraison expressed in the given thermal indice by the average temperature (and possibly photoperiod) at plus and minus five DOY around the considered phenological stage.

4.2.4 Results and discussion

Table 4.3 – Minimum phenological deviation and corresponding thermal indices for vineyards A, B and C. For each type of thermal index, the minimum phenological deviation is presented with the corresponding parameters, T_a : daily average temperature, T_b : base temperature, T_1 : minimum temperature threshold, T_2 : maximum temperature threshold, P : photoperiod.

	Index 1		Index 2		Index 3	
	1a	1b	2a	2b	3a	3b
Vineyard A	T_b=4°C 0.9	T_b=4°C 1.02	T_1=10°C T_2=29°C 1.27	T_1=10°C T_2=29°C 1.28	T_1=9°C T_2=30°C 1.51	T_1=10°C T_2=28°C 1.52
Vineyard B	T_b=0°C 0.38	T_b=0°C 0.36	T_1=5°C T_2=29°C 0.32	T_1=5°C T_2=29°C 0.28	T_1=5°C T_2=29°C 0.32	T_1=5°C T_2=29°C 0.28
Vineyard C	T_b=6°C 2.97	T_b=5°C 2.94	T_1=10°C T_2=28°C 2.52	T_1=10°C T_2=29°C 2.43	T_1=10°C T_2=28°C 2.45	T_1=10°C T_2=28°C 2.37

Independently on the thermal indices considered in Table 4.3, the phenological deviations are smaller for vineyard B than vineyard A and for vineyard A than vineyard C. Part of this statement may be due to the different number of years considered in each data set. The thermal index allowing a minimal deviation of bloom and veraison dates are not the same from one vineyard to another: index 1a with a base temperature of 4°C is selected for vineyard A, index 2b with temperature thresholds of 5°C and 29°C is selected for vineyard B and index 3b with temperature thresholds of 10°C and 28°C is selected for vineyard C. On the basis of the working hypothesis, this means that the local effect of temperature on grapevine development is different for the three vineyards. The effective growing daily average temperatures seem to be situated

above 4°C for vineyard A and do not seem to present a maximum. The base temperature is lower than those mostly found in the scientific literature (Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017). The effective growing daily average temperatures seem to be situated between 5 and 29°C for Vineyard B. The effective growing daily average temperatures seem to be situated between 10 and 28°C and temperatures exceeding these thresholds seem to be detrimental to grapevine development for Vineyard C. Upper thresholds are coherent with optimal growing temperatures around 25°C announced by numerous literature studies (Vasconcelos et al. 2009).
Non growing events as defined locally for each vineyard, i.e daily average temperatures exceeding local base or threshold temperatures, are present in each data set. Therefore, the selection of thermal indices is not due to a data set effect, for which extreme temperatures would be absent. Instead, variations in plant material (clone and rootstock), environment and cultural practices may explain these differences.

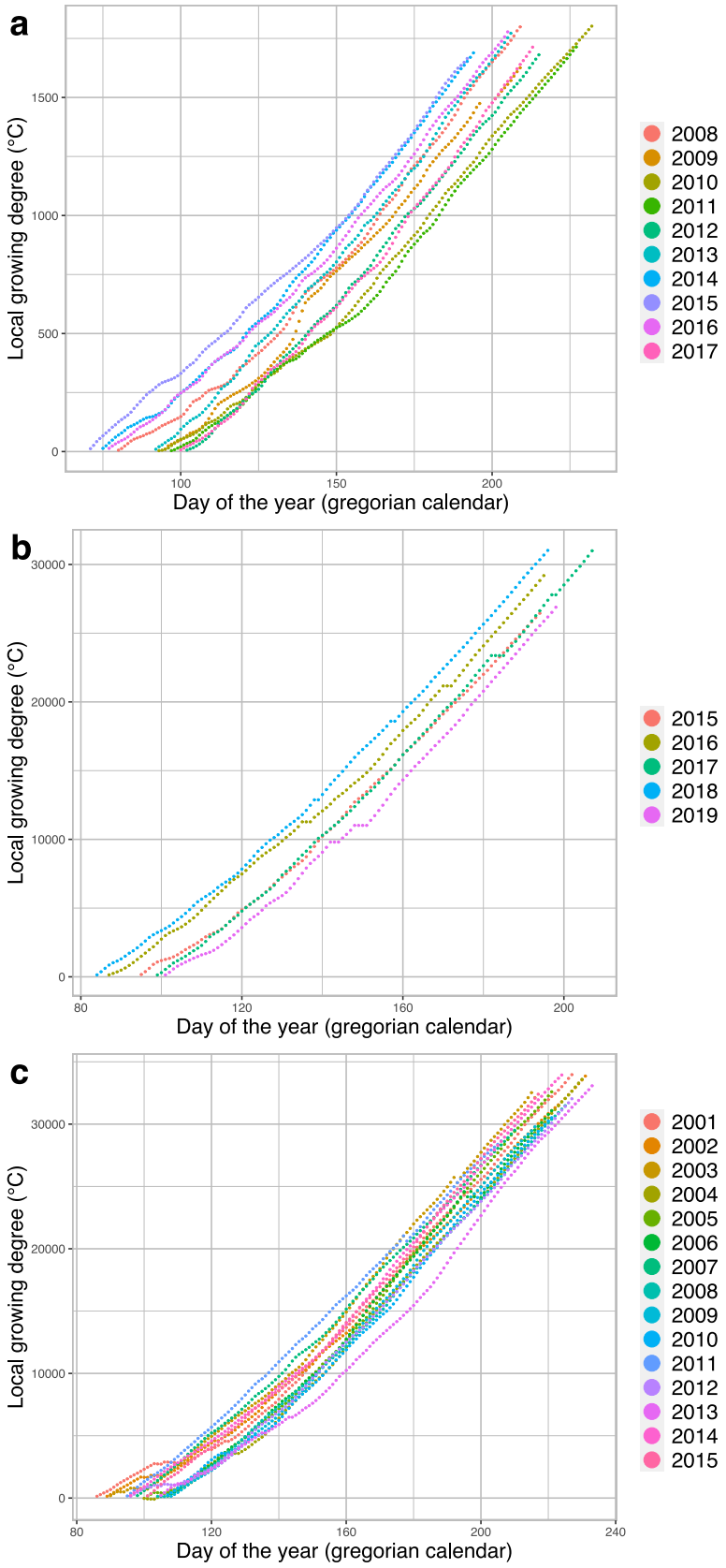

Figure 4.2 – Annual profiles of local thermal indices from budbreak to veraison for one block of (a) vineyard A, (b) vineyard B, (c) vineyard C as a function of time (DOY).

Fig. 4.2 shows that the calculated growing degrees have the same unit (°C) but not the same meaning: 1°C in the thermal index of vineyard A corresponds to less degrees felt in reality than 1°C in the thermal index of vineyard C. Moreover, the photoperiod is only used in the computation of the indices of vineyards B and C. As a result, the scales of growing degree are different between vineyards: from 0 to 1700 °C for vineyard A and from 0 to 32000°C for vineyards B and C. The profiles have the same shape, as they all come from a linear computation. However, differences can be observed between the years of each vineyard, in particular with slowing and speeding effects of the effective growing temperature. For example, slowing effects can be observed for 2019 in Fig. 4.2b and for 2013 in Fig. 4.2c. Speeding effects can be observed for 2008 in Fig. 4.2a. Regarding the years of the same block or the different vineyards for the same year, it should also be noted that profiles do not all start at the same time (depending on the budbreak date) and do not all have the same duration. These differences are illustrated by the computation of phenological deviation.

Table 4.4 – For each vineyard, average across the three blocks of the maximum deviation between the dates of flowering and veraison expressed in DOY and in DOY equivalents for the selected thermal index. In bold: DOY or DOY equivalents in the selected local index. In italics: average across the three blocks of the maximum deviation between the bloom or veraison dates expressed in the local thermal index, average day temperature around bloom or veraison and average photoperiod around bloom or veraison used to compute DOY equivalents.

	Average of the maximum deviation between the bloom dates by block (DOY)		Average of the maximum deviation between the veraison dates by block (DOY)	
	Gregorian calendar	Local thermal index	Gregorian calendar	Local thermal index
Vineyard A	**42**	**15** *1216.2°C, 18°4°C*	**40**	**10** *1172.8°C , 21.1°C*
Vineyard B	**19**	**3** *1889.5°C , 21.1°C , 13.7h*	**12**	**12** *14015.3°C , 24.5°C , 14.1h*
Vineyard C	**38**	**21** *14149.7°C , 14.6°C , 13.3h*	**27**	**24** *15456.6°C , 16.8°C , 13.7h*

The average DOY deviation between the dates of bloom and veraison are respectively smaller for all local thermal indices than for the Gregorian calendar (Table 4.4). According to the working hypothesis, this means that each local thermal index better reflects the physiological processes underlying grapevine phenology in the given vineyard. The local thermal indice allows a reduction of respectively 65% and 75% in the deviation between the dates of flowering and veraison of the blocks of vineyard A. The local thermal index provides a deviation of three DOY equivalents between the flowering dates of the blocks of vineyard B. This deviation seems acceptable regarding the operational needs for decision support of cultural practices. However, the respective deviations between the dates of bloom and veraison are only improved by 45% and 15% for the blocks of vineyard C. Thus, it can be considered that there exists another thermal index that would allow a better expression of the phenology of vineyard C. In that sense, other thermal indices should be explored in future work. In particular, it can be noticed in Table 4. 4 that the deviation between flowering dates is more improved by the local thermal indice than the deviation between veraison dates for vineyards B and C. This suggests that climate-related physiological processes may not be the same for the two phenological stages. Therefore, it might be relevant to test non-linear thermal indices to constitute a common timeline for the whole season of grapevine development (Sadras and Moran 2013; Parker, García de Cortázar-Atauri, et al. 2020). In addition, the computation of local indices on the basis of daily average temperatures implies that the physiological processes to be illustrated are sensitive to relatively long thermal episodes. However, it has been shown that short exposures to intense temperatures can have an impact on grapevine development (Gouot et al. 2019a). Therefore, it could be useful to include daily maximum and minimum temperatures or even hourly temperatures in the computation of new indices in order to better represent the thermal conditions of the day (Rienth, Torregrosa, et al. 2014) . In this study, the photoperiod was used as a means of reporting the duration of exposure to effective temperatures for grapevine development. This seemed indeed to improve the deviations between the bloom and veraison dates for almost all indices and all vineyards (cf. Table 4.3). However, temperature and light are considered as presenting combined effects. When data is available in the field, it could also be interesting to include radiation in the computation of local indices (Prats-Llinàs et al. 2020).

In this study, the working hypothesis was to consider that the collected dates of phenological stages faithfully represent a locally identical process from one year to another if expressed in a relevant local climatic index. However, not all plants in the same block reach these phenological stages completely synchronously. This makes it difficult to assess precisely when 50% of the block has actually reached a given stage. The deviations

observed between the bloom and veraison dates necessarily contain noise related to this difficulty. Moreover, it should be noted that, in this study as it is always the case operationally in the field, weather data is recorded using a single station for the three blocks of each vineyard. Therefore, it is not exactly representative of the meso- and microclimatic conditions actually experienced by the plants, according to environmental factors such as topography, elevation, exposition etc. This may also explain part of the irreducible noise that exists when comparing the dates of the phenological stages expressed in climatic indices. Finally, for operational reasons, this study only considers the beginning of the grapevine growing period, by stopping at veraison. It would obviously be beneficial to complete this work by considering other phenological stages and see if the selected thermal indices are still the same.

4.2.5 Conclusion

This study showed that local i.e. site-specific thermal indices can better show the consistency of the dates of bloom and veraison achievement between the different seasons experienced by the blocks of the same vineyard. By using these indices as a timeline to express the timing of any other explanatory variables consistently with grapevine phenology, a further objective of this work is to better understand the influence of environmental factors such as climate or cultural practices on grapevine development and yield. In addition, by focusing on operational data from commercial vineyards, this work opens up opportunities for site-specific local calibration of thermal indices for vineyard management purposes.

4.3 eGDD : an extension of the Growing Degree Days approach to compute a site-specific thermal index in viticulture using a constrained optimization method. Part I : Method development.

4.3.1 Details about the paper

4.3.1.1 Title and publication informations

The work presented hereafter will be submitted as a two-part scientific article to the journal Agriculture and Forest Meteorology. The first part of the article is presented in this section. It is entitled "eGDD : an extension of the Growing Degree Days approach to compute a site-specific thermal index in viticulture using a constrained optimization method. Part I : Method development".

4.3.1.2 Authors

C. Laurent[1,2,3,*], G. Le Moguédec[4], J. Taylor[3], T. Scholasch[1], B. Tisseyre[3],A. Metay[2]

[1] Fruition Sciences, 34000 Montpellier, France
[2] ABSYS, Univ. Montpellier, CIRAD, INRAE, Institut Agro, 34060 Montpellier, France
[3] ITAP, Univ Montpellier, INRAE, Institut Agro, 34060 Montpellier, France
[4]AMAP, Université Montpellier, INRAE, Cirad CNRS, IRD, 34000 Montpellier, France
[*] corresponding author : cecile@fruitionsciences.com

4.3.1.3 Abstract

Temperature corresponds to one of the main terroir factors influencing grapevine physiology and a fortiori its phenology. Therefore, thermal indices were developed to predict the achievement date of some key phenological stages such as budbreak, bloom and veraison. Thermal indices are also used to re-express time as a pre-processing of time series data. Such a use allows a synchronization of the time series according to a new timeline in order to compare grapevine performances of different blocks or years more consistently with grapevine phenology than the Gregorian calendar.
Among these indices, the Growing Degrees Days (GDD) approach is the most widely used. The GDD approach considers temperatures that are above a temperature threshold, named base temperature (T_b), effectively promoting grapevine development. The present paper first investigates the opportunity to re-examine the relationship between grapevine development and temperature that underlies the GDD approach

by integrating additional temperature thresholds. It then proposes a constrained optimization approach to site-specifically calibrate such a relationship on the basis of operational data and depending on the intended use of the thermal index (Prediction or Synchronization). Finally, it is also proposed to weight this relationship between grapevine development and temperature by the photoperiod. These proposals are included in a method that is called extended Growing Degree Day (eGDD).
In order to assess the relevance of the eGDD method, its results in terms of Prediction and Synchronization are compared to those of the Gregorian calendar and of the classical GDD approach. The GDD method is employed with the commonly used T_b of 10°C but also with a T_b calibrated with a similar constrained optimization method.

A data set from a vineyard located in the Bordeaux region is used to present the eGGD method in this paper. It could also be applied to data sets from other vineyards and even to other perennial or annual plants.

4.3.1.4 Highlights

— An original method, called Extended Growing Degree Day (eGDD) is proposed
— Site-specific thermal indices are computed using a constrained optimization approach
— The calibration process is driven according to Prediction or Synchronization purposes
— The eGDD method is designed to work with farm data and for operational purposes
— The eGDD method shows better Prediction and Synchronization results than the Growing Degree Days approach

4.3.1.5 Keywords

Grapevine Phenology, Temperature, Thermal Time, Time Series Data Analysis, Operational Data

4.3.2 Introduction

Temperature corresponds to one of the main terroir factors influencing grapevine physiology and a fortiori its phenology (Tonietto and Carbonneau 2004; Pagay and Collins 2017; Prats-Llinàs et al. 2020). Thus, temperature indicators are required both to understand grapevine response to temperature and to support operational decision-making in the field in line with grapevine phenology. Thermal indices were mainly developed to predict the achievement date of some key phenological stages such as budbreak, bloom and veraison (Prediction use). Most thermal indices designed to model grapevine phenology are based on the accumulation of heat units (Wang 1960; Cross and Zuber 1972; Grigorieva, Matzarakis, and Freitas 2010). Therefore, they are based on the hypothesis that temperature is the main factor that drives grapevine phenology and that temperature influence on grapevine phenology corresponds to a forcing effect. Among these indices, the Growing Degrees Days model (GDD) is the most widely used (Parker, De Cortázar-Atauri, et al. 2011; Fila et al. 2014; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017). The GDD approach considers temperatures that are above a temperature threshold, named base temperature (T_b), as effectively promoting grapevine development. When daily mean temperatures are above T_b, their difference with T_b is accumulated day after day to result in a timeline for the whole season. Other thermal indices such as the UniFORC and the UniPhen models have been proposed to model the temperature forcing effect by additionally considering optimal and critical temperature thresholds (T_o and T_c), above which grapevine development respectively slows down or becomes null (Chuine 2000; Parker, De Cortázar-Atauri, et al. 2011). These thermal indices require more parameters to be computed. Therefore, they require more data to be calibrated and are more prone to overfitting. The GDD is hence often recognized as a good trade-off between parsimony and efficiency for modelling grapevine phenology (Parker, De Cortázar-Atauri, et al. 2011; Fila et al. 2014). However, this conclusion was drawn from studies that only attempted to model one phenological stage at a time. This is frequently not the case in practice. For example, the GDD with a T_b of 10°C is often used to predict the date of budbreak, flowering and veraison at the same time. In this case, the temperature range taken into account when considering the three stages exceed and encompass the intervals of the three stages respectively (e.g. from -5°C to 35°C). This should encourage the additional consideration of T_o and T_c that are often not considered when studying only one phenological stage because they are in temperature ranges never observed in the field during the periods of interest (Molitor, Fraga, and Junk 2020). Moreover, the GDD approach does not take into account the fact that the same average daily

temperature at different dates for the same site or for the same date on two sites at different latitudes does not have the same significance for grapevine development (Mahmud et al. 2018).
Nevertheless a GDD thermal index is often recognized as better representing grapevine phenology than the Gregorian calendar. Therefore, it is used in a deviated way to re-express time as a pre-processing of time series data. Such a use allows a synchronization of the time series according to a new timeline in order to compare grapevine performances of different blocks or years more consistently with grapevine phenology than the Gregorian calendar. An example of it can be found in Guilpart, Metay, and Gary 2014 where the analysis of the effect of water stress on grapevine yield is performed using data from different blocks characterized by different weather time series. In Guilpart, Metay, and Gary 2014, time series are intended to be discretized into periods that are considered as explaining variable candidates in linear regressions. The time series are discretized into periods of 100 Degree.Days so as to analyze the effect of water stress on yield for blocks that are considered in the same stage of development. The purpose of this use, called Synchronization, is different from the Prediction Use. Indeed, it is no longer intended to accurately predict the date of achievement of the phenological stages in Gregorian days but rather to make the dates of achievement of the phenological stages match across blocks or years once expressed according to the GDD thermal index. When expressed according to this thermal index, the dates of budbreak achievement in different years are wanted to be as similar as possible and so are the dates of the other phenological stages. From now on in this document, such dates expressed in any thermal index will be called "score". The objective of the Synchronization is to minimize the dispersion of the scores for each phenological stage. Indeed, the more consistent the scores of each phenological stage, the more you can divide the time series into short periods and the finer the analysis of time series is likely to be. Therefore, the calibration criterion of any thermal index and a fortiori of the GDD should not be the same for Synchronization purposes as for Prediction purposes.
In parallel, for both the Prediction and the Synchronization uses, the interest of site-specifically calibrating the thermal index has been demonstrated (Nendel 2010; Caffarra and Eccel 2010; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017; de Rességuier et al. 2020; Laurent, Scholasch, et al. 2020)
Today, in a context of precision viticulture, an increasing number of vineyards are equipped with a physical or virtual weather station and field observations are more and/ or better traced. It is an opportunity to test and eventually improve the GDD approach with a site-specific calibration for a wide variety of vineyards. Such a site-specific computation is of real operational interest for commercial vineyards because it allows to better understand the site-specific performance of their own vines when research work may have been performed and are valid in different environmental conditions that cannot be applied for them. However, computing a site-specific thermal index for commercial vineyards involves analyzing operational data i.e. data that has been collected in a commercial vineyard according to its own needs, constraints and habits. Operational data is characterized by heterogeneous quality and parsimony in terms of the variables available for analysis. When performed, phenological observations mostly focus on three stages : budbreak, bloom and veraison. These observations are made at the block scale or at the vineyard scale, the later depending on the observation of reference blocks. Additionally, temperature data is often reduced to an average, minimum and maximum temperature per day. There is usually only one weather station for all the blocks of the vineyard. It can be a private station located on the vineyard or another station, located outside the vineyard and often further away. In order to identify a site-specific thermal index, this operational reality calls for a robust calibration method that is able to work with a few reference points per year (budbreak, bloom and veraison). It also implies a shift in the interpretation that can be made of any thermal index. Indeed, a thermal index should hence be interpreted as integrating three main sources of variability: i) the thermal determinism of grapevine development ii) spatial variations in the temperature conditions and iii) the quality of the data set. The data set quality refers to both the weather data representativity of temperature conditions really experienced by the considered vines and the precision in the observation of achievement dates for the different phenological stages.
In such a context, the present paper first investigates the opportunity to re-examine the relationship between grapevine development and temperature that underlies the GDD approach. It then proposes a constrained optimization approach to site-specifically calibrate such a relationship on the basis of operational data and depending on the intended use of the thermal index. Finally, it is also proposed to weight this relationship between grapevine development and temperature by the photoperiod in order to take into account the fact that the same average daily temperature at different dates for the same site or for the same date on two sites at different latitudes does not have the same significance for grapevine development. These proposals are included in a method that is called extended Growing Degree Day (eGDD). In order to assess the relevance of the eGDD method, its results in terms of Prediction and Synchronization are compared to those of the

Gregorian calendar and of the classical GDD approach. The GDD method is employed with the commonly used T_b of 10°C but also with a T_b calibrated with a similar constrained optimization method. A data set from a vineyard located in the Bordeaux region is used to present the eGGD method in this paper. However, it can obviously be applied to other data sets from other vineyards and even to other perennial or annual plants, by adapting the observed phenological stages or other advancement variables.

4.3.3 Materiel and Method

4.3.3.1 Data description

The data used to present the eGDD approach was collected from a commercial vineyard located in the Bordeaux region, France. 79 geographically adjacent blocks are considered for the study. The blocks are planted with different clones of Cabernet-Sauvignon, Merlot and Petit Verdot and on various rootstocks. They are characterized by soils of different nature, different density and row orientation but are managed with similar cultural practices.
For each year and each block, the dates of 50% achievement of budbreak, bloom and veraison have been visually assessed and recorded according to the Gregorian calendar. Daily mean temperatures have been recorded using a single private weather station for all the blocks. This weather station is located in the center of the vineyard.
All data was collected for 13 years between 2001-11 and 2014-15.

4.3.3.2 Theory about thermal indices

Although not presented as such in the literature, cumulative thermal indices can all be formalized as the integration over the season of a function linking grapevine development speed as evidenced by its phenology to temperature. Such a function is henceforth named Phenological Advancement Speed as a function of Temperature (PAST). The PAST function has no unit. It can be seen as a weight function of the Gregorian time (e.g. days) according to the daily or hourly mean temperature. It also corresponds to the slope of the thermal index (obtained by integration) when plotted against the Gregorian time. The construction of any cumulative thermal index defined as the integration over Gregorian time (expressed in days or hours) of a PAST function is transcribed in Eq. 4.1.

The following section presents the general concept of the PAST function with the examples of the Gregorian calendar, and of the GDD model. To do so, the PAST functions corresponding to each approach are presented in Fig. 4.3a and 4.3c and an example of their respective integration over the same temperature annual profile is presented in Fig. 4.3b and d. The data used for such an example corresponds to the same block (block 3) and to the same annual history of daily mean temperature (2003).

$$Thermal\ Index = \int_{season} PAST(t)\,\mathrm{d}t \tag{4.1}$$

with PAST the Phenological Advancement Speed as a function of Temperature
and t the time in Gregorian units

The Gregorian calendar is not affected by temperature conditions: whatever the temperature conditions, grapevine development is considered to happen at an unchanging pace when it is expressed in the Gregorian Calendar. Therefore, it can be considered that the Gregorian calendar corresponds to the integration over the season of a PAST function constantly equal to 1 as illustrated in Fig. 4.3a and b. In contrast, the GDD model considers that the pace at which grapevine development occurs depends on the temperature conditions. Below a base temperature (T_b), grapevine development is considered ineffective. Therefore, the PAST function is equal to zero for temperatures below T_b. Above T_b, the higher the temperatures, the faster grapevine development is considered to happen. Each additional degree in the daily mean temperature corresponds to the same advancement rate in grapevine phenology. Therefore, the PAST function corresponds to an increasing linear curve for temperature above T_b. In the GDD model, the PAST function corresponds to a non-ending curve for temperatures above T_b i.e. the maximum of the PAST function is not fixed but depends on the maximum mean temperature observed in the temperature history. The PAST function corresponding to the GDD with T_b equal to 10°C is illustrated in Fig. 4.3c and d. It can be noticed that the slope of the thermal index in Fig. 4.3d is different from day to day, according to the value of the PAST function for each

day. The GDD thermal index resulting from the integration over the season of such a PAST function should be interpreted as an accumulation of heat units (Degree.Days). The definition of heat units depends on the value chosen for T_b. Therefore, the scores of two GDD thermal indices computed with two different values for T_b have different meanings and should not be compared as such. For example, a score of 100 Degree.Days doesn't mean the same if it has been computed with a PAST function whose T_b was equal to 8°C or 12°C.

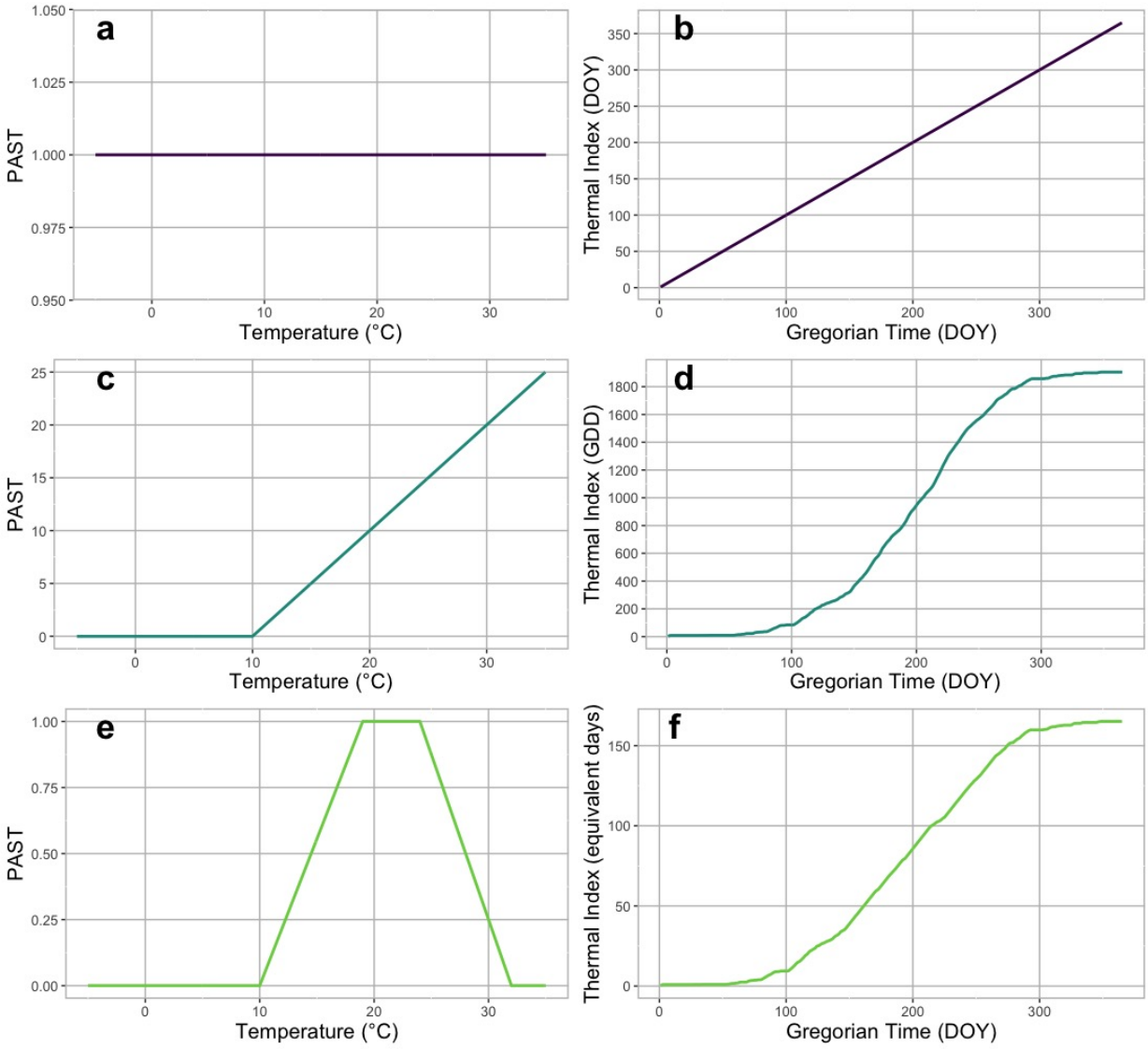

Figure 4.3 – Figure 1: Phenological Advancement Speed as a function of Temperature (PAST functions) and thermal indices corresponding to their integration over the 2003 history of daily mean temperature for block 3. Figures a and b refer to the Gregorian Calendar. Figures c and d refer to the Growing Degree Days (GDD) model with a base temperature of 10°C. Figures e and f refer to the Extended Growing Degree Days (eGDD) model with base, optimal and critical temperatures of 10, 19 , 25 and 32°C respectively.

4.3.3.3 Presentation of the eGDD approach

4.3.3.3.1 Construction of the PAST function for the eGDD model

As an extension of the GDD approach, the eGDD model also considers that the pace at which grapevine develops depends on the temperature conditions. In addition to T_b, the corresponding PAST function includes other temperature thresholds. These thresholds are drawn from ecophysiological literature (Pouget 1981; Vasconcelos et al. 2009; Keller, Tarara, and Mills 2010; Gouot et al. 2019b; Camargo-Alvarez et al. 2020) and correspond to a biophysical understanding of temperature effects. They can be found in other literature works under different mathematical formalisms (Chuine 2000; Zhou and Wang 2018; Molitor, Fraga, and Junk

2020). These additional thresholds aim at taking into account the fact that grapevine development is not indefinitely accelerating when temperature increases i.e. that it reaches a maximum. Therefore, the PAST function is composed of an increasing curve that reaches a maximum for optimal temperatures comprised in the interval $[T_o^1, T_o^2]$. The eGDD model also embeds the fact that some high temperatures may have a reducing effect on grapevine development. In consequence, the PAST function includes a decreasing slope until reaching zero at a given temperature threshold called critical temperature (T_c). The values taken by the PAST function are defined on [0,1]. An example of a PAST function in the eGDD model and of the related Thermal Index is given in Fig. 4.3e and f. These requirements can be formalized into multiple different functions with varying numbers of parameters. Given the effort of simplicity and robustness required by the parsimony of the operational data, it has been decided to consider linear slopes and to formalize the PAST function as a trapezoidal function. This function is coded as the subtraction of a first uniform cumulative distribution from a second one (Eq. 4.2 and 4.3).

$$f_1(T) = \begin{cases} 0 & if\ T < T_b \\ \frac{T-T_b}{T_o^1-T_b} & if\ T_b \leq T \leq T_o^1 \\ 1 & if\ T > T_o^1 \end{cases} \quad f_2(T) = \begin{cases} 0 & if\ T < T_o^2 \\ \frac{T-T_o^2}{T_c-T_o^2} & if\ T_o^2 \leq T \leq T_c \\ 1 & if\ T > T_c \end{cases} \tag{4.2}$$

$$PAST(t) = f_1(T) - f_2(T) \tag{4.3}$$

with t a gregorian time unit, T the mean temperature per Gregorian time unit, T_b the base temperature, $[T_o^1, T_o^2]$ the interval of optimal temperatures and T_c the critical temperature.

Therefore, the PAST function is controlled by four parameters that correspond to the temperature thresholds T_b, T_o^1, T_o^2 and T_c. $T_b \leq T_o^1 \leq T_o^2 \leq T_c$ is imposed. Depending on the values for each parameter, the PAST function can take various shapes that are described in Fig. 4.4 as nuances of the PAST function shape shown in Fig. 4.3e.

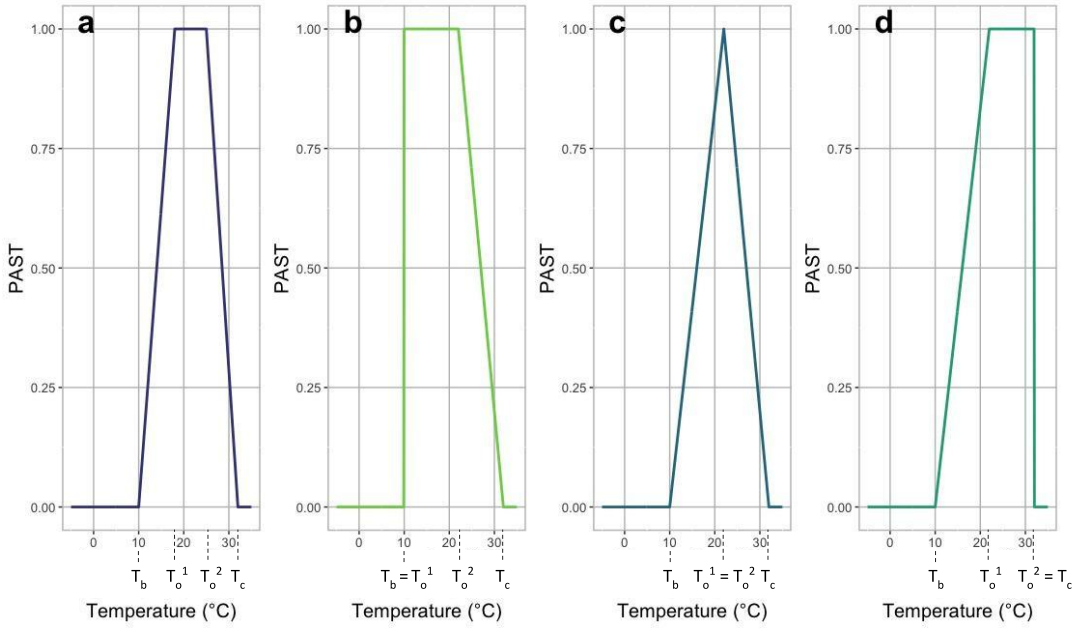

Figure 4.4 – Various shapes of the eGDD PAST function
with (a) $T_o^1 > T_o^2$,(b) $T_b = T_o^1$,(c) $T_o^1 = T_o^2$ and (d) $T_o^2 = T_b$.
T_b corresponds to the base temperature, $[T_o^1, T_o^2]$ to the interval of optimal temperatures and T_c to the critical temperature.

In contrast to the GDD model which doesn't accept a maximum for the PAST function, the PAST function used in the eGDD approach takes values equal to 1 at the temperature(s) allowing the maximum grapevine development. Therefore, the eGDD scores are to be interpreted as an equivalent of Gregorian time passed at an optimal temperature. This is different from Growing Degrees Days, which represent a heat accumulation without taking the optimal temperature(s) into account. For example, let's say that 23°C daily is one of the optimal temperatures as in Fig. 4.3e. In the case of the GDD model, a day at a daily mean temperature of 27°C counts more (by 4 Growing Degree Days) than a day at 23°C while it counts less in the case of the eGDD. In the eGDD case, a day at 27°C is considered to be equivalent to 75% of a Gregorian day spent at an optimal temperature. Nevertheless, as for the GDD model, two different eGDD thermal indices (computed with two different sets of temperature thresholds) have different meanings and should not be compared as such.

4.3.3.3.2 PAST function weighting by the photoperiod and integration over the season

In the eGDD approach, the PAST function is itself weighted by the photoperiod before being integrated over the season, as represented in Eq. 4.4 It allows the same daily mean temperature to have different weights according to its DOY and to the site latitude.

$$eGDD\ Thermal\ Index = \int_{season} PAST(t) * photoperiod(t)\,\mathrm{d}t \tag{4.4}$$

with PAST the Phenological Advancement Speed as a function of Temperature
and t the time in Gregorian units

Consequently, the unit of a eGDD thermal index corresponds to an equivalent of daylight hours spent at an optimal temperature. Such a unit is then called Thermally Optimal Daylight Hours (TODH).

4.3.3.3.3 Calibration of the PAST function for the eGDD model

The goal of the calibration is to set site-specific parameters T_b ,T_o^1,T_o^2 and T_c i.e. parameters that are adapted to the local conditions of a given site for which the eGDD thermal index will be used. Calibration is performed through a constrained optimization method. The optimal set of parameters is tuned according to the minimization of an optimization criterion. The optimization criterion is chosen according to the objective the eGDD method is used for (Prediction or Synchronization use).
In the Prediction case, the optimization criterion consists in a Mean Square Error (MSE) between predicted and observed dates of budbreak, bloom and veraison. The predicted dates of each phenological stage corresponds to the first day for which a specific score is reached. The specific score is set for each phenological stage by a nested optimization approach, hence with one parameter. It aims at identifying the score that allows the best prediction of the dates of achievement of budbreak, bloom and veraison respectively for a given PAST function and corresponding thermal index.
In the Synchronization case, the optimization criterion logically aims at synchronizing the different time series. It corresponds to the sum of inter-annual variance of budbreak, bloom and veraison scores, which are normalized according to the mean maximum score of all years.
The optimization criteria corresponding to the Prediction or Synchronization use are respectively noted P and S. For both criteria, a ponderation of the equation components corresponding to each phenological stage has been added as an option. It allows the user to drive the optimization towards the best results for a particular phenological stage. For example if the user may want to give priority to the results of the eGDD model because he/she has more confidence in the quality of the bloom observations than those of budbreak. The optimization criteria corresponding to the Prediction or Synchronization uses are respectively noted P and S and described in Eq. 4.5 and 4.6.

For a given site, the minimum value of the optimization criterion allows to select the optimal set of T_b, T_o^1, T_o^2 and T_c parameters. Fig. 4.5 presents an example of the values taken by S according to the values of T_b, T_o^1, T_o^2 with T_c being fixed for a given site. It corresponds to a view of the space of admissible parameters. A similar representation could be observed for P.

$$P = a\sum_{i=1}^{n}\frac{(\hat{d}_i^{bud} - d_i^{bud})^2}{n} + b\sum_{i=1}^{n}\frac{(\hat{d}_i^{blo} - d_i^{blo})^2}{n} + c\sum_{i=1}^{n}\frac{(\hat{d}_i^{ver} - d_i^{ver})^2}{n} \tag{4.5}$$

with n the number of considered years for a given site,
$d_i^{bud,blo,ver}$ and $\hat{d}_i^{bud,blo,ver}$
respectively the observed and predicted dates of budbreak, bloom and veraison,
a, b and c the ponderation for each phenological stage with $a + b + c = 1$

$$S = a\sum_{i=1}^{n}\frac{(\frac{s_i^{bud} - \bar{s}_i^{bud}}{s_{max}})^2}{n} + b\sum_{i=1}^{n}\frac{(\frac{s_i^{blo} - \bar{s}_i^{blo}}{s_{max}})^2}{n} + c\sum_{i=1}^{n}\frac{(\frac{s_i^{ver} - \bar{s}_i^{ver}}{s_{max}})^2}{n} \tag{4.6}$$

with n the number of considered years for a given site,
$s_i^{bud,blo,ver}$ and $\bar{s}_i^{bud,blo,ver}$ the observed and predicted scores,
s_{max} the mean maximal score for all years,
a, b and c the ponderation for each phenological stage with $a + b + c = 1$

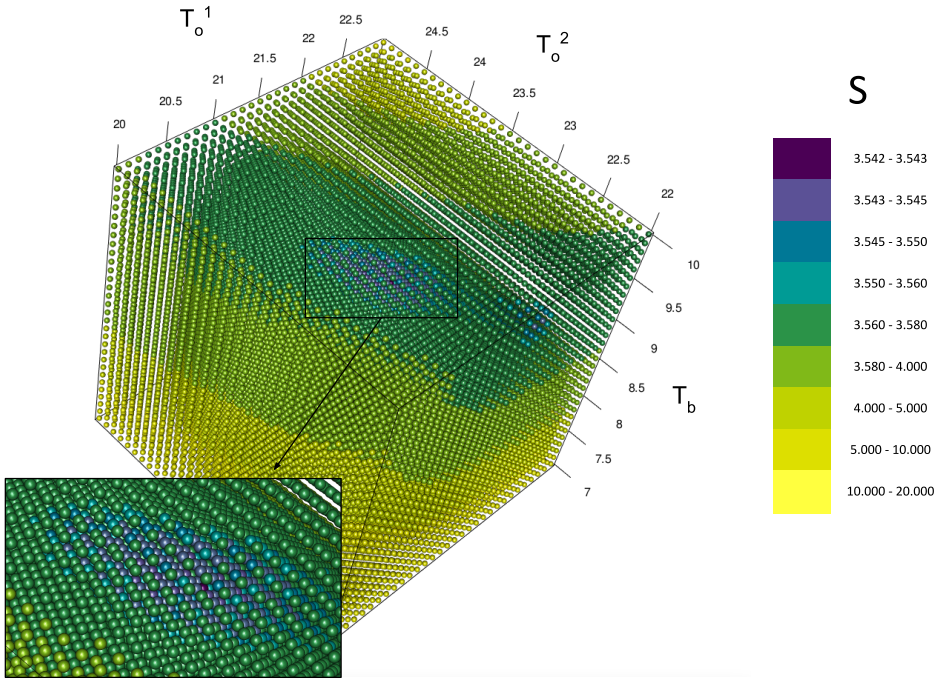

Figure 4.5 – Values taken by the Synchronization Optimization Criterion (S) used in the Extended Growing Degree Days (eGDD) method according to the values of the base temperaure (T_b) and optimal temperatures (T_o^1 and T_o^2) for block 3. In this example, the critical temperature (T_c) is fixed at 35°C for an easy visualization but it is also optimized when running the eGDD method.

The exploration of the space of admissible parameters aiming at finding the lowest P or S values was performed with a genetic algorithm (Sivanandam and Deepa, 2008 ; Srucca, 2013) to identify the subspace including the lowest values. A Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Fletcher and Reeves, 1964) was then used to explore the resulting subspace to find optimal value of P or S.

4.3.3.4 Evaluation of the eGDD method

4.3.3.4.1 Presentation of the use case

The eGDD method was applied to each block of the data set. A set of T_b, T_o^1, T_o^2 and T_c parameters was optimized on the basis of each block. During the calibration procedure, T_b,was constrained to plausible

values from -5 to 15°C, T_o^1, and T_o^2,from 15 to 25°C and T_c from 28 and 35°C. Equal weights were used for all phenological stages.

4.3.3.4.2 Comparison references and criteria

For both Prediction and Synchronization uses, the eGDD method was compared to i) the Gregorian calendar, ii) the GDD approach with a base temperature of 10°C (GDD_{10} approach) and iii) the GDD method with a base temperature that has been optimized according to the same approach than in the eGDD method (GDD_{opt} approach). A comparison criterion was defined for each use.

For the Prediction use, the four calendars were compared according to a cross-validated Root Mean Square Error (RMSE) between predicted and observed dates of achievement of budbreak, bloom and veraison respectively.

For the Synchronization use, the four calendars were compared according to the highest number of periods that could be discretized according to corresponding scores (or dates for Gregorian Calendar) in the one-year time series so that budbreak, bloom and veraison in all years respectively occur within the same period. The more synchronized the phenological stages are between years, the shorter the periods and the higher the number of periods that can be discretized. Short and precise periods, which can eventually be aggregated, were assumed as determining for the success of any subsequent time series analysis.
The maximum number of periods to be discretized in time series was calculated with an optimization process at one parameter. This optimization aimed at identifying the minimum duration of a period which, once the time series divided by this period duration, allowed budbreak, bloom and veraison in all years respectively to occur within the same period but different between the three. The maximum number of discretized periods was obtained for each site by dividing the time series by the minimum period duration. For a given site, all time series were truncated according to the shortest time series to calculate a number of periods comparable between series while dividing them by a fixed period duration.

In both cases, a cross-validation was performed. It led to the exclusion of all the data corresponding to one year for each repetition. The number of repetitions was equal to the number of years in the data set. The results of the cross validation were expressed as means over all the repetitions.

For each phenological stage, the mean and extreme values of the results over the 79 blocks of the data set are presented in this paper. To further illustrate such results, 3 blocks have been chosen as examples so as to represent the specific combinations of optimized values for T_b, T_o^1, T_o^2 and T_c that could be observed. These blocks were selected amongst the 79 according to their position in the individuals cloud of a principal components analysis (PCA) using the T_b, T_o^1, T_o^2 and T_c values as variables and blocks as individuals. One PCA was computed for each use, Prediction or Synchronization.

4.3.4 Results

4.3.4.1 Implementation of the eGDD method

4.3.4.1.1 Description of the PAST function, the set of optimized parameters and the related thermal indices for a given site

For each site (here, each block) the results of the eGDD method include a set of optimized values for T_b, T_o^1, T_o^2 and T_c parameters that allows defining a PAST function and a thermal index computed for each year of the analysed data set. The minimum value of the optimization criterion that allowed the selection of these parameters is also given. An example of graphical outputs of the eGDD method is shown in Fig. 4. 6. It corresponds to a Synchronization use on the eGDD method for one of the block (block 3) and using the 13 years of the data sets for the calibration process. In this case, the optimized values of T_b, T_o^1, T_o^2 and T_c were of 9.6, 21.1, 23.1 and 34.9°C respectively. Therefore, the shape of the PAST function presented in Fig. 4.6a was similar to the one presented in Fig. 4.4a. The thermal index resulting from the integration of the PAST function over the temperature annual history of 2003 is presented in Fig. 4b. It started from 0 TODH on the first day of the year and reached its maximum value of 2090 TODH on DOY 358. In the year 2003,

the temperature conditions that had been experienced by grapevine in block were considered as equivalent to 2090 hours spent in daylight and at optimal temperatures i.e. between 21.1 and 23.1°C. In addition, the thermal index increase i.e. the accumulation of TODH was not regular during the season. It can be seen in Fig. 4b that the slope of the thermal index plotted against the Gregorian calendar is not constant throughout the season. The slope was null at the beginning of the season, due to daily mean temperature inferior to the Tb of 9.6°C. It progressively increased from DOY 55 to DOY 200 in relation with the increase of daily mean temperatures and photoperiod during the spring and beginning of the summer. It remained constant for most of the days until DOY 240, in relation with most daily mean temperatures being comprised between 21.1 and 23.1°C. Some decreases in the slope could be noticed, for example on DOY 214 to 225 in relation with daily mean temperatures being above 23.1°C. Finally, the slope decreased from DOY 240 onward in relation with the photoperiod decrease and daily mean temperatures also decreasing until being inferior to 9.6°C. The thermal indices corresponding to the 13 years of the data set are shown in Fig. 4.6c. The curves have similar shapes with differences mainly observed during the periods between DOY 80 and 160 and after DOY 250. The periods of spring and early summer then of autumn, were the periods that allowed differentiating the years from a thermal point of view for this given block. It could be noticed that the different thermal indices reached different maximum values.

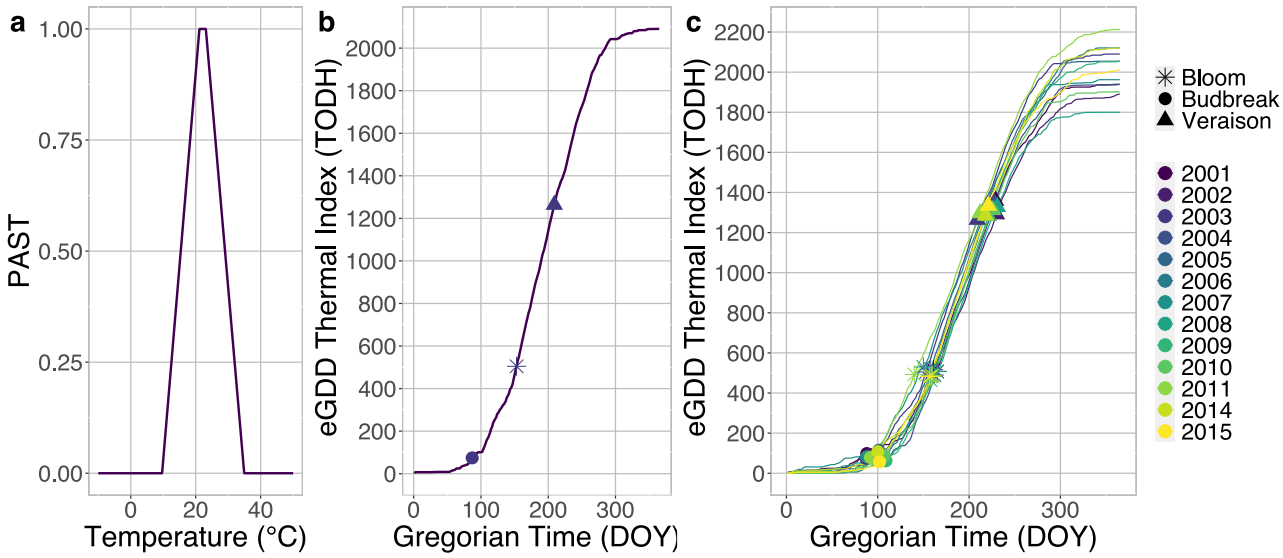

Figure 4.6 – Example of the outputs of the Extended Growing Degree Days (eGDD) method for one block (block 3). Figure a corresponds to the optimized Phenological Advancement Speed as a function of Temperature (PAST function). Figure b corresponds to the eGDD thermal index obtained by integration of the PAST function on the 2003 history of daily mean temperature. It is plotted against the Gregorian calendar in Day of the Year (DOY). Figure c corresponds to the eGDD thermal indices obtained by integration of the PAST function on th daily mean temperature history of all the years of the data set. They are also plotted against the Gregorian calendar in DOY. Achievement dates of budbreak, bloom and veraison expressed in the eGDD thermal index and in the Gregorian calendar are shown as references.

4.3.4.1.2 Overview of the eGDD outputs for the case study

The values obtained for the calibration criterion differed between blocks but were consistent between the Prediction and Synchronization uses

Applied to the 79 blocks of the use case, the eGDD method resulted in the selection of parameters sets for an optimization criterion value ranging from 33.0 to 83.5 with a mean of 59.0 for the Prediction use and from 3.5 to 10.4 with a mean of 6.0 for the Synchronization use.

In addition, the blocks that obtained a small value of optimization criterion in the Prediction case (P criterion) also obtained a small value in the Synchronization case (S criterion). For example, in Table 4.5, the blocks 1 to 3 presented a decreasing order in their P values (62.5, 46.3 and 33.0 respectively) and also in their S values (6.4, 5.6 and 3.5 respectively). Thus, the blocks presenting the best results after the calibration process are generally the same for the Prediction and Synchronization uses.

Table 4.5 – Values of the optimization criterion obtained with the Extended Growing Degree Days (eGDD) method for the Prediction and Synchronization uses (respectively P and S) and corresponding parameters for three individual examples (Blocks 1, 2, 3). The optimized parameters correspond to the base temperature (T_b), optimal temperatures (T_o^1 and T_o^2) and critical temperature (T_c).

Block	Prediction					Synchronization				
	P	T_b	T_o^1	T_o^2	T_c	S	T_b	T_o^1	T_o^2	T_c
1	62.5	9.6	20.9	21.6	33.1	6.4	10.3	20.6	20.6	35.0
2	46.3	9.0	20.7	20.7	34.7	5.6	11.2	20.0	20.0	32.6
3	33.0	7.6	23.1	23.1	35.0	3.5	9.6	21.1	23.1	34.9

The Prediction use resulted in parameters values that were more dispersed across the 79 blocks than the Synchronization use

Across the 79 blocks, T_b optimized values ranged from -5 to 9.7°C with a mean of 6.2°C for the Prediction use and from 8.6 to 11.6 °C with a mean of 10.1°C for the Synchronization use (Fig 4.7a). T_o^1 optimized values ranged from 20.5 to 25 °C with a mean of 21.5°C for the Prediction use and from 18.5 to 21.3°C with a mean of 20.6°C for the Synchronization use. (Fig 4.7b). T_o^2 optimized values ranged from 20.7 to 25 °C with a mean of 22.6°C for the Prediction use and from 20 to 25°C with a mean of 21.1°C for the Synchronization use (Fig 4.7c). T_c optimized values ranged from 32.3 to 35 °C with a mean of 34.7°C for the Prediction use and from 30.8 to 35°C with a mean of 34.5°C for the Synchronization use. (Fig 4.7d). In addition, a group of 6 blocks differed from the whole in the case of Prediction, with a different type of parameters combination : T_b was around -5°C while the majority had T_b values between 6.5 and 9.8 °C and $T_o^1$and T_o^2 were around 25°C while the majority obtained values between 20.5 and 21.5°C for T_o^1 and between 20.7 and 23.1°C for T_o^2. This small group didn't show differences with the majority regarding the $T_c$values.

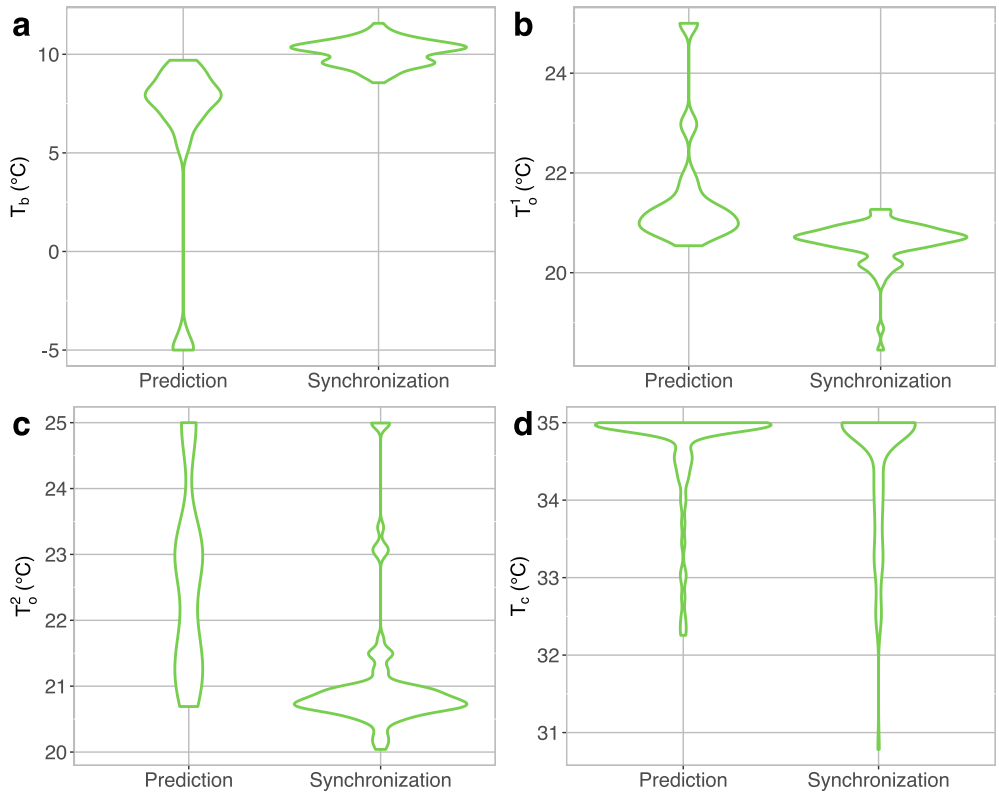

Figure 4.7 – Distribution over the 79 blocks of the values of the base temperature (T_b), optimal temperatures (T_o^1 and T_o^2) and critical temperature (T_c) of the Phenological Advancement Speed as a function of Temperature obtained with the Extended Growing Degree Days (eGDD) method for the Prediction and Synchronization uses respectively.

Different parameters combinations were observed between the 79 blocks and they were not consistent between the Prediction and the Synchronization uses

The mean difference between T_b and T_o^1 is 10.5°C for the Synchronization use and 15.3°C for the Prediction use. Thus, the ascending part of the PAST function seemed more stretched for the Prediction than the Synchronization use. In addition, the mean difference between $T_o^1$and T_o^2 is 0.5°C for the Synchronization use and 1.1°C for the Prediction use. Thus, the range of optimal temperatures is very small in the case of the Synchronization use, resulting in an almost triangular PAST function as in Fig. 4.4c. In the case of the Prediction use, the interval of optimal temperatures is wider, resulting in a trapezoidal PAST function as in Fig. 4.4a.

However, it can be seen in Table 4.5 that different parameters combinations stood behind the means mentioned in the previous paragraph. In particular, the blocks that had the smallest T_b values were not necessarily the ones that had some of the smallest values for the other parameters and vice versa. For example, block 3 had a T_b of 7.6°C but a T_o^1 of 23.1°C while block 1 had a T_b of 7.6°C but a T_o^1 of 20.9°C. The same observation could be made for the use of Synchronization. It can also be noted that trapezoidal (as in Fig. 4.4a) and triangular (as in Fig. 4.4c or 4.4d) shapes of the PAST function were observed across the 79 blocks for both uses. Note however that triangular shapes as in Fig. 4.4b, where T_b is equal to T_o^1, were not observed. In addition, the same block did not display the same PAST function shape for the Prediction and Synchronization uses. For example, some blocks, e.g. block 1, showed a trapezoidal shape for the Prediction use but a triangular shape for the Synchronization use. Conversely, some blocks, e.g. block 3, showed a triangular shape for the Prediction use and a trapezoidal shape for the Synchronization use.

4.3.4.2 Evaluation of the eGDD method

4.3.4.2.1 The eGDD method is the best trade-off to predict the achievement dates of all the three budbreak, bloom and veraison

For both Prediction and Synchronization uses, the eGDD method was compared to i) the Gregorian calendar, ii) the GDD_{10} approach and iii) the GDD_{opt} approach. The optimized values of T_b in the GDD_{opt} approach were around -4.9°C for all blocks in the Prediction use and ranged from 7.5 to 8.4°C in the Synchronization use.

For the Prediction use, the eGDD method was compared to the other approaches according to the distribution across the 79 blocks of the cross-validated RMSE between observed and predicted dates (days) for budbreak, bloom and veraison.

Regarding budbreak prediction, the cross-validated RMSE between observed and predicted dates ranged from 3. 6 to 8.5 days with a mean of 6.1 days for the Gregorian calendar, from 6.0 to 11.4 days with a mean of 8.9 days for the GDD_{10}approach, from 2.7 to 7.7 days with a mean of 4.4 days for the GDD_{opt} approach and from 3.0 to 7.6 days with a mean of 5.3 days for the eGDD approach (Fig 4.8a).

Regarding bloom prediction, the cross-validated RMSE between observed and predicted dates ranged from 6.1 to 7.9 days with a mean of 7.1 days for the Gregorian calendar, from 2.1 to 4.1 days with a mean of 2.9 days for the GDD_{10} approach, from 5.0 to 7.4 days with a mean of 6.3 days for the GDD_{opt} approach and from 1.7 to 5.5 days with a mean of 3.3 days for the eGDD approach (Fig 4.8b).

Regarding veraison prediction, the cross-validated RMSE between observed and predicted dates ranged from 5.3 to 7.6 days with a mean of 6.3 days for the Gregorian calendar, from 4.5 to 7.8 days with a mean of 6.3 days for the GDD_{10} approach equal to 10°C, from 4.9 to 7.6 days with a mean of 6.1 days for the GDD_{opt} approach and from 2.4 to 5.7 days with a mean of 3.8 days for the eGDD approach (Fig 4.8c).

Thus, the eGDD method showed the lowest cross-validated RMSEs in the case of veraison and the second lowest in the case of budbreak and bloom, with a difference of less than one day from the best RMSE. The eGDD method appeared as the best trade-off to predict the achievement dates of all the three budbreak, bloom and veraison. In addition, the GDD approaches did not always show better RMSEs than the Gregorian calendar : only the GDD_{opt} approach displayed lower RMSEs than the Gregorian calendar in the case of budbreak, only the GDD_{10} approach showed lower RMSEs in the case of bloom and both GDD approaches presented higher RMSEs than the Gregorian calendar in the case of veraison.

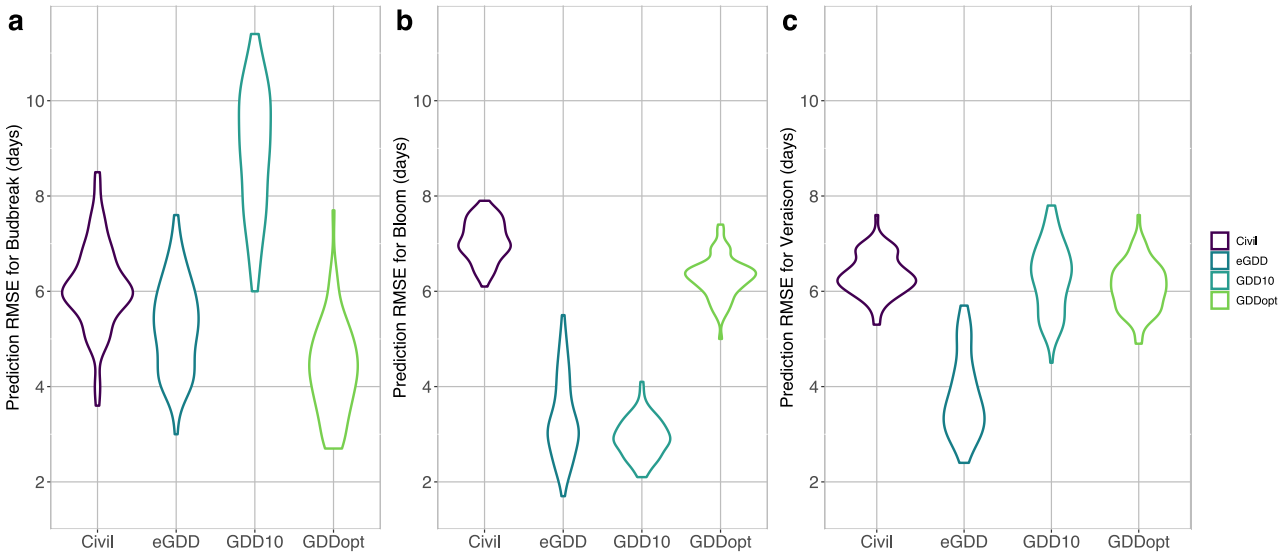

Figure 4.8 – Distribution over the 79 blocks of the cross-validated Prediction Root Mean Square Error (RMSE) for budbreak (a), bloom (b) and veraison (c) obtained with the Gregorian calendar approach (Civil), the Growing Degree Days approach with base temperature of 10°C (GDD_{10}), the Growing Degree Days approach with an optimized base temperature (GDD_{opt}) and the Extended Growing Degree Days method with a Prediction use (eGDD).

4.3.4.2.2 The method presented the best results in general for the Synchronization use but they were more scattered between blocks

For the Synchronization use, the results of the eGDD method are plotted in Fig. 4.9 as a distribution across the 79 blocks of the cross-validated maximum number of periods that could be discretized for each site (block) with budbreak, bloom and veraison of all years respectively occurring within the same period. The results of the three other approaches are also shown. The cross-validated maximum number of periods ranged from 6 to 11 with a mean of 8 for the Gregorian calendar. It ranged from 4 to 9 with a mean of 6 for the GDD_{10} approach. It ranged from 5 to 9 with a mean of 7 for the GDD_{opt} approach. Finally, it ranged from 8 to 16 with a mean of 12 for the eGDD model.

Thus, both GDD approaches displayed a lower number of discretized periods than the civil calendar. Only 17 out of the 79 blocks showed a higher number of discretized periods for the GDD_{opt} approach than the Gregorian Calendar. The eGDD method presented the highest number of discretized periods. Only 3 out of 79 blocks presented a higher number of discretized periods for the Gregorian calendar than the eGDD approach. However, the number of discretized periods was more scattered across the 79 blocks for the eGDD approach than for the others. The improvement between the eGDD results and the best results of the other three approaches was of 3.5 periods on average, with a standard deviation of 2.1.

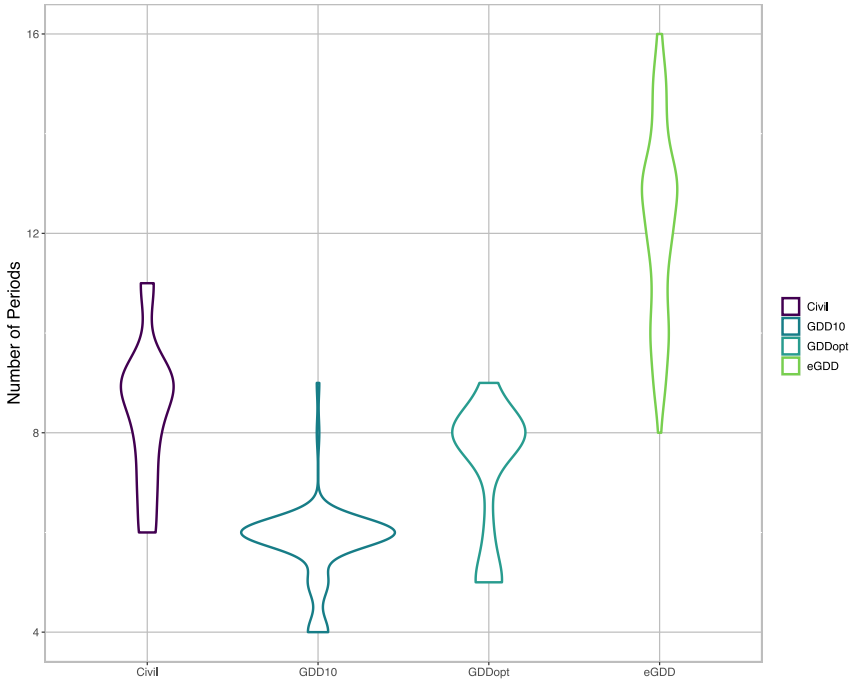

Figure 4.9 – Distribution over the 79 blocks of the cross-validated maximum number of periods that are discretized for each block with budbreak, bloom and veraison of all years respectively occurring within the same period for the Gregorian calendar approach (Civil), the GDD approach with base temperature of 10°C (GDD_{10}), the GDD approach with an optimized base temperature (GDD_{opt}) and the eGDD approach with a Synchronization use (eGDD).

For informative purposes, the conversion of the duration of each period into days varied according to the period position within the year and to the thermal index and the year considered. For the year 2003, such a period corresponded to 105 days at the beginning of the season and up to 8 days during the summer with a mean of 11.2 days without counting the first period of the year.

4.3.5 Discussion

4.3.5.1 The eGDD method proposes four improvements compared to the GDD approach

4.3.5.1.1 The eGDD method considers additional temperature thresholds to represent the relationship between grapevine development and temperature through the PAST function

Consistently with the ecophysiology literature, the eGDD method proposes to consider i) a base temperature from which grapevine starts developing but also ii) a set of optimal temperatures from which grapevine has reached its maximum development speed and iii) a critical temperature from which grapevine develops at lower or no speed (Pouget 1981; Gouot et al. 2019b; Yan and Hunt 1999; Vasconcelos et al. 2009; Keller, Tarara, and Mills 2010; Garcia De Cortazar Atauri et al. 2017; Camargo-Alvarez et al. 2020).

After calibration, the optimized parameters found by the eGDD method were consistent with others literature results, taking into account that the eGDD method uses daily mean temperatures. The T_b values were in the range of what was found by Zapata, Salazar-Gutierrez, et al. 2017 for 17 varieties : 6.1 to 8.4°C for budbreak, from 7.2 to 10.5°C for bloom and from 9.4 to 12.8°C for veraison. The T_b values found by Parker, De Cortázar-Atauri, et al. 2011 were lower (between 0° and 3°C) for bloom and veraison but this difference is explained by the fact that they used a GDD approach initiating from DOY 50 to 100. The T_b obtained with the eGDD method were also in the range found by Molitor, Fraga, and Junk 2020 and so were T_o^1 and T_o^2 values although they tended to be higher by a couple of degrees. The T_o^1 and T_o^2 values are also in the range of optimal temperatures reported for various physiological processes from flower induction

to berry growth and maturation (Buttrose 1970; Staudt 1982; Ebadi, May, and Coombe 1996; Dokoozlian 2000; Keller, Tarara, and Mills 2010; Pagay and Collins 2017). Finally, Pagay and Collins 2017 also reported critical temperatures above 35°C.

In addition, it should be noted that a small set of blocks presented surprising values for some temperature thresholds. For example, some blocks presented T_b values around -5°C for the Prediction use (Fig. 4.8). This did clearly not represent realistic values from an ecophysiological point of view. However, such values allowed a certain weighting of the daily mean temperatures actually recorded in the field i.e. a certain slope of the PAST function between T_b and T_o^1 This weighting obtained the best results during the optimization and likely led to the selection of surprising values for the temperature thresholds.

4.3.5.1.2 The eGDD method proposes a site-specific calibration of the PAST function using a constrained optimization approach

The eGDD method proposes a constrained optimization approach allowing a site-specific calibration of the temperature thresholds in the PAST function from operational data. The interest of site-specifically calibrating the thermal index has been demonstrated to take into account the diversity of factors that, at a local scale, can modulate the observed response of grapevine development to temperature (Nendel 2010; Caffarra and Eccel 2010; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017; de Rességuier et al. 2020; de Rességuier et al. 2020). The predominant sensitivity of the thermal index calibration has also been highlighted by (Fila et al. 2014). In coherence with these studies, the site-specifically optimized parameters and their combinations found with the eGDD method were different between blocks, with a higher dispersion in the case of Prediction.

4.3.5.1.3 The eGDD method both addresses the two different objectives of Prediction and Synchronization in its calibration process

The objective of timing phenological stages with consistent scores (i.e. dates expressed in the thermal index) across years (Synchronization) is different from that of accurately predicting dates (in the Gregorian calendar) for phenological stages (Prediction). This difference has not been formalized in the literature but it can be understood by looking at the graph of eGDD thermal indices according to Gregorian time for a given block as shown in Fig. 4.6c. In the Synchronization case, the slope of the curve is to be minimized so as to minimize the dispersion of the scores of budbreak, bloom and veraison respectively along the y-axis. In the Prediction case, the slope of the curve is to be maximized so as to minimize the difference between the predicted and observed dates of the three phenological stages along the x-axis. This difference justified the use of two different optimization criteria for the calibration process. As a consequence, the calibration processes resulted in different optimized parameters sets for the Prediction and Synchronization uses. However, the blocks results were consistent for both uses : a block such as block 3 which showed low prediction errors also showed good synchronization accuracies and inversely.

4.3.5.1.4 The eGDD method integrates the modulation of the PAST function by the photoperiod in the construction of the thermal index

The eGDD method proposes to recognize that a daily mean temperature may not have the same effect on grapevine development according to the considered dates i.e. that grapevine thermal determinism may evolve during the season. This requirement was motivated by the fact that the eGDD method aims at synchronizing the scores or predicting the dates of the three phenological stages with the same parameters set. The eGDD method also proposes to recognize that grapevine thermal determinism may differ between latitudes, with different enlightenment modalities (Li-Mallet, Rabot, and Geny 2016). Therefore, the eGDD method includes a weighting of the PAST function by the photoperiod. As a consequence, the unit of an eGDD thermal index is not expressed in thermally optimal days but daylight hours (TODH). It allowed to improve the eGDD prediction and synchronization results compared to the integration over the season of the PAST function only (data not shown).

4.3.5.2 The eGDD method allowed improved predictions and synchronizations compared to the GDD approach

4.3.5.2.1 The eGDD method showed improved Prediction results for budbreak, bloom and veraison

Regarding budbreak prediction, only the GDD_{opt} approach and the eGDD method showed lower cross-validated RMSEs than the Gregorian Calendar. Using the fixed date of 1st of January to initiate the computation of the thermal indices, these results highlighted the interest of the site-specific calibration of the parameters, especially T_b. The RMSEs of the eGDD method were slightly lower but remained close to those of the GDD_{opt} approach. In average across the 79 blocks, the RMSEs of the eGDD approach were 0.9 day higher than those of the GDD_{opt} approach. Thus, the parameters T_o^1, T_o^2 and T_c seemed to play a role of minor importance in the case of the eGDD method. This could easily be understood because of the range of temperatures commonly observed during the period around budbreak. The weighting with the photoperiod didn't seem to improve the budbreak prediction either.

Regarding bloom prediction, all thermal indices showed lower cross-validated RMSEs than the civil calendar. The RMSEs of the eGDD method were slightly lower but remained close to those of the GDD_{10} approach. In average across the 79 blocks, the RMSEs of the eGDD approach were 0.4 day higher than the GDD_{10} approach. Thus, a GDD approach with a temperature around 10°C seems to be sufficient for the prediction of flowering. Given the commonly observed temperatures during the period around flowering and the values of the optimized T_b, the eGDD model can indeed be simplified as such for the blooming period.

Regarding the veraison prediction, the RMSEs of both GDD approaches were comparable to the RMSEs of the Gregorian calendar. The eGDD method showed the lowest cross-validated RMSEs for veraison. In average, it resulted in RMSEs lower by 2.9 days than the lowest RMSEs of the three other approaches. Therefore, the combination of improvements only proposed by the eGDD method seemed to play an important role in the improvement of the veraison prediction.

Regarding the prediction of the achievement dates of the three phenological stages with the same thermal index (same approach and same parametrization), the eGDD method showed the best results.

Thus, if some construction choices did not seem relevant for the prediction of budbreak or bloom (e.g. optimal and critical temperature thresholds or weighting by photoperiod), their contribution becomes apparent to balance the results of the eGDD index over the three phenological stages and thus lead to the best trade-off for predicting the three at the same time.

4.3.5.2.2 The eGDD method showed improved the Synchronization of time series according to budbreak, bloom and veraison scores

Regarding the Synchronization use, the GDD_{opt} approach showed slightly better results than the GDD_{10} approach. However, both GDD models didn't result in better discretization results than the Gregorian Calendar. This shows that using the GDD model, even by calibrating the temperature with a criterion adapted to the synchronization needs, is not sufficient to improve the discretization made in the empirical case, based on the Gregorian calendar. The eGDD method showed the best discretization of time series. Again, this demonstrated the relevance of the combination of the four original proposals included in the eGDD method to improve the synchronization of time series based on grapevine phenology. Compared to the Gregorian calendar, the GDD methods allowed to reduce the dispersion of the numbers of discretized periods among the different blocks. This shows some consistency in relying on temperature to monitor each specific block. The eGDD method resulted in an increased dispersion but with a major part of the 79 blocks results were higher than the other three approaches. This showed that the proposals integrated into the eGDD method have led to an overall improvement in the results but that the potential for improvement is higher for some blocks than for others.

4.3.5.3 The results of the eGDD method, as of other thermal index approaches, should be interpreted with caution

4.3.5.3.1 The eGDD method aims at identifying a trade-off between the thermal determinism of budbreak, bloom and veraison

For the Prediction use, the RMSE presented for the eGDD method were better for flowering and veraison than for budbreak (except for a group of 6 blocks).These results corresponded to the minimum value of P,

obtained by summing the prediction RMSE of each phenological stage. However, another local minimum of P favoring the budbreak RMSE was generally found for all blocks. It corresponded to very low T_b values (-15°C) as well as to very high and close T_o^1 and T_o^2 values (25-28°C), which introduced another relationship between the weights given by the PAST function to the temperatures observed during the budburst period. This shows that the way temperature data should be considered to predict budbreak is clearly different from the one of bloom and veraison. In this way, it should be remembered that the eGDD method proposes to consider a PAST function as a trade-off between the three determinisms for practical purposes. Therefore, the eGDD optimized parameters should be interpreted accordingly and not directly compared with stage-specific literature results. Regarding the Synchronization use, this point was less apparent because the budbreak scores were low (at the beginning of the season), so their synchronization was not a major problem compared to bloom and veraison.

4.3.5.3.2 The eGDD thermal index includes the effects of the thermal determinism of grapevine phenology, of temperature spatial variability and of data quality

The differences in the optimized parameters values between blocks could be directly interpreted as differences in the thermal determinism of grapevine phenology due to a block-specific interaction of various environmental factors. However, the PAST function that underlies any thermal index incorporates not only the grapevine response to the forcing effect of temperature but also a site-specific, possibly non-linear, bias due to the spatial variability of the temperature and to data quality (temperature and phenological observations). Therefore, the optimized parameters T_b, T_o^1, T_o^2 and T_c and the resulting PAST function are not to be compared as such with results of purely ecophysiological experiments. For example, given that there is only one weather station for all blocks, the fact that some blocks are characterized by a lower T_b than others may simply mean that they are located in a warmer area, which is not well represented by the temperature data. This may also be partly due to the variable quality of phenological observations in the field (which is not a trivial operation). The differences in the Prediction or Synchronization results of the eGDD method between blocks could be partly explained by the fact that the forcing effect of temperature is not the only factor influencing grapevine phenology and that other influencing factors (other temperature effects, water regime, cultural practices, etc.) vary between blocks in the data set. A priori from this data set, this is not related to the grape variety. Again, some part of the Prediction or Synchronization results variability may also be explained by the data quality.

In this way, the eGDD model seems more dedicated to practical use and for a better local knowledge inferred from field data than to generate proper knowledge on grapevine thermal determinism. However, the finer the data that is used to run the eGDD method, the more likely its results are assumed to represent the reality of grapevine physiology. For example, using hourly mean temperatures instead of daily mean temperatures results in finding parameters much closer to the temperature thresholds announced in the ecophysiological literature (data not shown). The eGDD method could also be used not for the whole season but for a portion of the season, to adapt the curve to the specific determinism of a phenological stage when studying it in a wide range of temperature conditions (so as to encompass all temperature thresholds).

4.3.5.3.3 The eGDD method is designed to work with farm data and for operational purposes

Due to logistical reasons, only a few phenological stages are monitored in commercial vineyards. They generally relate to budbreak, bloom and veraison that represent key stages in grapevine phenology for growers. In terms of constrained optimization, this means that the calibration of the PAST function is performed with at most three points to fit the functional form. Therefore, a trade-off had to be found between the quality of representation of a grapevine physiological response to temperature and the robustness of its calibration. For example, a smoothed curve with 4 temperature thresholds would have required 8 parameters to be optimized, which may be too many with regard to the volume and the quality of operational data generally encountered in commercial situations.

In addition, not all temperatures are represented by the same population over the year. Some daily average temperatures like 15°C are much more represented than daily average temperatures of 0 or 28°C. This means that the section of the PAST function around 15°C is fitted with more data than the section around 0°C or 28°C. Therefore, the parts corresponding to smaller volumes of data should not be calibrated with too much freedom, otherwise each optimization would lead to a different result.

These requirements justified the choice of the function given in Eq. 4.2 and 4.3 of trapezoidal inspiration and controlled with four parameters. More complex functions, i.e. with more parameters, would likely lead to a non converging optimization problem on the basis of operational data. Thus, the results of the eGDD

method should be considered as simplified results with respect to grapevine physiological reality but yet realistic and robust enough to support decision making in operational conditions.

4.3.5.4 A thermal index is always an objective-oriented model to understand the thermal determinism of grapevine phenology

It remains conceptually debatable whether the Prediction or Synchronization use is the most appropriate approach to study the thermal determinism of grapevine phenology. There is indeed no absolute reference in indicators, methods and criteria aiming at studying the thermal determinism of grapevine phenology through the construction of a thermal index and at evaluating the relevance of such an index. In this way, the precise values obtained with the eGDD method for each parameter in each case study should not be understood as accurately describing the reality of grapevine thermal determinism. Instead, they should be considered different understanding frameworks of grapevine thermal determinism given different objectives (synchronizing thermal scores or predicting Gregorian dates). Therefore, the agronomic hypotheses driving the thermal index construction should be transparent and the corresponding results have to be interpreted consciously. As many case studies can be covered by the eGDD method, special attention should be paid to adapting the conclusions that are drawn to the data that have been used to run it (e.g. daily or hourly temperature data have different meanings) and to the purpose of the eGDD method use (Prediction or Synchronization). Note however that these precautions are also valid for the classical GDD method, whose use has sometimes become an unquestioned habit.

4.3.6 Conclusion and perspectives

The eGDD method offers four improvements with respect to the GDD approach : i) it considers additional thresholds for the relationship between grapevine development and temperature, ii) it uses a constrained optimization approach to site-specifically calibrate these thresholds, iii) it drives the calibration process according to the objective use of the thermal index (Prediction or Synchronization) and iv) it proposes an integration of the resulting PAST function over the growing season combined with the photoperiod. For both the Prediction and Synchronization uses on a commercial vineyard in the Bordeaux region, the overall performances of the eGDD method were shown better than the Gregorian calendar or the GDD method. This validates the construction choices the method integrates. In addition, these results are to be consolidated by processing other data sets in other locations so as to test more differences between the specific conditions of each site that the method is designed to fit. Finally, the eGDD method has been designed to be adaptable to many case studies. Plenty of definitions of a site could be used with the eGDD method. For example, a site could represent a vineyard, a group of blocks that are supposed to have a similar operational response to temperature. In parallel, different time steps can be used for the temperature data (e.g. hours) so as to make the most of each operational data set. The eGDD method could be used for other variables monitoring grapevine development (e.g. biomass or shoot elongation measurements) or for other crops, in particular perennial ones.

4.3.7 Acknowledgements

This work was supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004.

4.4 eGDD : an extension of the Growing Degree Days approach to compute a site-specific thermal index in viticulture using a constrained optimization method. Part II : Method validation

4.4.1 Details about the paper

4.4.1.1 Title and publication informations

The work presented hereafter will be submitted as a two-part scientific article to the journal Agriculture and Forest Meteorology. The first part of the article was presented in section 3 and the second part is subsequently presented in this section. It is entitled “eGDD : an extension of the Growing Degree Days

approach to compute a site-specific thermal index in viticulture using a constrained optimization method. Part II : Method validation".

4.4.1.2 Authors

C. Laurent[1,2,3,*], G. Le Moguédec[4], J. Taylor[3], T. Scholasch[1], B. Tisseyre[3],A. Metay[2]

[1] Fruition Sciences, 34000 Montpellier, France
[2] ABSYS, Univ. Montpellier, CIRAD, INRAE, Institut Agro, 34060 Montpellier, France
[3] ITAP, Univ Montpellier, INRAE, Institut Agro, 34060 Montpellier, France
[4]AMAP, Université Montpellier, INRAE, Cirad CNRS, IRD, 34000 Montpellier, France
[*] corresponding author : cecile@fruitionsciences.com

4.4.1.3 Abstract

Temperature is one of the main factors influencing grapevine physiology and a fortiori phenology. So far, numerous works have used a thermal index to predict the date of achievement of key phenological stages (Prediction) or as a timeline to synchronize data so that further analyses can be performed consistently with and according to grapevine phenology (Synchronization). Amongst these thermal indices, the Growing Degree Days approach is the most widely used. For both intended uses (Prediction or Synchronization), additional works have shown the interest of site-specifically calibrating thermal indices. The assumption underlying such a calibration is that the interaction of site-specific factors that influence grapevine response to temperature (pedo-climate, topography, cultural practices etc.) and data representativity of actual field conditions should be taken into account when interpreting thermal indices in the field.
Considering these assumptions led to propose a new method called Extended Growing Degree Days (eGDD). It provides a thermal index that is built by integrating phenological advancement speed as a function of temperature. The parameters of this function are site-specifically calibrated using a constrained optimization approach with ancillary data.
In this paper, the relevance of site-specifically computing a thermal index is investigated using data from three vineyards, respectively located in the Bordeaux region (France), California (USA) and Israel. Two thermal indices were computed with the eGDD method for each block of each vineyard, for a prediction and synchronization purpose respectively. The Prediction or Synchronization results of the corresponding thermal indices are compared to the results of a thermal index classically computed with the Growing Degrees Days approach.
The thermal indices obtained with the eGDD method were based on similar combinations of parameters between blocks of the same vineyard but different combinations between blocks of different vineyards. Their results for both Prediction and Synchronization purposes were found to be better than those of the classical thermal index (GDD). Such results support the assumption that thermal indices should reflect the site-specific interaction of various factors and confirm the interest of a site-specific calibration.

4.4.1.4 Keywords

Grapevine Phenology, Temperature, Thermal Time, Time Series Data Analysis, Operational Data

4.4.2 Introduction

Temperature is one of the main factors influencing grapevine physiology and a fortiori phenology (Tonietto and Carbonneau 2004; Pagay and Collins 2017; Prats-Llinàs et al. 2020). So far, numerous works have used a thermal index to predict the date of achievement of key phenological stages (Chuine 2000; Parker, De Cortázar-Atauri, et al. 2011; Fila et al. 2014; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017; Molitor, Fraga, and Junk 2020) or as a timeline to synchronize time series data sets with grapevine phenology so that further analyses can be performed consistently with and according to grapevine phenology (Guilpart, Metay, and Gary 2014). Among these thermal indices, the Growing Degree Days approach is the most widely used to achieve these Prediction and Synchronization purposes. Furthermore, for both intended uses, additional works have shown the interest of site-specifically calibrating thermal indices (Nendel 2010; Caffarra and Eccel 2010; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017; de Rességuier et al. 2020; Laurent, Scholasch, et al. 2020). The assumption underlying such a calibration is that the interaction of site-specific factors that influence grapevine response to temperature

(pedo-climate, topography, cultural practices etc.) and data representativity of actual field conditions should be taken into account when interpreting thermal indices in the field.

To address this issue, a new method called Extended Growing Degree Days (eGDD) has been proposed. The eGDD method i) includes additional thresholds to the conventional base temperature for the relationship between grapevine development and temperature, ii) uses a constrained optimization approach to site-specifically calibrate these thresholds with available data sets, iii) drives the calibration process according to the objective use of the thermal index (Prediction or Synchronization) with a specific optimization criterion for each use and iv) provides an integration of the resulting relationship between grapevine development and temperature over the growing season combined with photoperiod.

The eGDD method has been built to work with farm data and in various operational contexts, implying different pedo-climatic environments and vineyard managements. The first part of this paper presented the development of the method on the basis of a single farm data set. The objective of the second part hereby aims at extending the method validation to other farm data sets and thus to other operational contexts. To do so, the eGDD method was applied to the data set of three different commercial vineyards respectively situated in the Napa Valley (California, USA), Israel and the Bordeaux region (France). These vineyards differ in terms of pedo-climatic environment, latitude, topography and altimetry but also in terms of vineyard management. The data sets are also characterized by different volume (number of years) and quality (in relation with measurement, sampling and traceability issues). The eGDD method was applied to 3 blocks of each vineyard for both a Prediction and a Synchronization use.

It was first verified that the eGDD method resulted in better prediction errors and synchronization accuracies than the Gregorian calendar or the classical GDD approach. The prediction error was estimated through a Root Mean Square Error between observed and predicted dates. The synchronization accuracy was assessed through the highest number of periods that could be discretized in each time series while ensuring that budbreak, bloom and veraison respectively occurred in the same period for all years of the same block. In a second time, it was investigated whether the eGDD method offers greater value if it is performed at the block or vineyard scale. To do so, the temperature thresholds optimized with the eGDD method were compared between blocks of each vineyard and between vineyards. Finally, studying the differences in the temperature thresholds across the blocks and vineyards corresponding to different operational contexts also provided the opportunity to deepen the interpretation that can be made of the thermal indices computed with the eGDD method.

4.4.3 Materiel and Method

4.4.3.1 Data description

Data was collected from three commercial vineyards situated in the Napa Valley (California, USA), Israel and the Bordeaux region (France). They are respectively noted vineyard A, B and C in this paper. For each vineyard, three blocks planted with Cabernet Sauvignon were studied. These blocks are respectively noted blocks 1, 2 and 3. The main characteristics of the three vineyards are presented in Table 4.6.

Table 4.6 – Characteristics of the three blocks of vineyards A, B and C.

	Vineyard A	Vineyard B	Vineyard C
Localization	California, USA	Israel	Bordeaux, France
Latitude	38	32	45
Altitude	210	770	20
Type of climate	semi-arid	semi-arid	oceanic
Area per block (hectare	0.5 to 2	1.2	1.1 to 1.4
Plantation date	1994	from 2005 to 2011	from 1985 to 1986
Plantation density (stocks/hectare)	2220	2500	8700
Rootstocks	1103P or 110R	101-14 MGt or none	101-14MGt or 3309C
Irrigation	yes	yes	no
Location of weather station	within the vineyard	2km away	within the vineyard

Data of vineyards A, B and C were available respectively from 2008 to 2017 (10 seasons), from to 2015 to 2019 (5 seasons) and for the years 2001-11 to 2014-15 (13 seasons). For each vineyard, daily average temperatures have been recorded using a commercial weather station located within or near the vineyard.

For each block, the dates of budbreak, bloom and veraison have been collected according to the Gregorian calendar at 50% of achievement.

4.4.3.2 Computation of the thermal indices

For each block of the 3 vineyards, 3 thermal indices were computed. The first thermal index corresponded to the Growing Degree Days approach with a base temperature (T_b) equal to 10°C. This approach is noted GDD_{10} henceforth. The second thermal index corresponded to the Growing Degree Days approach with an optimized T_b. This approach is noted GDD_{opt} henceforth. The third thermal index corresponded to the eGDD method. It led to the optimization of four temperature thresholds : T_b, the two bounds of the range of optimal temperatures (T_o^1 and T_o^2) and a critical temperature (T_c). The optimization of T_b, in the GDD approach and of T_b, T_o^1, T_o^2 and T_c in the eGDD method were performed for both objective use (Prediction and Synchronization) according to the criteria that were described in the first part of this paper.

4.4.3.3 Evaluation of the prediction error and synchronization accuracy allowed by the eGDD method

For each block of the 3 vineyards and for both a Prediction and a Synchronization use, the eGDD method was compared to i) the Gregorian calendar, ii) the GDD_{10} approach and iii) the GDD_{opt} approach. A comparison criterion was defined for each use.
the Prediction use, the comparison was performed according to a cross-validated Root Mean Square Error (RMSE) between predicted and observed dates of achievement of budbreak, bloom and veraison respectively.

For the Synchronization use, the comparison was performed according to the highest cross-validated number of periods that could be discretized in the one-year time series. The time step used to discretize the time series was expressed in the unit of each thermal index i.e. Growing Degree Days for the GDD approach and Thermally Optimal Days (TODH) for the eGDD method. Gregorian days were used in the case of the Gregorian calendar. The highest number of discretized periods in each time series was obtained by minimizing the time step that allowed budbreak, bloom and veraison scores or dates to respectively occur within the same period for all years of each block. The periods were also constrained to be different between the three phenological stages.

In both cases, the cross-validation led to the exclusion of all the data corresponding to one year for each repetition. The number of repetitions was equal to the number of years in the data set. The results of the cross validation were expressed as means over all the repetitions.

4.4.3.4 Comparison of the optimized temperature thresholds obtained with the eGDD method

For both Prediction and Synchronization uses, the block-specific results of the eGDD method were compared according to the value of each optimized temperature threshold, in relation with the daily mean temperatures actually observed in each vineyard. A principal component analysis (PCA) was then performed using the four optimized temperature thresholds as variables to differentiate the 9 blocks as individuals.

4.4.4 Results

4.4.4.1 Validation of the eGDD method

4.4.4.1.1 The eGDD method resulted in amongst the lowest prediction errors

Regarding budbreak prediction (Fig. 4.10a), the RMSEs between observed dates and the predicted dates obtained with the GDDopt and the eGDD approaches were close and lower than the RMSEs corresponding to the Gregorian Calendar and the GDD_{10} approach each vineyard respectively.

Regarding bloom prediction (Fig 4.10a), the RMSEs between observed dates and the predicted dates obtained with the eGDD method were lower than the RMSEs corresponding to the Gregorian Calendar, GDD_{10} and GDD_{opt} approaches for vineyards A and B respectively. For vineyard C, the RMSEs obtained with the eGDD method were similar to the ones of the GDD_{10} approach but lower than the other two approaches.

Regarding veraison prediction (Fig. 4.10c), the RMSE between predicted and observed dates obtained with the eGDD method were lower than the RMSEs of the other approaches for vineyard C. They were

one-day-higher than the GDD_{opt} approach for vineyard A. In the case of vineyard B, the RMSEs obtained with the eGDD method were similar to those obtained with the Gregorian calendar for blocks 1 and 2 but lower than those of the other two approaches. The eGDD showed the lowest RMSEs in the case of block 3.

Thus, for each vineyard and each phenological stage, the eGDD method was among the best approaches regarding prediction. As a trade-off regarding the prediction of all phenological stages for the three vineyards, the eGDD method performed better than the Gregorian, GDD_{10} and GDD_{opt} approaches.

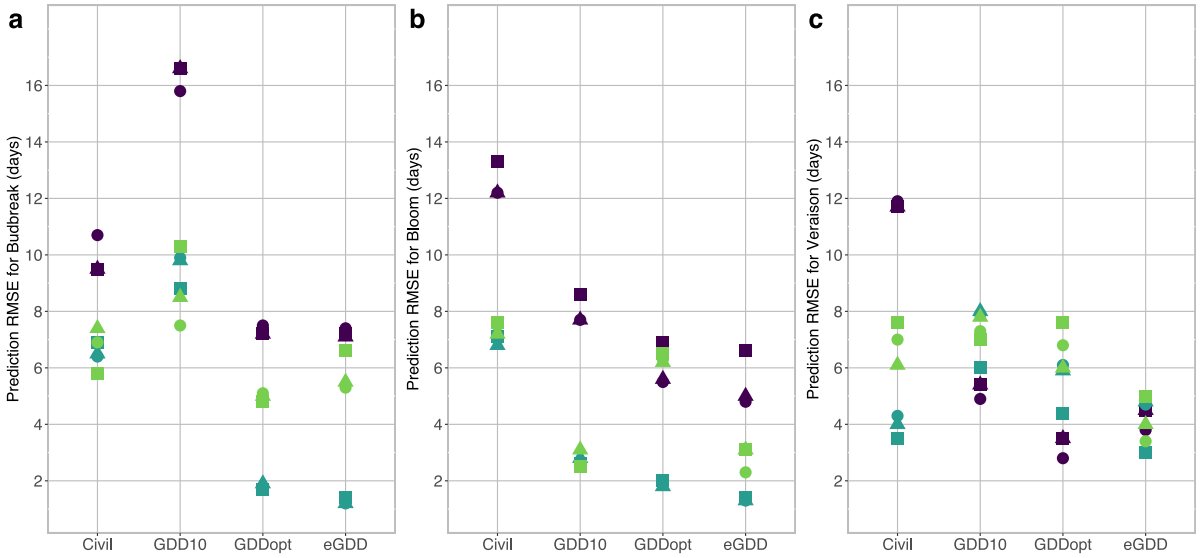

Figure 4.10 – Cross-validated Prediction Root Mean Square Error (RMSE) expressed in days for budbreak (a), bloom (b) and veraison (c) obtained with the Gregorian calendar approach (Civil), the GDD approach with base temperature of 10°C (GDD_{10}), the GDD approach with an optimized base temperature (GDD_{opt}) and the eGDD approach with a Prediction use (eGDD) for each block of vineyards A, B and C.

4.4.4.1.2 The eGDD method resulted in among the most accurate synchronizations of time series

Regarding the Synchronization use (Fig. 4.11), the eGDD method allowed the highest number of discretized periods for vineyards B and C. This result corresponded to a significant improvement compared to the best of the three other approaches : the number of periods was increased by 11 on average for vineyard B and by 6 on average for vineyard C. However, this success corresponded to very different values : the times series of vineyard B could be discretized into 28 periods on average whereas they could be discretized into 12 periods for vineyard C.

In the case of vineyard A, the GDD_{opt} approach resulted in a number of periods higher by 1 or 2 than the eGDD method for blocks 1 and 2. The eGDD method performed the best synchronization of time series for block 3. On average, the eGDD method allowed the time series of vineyard A to be discretized into 7 periods.

As a trade-off regarding the times series synchronization for the three vineyards, the eGDD method performed better than the Gregorian, GDD_{10} and GDD_{opt} approaches.

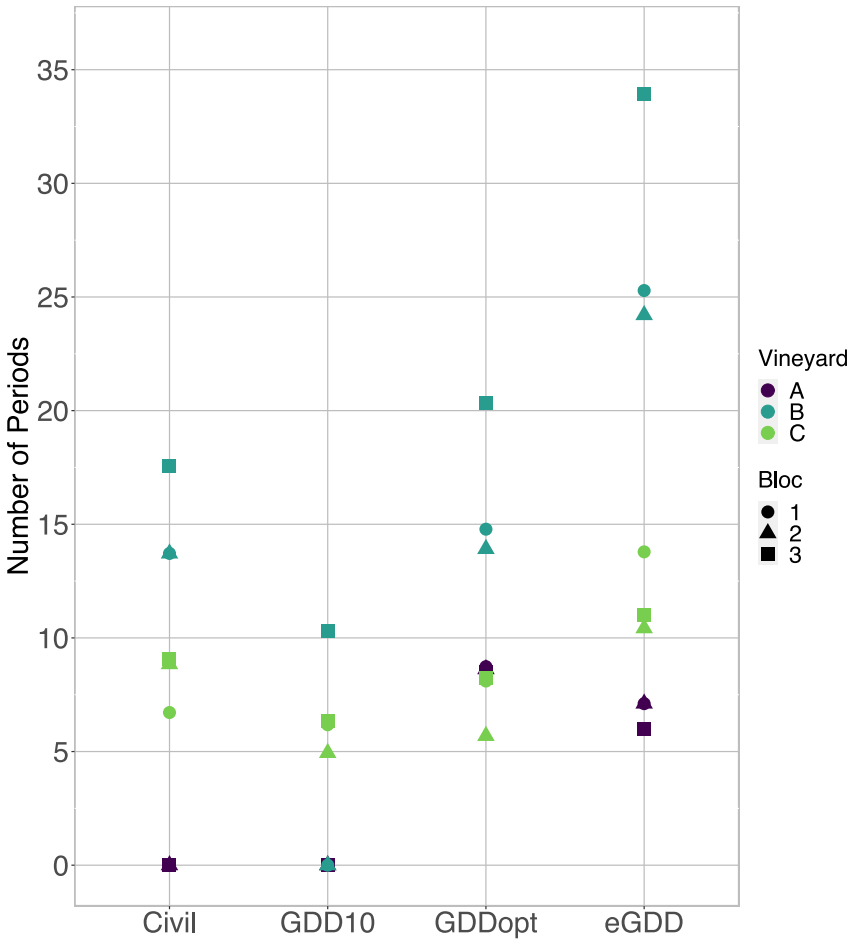

Figure 4.11 – Cross-validated maximum number of periods that are discretized for each block with budbreak, bloom and veraison of all years respectively occuring within the same period for the Gregorian calendar approach (Civil), the GDD approach with base temperature of 10°C (GDD_{10}), the GDD approach with an optimized base temperature (GDD_{opt}) and the eGDD approach with Synchronization use (eGDD) for each block of vineyards A, B and C. No discretization could allow the three phenological stages of all years to respectively occur within the same period for the Gregorian calendar and the GDD_{10}) approach in the case of vineyard A.

4.4.4.2 Investigation of the eGDD results for different blocks within the same vineyard and between vineyards

4.4.4.2.1 The three vineyards experienced different distributions of mean daily temperature over their history

The daily mean temperature distribution across all the years of the data set of each vineyard is given in Fig. 4.12. Vineyard A experienced high daily mean temperatures with a mean of 15.7°C and a maximum at 32.8°C. The daily mean temperatures were uniformly distributed within the second and third quartile i.e. 11.0 and 19.9°C. Vineyard B also experienced generally high daily mean temperatures with a mean of 15.2°C and a maximum of 30.8°C. However, daily mean temperatures around 8.0°C and around 22.0°C are more frequently recorded than temperatures in the intermediate interval. Vineyard C experienced lower daily mean temperatures in general with a mean of 13.4°C and a maximum of 29.6°C. The temperatures were uniformly distributed between the first and third quartile i.e. 9.1 and 18.1°C.
Vineyard A didn't experience negative daily mean temperatures while vineyards B and C experienced up to -3.4°C and -3.7°C respectively. Amongst the three vineyards, vineyard B experienced the wider range of daily mean temperatures.

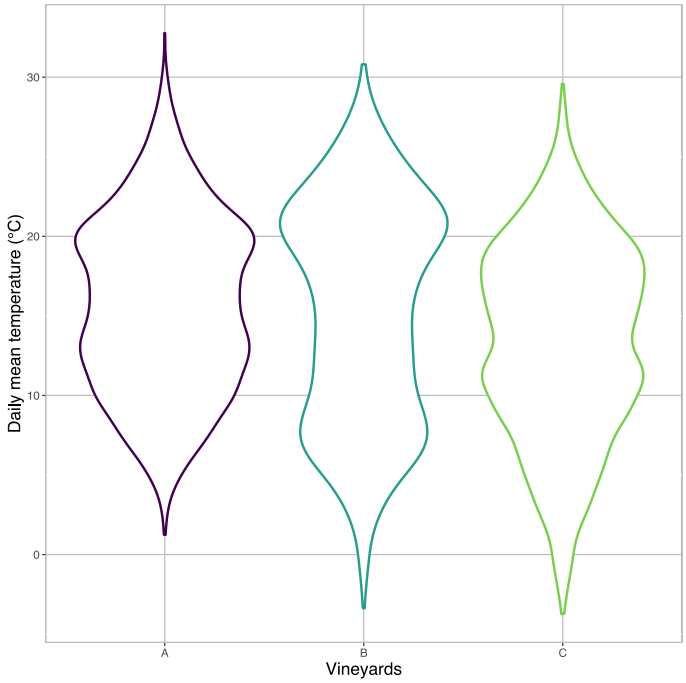

Figure 4.12 – Distribution of the daily mean temperature across the whole history of vineyards A, B and C data sets (respectively 10, 5 and 13 years)

4.4.4.2.2 The eGDD method resulted in a vineyard-consistent optimization of temperature thresholds

Regarding the optimization of T_b, the eGDD method with a Prediction use resulted in mean values of 7.6, 2.4, -5.4°C for vineyards A, B and C respectively (Fig. 4.13a). The eGDD method with a Synchronization use resulted in mean values of 10.6, 3.5 and -11.6°C for vineyards A, B and C respectively (Fig. 4.13b).

Regarding the optimization of T_o^1 and T_o^2 for both Prediction and Synchronization uses, vineyards B and C obtained close values while vineyard A obtained higher values. The mean values of T_o^1 for vineyards A, B and C were respectively 29.7, 21.9, 20.8°C for the Prediction use (Fig. 4.13a) and respectively 28.8°C, 20.8 and 19.8°C for the Synchronization use. The mean values of T_o^1 for vineyards A, B and C were respectively 30.0, 21.9, 21.6°C for the Prediction use and respectively 28.9, 21.0 and 20.6°C for the Synchronization use (Fig. 4.13b)

Regarding the optimization of T_c the vineyards were differently ranged between the Prediction and the Synchronization uses. For the Prediction use, vineyards A and C obtained higher values than vineyard B : the mean values of T_c were respectively 35.0, 32.9 and 34.9 °C for vineyards A, B and C. For the Synchronization use, (Fig. 4.13a) vineyards B and C obtained higher values than vineyard A : the mean values of T_c across the three blocks were respectively 31.7, 34.7 and 33.9°C (Fig. 4.13b).

Thus, for both Prediction and Synchronization, the optimized values of T_b allowed to differentiate the three vineyards. The optimized values of T_o^1 and T_o^2 allowed to differentiate vineyard A from vineyards B and C. The results were less vineyard-consistent for T_c and depended on the Prediction or Synchronization use.

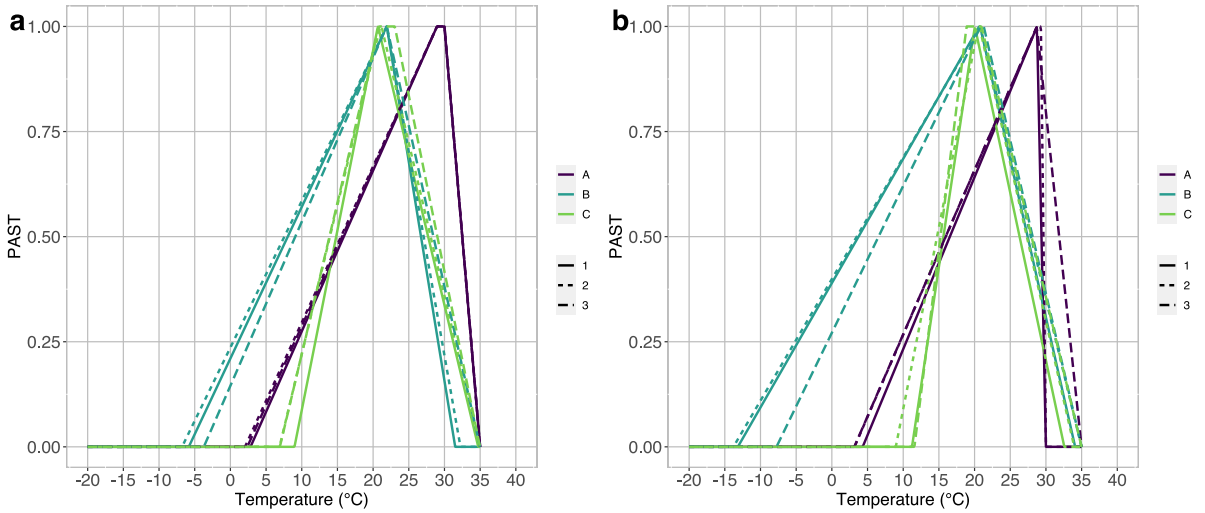

Figure 4.13 – Profiles of the PAST function including the optimized values of the base temperature (T_b), optimal temperatures (T_o^1 and T_o^2) and critical temperature (T_c) obtained with the eGDD method for each of the data set for the Prediction (a) and Synchronization (b) uses.

4.4.4.2.3 The eGDD method resulted in vineyard-consistent and vineyard-specific sets of optimized temperature thresholds

The results of the PCA using the four temperature thresholds as variables and the 9 blocks as individuals are shown in Fig. 4.14. The two first components represented 94.4% of the 9 blocks variability according to their temperature threshold values for the Prediction use and 88.8% for the Synchronization use.
The joint analysis of the PCA principal plane and the correlation circle differentiated blocks characterized by high T_o^1 and T_o^2 (blocks belonging to vineyard A mainly) from blocks characterized by lower T_o^1 and T_o^2 (blocks belonging to vineyard B and C). This was true for both Prediction and Synchronization uses.

The blocks with lower T_o^1 and T_o^2 were separated into two categories. Blocks with high T_b and T_c values (vineyard C) were distinguished from blocks with lower T_b and T_c values (vineyard B mainly with the exception of block 3) in the Prediction use. In contrast, blocks with high T_b and low T_c values (vineyard C) were distinguished from blocks with lower T_b and higher T_b values in the Synchronization case.

Thus, for both Prediction and Synchronization of the eGDD method the blocks of the same vineyard obtained very similar combinations of temperature thresholds and these combinations were different between the 3 vineyards. Only one block of the vineyard A differed from the others in the Synchronization case. Therefore, the set of temperature thresholds were vineyard-specific.

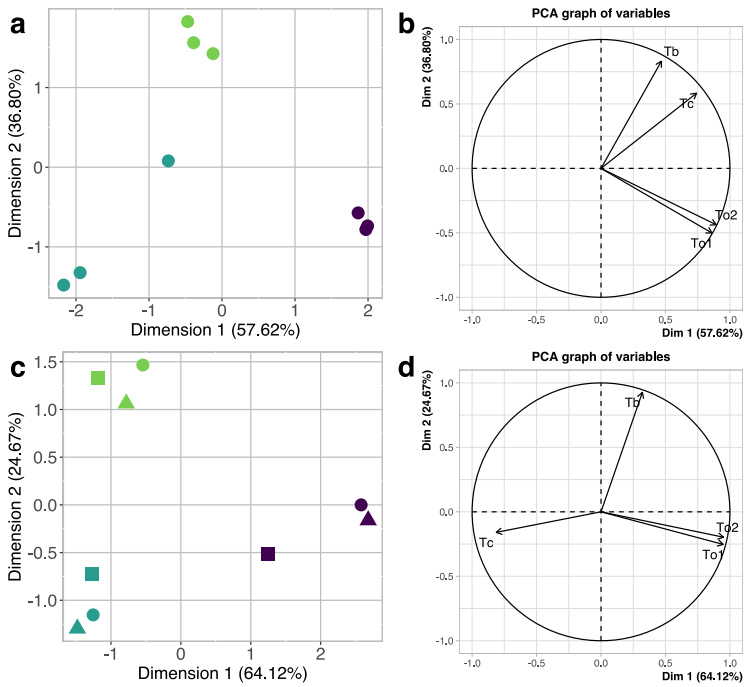

Figure 4.14 – Scatter plots (a and c) and correlation circles (b and d) of components 1 and 2 of the Principal Component Analysis of the results respectively obtained with a Prediction and a Synchronization use of the eGDD method.

4.4.5 Discussion

4.4.5.1 The eGDD method resulted in better and site-specific prediction and synchronization

4.4.5.1.1 The eGDD method resulted in better prediction and synchronization

For each vineyard, the eGDD method was found to generally offer the best results across the 3 blocks of the data set both for a Prediction and a Synchronization use. Therefore, this result justifies the use of the eGDD method over the traditional GDD approach and the Gregorian calendar, whatever the objective of use. In addition, for the 9 blocks of the study, the Prediction RMSE ranged from 1.2 to 7. 4 days for budbreak, from 1.3 to 6.6 days for bloom and from 3.0 to 5.0 days for veraison. These results seemed relevant in an operational context to support the decision-making of vineyard operations. The number of periods that could be discretized over the temperature time series ranged from 6 to 34. 6 periods to divide an entire year might seem a somewhat rough discretization but this was found in the single case of vineyard A and GDD_{opt} barely performed better while the other two approaches didn't allow any discretization. This less accurate synchronization for vineyard A may be partly explained by a greater difference between the blocks of vineyard A (blocks 1 and 2 versus 3) than between the blocks of the other vineyards. The results found for vineyards B and C seemed to be suitable for a discretization of the year that would allow a fine analysis of the timing of environmental influences on grapevine according to their phenology.

4.4.5.1.2 The eGDD method resulted in site-specific prediction and synchronization

The RMSEs obtained with the eGDD method were consistent between the blocks of each vineyard for each phenological stage. The RMSEs of the eGDD method were different between vineyards for budbreak and bloom but not for veraison. For budbreak and bloom, the vineyard order in terms of prediction error was consistent for the three phenological stages : vineyard B performed better than vineyard C which in turn performed better than vineyard A. Part of this difference may be due to the number of years in the data set of each vineyard . However, the same order was not necessarily found for the GDD approaches and vineyard C, which corresponded to 13 years of history, performed better than vineyard A, which had 10 years of history. Regarding the Synchronization use, the same vineyard order was observed: vineyard B performed better than vineyard C which in turn performed better than vineyard A. Again, only part of this difference could be attributed to the difference of history volume since the same order was not necessarily found for the GDD approaches and vineyard C performed better than vineyard A.

Thus, the Prediction and the Synchronization results obtained with the eGDD method were found vineyard-specific. They were also found block-specific in the case of vineyard B and C i.e. the same blocks always performed better within each of the three vineyards.

4.4.5.1.3 Inter-vineyard variations in the eGDD results were likely to include determinants related to grapevine agronomy and to data quality

The prediction errors and the synchronization accuracies obtained with the eGDD method were found site-specific. They might also be explained by the number of years in the studied history of each vineyard. In addition, several hypotheses could explain the poorer results of the eGDD method for vineyard A (especially blocks 1 and 2) than for vineyards B and C. First, a problem with the consistency of the vineyard phenology across years can explain the site-specific character of the eGDD results, perhaps due to inappropriate management practices or to atypical vintage conditions for several years. Second, the quality of temperature data and of phenological observations could also explain part of the problem. Thus, the site-specific character of the eGDD results is also likely to include agronomic determinants related to vine physiology and determinants related to the quality of the data studied.

4.4.5.2 The eGDD method resulted in site-specific thermal indices

4.4.5.2.1 The eGDD method resulted in vineyard-specific and block-specific thermal indices

The eGDD method was applied to each block of the data set in order to calibrate temperature thresholds that were best suited for prediction or synchronization purposes. For both a Prediction and a Synchronization use, the optimized temperature thresholds showed higher inter-vineyard variations than inter-block variation for the same vineyard. In this respect, the thermal indices can be defined as vineyard-specific.
Moreover, the variations between the temperature thresholds of the three blocks of each vineyard also showed significant variations in the sense that they have resulted in important differences in the value of the optimization criterion that allowed their selection (data not shown). Allocating exactly the same thresholds to the three blocks would not have resulted in such low prediction errors and synchronization accuracies for each of the corresponding thermal indices. In this respect, the thermal indices can also be defined as block-specific.

4.4.5.2.2 The site-specific thermal indices integrate the effects of the thermal determinism of grapevine phenology, temperature spatial variations as well as the quality of temperature data and of phenological observations

In this study, the temperature data used for the three blocks of each vineyard was provided by the same weather station. For each vineyard, the three blocks were geographically adjacent so that they may have been visited by the same operator and with the same protocol for phenological observations, at least within each year. Therefore, the quality of temperature and phenological observations were hypothesized as being equivalent for the three blocks of each vineyard respectively. Under this hypothesis, the inter-block variations in the optimized temperature thresholds could highlight variations in the thermal determinism of grapevine phenology as well as temperature variations between blocks that were not captured by the common weather station. For example, the fact that block 1 of vineyard C has a higher optimized T_b in Fig. 4.13a than the two other blocks of the same vineyard may imply that it has higher heat requirements to start developing or that it corresponds to an area that is colder than the area in which the weather station is localized. In an operational context, there is no way to know which of these options is the most plausible.
In contrast, the inter-vineyard variations in the optimized temperature thresholds were of higher magnitude which surely exceeds the spatial temperature differences between adjacent blocks. Thus, inter-vineyard variations were more likely to highlight site-specific differences in the thermal determinism of grapevine phenology, in relation with the different environment characteristics and vineyard management of each vineyard. To limit the contribution of plant material to this so-defined thermal determinism, the same variety (Cabernet-Sauvignon) was considered but variations due to clone and rootstock differences could also be observed.

4.4.5.2.3 Inter-vineyard variations of temperature thresholds can be interpreted as variations in the thermal determinism of grapevine phenology with care for the temperatures commonly recorded on site

The optimized temperature threshold values should not be interpreted as absolute values but rather as filters to select and weight the temperatures recorded in the data set that are relevant for grapevine development. For example, a T_b of -13°C as observed in Fig. 4.13a for vineyard B may mean that these vines actually

start to develop from -13°C or that very few temperatures between -10°C and the true physiological threshold temperature have been recorded. In this second case, the value of T_b influences especially the increasing slope between T_b and T_o^1which conditions the relative weights of the temperatures commonly observed between 8 and 21°C (Fig. 4.12). Given that no temperature below -4°C has been recorded in Vineyard B (Fig. 4.12), the second option is assumed. It seemed also for valid for vineyard B. On the contrary, since vineyard C experienced relatively frequent temperatures between -3 and 10°C, the T_b values of about 8°C and 12°C, for Prediction and Synchronization uses respectively, can be directly interpreted from a physiological angle. This reasoning can also apply for the values of T_o^2 and mostly of T_c.

With these interpretative precautions, the results of the study could hardly be compared to the results of ecophysiological works except for vineyard C (Pouget 1981; Pagay and Collins 2017; Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017; Molitor, Fraga, and Junk 2020). However, they were likely to make an agronomic sense. It can be noticed in Fig. 4.12 that vineyard A, which showed on average the warmest temperatures, also presented the highest T_o^1 and T_o^2 around 29-30°C. Vineyard B, which also experienced similar temperatures in the fourth quartile but presented lower temperatures on average, presented lower T_o^1 and T_o^2 in line with Vineyard C. Thus, vineyard A seems to be the most adapted to endure high average daily temperatures, but it undergoes a slower phenological advancement than vineyards B and C for the most frequently observed temperatures, between 10 and 22°C. Furthermore, the increasing slope bounded by T_b and T_o^1 is steeper for vineyard C than A and B. Thus, for this temperature interval, each additional unit of temperature results in a higher development gain for vineyard C than for the other two.Thus, vineyard C would be adapted to short growing seasons with relatively low temperatures compared to the vineyards A and B. Vineyard B would be adapted to relatively longer growing seasons but with similar temperatures. Vineyard A would be adapted to longer and relatively warmer growing seasons. These differences should be interpreted as resulting from the interaction of grapevine internal factors, environmental influences and vineyard management.

4.4.6 Conclusion

The eGDD method was proposed to compute site-specific thermal indices for the purposes of predicting the achievement dates of key phenological stages ro to synchronize time series data according to a timeline consistent with grapevine phenology. The development of the eGDD method was presented in the first part of this paper. In the present second part, the eGDD method was shown to allow for better predictive or synchronizing results for three case commercial vineyards corresponding to different operational contexts. The thermal indices obtained with the eGDD method were consistent between blocks of the same vineyard but different between vineyards. Therefore, the vineyard scale seemed to allow the greatest value when implementing the eGDD method. However, this statement should be modulated when different varieties are studied. Finally, the three case studies provided further insight on the way eGDD results should be interpreted.

4.5 General conclusion of the chapter

The research work presented in this chapter has generated several scientific contributions to the related literature. First, it formalized the fact that the same thermal indices were used for different purposes, either Prediction or Synchronization, in the literature. Second, it presented a generalized formalization of the construction of a cumulative thermal index with the introduction of a Phenological Advancement Speed as a function of Temperature (PAST function). Third, it participated in proving the necessity of site-specifically calibrating thermal indices in order to account for the interactions of numerous site-specific factors that influence the way grapevine development should be monitored with temperature data, especially in an operational context such as in commercial vineyards.

These contributions were integrated into an operable method called Extended Growing Degree Days (eGGDD), which allows the computation of site-specific thermal indices in accordance with a Prediction or Synchronization use from operational data. These thermal indices are expressed in Thermally Optimal Daylight Hours (TODH) and can rigorously be compared from an agronomical (not numerical) point of view between years and sites. The eGDD method was evaluated for 3 different operational contexts, corresponding to three vineyards with very different pedo-climatic and latitude conditions. The eGDD showed better results than the classical Growing Degree Days approach and the Gregorian Calendar as well as site-specific, either

at the vineyard or at the block scales.

Finally, by interpreting the results of the eGDD method, this research work also contributed to the interpretation of any thermal index, especially when computed on the basis of operational data. Such a thermal index is composed by an intimate interaction between i) grapevine physiological reality in terms of thermal determinism, ii) spatial temperature variations that cannot be exhaustively captured by physical or even virtual stations and iii) data quality. This is not a problem because a thermal index is systematically composed as such in practice. However, this must be considered in order to properly interpret thermal indices and eventually work on limiting the contribution of a particular source of variation in the data sets.

Chapter 5

Handling the temporality of yield development through the analysis of time series of weather data

5.1 Intention note

Time series of weather data are commonly found in farm data sets and a fortiori in the data sets of commercial vineyards. These weather variables are of primary importance for grape yield development as they influence grapevine system functioning including yield development, directly or indirectly by interacting with other variables (e.g. plant water status), without being retrofluenced by yield development. Moreover, time series of weather data are available at a daily time step (or a more precise time step) for most commercial vineyards. For these two reasons, weather data analysis was chosen to address the temporality of grape yield development and in particular to study the trajectory effects it incorporates. **Thus, the objective of this chapter was to develop an analytical framework which would allow the leveraging of time series data with the perspective of using the resulting weather indicators as covariates of a yield forecasting model.** It corresponds to the second analytical step of the grape yield forecasting method, as shown in Fig. 5.1

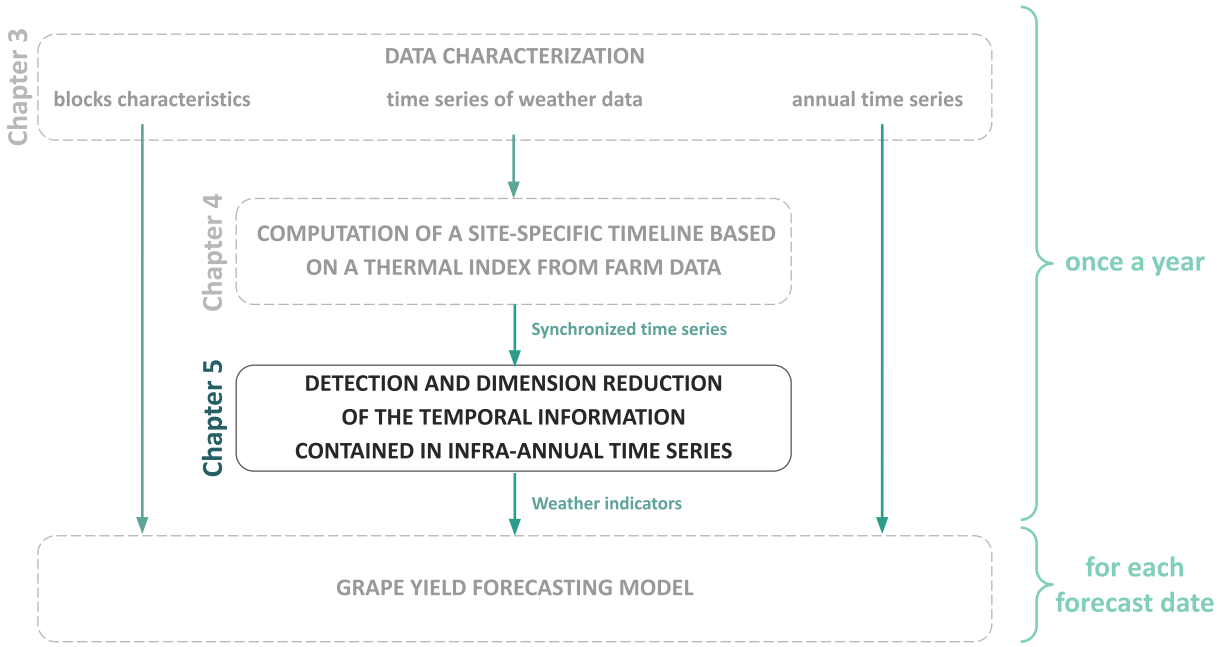

Figure 5.1 – Positioning of chapter 5 in the framework of the yield forecasting method

Time series are defined as a set of observations sequentially organized in time as a realization of a stochastic process (Brockwell and Davis 2009). Therefore, the observations can not be considered as independent covariates and analyzed through classical analyses such as multivariate linear regressions. To circumvent this issue, most literature studies have focused on using weather variables at a few known key phenological stages (Buttrose 1974; Pouget 1981) or time steps (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017), which can be considered as independent. However, these classical approaches have limitations: i) they depend on choices of climate variables and timing, ii) it is often necessary to suppress data or to analyse only parts of a time series and iii) times series autocorrelation (observations correlated over time) is not considered. Therefore, information about the weather influence on yield development may be potentially missed. **A first working hypothesis was hence that time series of weather data could reveal information of high potential for yield modelling when appropriately processed and analysed.** For this purpose, time series of weather data could be treated as functional covariates of a yield model. However, their integration as such in a complex multivariate model, involving additional covariates and training on operational data volumes, would risk disrupting the model training and results due to estimation issues. Therefore, the first scientific question of this research axis was how to detect and reduce the dimension of the information relevant to yield development contained in time series of weather data ? In addition, the weather influence on yield development is assumed to include both general and site-specific influences resulting from unique interactions between numerous site-specific factors in each vineyard. **Therefore, the second working hypothesis was that a site-specific analysis of the time series data could detect weather covariates that are locally more relevant to model yield development.** Hence, a second scientific question for this research axis was how to site-specifically adapt the analysis of time series of weather data ? **Finally, a third working hypothesis was that operational weather data could support such site-specific analyses of time series and still present valuable results in terms of detection of operational weather indicators for yield development modelling.** Therefore, the third scientific question of this research axis was how to account for the characteristics of farm data into the site-specific analysis of times series of weather data ? A Bayesian functional Linear regression with Sparse Step functions (Grollemund et al. 2019) aims at providing a Bayesian estimate of the support of the coefficient function of a functional linear regression on a scalar response. In other words, it aims at detecting the periods when a functional covariate, such as a weather time series taken as a function of time, influences a scalar response such as grape yield. Therefore, such a method could be used to analyze the time series of weather data appropriately with regards to their characteristics and extract relevant weather indicators (weather variables on given periods) for yield elaboration with reduced dimensions.

The first part of this chapter was dedicated to the evaluation of the BLiSS method. It aimed at testing the relevance of the BLiSS method in highlighting already known results and identifying new critical time periods in terms of weather influence on grape yield elaboration. Therefore, it was performed with the mean number of bunches/bunches per stock as a response variable to benefit from a higher number of references mentioning precisely timed periods of temperature influence than if grape yield was used as a response. **The second part of the chapter was dedicated to the use of the Bliss method in the context of this PhD thesis: in a site-specific way and from farm data. Therefore, it mobilizes the eGDD method presented in the previous chapter to preliminary synchronize the time series data.**

5.2 Evaluation of a functional Bayesian method to analyse time series data in precision viticulture

5.2.1 Details about the paper

5.2.1.1 Title and publication informations

This section has been presented as an oral communcation at the 12th European Conference on Precision Agriculture (2019, Montpellier, France). It is entitled "Evaluation of a functional Bayesian method to analyse time series data in precision viticulture".

5.2.1.2 Authors

C. Laurent[1,2,3,*], M. Baragatti[4], J. Taylor[3], T. Scholasch[1], A. Metay[2], B. Tisseyre[3]

[1] Fruition Sciences, 34000 Montpellier, France
[2] ABSYS, Univ. Montpellier, CIRAD, INRAE, Institut Agro, 34060 Montpellier, France
[3] ITAP, Univ Montpellier, INRAE, Institut Agro, 34060 Montpellier, France
[4] MISTEA, Univ Montpellier, Institut Agro, INRAE, 2 Place Pierre Viala, 34060 Montpellier, France
[*] corresponding author : cecile@fruitionsciences.com

5.2.1.3 Abstract

In precision agriculture, most studies focus on spatial crop variability whereas temporal variability and its role in decision-making is equally important. The classical methods for temporal analysis have limitations, potentially resulting in information loss. A novel method based on a Bayesian functional Linear regression with Sparse Steps functions (BLiSS method) is evaluated in this paper to investigate continuous influence analysis when working with time series data. The example of the influence of temperature on the number of bunches per vine during the year before harvest was considered as an example application. The evaluation of the BLiSS results was done by comparing identified critical time periods with traditional viticulture knowledge in the literature. It showed the relevance of the BLiSS method, highlighting already known results and identifying new critical time periods for yield elaboration.

5.2.1.4 Keywords

BLiSS method, Temporal analysis, Temperature, Yield, Number of bunches

5.2.2 Introduction

In precision agriculture, most studies have focused on spatial crop variability (Oliver, Bishop, and Marchant 2013) whereas temporal variability and its role in decision-making is equally important. Indeed, most indicators of production yield and quality are time-dependant. Understanding their evolution during the crop growth cycle and the factors that drive them is often an issue for better production management. This implies understanding the pattern of influencing factors over a whole production period, as well as the correlations between influences.

Historically, 'temporal' studies have focused on using climate variables at a few known key phenological

stages or time steps (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017). This has been done to simplify the analysis and permit the use of classical methods, such as least squares linear regression. However, these classical approaches have limitations: i) they depend on choices of climate variables and timing, ii) it is often necessary to suppress data or to analyse only parts of a time series, and iii) times series temporality (observations correlated over time) is not considered. Therefore, information about temporal crop physiological regulation is potentially missed.

To advance temporal analysis in the crop production domain, this paper evaluates the potentialities of a new approach to study quantitative time series data. It uses a Bayesian functional Linear regression with Sparse Steps functions (BLiSS method, Grollemund et al. 2019). The principle of this method is to analyze the complete history of an explanatory variable to identify the periods during which it has an impact on the explained variable. It takes into account the data correlation over time i.e. it allows the analysis of a period impact according to the impact of others periods. By overcoming the classical approaches limitations, the BLiSS method could result in the discovery of new periods of influence of an explanatory variable on an explained variable in many agricultural sectors.

The aim of this paper is to test the potentialities of the BLiSS method on a study case that is well understood in the literature. To this end, the impact of temperature history in the year before harvest on the number of bunches per vine has been chosen as a study case. Indeed, grape yield elaboration is an interesting case study in temporal terms as grapevine is a perennial crop, meaning its yield determining process covers two growth cycles (Howell 2001; Vasconcelos et al. 2009). These two years are noted n-1 and n, the year n being the year of harvest (Fig. 5.2). Thus, at least two growth cycles overlap and there is an interdependency between the different physiological processes leading to memory effects in vine physiology (Ravaz and Sicard 1911; Keller, Tarara, and Mills 2010).

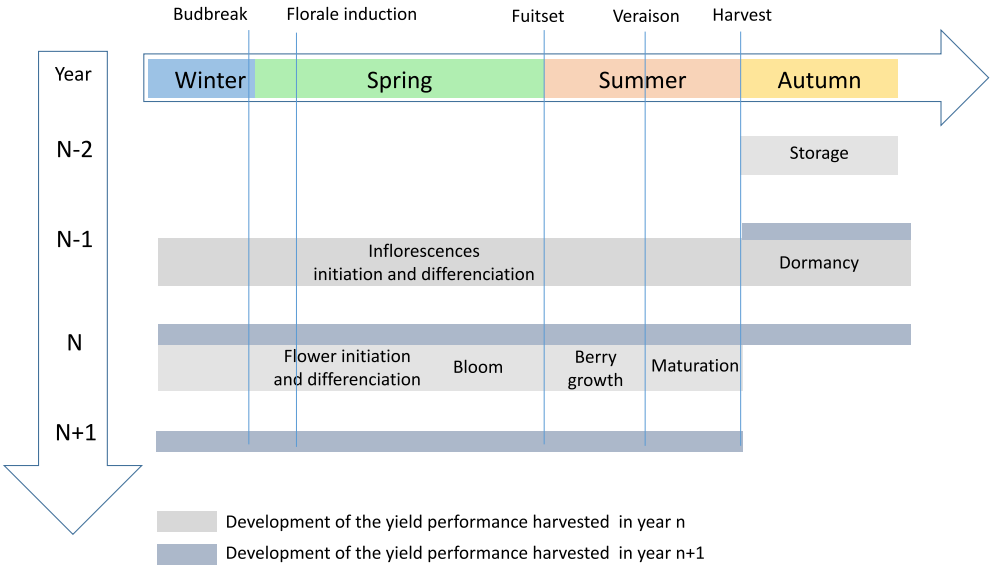

Figure 5.2 – Key phenological stages and the temporal pattern of the grapevine yield determining process over two seasons (aligned for the northern hemisphere

The number of bunches per vine (BN) is a major yield component, said to explain up to 60% of the final yield variability (Clingeleffer et al. 2001). However, understanding the timing of its elaboration and its sensitivity to external factors such as temperature is complicated by its interdependency with other physiological processes over time. Therefore, there is a challenge to determine all the periods during which any explanatory variable, such as temperature, has an impact on BN taking into account that the effects of these periods are correlated to many others periods' multivariable impacts. To do this, classical approaches have been carried out using key phenological stages (Buttrose 1974; Pouget 1981) or time steps (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017). They have shown that temperature is one of the major factors driving grape fruitfulness (Buttrose 1974; Srinivasan and Mullins 1981; Petrie and Clingeleffer 2005). In the literature, periods of phenological development during year n-1 for one cycle are referred to by their concomitance with the phenological stages of year n. For example, the development of inflorescence primordia linked to the yield in year n starts with budbreak in year n-1 (Vasconcelos et al. 2009). Following this notation,

a few hours of elevated temperature around budbreak of year n-1 should be sufficient to positively impact inflorescences or bunches number per vine in year n (Buttrose 1974; Pouget 1981; Srinivasan and Mullins 1981; Lobell et al. 2006) . Critical stages have also been highlighted around bloom (Durquety, Naude, and Blanchard 1982; Molitor and Keller 2017) of year n-1. Molitor and Keller 2017 found that high maximum temperatures around veraison and low average temperatures during maturation of year n-1 favoured high yield. As a study case, this paper aims to assess whether using the BLiSS method to analyse the impact of the whole history of temperature in year n-1 on BN permits the identification of critical periods previously described and well understood within literature. It also tests the potential detection of previously unknown critical periods.

5.2.3 Material and Method

5.2.3.1 Data description

5.2.3.1.1 Bunch number per vine

Data were collected from a commercial Cabernet-Sauvignon (Vitis vinifera cv. Cabernet Sauvignon) vineyard field (1 ha) in the Bordeaux region, France. The vines are short-pruned, trellised, non-irrigated, partially grass covered in the inter-row and planted at a density of 8696 vines/ha. The BN was determined after fruitset on the same 100 vines in the vineyard from 2007 to 2018, with the exception of 2014. Note that depending on the seasonal conditions, bunch thinning is often performed in this vineyard, such that the final harvest differs from the amount of fruit set. The occurrence date of the principal phenological stages (budbreak, bloom, fruitset, veraison) was recorded by the vineyard manager according to the Gregorian calendar from 2006 to 2018.

5.2.3.1.2 Temperature data

Temperature data were collected by a local weather station (provided by DE.MET.E.R, Villenave d'Ornon, France) located in the wine estate from 2006 to 2018. Multiple measurements were observed, however, only daily minimum (Tmin) and average (Tavg) temperature (°C) are considered in this analysis as their impact on grape yield is well documented in the literature. The Tmin and Tavg influences on BN have only been studied during the growing season of year n-1 (from 1st of March to 25th of October). The potential impact of winter or the growing season of year n was not investigated in this preliminary study, but it could be included following the same steps. To reduce processing time, temperature was aggregated into 10-day periods. Therefore, the Tmin corresponded to the minimum daily temperature during each of these periods. The Tavg corresponded to the average of daily temperatures over the 10-day periods. It is expected that better software and increased computing power will remove the need for this pre-processing step in the near future.

5.2.3.2 The BLiSS method

5.2.3.2.1 Theory

The BLiSS method (Grollemund et al. 2019) allows quantitative data (variable Y) to be explained by functional data (variable X) using a linear functional regression model.

$$Y = \alpha + \int_{\tau} \beta(t)Xt)\mathrm{d}t + \epsilon \tag{5.1}$$

Where Y is the scalar response variable, X the explanatory variable, $\beta(t)$ the parameter function to be estimated and a residual error ϵ supposed to follow a normal distribution $\mathcal{N}(0, \sigma^2)$. In this paper, Y corresponds to BN and X corresponds to Tmin or Tavg. BLiSS is a Bayesian method: it supposes that the parameter $\beta(t)$ is random, formulates an a priori estimation of the $\beta(t)$ distribution and updates it with the data to produce an a posteriori estimation of the $\beta(t)$ distribution. The BLiSS method investigates the estimated distribution support, i.e. all the t instants for which $\beta(t)$ differs from 0. Thus, the BLiSS method delivers two estimators: one of the $\beta(t)$ function and another of the $\beta(t)$ function support (here time). The latter allows periods during which X has an impact on Y to be identified.

The output of BLiSS is best represented in a graphical form (e.g. Fig. 5.3). By way of explanation, Fig. 5.3 features the a posteriori distribution of the BLiSS estimator of $\beta(t)$ as a line estimated over time. If the $\beta(t)$ estimator is null (= 0) during a certain period, it means that the variable X is not related to Y. If it has a positive value during a given period of time, X is positively correlated to Y. In contrast, if it has a negative value during a period of time, X is negatively correlated to Y. The duration of the detected periods corresponds to their time length (distance on the x-axis). The magnitude of the detected periods corresponds to their estimated $\beta(t)$ value (distance on the y-axis). The higher the magnitude, the greater the impact X has on Y. The BLiSS method also allows the correlation between X and Y to be estimated. Practically, it corresponds to the influencing period detection confidence. The color code is a visual way to represent it. The darker the color is (red on the extreme case), the more confident the correlation between X and Y during this period. In contrast, the lighter the color is (white on the extreme case), the less confident the correlation.

5.2.3.2.2 Implementation

Each individual corresponds to a BN per year. Therefore, the data set was composed of 5368 individuals to be analyzed. The analysis was performed using the package bliss version 1.0.0 (Grollemund et al. 2019) in R 3.5.1. The Tmin and Tavg time series were analyzed independently.

5.2.3.2.3 Results interpretation

Assuming that the ecophysiological results reported in the literature and reviewed in the introduction represent the diversity of the classical method applications, the BLiSS method results were evaluated by comparison with already identified critical time periods. The literature review focused on the temperature influence on the number of bunches per vine and on production yield during the year n-1.

5.2.4 Results

5.2.4.1 The Bliss method allows to detect periods of temperature influence on the bunch development

Fig. 5.3a shows the influence of daily minimum temperature during the year n-1 on the number of bunches per vine (BN). Four periods of influence were detected: (i) from mid-March to late April, (ii) from mid-July to early August, (iii) from early August to late September and (iv) from late September to mid-October. The detection of each of these periods had a high degree of confidence (dark colour). Three out of these four periods (i,ii and iv) were positively correlated to the BN i.e. elevated Tmin during these periods favored a high BN. In contrast, high Tmin during the (iii) period favored a lower BN. The duration of periods (i) and (iii) exceeded one month whereas the period (ii) and (iv) lasted between two and three weeks. The absolute magnitude for each period gave an indication of the intensity of the temperature impact on BN. It was 6 for periods (ii) and (iv) and 1.5 and 3 for periods (i) and (iii) respectively .

Fig. 5.3b shows the influence of daily average temperature on BN during year n-1. Three periods of influence were detected: (v) from late April to early June, (vi) from late June to late July and (vii) from late July to early September. The detection of each period had a lesser degree of confidence (lighter yellow) than periods (i) and (iii) (Fig. 5.3a). Two periods, (v) and (vi), were positively correlated to BN, meaning high Tavg during these periods favored a higher BN. In contrast, high Tavg during period (vii) reduced BN. The duration of these three periods were between one month and a month and a half. The absolute magnitude of periods (v), (vi) and (vi) were 9, 2 and 11 respectively.

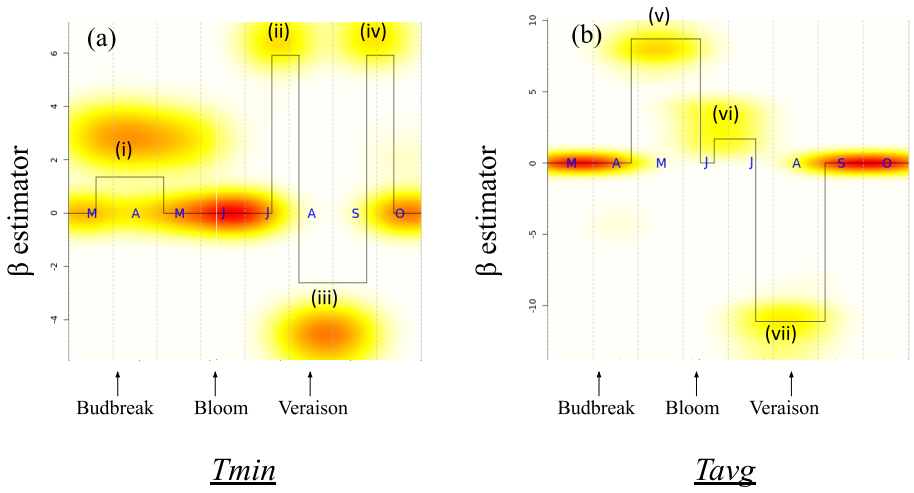

Figure 5.3 – Analysis of possible periods of influence respectively of (A) daily minimal temperature (Tmin) and (B) average temperature (Tavg) on the number of bunches per vine from 1st of March to 25th of October of year n-1. The black line indicates the β estimator distribution. The colour gradient from white to red illustrates the density probability function of the β estimator distribution. The x axis labels M,A,M,J,J,A,S,O represent March, April, May, June, July, August, September, October respectively. Average dates of budbreak, bloom and veraison recorded on the commercial field from 2006 to 2018 vintages are indicated.

5.2.4.2 The BLiSS results are in agreement with literature

Period (i) (Fig. 5.3a) was associated with the budbreak period in the Bordeaux region. It is in agreement with the results of Pouget 1981 on the number of inflorescences per vine under experimental conditions and the findings of Lobell et al. 2006 when studying night temperature around budbreak in California. Period (v) corresponded to the bloom period in the Bordeaux region. This agrees with the results of Durquety, Naude, and Blanchard 1982 on BN for Petit Manseng and the study of Molitor and Keller 2017 on the impact of minimal temperature on production yield for Müller-Thurgau and Riesling in Alsace (North-East of France). In the latter case, the correlation between minimum temperature around bloom and yield was tested and found positive, as observed in period (v) (Fig. 5.3b). Period (ii) coincided with veraison in the Bordeaux region. Only Molitor and Keller 2017 reported this period studying daily Tmax effects on yield for Müller-Thurgau. The (iii) and (vii) periods are associated with the maturation period in the Bordeaux region. Molitor and Keller 2017 also reported a negative impact of minimum temperature on future yield potential during this period.

5.2.4.3 The BLiSS results allow further exploration of the data

Periods (iv) and (vi) have not been previously cited in the literature but they had a good degree of certainty according to the BLiSS method. They respectively correspond to a post-harvest and a fruitset period. Since the method correctly identified known periods of influence (previous section), the hypothesis is that these periods are also influential on BN. This shows the potential ability of BLiSS to highlight unknown periods of influence of temperature on BN. Unlike the classical method results, for which the duration of the resulting periods is determined by the pre-analysis choices, the periods detected by the BLiSS method were characterized by different durations. It is also interesting to note that the temperature influence during summer was characterized by several periods in Fig. 5.3 whereas Molitor and Keller 2017 only reported one period of long duration. The BLiSS method had also refined the duration of known key influencing

periods. The proposed alterations in duration and the identification of new periods of influence now need to be properly validated in a viticulture context. If true, these results have increased the knowledge and the temporal resolution of environmental effects on yield and consequent potential management.

5.2.5 Discussion

5.2.5.1 The BLiSS method : a novel and relevant method to analyse time series data

The BLiSS method allowed the detection of already confirmed periods of daily minimal and average temperature on BN and potential vineyard yield. This point validated the use of this method to analyze time series of agricultural data. Moreover, the method allowed the detection of previously unknown periods of influence or the refinement of the duration of known key periods. This demonstrated the advantages of a linear functional model in a Bayesian framework for the temporal analysis of agricultural data.

5.2.5.2 Comment on the interpretation of the BLiSS results

Each period that was detected with the BLiSS method was characterized by a magnitude, a duration and a degree of confidence. These three criteria can be used to compare the different periods. However, classifying the periods by decreasing order of influence required an integration of the information linked to both the degree of confidence and the magnitude. This may not be simple in every case. For example, a high magnitude but medium degree of confidence period cannot be prioritized in comparison with a high degree of confidence but medium magnitude period. Thus, the interpretation of the BLiSS results should be decomposed into the three criteria first and an expert knowledge of the field of study should then balance the interpretation. Clearer interpretation of the BLiSS outputs is an area for further development.

5.2.5.3 Potentialities for various study field analysis using the BLiSS

The BLiSS method allows further exploration of many types of data sets. Any variable characterized by time series data, which is a common case in agriculture, can be analyzed with the BLiSS method. The method could also be applied to other variables evolving over different continuous supports such as wavelength or spatial coordinates.

While two univariate cases were presented here, several explanatory variables could be studied in a multivariable analysis to identify their potential combined effects as well as their individual (univariate) effects. For example, Tmin and Tavg have simultaneously been considered in a bivariate BLiSS analysis to distinguish their proper effects on BN (data not shown). However, there are some limitations to the current BLiSS method. Firstly, the computation time has to be taken into account when designing the analysis. For example, the results presented in Fig. 5.3 (derived from 5368 individuals and using 27 10-day climatic periods) took approximately 4 hours on a machine of 4 CPU and 5 Gb memory. Secondly, it is not possible to know the portion of the Y variability explained by X at the moment. It is also not yet possible to include qualitative explanatory variables but this is planned to be implemented soon. In addition, it is important to remember that the BLiSS method is subject to the same limitations that any other linear analysis, such as disturbance due to the explanatory variables correlation.

Finally, the BLiSS method presents another option which has not been tested in this paper. Expert knowledge can be taken into account in the prior formulation. In practice, this can help the method to detect relevant periods using a priori literature results or field experience, in the case of a small number of data for example.

5.2.6 Conclusion

The BLiSS method combines the advantages of both functional and Bayesian models to perform advanced temporal variability analysis. Functional analysis allowed a more complete and objective analysis of a data set, taking into account the explanatory variables histories and allowing detection of periods of influence in the time series. Bayesian analysis allowed better uncertainty management using conditional probability. This approach can be used for further data exploration in many different fields, in agriculture or others. It is an interesting method to better study the effect of high temporal resolution variables, such as those generated in precision agriculture. The ability to correctly identify and manage key phenological stages in agriculture is an area that has often been overlooked in precision agriculture but is needed to improve the temporal-resolution of decision-making on farm.

5.2.7 Acknowledgements

The authors fully acknowledge the commercial estate that provided the data as well as the authors of the BLiSS method who spent time explaining its application.

5.3 Identifying operational weather indicators of grapevine yield development by exploring time series of farm data synchronized according to an eGDD thermal index with a linear functional Bayesian method

5.3.1 Details about the paper

5.3.1.1 Title and publication informations

This section is entitled "Identifying operational weather indicators of grapevine yield development by exploring time series of farm data synchronized with a eGDD thermal index and using a linear functional Bayesian method". It is intended to be it is intended to be submitted to a thematic journal in viticulture or to a journal like Agriculture and Forest Meteorology.

5.3.1.2 Authors

C. Laurent[1,2,3,*], G. Le Moguédec[4], J. Taylor[3], T. Scholasch[1], B. Tisseyre[3],A. Metay[2]

[1] Fruition Sciences, 34000 Montpellier, France
[2] ABSYS, Univ. Montpellier, CIRAD, INRAE, Institut Agro, 34060 Montpellier, France
[3] ITAP, Univ Montpellier, INRAE, Institut Agro, 34060 Montpellier, France
[4]AMAP, Université Montpellier, INRAE, Cirad CNRS, IRD, 34000 Montpellier, France
[*] corresponding author : cecile@fruitionsciences.com

5.3.1.3 Abstract

Climate influence on grapevine yield development is prevalent and this influence is only expected to increase with climate change. Moreover, weather time series data are commonly encountered in the data sets of commercial vineyards, even when data from other influencing variables might be missing. Therefore, time series of weather data should be leveraged as much as possible to allow for yield forecasting in an operational context. However, the climate influence on yield development includes both general and site-specific influences resulting from unique interactions between numerous site-specific factors in each vineyard.

In such a context, this paper investigates the opportunity to site-specifically detect weather influences on grapevine yield development in the two seasons leading up to a harvest (years n-1 and n). It is based on the study of three commercial vineyards respectively situated in the Bordeaux region (France), California (USA) and Israel. For each vineyard, times series of temperature and rain data were synchronized according to a site-specific thermal index, which was computed using a new method called Extended Growing Degree Days (eGDD). Synchronized time series were then analyzed with the Bayesian functional Linear regression with Sparse Steps functions (BLiSS) method in order to respectively detect periods of influence on grapevine yield development.

Common periods of weather influence on yield development corresponding to floral initiation or differentiation were found for all vineyards but with differences in their timing, duration and correlation direction with yield. Other periods, particularly in year n-1, were found whereas they were not referred to in literature. These results confirm the interest to optimize the use of local time series climate data to detect potential new climate indicators of yield development that may be site-specific or global in nature. Such indicators may subsequently be used for further analysis aimed at forecasting grapevine yield with the assumption that the site-specific indicators will enhance the local yield forecasting.

5.3.1.4 Keywords

extended Growing Degree Days (eGDD), Bayesian functional Linear regression with Sparse Steps functions (BLiSS), Commercial Vineyards, Operational data, Viticulture, Phenology, Spatio-temporal analysis

5.3.2 Introduction

Climate influence on grapevine yield development is prevalent and this influence is only expected to increase with climate change. In particular, temperature plays an important role in defining the yield potential. Rain, through water availability, is one of the main yield limiting factors (Van Ittersum et al. 2013). However, rain influence is decreased when the vineyard is irrigated. In addition, both temperature and rain may have a reducing influence on yield development during extreme events. Temperature and water availability are known to be particularly influential during periods of increased sensitivity in yield development(Ojeda, Deloire, and Carbonneau 2001; Petrie and Clingeleffer 2005; Keller, Tarara, and Mills 2010; Guilpart, Metay, and Gary 2014; Pagay and Collins 2017; Triolo et al. 2019). These periods of sensitivity are related to the successive implementation of yield components (Laurent, Oger, et al. 2021). Moreover, weather time series data are commonly encountered in the data sets of commercial vineyards, even when data from other influencing variables might be missing. Therefore, time series of weather data should be leveraged as much as possible to allow for yield forecasting in an operational context. However, the characteristics of the time series data call for specific methods of analysis. Indeed, time series are defined as sets of observations sequentially organized in time as a realization of a stochastic process (Brockwell and Davis 2009). Therefore, the observations can not be considered as independent covariates to be analyzed with classical methods such as multivariate linear regressions. To circumvent this issue, most literature studies have focused on using weather variables at a few known key phenological stages (Buttrose 1974; Pouget 1981) or time steps (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017) which can be considered as independent. However, these classical approaches have limitations: i) they depend on the subjective choice of a few periods (timing and duration) to be analyzed, ii) it is often necessary to analyze only parts of a time series when they contain a lot of available observations and iii) times series autocorrelation (observations correlated over time) is often not considered. Therefore, information about the weather influence on yield development is potentially missed. To address this issue, a Bayesian functional Linear regression with Sparse Step functions (BLiSS, Grollemund et al. 2019) has been shown to be a relevant method to explore time series of weather data taken as functional covariates in order to identify key periods when the weather influences yield elaboration (Laurent, Baragatti, et al. 2019). In addition, the interest to synchronize the time series according to a site-specific index has been shown to study environmental influence on grapevine performances consistently with and according to grapevine phenology (Laurent, Scholasch, et al. 2020). To address this issue, a method called Extended Growing Degree Days (eGDD) has been proposed to compute site-specific thermal indices to be used as timelines to synchronize time series according to budbreak, bloom and veraison. Therefore, it is hypothesized that the BLiSS analysis of time series data synchronized with the eGDD method will allow a good exploration of climate influence on grape yield development through the analysis of time series of weather data.

However, time series of farm data present some limitations : they are characterized by heterogeneous measurement quality, the sampling design is not optimal since it is often designed for other purposes than the current analysis, data sets present overlapping and missing data etc. Therefore, it is assumed that the volume of analyzed data and the use of proper statistics can compensate for these limitations and still lead to the detection of relevant results, here in terms of weather influence on yield development.

In such a context, this paper aims at validating the ability of the analytical framework including the eGDD and the BLiSS methods to improve the analysis of time series of weather data by extracting and reducing the dimension of the information relevant to the considered response variable, here yield development. It also investigates the ability of such an analytical framework to be adapted to different farm data sets. Therefore, the case studies of three commercial vineyards respectively situated in the Bordeaux region (France), California (USA) and Israel are considered.

5.3.3 Material and Method

5.3.3.1 Data description

Data was collected from three commercial vineyards situated in the Napa Valley (California, USA), Israel and the Bordeaux region (France). They are respectively noted vineyard A, B and C in this paper. Vineyards A and B were composed of different ranches i.e. different groups of blocks spaced a few kilometers apart. Both vineyards were irrigated. Vineyard C was composed of only one block and was rain-fed (Table 5.1).
For each vineyard, the achievement dates of budbreak, bloom and veraison were recorded according to the

Gregorian calendar at 50% of achievement.

Vineyard A was divided into 4 ranches. Each ranch was equipped with a weather station and respectively counted 3, 20, 5 and 5 blocks. Yield and phenological observations were recorded from 2008 to 2018 for each block. Temperature data was recorded at a daily time step respectively from 2008 to 2018, from 2007 to 2018, from 2012 to 2018 and from 2010 to 2018 for each weather station. The years when phenological and yield observations were made could be different from one block to another. Therefore, vineyard A data set is considered as unbalanced in terms of years and blocks.

Vineyard B was divided into 3 ranches with a weather station for the whole vineyard. Each ranch respectively counted 58, 32, 42 blocks with yield observations among which 6, 17, 15 blocks had phenological observations. Yield and phenological observations were recorded from 2000 to 2019. Temperature was recorded at a daily time step in 1999-2012 and 2014-2019. The years when phenological and yield observations were made could be different from one block to another. Therefore, vineyard B data set is considered as unbalanced in terms of years and blocks.

Vineyard C counted 79 blocks in a single ranch. All blocks had phenological and yield observations for the years 2002-11 to 2014-15. Weather data was recorded in 2001-11 to 2014-15. The blocks presented phenological and yield observations for the same number of years and the same years.

The main characteristics of the data sets of the three vineyards are summarized in Table 5.1.

Table 5.1 – Characteristics of the three vineyards A, B and C.

	Vineyard A	Vineyard B	Vineyard C
Localization	California USA	Israel	Bordeaux, France
Latitude	38	32	45
Type of climate	semi-arid	semi-arid	oceanic
Irrigation	yes	yes	no
Varieties	Cabernet-Sauvignon Merlot Petit Verdot	Cabernet-Sauvignon Merlot Syrah	Cabernet-Sauvignon Merlot Petit Verdot
Number of ranches	4	3	1
Number of weather stations	4	1	1
Years of weather data	from 2008 to 2018 from 2007 to 2018 from 2012 to 2018 from 2010 to 2018	from 2008 to 2019	from 2001 to 2011 and from 2014 to 2015
Number of blocks with phenological observations per ranch	3,20,5,5 (33 in total)	6,17,15 (38 in total)	79
Mean number of years for phenological observations per block	7.5	4.5	13
Number of blocks with phenological observations per ranch	3,20,3,5 (36 in total)	58,32,42 (132 in total)	79
Mean number of years for yiled observations per block	5.6	5.2	13

5.3.3.2 Theory

5.3.3.2.1 Theory about the Extended Growing Degree days (eGDD) method

The eGDD method, developed in chapter 4, computes site-specific thermal indices by integrating Phenological Advancement Speed as a function of Temperature (PAST function). This PAST function represents the operational relationship that links farm temperature data to the vine response in terms of phenology. It includes four temperature thresholds that respectively represent the base temperature from which the vine starts developing (T_b), two optimal temperatures between which the vine develops at its highest speed (T_o^1 and T_o^2) and a critical temperature T_c from which the vine stops developing (Fig. 5.4).

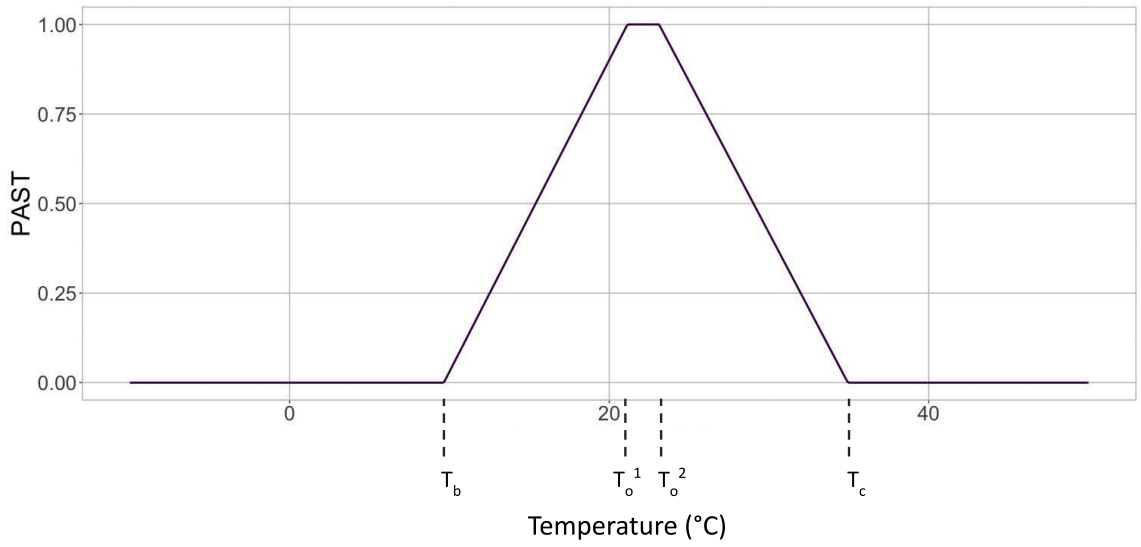

Figure 5.4 – Example of PAST function obtained with the eGDD method. T_b corresponds to the base temperature, T_o^1 and T_o^2 to the bounds of the interval of optimal temperatures, and T_c to the critical temperature.

These temperature thresholds are site-specifically optimized using a constrained optimization approach. The optimization criterion is designed to serve the purpose of Prediction of the achievement date of phenological stages or of Synchronization of time series of data based on the vine phenology. In the second case, the criterion to be minimized relates to the respective variance of the dates of budbreak, bloom and veraison of all the years for a site when they are expressed in a thermal index and normalized according to the mean length of the time series (Eq. 5.1)

$$S = a\sum_{i=1}^{n}\frac{(\frac{s_i^{bud}-\bar{s}_i^{bud}}{s_{max}})^2}{n} + b\sum_{i=1}^{n}\frac{(\frac{s_i^{blo}-\bar{s}_i^{blo}}{s_{max}})^2}{n} + c\sum_{i=1}^{n}\frac{(\frac{s_i^{ver}-\bar{s}_i^{ver}}{s_{max}})^2}{n} \tag{5.2}$$

with n the number of considered years for a given site,
$s_i^{bud,blo,ver}$ and $\bar{s}_i^{bud,blo,ver}$ the observed and predicted scores,
s_{max} the mean maximal score for all years,
a, b and c the ponderation for each phenological stage with $a+b+c=1$

The site-specifically optimized PAST function is then weighted by the photoperiod and integrated over the season to result in a thermal index for each year for the given site as in Eq. 5.2.

$$eGDD\ Thermal\ Index = \int_{season} PAST(t) * photoperiod(t)\,\mathrm{d}t \tag{5.3}$$

with PAST the Phenological Advancement Speed as a function of Temperature
and t the time in Gregorian units

The resulting thermal indices are expressed in Thermally Optimal Daylight Hours (TODH).

5.3.3.2.2 Theory about the Bayesian functional Linear regression with Sparse Step functions (BLiSS method)

A functional linear model relates a time series of data taken as a functional covariate x_j to a scalar response variable y. In this paper, x refers to a time series of temperature data taken as a functional covariate and y to the yield response (Eq. 5.3).

$$\widehat{y} = \mu + \int_{\tau} \widehat{\beta(t)} x(t) \mathrm{d}t \tag{5.4}$$

where $\widehat{y}$ is the response variable y, τ is an interval of $\mathbb{R}$,μ is the intercept, x is the functional covariates with its coefficient functions β

The BLiSS method (Grollemund et al. 2019) proposes a Bayesian approach to estimate the β function and most importantly its support (e.g. time). The BLiSS method is based on a hierarchical Bayesian model. In this model, the support of the coefficient function is taken as a union of possibly overlapping intervals, whose number is controlled by a hyperparameter called K. Each interval is defined by two parameters: its position (its center) and its half-length. The prior associated with the position parameter corresponds to a uniform law over the entire time series and the prior of the length parameter is an exponential law. Given these intervals, the functional linear model becomes a multiple linear model involving the partial integrals of the coefficient function over the intervals as covariates as in Eq. 5.4.

$$\widehat{y_i} = \mu + \sum_{k=1}^{K} b_k x_i(I_k) \text{ where } x_i(I_k) = \frac{1}{|I_k|} \int_{I_k} x_i(t) dt \tag{5.5}$$

where μ is the intercept, x is the functional covariate, I_k a given interval and b_k the related coefficient.

In this way, the BLiSS method leads to the detection of periods during which a covariate (e.g a weather variable) influences a quantitative response variable (e.g. yield performance). These periods correspond to the intervals I_k during which the BliSS estimator takes non-null values. The sign of the b_k coefficient indicates whether the covariate is negatively or positively correlated to the response variable during the correlation direction during each time interval I_k The number of detected periods is constrained by K. In parallel, the probability for a given time to be in the β function support is established. Its posterior distribution provides an assessment of the reliability with which the intervals I_k are detected.

5.3.3.3 Analytic strategy

The analytical framework proposed in this paper includes three steps, in a chronological order : the implementation of the eGDD method in order to obtain synchronized time series, the optimization of discretization time step and the implementation of the BLiSS method to detect periods of influence on yield development. They are summarized in Fig. 5.5.

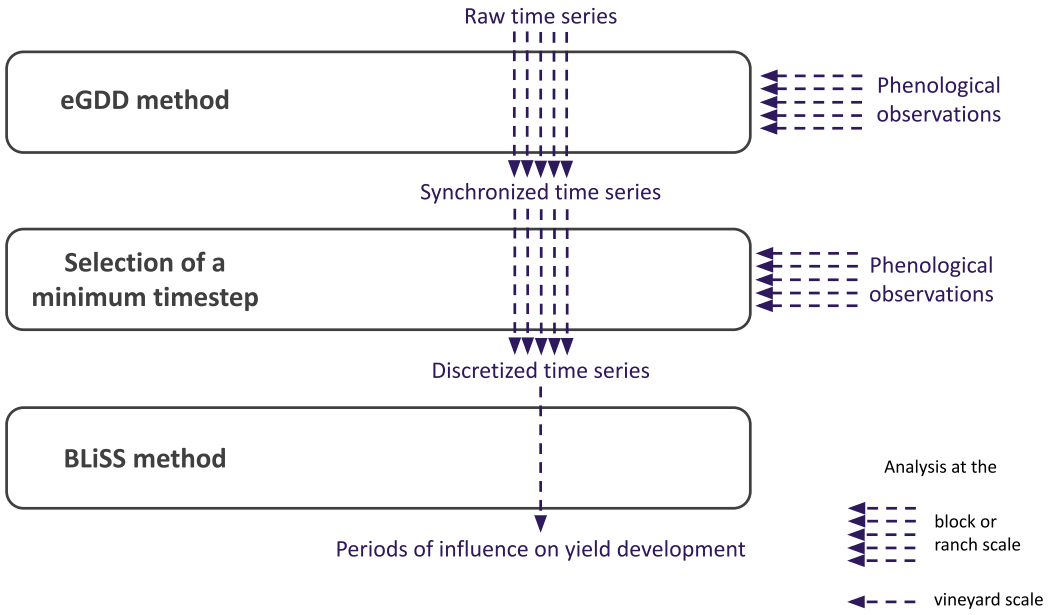

Figure 5.5 – Description of the proposed framework aiming at identifying periods of influence on yield development for each vineyard. Time series of weather data are synchronized according to thermal indices computed with the extended Growing Degree Days approach (eGDD method). Then, the synchronized time series are discretized according to an optimized time step and they are analyzed with the BLiSS method.

The eGDD method is preliminary used to compute a timeline consistent with grapevine phenology according to which time series will be expressed. This synchronization of time series is needed to detect unequivocally periods of the weather influence on yield development thanks to the BLiSS method. Indeed, the BLISS requires each time series to be discretized at a given time step as an input parameter. This time step corresponds to the minimal time step according to which the time series will be passed into a functional data i.e. the maximum number of basis functions whose linear combination will lead to the functional data. The time step is named discretization time step henceforth. The discretization time step was defined according to the eGDD thermal indices. For example, time series expressed according to a given eGDD thermal index could be discretized into periods of 200 TODH. To further the synchronization of the time series according to grapevine phenology, the discretization time step should be optimized so that the respective scores of budbreak, bloom and veraison are included at best in the same discretized period or at least in two discretized period for all the years and the blocks of the vineyard.

A minimum of 5 years of phenological and weather data has been empirically identified to ensure a correct implementation of the eGDD method (convergence of the optimization problem). Consequently, it was possible to apply the eGDD method at the block, ranch or vineyard scale. In contrast, the BLiSS method implementation requires the largest data set possible to limit estimation problems. Therefore, it requires to be applied at the vineyard scale. Thus, the eGDD method was applied at the finer spatial scale possible, depending on the available data : blocks for vineyard A and C, varieties per ranch for vineyard B. In this case, the optimization of a discretization time step for each block or variety per ranch. For example, for vineyard A, all time series will be discretized into 17 periods but these periods will last 300 TODH for block 1 and 320 TODH for block 2. This corresponds to the initial hypothesis that each block has its own rhythm i.e. its own phenology.

The blocks of vineyard B presented phenological observations for a different number of years among which years also varied. In other words, over the whole data set, years were represented by different numbers of phenological data. This was likely to disrupt the eGDD analysis because it would have been driven by the most represented years. Therefore, a weight corresponding to the inverse of the number of observations in the same year was attributed to each phenological observation so that all years have the same weight in the optimization of the PAST function.

5.3.3.4 Analytic strategy

Step 1: implementation of the eGDD method

The eGDD method with Synchronization option (cf. Eq. 5.2) was employed to compute site-specific indices. One eGDD thermal index was computed for each block of vineyards A and C. Regarding vineyard B, some blocks of vineyard B presented a low number of years with phenological observations which prevented the eGDD method to be applied at the block scale. Therefore, one eGDD thermal index was computed by variety for each ranch of vineyard B.
In the case of vineyard B, each combination of variety and ranch included several blocks to which different years of phenological data were respectively associated. Therefore, a weighting was applied to the phenological observations of vineyard B to manage the year imbalance within each group of blocks (variety per ranch).

Step 2: Discretization of the weather data time series

For each block (or ranch for vineyard B) and each year, the timing of the daily mean, maximum and minimum temperature as well as rain time series was reexpressed according to the corresponding eGDD thermal index.
A discretization time step is optimized (minimized) in a block or ranch-specific way with the constraint that the respective scores of budbreak, bloom and veraison are included at best in the same period or at least in two consecutive sections for all the years and the blocks of the vineyard.
Each time series was then discretized according to its site-specific time step by averaging the mean, minimum and maximum daily temperature over each period for the two years before harvest (noted years n-1 and n).

Step 3: Implementation of the BLiSS method

For each vineyard, the discretized time series of all blocks and years were regressed to the yield data using the BLiSS method. The K hyperparameter, which defines the number of influence periods searched for in the time series is tuned using a Bayesian selection approach based on a Bayesian Information Criterion (BIC).

5.3.3.5 General assessment of the framework

The detected periods of influence were compared with literature results about weather influence on yield development and between the three vineyards according to the timing, duration and correlation direction with the yield response.

5.3.4 Results

5.3.4.1 The three vineyards were characterized by different temperature profiles

The profile of daily mean temperature across all the years of each vineyard is given in Fig. 5.6. The range of daily mean temperatures observed in vineyards A and B are similar, approximately from 3 to 33°C. However, vineyard A seemed to experience mildest winters, with daily mean temperatures around 10°C between DOY 0 and 100. The climate of vineyard C was found cooler, with a range of daily mean temperatures approximately from 0 to 28°C. The daily mean temperatures observed in vineyard A showed a higher inter-annual variability than those observed in vineyards B and C, especially from DOY 0 to 250.

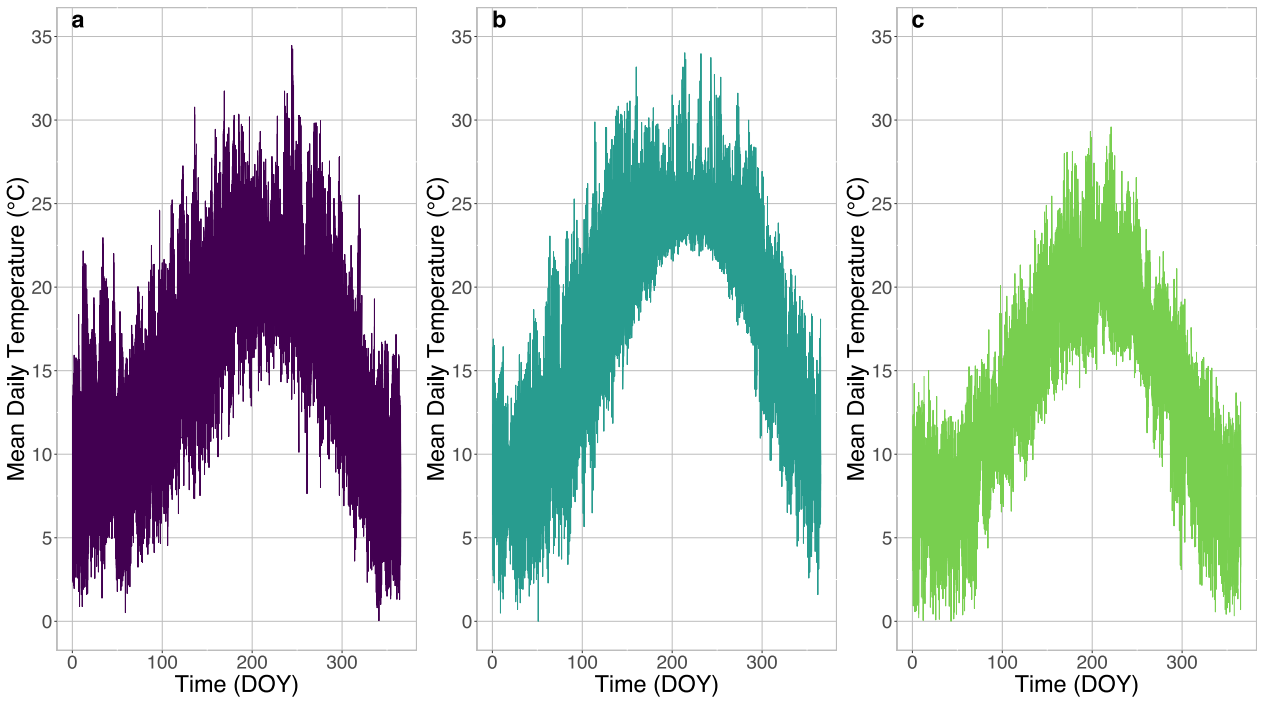

Figure 5.6 – Daily mean temperature data in vineyard A, B and C (a, b and c). A single weather station of vineyard A was taken as an example of the four weather stations.

5.3.4.2 The three vineyards obtained different site-specific eGDD thermal indices

Fig. 5.7 presents the PAST functions obtained with the eGDD method. They are described according to the values of T_b, T_o^1, T_o^2 and T_c that were optimized block by block for vineyards A and C and by ranch for vineyard B. First, it can be observed that the eGDD thermal indices of vineyards A and B are more dispersed than those of vineyard C. For vineyard A, the PAST functions showed a large range of values for each temperature threshold : from -4° to 13°C for T_b, 18 to 25°C for T_o^1 and T_o^2 and 25°C to 50°C for T_c. These values are consistent between the blocks of the same ranch and the slope between T_b and T_o^1 is relatively constant for the whole vineyard. The optimal temperatures, i.e. $[T_o^1, T_o^2]$ interval, range from a single value to a range of 4°C according to the blocks. The values of T_c and in consequence the slope between T_o^2 and T_c vary significantly between blocks : the decrease of phenological advancement speed for temperatures above 25°C vary between bloks. For vineyard B, the values of the temperature thresholds range from -10 to 6°C, 21 to 32°C for T_o^1 and T_o^2 respectively and from 30 to 41°C for T_c. The slope between T_b and T_o^1as well as between T_o^2 and T_c is different between the different ranches. For vineyard C, the values of temperature thresholds range from 8 and 11°C for T_b, from 19 to 21 (with an exception at 27°C) for T_o^1 and T_o^2 and from 31 to 45°C for T_c. The values of T_b, $T_o^1$and T_o^2as well as the slope between T_b and T_o^1 are very similar between blocks while the values of T_c and therefore the slope between $T_o^2$and T_c present significant differences between the blocks. However, this part of the PAST function relates to few temperature observations actually recorded in the field (Fig. 5.6).

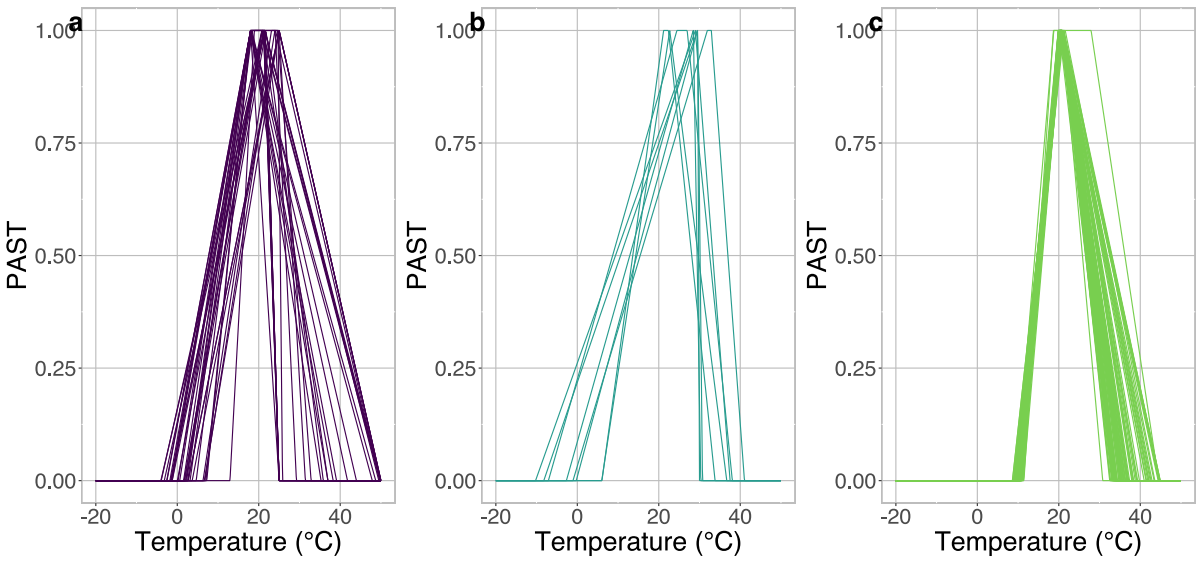

Figure 5.7 – Phenological Advancement Speed as a function a Temperature (PAST function) computed with the eGDD method for each block of vineyards A and C (respectively a and c) and for each ranch of vineyard B (b).

5.3.4.3 The three vineyards obtained different discretization time steps

Table 5.2 – Results of the time series discretization for vineyard A, B and C.

Vineyard	Number of discretized periods over the years n-1 and n	Mean time step (in TODH)	Number of blocks/ranches excluded from the BLiSS analysis	Time period rank corresponding to Budbreak	Time period rank corresponding to Bloom	Time period rank corresponding to Veraison	Time period rank corresponding to Harvest
A	17	354.2	10	2, 3	4, 5	7	9,10
B	18	252.4	0	1	2,3	6,7	8,9,10,11
C	19	191.5	7	1	2,3	7	9,10

Table 5.2 presents the results of the optimization of the discretization time steps for each vineyard. This process imposed the exclusion of some blocks for vineyards A and C because the periods into which budbreak, bloom and veraison were positioned showed different combinations (with more than two periods of difference). For example, a block of vineyard C whose budbreak would happen in a period of rank 2, bloom in a period of rank 5 and veraison in a period of rank 7 would be excluded. The exclusion mainly concerned blocks whose variety was in minority in the vineyard i.e. blocks planted with another variety than Cabernet-Sauvignon for vineyard A and blocks planted with Petit-Verdot for vineyard C. No exclusion was performed for vineyard B since the eGDD thermal indices were computed at the ranch scale, preventing such a fine selection.

The discretization time step differed between vineyards in accordance with the difference of eGDD thermal indices. However, they allowed a similar number of discretized periods over the time series of weather data. It is to be noticed that the position of budbreak, bloom and veraison is more consistent for vineyard C than for the other two vineyards. The veraison is consistently timed between vineyards : it is always positioned in the 6th or 7th rank. However, budbreak and bloom are respectively positioned in the 1st and 2nd or 3rd periods for vineyard B and C while they are positioned in later periods for vineyard A.

5.3.4.4 The three vineyards were characterized by different periods of weather influence on yield development

Fig. 5.8 shows the results of the BLiSS analysis of the discretized time series of daily mean temperatures for the three vineyards. The results correspond to the detection of periods when daily mean temperature influences yield elaboration. The timing and duration (expressed in discretized periods) of the detected periods as well as their correlation direction (sign of the BLiSS estimator) are to be interpreted. The values

taken by the BLiSS estimator are not to be interpreted between vineyards but only relatively between periods of influence for each vineyard. The color gradient corresponds to the distribution of the posterior distribution of the β estimator. It is to be interpreted as a confidence indicator for the detection of influence periods thanks to the BLiSS estimator.

First, it can be noticed that the confidence in the estimation of the β coefficient is lower for vineyard A than for vineyards B and C i.e. periods of influence are more strongly detected for vineyards B and C than for vineyard A, surely in relation with the number of analyzed individuals.
For vineyard A, only one period of influence can be reliably identified for periods 12 to 14. Two other periods can be presumed from periods 1 to 5 and 9 to 11. Regarding the period from period 12 to 14, the value of the BLiSS estimator is positive i.e. the daily mean temperature observed during this period is positively correlated with the yield performance (the higher the temperature, the higher the yield).

For vineyard B, 4 periods of influence can be identified respectively from periods 1 to 2, 7 to 9, 10 to 14 and 17 to 18. A fifth period could even be detected in period 16 although it has not been selected by the sparse step of the BLiSS estimator. The 2nd and 4th period are positively correlated with the yield performance while the 1st, the 3rd and the 5th period are negatively correlated with it.

For vineyard C, 4 periods of influence are detected, respectively from periods 1 to 2, 6 to 8, 12 to 14 and 18 to 19. The 2nd and 3rd periods are positively correlated with yield performance while the1st and 4th are negatively correlated with it.

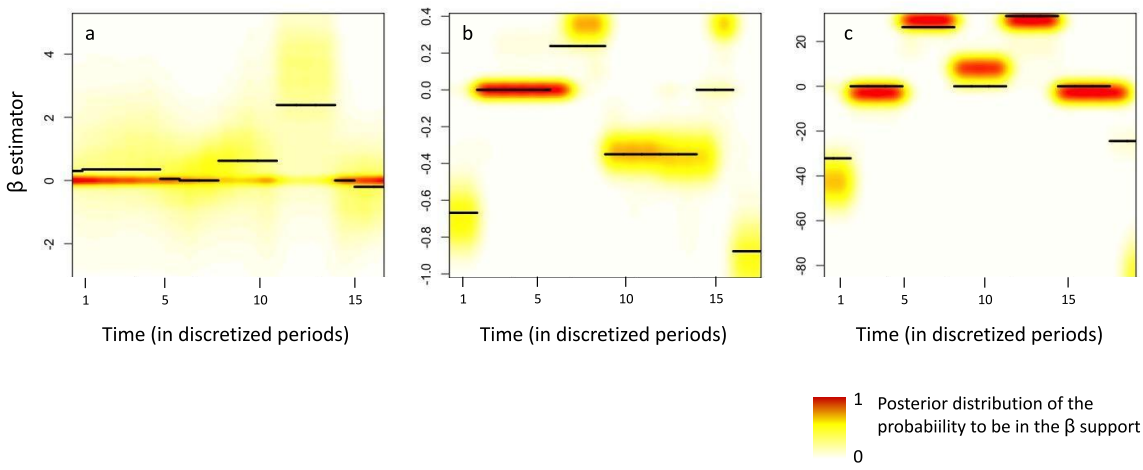

Figure 5.8 – BLiSS estimation for the synchronized time series of averaged daily mean temperature data for vineyards A, B and C (respectively a, b and c). The discretized periods that graduate the X-axis correspond to a segmentation (discretization) of the site-specific eGDD thermal indices that were used as a timeline to express the temperature time series. Positive, null or negative values of the estimator on the Y-axis indicate that the daily mean temperature promotes, doesn't affect or hinders yield development during the considered period.

The results of the BLISS analysis of discretized time series of daily mean, minimum and maximum temperatures are summarized in relation with the vine phenology of the three vineyards in Table 5.3.

Table 5.3 – Timing, duration and direction of correlation with the yield response (- : negative, + : positive) of the periods of influence detected with the BLiSS method for the time series of daily mean, maximum and minimum temperature and rain data of each vineyard. The gradient of green colors respectively represents the periods of budbreak, bloom and veraison in year n-1 and year n for each vineyard. The grey cells do not correspond to periods because the optimization of the time step to discretize the time series data resulted in a smaller number of periods.

		Periods number over the year n-1										Periods number over year n								
Vineyard	Temperature indicator	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	Tmean												+	+	+					
A	Tmin	-	-	-					-	+	+	+								
A	Tmax		+	+	+	+														
A	Rain			+	+	+														
B	Tmean	-	-						+	+	-	-	-	-			+	-	-	
B	Tmin				+	-	-	-	+		-							-	-	
B	Tmax	-												-	-	+	+		-	
B	Rain	+	+	+					-	-										
C	Tmean	-	-				+	+	+				+	+	+				-	-
C	Tmin			+	+	+	+	+	+		-	-					+	+	+	-
C	Tmax						+	+	+		+	+	+				+	+	+	-
C	Rain		+	+	+	+	+						-	+	+					

5.3.5 Discussion

5.3.5.1 The analytical framework allowed a site-specific reduction of the information contained in the weather time series

5.3.5.1.1 The results of the different analytical stages were found coherent with literature but site-specific

Regarding stage 1, the eGDD method resulted in a large range of values for each optimized temperature threshold for vineyards A and B. In particular, most of the values obtained for T_b were between 4 and -4°C. In contrast, vineyard C presented higher values for T_b. Other studies have proposed a range from 5 to 12°C for T_b (Zapata, Salazar-Gutierrez, et al. 2017; Molitor, Fraga, and Junk 2020). These results were obtained in regions with cool climates and are coherent with the results of vineyard C, which experienced the coldest winters among the three vineyards. Besides, Parker, De Cortázar-Atauri, et al. 2011 reported T_b values between 0 and 4°C when the thermal index computation was initiated later in the season (around DOY 60). Such an initiation could represent the thermal conditions experienced in vineyards A and B during winter, where no daily mean temperatures below 2°C was observed. Therefore, the T_b values obtained with the eGDD in the case of vineyards A and B are not to be commented as such but rather according to the slope of the increasing part of the PAST function they imply. This slope corresponds to the relative weight associated with temperatures actually observed in the field. In this way, low values of T_b allow temperatures between 5 and 20 °C, i.e. most of the recorded temperatures, to have a closer weight in the definition of the thermal indices than higher values. This relative weight led to the minimization of the optimization criterion (Eq. 5.2) by the optimization process. The same logic could be applied to interpret the values of the other temperature thresholds.

Regarding stage 3, sensitive periods of yield development to temperature highlighted by previous experimental works were also detected by the BLiSS method (Table 5.3). However, they differed between vineyards in terms of temperature variable (Tmean, Tmin or Tmax) as well as precise timing, duration and direction of correlation with yield.

For example, the period around budbreak and bloom of year n-1 was detected for all vineyards. This period of floral induction and initiation was shown positively correlated to yield by literature (Buttrose 1969; Buttrose 1970; Pagay and Collins 2017; Molitor and Keller 2017). However, Tmean was found negatively correlated to yield during this period for vineyards B and C. In the case of vineyard A, Tmin was negatively correlated to yield while Tmax was positively correlated to yield. In addition, only budbreak was included in this period of influence for vineyard C while bloom was also included in the case of vineyards A and B.

The period comprising budbreak and bloom of year n was also detected as a period of temperature influence for all vineyards. It could correspond to floral differentiation as highlighted by previous studies (Pouget 1981; Dunn and Martin 2004; Clingeleffer et al. 2001; Jones, Menary, and Wilson 2009; Keller, Tarara, and Mills 2010). However, it was found positively correlated to yield for vineyard A and negatively correlated for the two other vineyards. The timing and duration of this period also differed between vineyards. The period of influence seemed to start earlier, from the budbreak period, for vineyards A and B while it was centered on bloom only for vineyard C. A short period of negative correlation between yield and temperature was also detected close to harvest for vineyard B and C but not for vineyard B. It could be concomitant with sugar accumulation for Tmean during the two periods before harvest and coherent with the assumption of berry dehydration in relation with a high vapor pressure deficit for Tmax during the last period before harvest (Keller, Smith, and Bondada 2006).

For all vineyards, rain influence was mainly detected around bloom of year n-1. Vineyards A and B were not irrigated during this period, so this seems consistent with the incidence of water stress effects found by Guilpart, Metay, and Gary 2014.
In parallel, periods of temperature influence that were not referred to in literature were also site-specifically detected. For example, the minimum daily temperature was found positively correlated to yield during a long period from bloom to veraison of year n-1 for vineyard C while this period was divided into three different one for vineyard B, the 1st and 3rd being negatively correlated to yield. Such an influence was not detected for vineyard A.

5.3.5.1.2 Weather indicators were precisely defined as periods of weather influence on yield

The analytic framework proposed in this paper allows the learning of the timing and duration of the influence periods based on the data while it would have been decided before a classical analysis. To encourage the detection of precisely defined periods, the time series of data were synchronized according to a eGDD thermal index so as to reduce the phenological shift between the analyzed blocks and years. Moreover, in the BLiSS method, the prior used for the half-length of each period of influence corresponded to an exponential law, which encouraged the detection of periods of a parsimonious duration. In this way, periods corresponding up to only one period could be detected, for example for the minimum daily temperature after veraison of year n-1 for vineyard B (Table 5.3). At this time of the year, a period expressed according to an eGDD thermal index corresponds to a period ranging from about ten to fifteen days, which is the finest time step that could be evidenced by Molitor and Keller 2017 with a Windows Pane approach. At this period of the year, such a period would also be equivalent to 100-150 Growing Degree Days, which is often the smallest time step explored in classical analyses (Guilpart, Metay, and Gary 2014).

5.3.5.2 The proposed analytical framework can be deployed on farm data sets

The analytical framework of weather time series proposed in this study was tested on the data sets of three commercial vineyards, which i) experienced different conditions of yield development (environment, cultivation practices, etc.), ii) are characterized by different organizations (e.g. one or more ranches) and iii) also correspond to different data sets quality. Thus, the applicability of the proposed analytical framework was tested according to different criteria that are representative of most issues of farm data analysis.

5.3.5.2.1 Small data sets can still be analyzed

The number of individuals involved in the Bliss analysis, i.e. the number of time series per plot and per year that could be synchronized and discretized at the vineyard scale, has a strong impact on the results in

terms of the number of periods detected and the precision with which they are detected. Thus the results obtained for vineyard B and C are more significant than for vineyard A : more periods were detected and the Bliss estimator fitted well with the center of the posterior distribution of the probability of any instant t to be in the support of the β estimator function. However, the small number of individuals analyzed for vineyard A would have prevented the analysis of the time series with such a discretization in a frequentist framework because of estimation problems (17 periods for only 140 individuals). In contrast, the Bayesian approach included in the BLiSS method still allowed the analysis of the data set and provided information on the uncertainty that characterized it. Thus, Table 5.3 only lists the periods of influence that were unequivocally detected for vineyard A but an expert analysis of the results could have allowed more periods to be selected.

5.3.5.2.2 The definition of a site can be adapted to the vineyard organization and data volume

The site-specific analysis proposed in this work is based on the assumption that there is a consistent vineyard effect over time. This assumption determines the quality of the final results : the more consistent the vineyard effect, the more reliable the period detection with BLiSS is fostered. Therefore, this assumption must be validated at each step of the analysis. First, the computation of a GDD thermal index required a minimum volume of data related to weather and phenological observations for a given site. A limitation in the quality of the results appeared when less than 5 years were considered for a site (data not shown). Therefore, it was sometimes necessary to balance the desired scale of application of the eGGD method with the scale that allowed for a sufficient volume of data. For example, there was not enough data to calculate the eGDD indices at the block scale for vineyard B. However, it is known from previous work that the most valuable scale of application of the eGDD method is the vineyard, taking into account the grape variety. Therefore, applying the eGDD method at the scale of all the blocks of a same ranch and of a same grape variety was assumed to be an appropriate trade-off. Secondly, the BLiSS analysis could not be performed at the block or even at the ranch level due to the amount of data available. Therefore, it was performed at the vineyard scale, assuming a certain consistency of the effects of the environment, cultural practices, etc. between the ranches of the same vineyard. However, the validity of this assumption required the exclusion of some blocks whose phenology was really different from the majority of the blocks or whose grape variety was poorly represented in the vineyard.

5.3.5.2.3 Unbalanced samples are supported

For a variety of reasons ranging from climatic hazards to logistical failures, the number of individuals (observations per year and per block) can commonly vary between blocks or years in farm data sets. This issue can lead to an unbalanced sampling of site and year effects within the analysis and to non-robust conclusions because potentially driven by a small number of individuals. Regarding the site effect, numerous precautions were undertaken in the whole analytical framework to assume a constant site effect at the vineyard scale. Therefore, the imbalance in the number of individuals representing each block (i.e. years per block) was not considered to be a major issue. Each block was indeed considered to be a realization of the same vineyard-specific pattern.

However, the reverse side of this unbalanced sampling is that the years were also represented with a different number of individuals (i.e. blocks per year). The analysis of the climate effect on yield elaboration being inherently prone to incorporate year effects, this unbalanced year sampling was considered to be a red-flag issue. This was especially true for phenological observations hence for the implementation of the eGDD method. As it was not possible to include a random vintage effect in the analysis, the individuals were weighted by the inverse of the number of blocks for their corresponding year to balance the different years. In the case a vineyard B, which presented the more unbalanced year sampling, performing the eGDD analysis at the ranch scale was also a way to gather more individuals representing the same years.

5.3.5.2.4 Different analytical designs can be supported but call for an operational interpretation of the results

Vineyards A and B were both split into several ranches whereas vineyard C only counted only one ranch but with a higher number of blocks. Each ranch of vineyard A was equipped with its own weather station while the same weather station covered all the ranches of vineyard B. Vineyard C was also equipped with a single weather station. All these options were workable with the proposed analytical framework but required to be taken into account when interpreting the results.

Regarding the eGDD method, a eGDD thermal index is assumed to integrate i) physiological variations in the vine response temperature depending on plant factors and also on environmental factors, ii) spatial

variations of temperature conditions and iii) the quality of phenological observations and weather data (Laurent, Scholasch, et al. 2020). In the case of vineyard A, the eGDD method is applied at the block scale with a weather station being close to each block. Therefore, it can be assumed that the difference of eGDD thermal indices between the blocks of a same vineyard mainly correspond to differences in the vine response to temperature or to the data quality rather than to spatial variations of temperature conditions. This hypothesis is reinforced by the fact that a clear consistency was observed between the eGDD indices of three of the four ranches of vineyard A, the last ranch being the most spatially extended and comprising the largest number of blocks. In the case of vineyard B and C, spatial variations of temperature conditions may play a greater role in the differences of eGDD thermal indices. In addition, for vineyard B, the eGDD method is applied at the ranch scale. Therefore, the resulting indices are to be interpreted as a trade-off between different vine responses to temperature (usually apprehended at the block scale). The clustering of blocks according to their ranch but also to their variety should help the trade-off not to hide very different realities. Otherwise the constrained optimization problem of the eGDD method would have difficulties to converge and would tend to obtain PAST functions with very close temperature thresholds so as to cumulate very few heat units.

Regarding the BLiSS method, site-specific periods of climate influence on yield elaboration were detected. These site-specific do not necessarily mean that the physiological mechanisms of yield elaboration are different. The site-specific conditions actually recorded in each vineyard may indeed explain part of this. For example, the fact that one of the vineyards is subject to a particular climate may explain why certain periods are not identified, because weather influence never turns to determining or limiting during this period, whereas it is for other vineyards. They can also be explained by the data representativity of the real conditions experienced in the vineyard, either due to spatial variations of the temperature conditions or other variations of other environmental factors or of cultural practices. In addition, the precision with which the periods are detected, especially with respect to the periods duration, is also determined by variations of the vineyard-specific effect among the ranches and blocks of the data sets. Even though the phenological shift between blocks is assumed to have been significantly corrected, there could exist other sources of variation for example differences in sensitivity to certain thermal influences due to the grape variety or a particular cultural practice that would not be verified at the whole vineyard scale. Finally, this could also be explained by the linearity of the BLiSS model, which implies the monotony of the effects of each period of influence. In other words, for a given period of influence, temperature can not be considered to have an optimum effect as suggested by the Arrhenius law and numerous studies on temperature effect on living organisms. Considering periods of influence whose effect could be common to all vineyards, this issue could lead to a same period being positively or negatively correlated to yield depending on the temperatures actually observed in each vineyard. For example, this could explain the differences of correlation direction of the daily minimum temperature with yield during the winter period between years n-1 and n. Vineyard A was indeed the vineyard which experiences the mildest winters (Fig. 5.6) and it was also the vineyard for which the minimum temperature was found positively correlated to the yield during the winter period between years n-1 and n (Table 5.3).

5.3.6 Conclusion

This study proposed an analytical framework combining two statistical methods : the eGDD and the BLiSS ones. It aimed at leveraging farm data sets to the maximum by implementing site-specific analysis of time-series data. The influence of climate on grapevine yield development in three different commercial vineyards was chosen as a case study. This showed the potential of the method in terms of i) a more comprehensive and less subjective analysis of time series data in order to extract weather indicators with reduced dimensions and ii) feasibility when working with farm data. The results of such analyses should be carefully interpreted since they integrate numerous determinisms in relation with the operational reality of commercial vineyards. Such analyses are also of interest to commercial vineyards as they give them guidelines to operationally interpret their own data to better understand their own vineyards. They could be applied to other crops, especially perennials one and could also relate to other time series data and response variables. However the proposed analytical framework is clearly an exploratory approach method. Its results should then be validated by quantified and more robust methods accounting for uncertainty.

5.4 General conclusion of the chapter

The work presented in this chapter has generated several scientific contributions to the related literature. First, it confirmed the potentialities of better leveraging time series of weather data for the purpose of better understanding and modelling yield elaboration. It proposed an analytical framework to improve the exploration of time series data and extract relevant indicators of reduced dimensions. **This framework includes the eGDD method to synchronize time series on the basis of grapevine phenology and the BLiSS method to regress the time series against the yield response in order to identify periods when a variable is particularly influential.** The growing season of years n-1 and n generally included three periods of weather influence with site-specific variations : a first period included budbreak and bloom, a second period surrounded veraison and a third period referred to the post-harvest season (year n-1) or the time just before the harvest (year n). **The interest of site-specifically performing such analyses was evidenced** by the refining of the influential variable (Tmin, Tmax, Tmean or Rain), timing, duration and the direction of correlation with yield for each period in each vineyard. **To empower it, the feasibility of the analytical framework on the basis of three farm data sets was demonstrated.** These three case studies referred to vineyards localized in France, California and Israel and illustrated a wide variety of site-specific situations.

Chapter 6

Discussion and perspectives

6.1 Looking back at the results of the PhD project

The PhD project focused on the construction of a yield forecasting method in viticulture for operational purposes and using farm data. This construction approach is summarized in Fig. 6.1

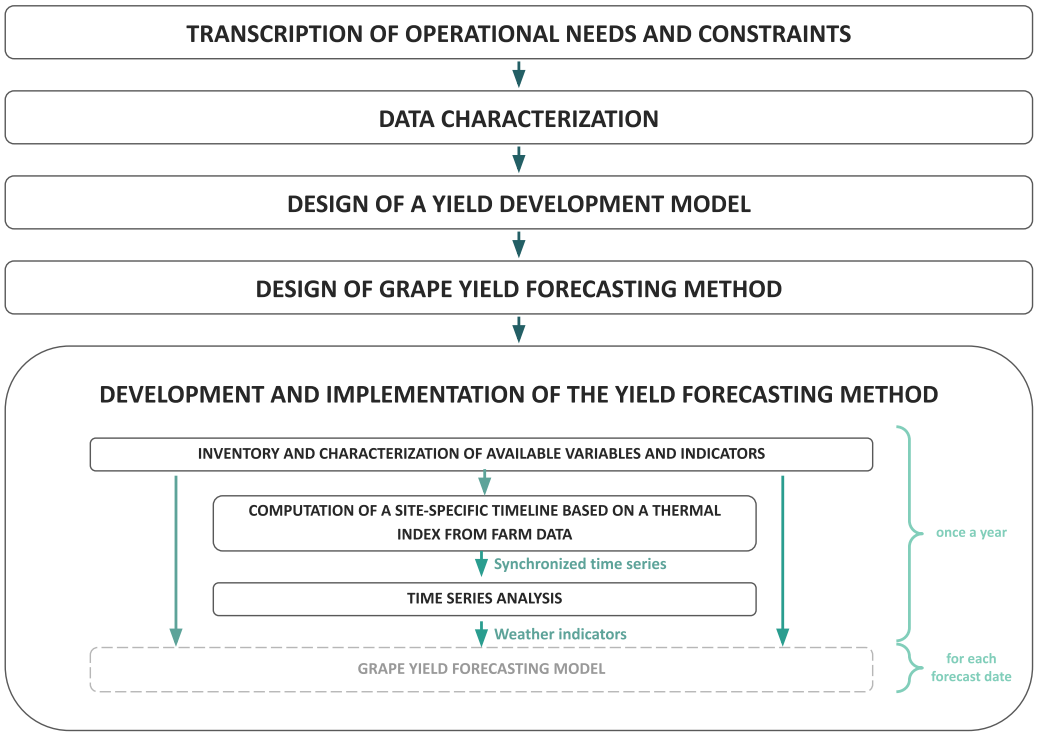

Figure 6.1 – Procedure for the construction of a yield forecasting method in viticulture for operational purposes and using farm data

First, a transcription of the operational context for grape yield forecasting into scientific issues and challenges was performed (chapter 2). It was based on the consolidation of feedback from the commercial vineyards /wineries and referred to the operational context as the needs, constraints and data of commercial vineyards. This transcription required the clarification of certain notions such as the difference between yield estimation, prediction and forecast to specify the goal of the study. Hence, a general framework for yield estimation, prediction and forecasting was proposed. It allowed the positioning of literature proposals and helped to identify gaps that might explain why the research has not yet met field expectations in terms of yield forecasting.

Then a characterization process of farm data was proposed from the study of farm data sets, which involved exchanges with the vineyards/wineries which shared them. The characterization of farm data sets implied two data categories : i) essential variables and indicators, which are commonly found in farm data sets and ii) optional variables and indicators, which are found in some data sets but not in others.
A conceptual model of yield development was designed on the basis of essential variables. It implied the definition of a site-specific system to be studied as a data-driven black box. Input variables to this model correspond to essential variables and the output variable corresponds to the final yield performance as it is estimated in each vineyard/winery at the block level. Combining the operational needs and constraints with the yield development model, a yield forecasting method was designed. It includes several forecasts to be set at different judicious dates. Each forecast corresponds to a specific model based on observations that were already collected. All corresponding models are valid for the considered vineyard/winery only since they are based on its available variables and indicators and a vineyard-specific black box is trained to link these variables to the yield response.

The concrete development and implementation of this method was then presented. It includes three steps among which two were presented in this manuscript. These two steps focused on time series of weather data. Therefore, time series of weather data indeed presented the greatest unexplored potential of information and the greatest opportunity in terms of improving yield forecasting.

As a first step, an original method was developed to compute site-specific thermal indices. It is called Extended Growing Degree Days (eGDD) and it involves a constrained optimization method to site-specifically calibrate the relationship to be considered between temperature and grapevine response in terms of temperature. Such thermal indices were used as timelines to synchronize time series of weather data according to grapevine phenology. The eGDD method additionally proposed the option of calibrating a thermal index to predict the achievement date of key phenological stages. The eGDD method was shown to offer better results in terms of synchronization and prediction than the Gregorian calendar and the classic Growing Degree Days approach. The results showed the interest of the approach for three commercial vineyards that corresponded to three very different operational contexts. **An analytical framework was then developed for processing time series in order to identify weather influence on yield development and to consequently formulate parsimonious weather indicators.** Such an analytical framework was shown to detect site-specific indicators. It was able to work from time series of farm data.

Finally, the resulting weather indicators and other essential indicators are to be incorporated as covariates under a Bayesian framework in a yield forecasting model. This is the main perspective of the PhD project for further work. It will be presented in section 4 of this chapter.

6.2 Discussing the response of the PhD project to the issues and challenges identified for grape yield forecasting in an operational context

Concluding from the confrontation of literature proposals with operational needs and constraints for yield forecasting in an operational context, criteria of operational relevance, temporality and site-specificity were identified as challenges for yield forecasting in an operational context (chapter 2). These challenges have been invested through the responses to the four research questions, from the design of a yield development model and yield forecasting method to their concrete implementation involving proper statistical approaches. This is summarized in Fig. 6.2

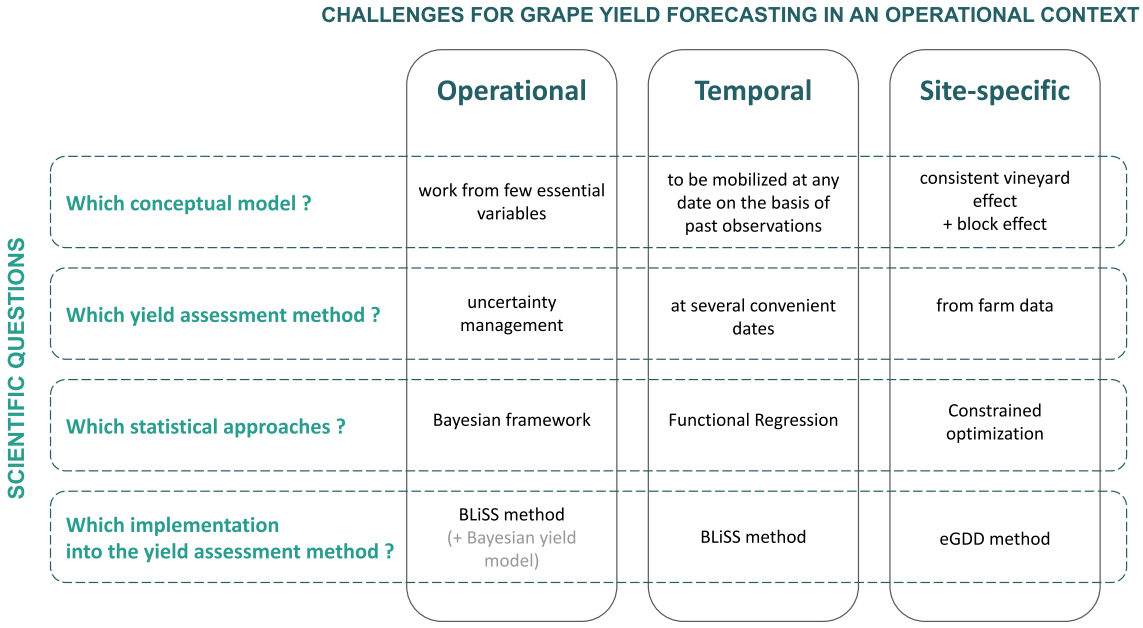

Figure 6.2 – Positioning of the scientific contributions of the PhD project to the issues and challenges identified for grape yield forecasting in an operational context. The light grey color refers to further work perspectives.

6.2.1 Aiming at operational relevance

Aiming at operational relevance first implied to leverage existing data before working on leveraging future farm data which would be collected according to other measurement and sampling protocols, as often proposed in the scientific literature. Hence, the focus was put on accessible data, most often present in the farm data sets, called essential data, and amongst them, weather time series. Thus, it was assumed that the return on investment for weather time series would be greater than for other data since climate influence on yield development is prevalent and that these data were among the less leveraged because of their specific characteristics.

However, detecting weather influences on yield development from the analysis of weather time series is not trivial (Parkes et al. 2019). Weather influence is indeed a complex influence with high inherent variability, which includes trajectory and memory effects. Therefore, weather influence on yield development is challenging to decompose. **The weather influence is all the more difficult to capture as the data observed in any farm data set only represents a small portion of the possible realizations of weather influence on yield.** Therefore, it is possible that the next realizations will be different from what was ever observed in the previous data set, especially in the context of climate change. Moreover, a farm is a permanent experimentation center with numerous agronomic and logistical constraints. Again, vineyard management will have to be adapted to cope with climate change (Naulleau et al. 2020). Therefore, data from different years and blocks can not be compared with the assumption of equal conditions because cultural practices may slightly change. However, a certain spatial consistency of the vineyard management is assumed at the vineyard scale because it corresponds to the scale at which most operations are decided. Finally, the quality of farm data set and its representativity of the conditions which really happen in the field is heterogeneous from one vineyard to another and always uncertain. In this way, the inherent uncertainty of yield development in relation with weather influence and the uncertainty arising from the operational context and farm data call for a strong uncertainty management. This uncertainty management is required for the analysis of weather time series but also in the modelling approach for yield assessment. **In this PhD project, uncertainty management was achieved by using the Bayesian statistical framework.** It was used in the BLiSS method to detect parsimonious periods of weather influence on yield development. It should also be used to design and train a yield forecasting model based on weather and other essential variables. This last point is the main perspective of the PhD project for further work.

6.2.2 Taking into account the temporality of yield development

Accounting for yield development temporality implied mobilizing an evolutive conceptual model of yield development. This was not achieved by designing a properly dynamic model like mechanistic ones (Cola et al. 2014; Nogueira Júnior et al. 2018) but rather a model ready to take different covariates as input parameters depending on the date when it is mobilized.

Accounting for yield development temporality also meant accounting for it in the delivery of the yield forecast so that it can serve decision making at the appropriate dates. **These dates were defined as following a period of high influence on yield elaboration (cf. chapter 5) or when a cultural operation needs to be decided.** Such decisions may concern the canopy management, the date of the first irrigation, the date and dose of nitrogen fertilization or the bunch thinning intensity (Keller, Mills, et al. 2005; Reynolds and Vanden Heuvel 2009; Intrigliolo and Castel 2009; Metay et al. 2015). To the best of the authors knowledge, no research proposal had been made to improve the timing for the yield assessment release. **Following this outline, the yield forecasting will be supported by a model settled for a given date on the basis of the available variables and valid for a given vineyard. Each forecast will be based on the already collected data, this is why the candidate variables may not be the same from one model to another according to the forecast date.** For example, new observations of yield components could be added as covariates along their progressive implementation. It also concerned the consideration of past periods of weather influence as the season progressed. Yet, the modelling process will remain the same. It is discussed in the section 4 of this chapter. It is important to note the difference with another approach that would have corresponded to the forecast of covariables as a preliminary processing before the yield forecasting step. In this case, the yield model would have sensitively been the same but the proportion of observed and forecast indicators would have varied between dates for the forecast release. For example, the yield forecasting model could take a temperature forecast as covariate. This possibility was discarded because it implied an even more advanced uncertainty management, which seemed limited by the size of the farm data sets. Nevertheless, it could constitute an area for further improvement of the yield forecasting method.

The identification of periods of weather influence on yield development required the leveraging of weather time series. This was achieved using a functional approach. It was implemented in the BLiSS method that also allowed to reduce the dimension contained in time series thanks to a parsimonious estimation of the number, timing and duration of periods of weather influence on yield. To the best of the authors knowledge, only a few studies proposed to investigate time series of weather data in order to identify original weather indicators (Molitor and Keller 2017; Sirsat et al. 2019). However, the BLiSS method did not account for the interactions between these periods of influence within so-called trajectory effects in the vine physiology. This will be considered as a perspective for further work, in relation with the Bayesian yield forecasting model. Indeed, such numerous interactions could not be estimated by pairs as covariates in such a model. However, trajectory effects could be considered at the scale of the whole yield development cycle by setting random effects within the yield forecasting model. This is particularly coherent in a Bayesian approach which already manages parameter distributions.

6.2.3 Proposing a site-specific approach of yield development

The site-specific approach of yield development was considered according to a vineyard effect which was assumed constant and which was refined by block effects (Fig. 3.1). In other words, each block was assumed to be a realization of the same stochastic vineyard effect, with small variation around it. Such effects could only be informed by leveraging vineyard and block-specific data i.e. farm data. This implied analyzing different indicators for the same variables from one vineyard to another. For example, yield performances were respectively expressed in tons/acre, tons/dunam or hectoliter/hectare in vineyards A, B and C. Therefore, a special attention was to the characteristics of farm data during each analysis implementation. This point is addressed in the third section of this chapter. This is an original contribution to the scientific literature because the only studies that have considered farm data focused on empirical models for yield assessment (Clingeleffer et al. 2001; De la Fuente et al. 2015). Besides, a first intention was to enable the hypothesis of a constant vineyard effect in analyses of data from different years and different blocks. This involved limiting the phenological shift between such data by re-expressing the timeline of each site (block or ranch) according to a thermal index. This site-specific calibration of such thermal indices was achieved using a Constrained Optimization approach. It was implemented in the original eGDD method.

Moreover, it is important to note that the site-specific conceptualization of yield development goes along

with a data-driven modelling of a vineyard-specific block box. Further detailing on the site-specific effect is not sought because it is considered that neither the data nor the knowledge (given all the possibilities of integrative operational contexts) are sufficient to do so. However, such a site-specific modelling approach aimed at identifying particular effects of input variables on yield development. For example, it is interesting to identify periods of weather influence on yield development to better understand it from an agronomic point of view. It is also interesting from a practical point of view in order to determine when and how it is relevant to take action in the vineyard. This interpretability request also justified the fact that artificial intelligence methods were not investigated, beyond the limitation due to the size of farm data sets.
However, only using essential variables to model yield raises questions about the precision of the yield forecast that can be achieved. Other variables in relation with the soil, the cultural practices or the plant water status are also known to strongly determine yield development (Ojeda, Deloire, and Carbonneau 2001; Reynolds and Vanden Heuvel 2009; Ripoche et al. 2011; Guerra and Steenwerth 2012; Guilpart, Metay, and Gary 2014; Guilpart, Metay, and Gary 2014) and their inclusion could significantly improve the yield forecasting precision. This question has not been fully answered yet and constitutes an area for further work. It should be addressed in two steps, first testing if the yield forecast is better than what is actually done in the field and then testing the best results that can be expected from the yield forecasting approach based on essential variables, before including optional ones.

6.3 Discussing the response of the PhD project to valuing farm data to perform agronomic research with operational purposes

6.3.1 Handling the characteristics of farm data

Farm data is the most representative data of any given vineyard because it is site-specifically collected. However, this definition inversely reflects the limitations of such a benefit. Indeed, farm data is collected according to the operational constraints. This results in a number of characteristics that need to be taken into account by an adequate statistical analysis (Marchand et al. 2014).

Farm data is often characterized by a parsimonious number of variables. This requires the design of appropriate models and the maximum leveraging of available variables, such as time series data. Although larger than experimental data sets, farm data still correspond to small data sets. This encouraged simplicity and parsimony for improved robustness in all the analyses of this thesis project. For example, a more realistic version of the eGDD method would have been to consider a smoothed and more free PAST function, whose shape would have evolved over the season to represent the different thermal determinism of the successive phenological stages i.e. at least budbreak, bloom, veraison. However, this would have required more parameters to be optimized under a functional approach. It would not have been possible to fit such a model on the basis of the three phenological stages that are generally observed during the year. Similarly, the BLiSS method was employed to reduce the information contained in time series so as to reduce the incidence of estimation problems in the further yield forecasting model. The option of directly integrating functional covariates into such a model was considered and tested. However, it systematically led to over-fitted models on the basis of the studied farm data sets. In addition, the Bayesian framework also fostered the analysis of small data sets (Che and Xu 2010).

Farm data sets also include interdependencies due to the fact that the same blocks are studied over time. Therefore, it is necessary to consider block and/or ranch as well as year effects in the analyses. Moreover, some data may be missing for a wide variety of operational reasons, which creates an unbalanced sampling between the different blocks and years. This was the case of vineyards A and B. The unbalanced block sampling was not considered to be an issue under the assumption of a constant vineyard effect. Each block was indeed considered to be a realization of the same vineyard-specific pattern. In contrast, the unbalanced year sampling was considered to be a red-flag issue due to the objective of taking the temporality of yield development into account. It was handled by weighting observations according to the number of observations for each year for the eGDD method. It could be handled by a year random effect in the further grape yield forecasting effect. Finally the quality of farm data is heterogeneous, in relation with measurement, sampling and traceability issues varying from one vineyard to another or even across years in the same vineyard. The

resulting potential noise must be taken into account in order not to mistake the decision-making process in the field. In this PhD project, the Bayesian framework was used to handle uncertainty arising from farm data quality in order to deliver information whose uncertainty was explicitly specified (Che and Xu 2010). Being explicit about uncertainty is also a way to challenge farmers : if they are not satisfied with the interpretation they can make out of their data, then they may have an interest in changing the way they collect the data. Therefore, such an approach is likely to encourage the involvement of farmers into a soft improvement loop.

6.3.2 Interpreting the results of analyses from farm data

The results of a site-specific analysis based on farm data are to be understood as an integration of the influence of several components. First, it includes the vine answer to the covariates that are being studied. For example, i) the eGDD thermal index integrated some thermal determinism of grapevine phenology site-specifically and ii) the periods detected with the BLiSS method represented periods during which a weather variable has a non negligible effect on yield development.

However, the results also integrate other sources of variability. The data representativity of the conditions that actually experienced in the field was one. In chapter 4, the eGDD thermal index was computed from the same weather data (only one weather station for the ranch or the vineyard). However, this data could reliably represent the weather conditions of some blocks and present a bias for other blocks. In the second case, the eGDD thermal index computed would also integrate this bias. Another example is the posterior distribution of the probability of any instant t to be in the support of the BLiSS estimator of the β function, when analyzing the periods of weather influence on yield development in chapter 5. This distribution could well represent the difference of the vine response to the weather variables between blocks. It could also represent the fact that the same value of the weather covariate could lead to different vine response because this value actually corresponds to different field conditions between blocks. Another source of variability is the uncertain quality of farm data sets, as presented in the previous section.

In this way, there is a shift between the way that experimental results and operational results from the analysis of farm data sets can be interpreted. This is not an issue when identified because this is how the farmers have to interpret their data in the field. However, there is an interest in formalizing and handling this uncertainty which is usually managed in an expert way by farmers. This would support the field decision-making in a more reproducible and traceable way. This can be achieved using a forecasting approach under a Bayesian framework.

6.4 How will the results of the PhD project be integrated into a grape yield forecasting model ?

Grape yield will be modeled using the hierarchical Bayesian model summarized in Fig. 6.3. This implies that all the model parameters are considered as random variables for which a probability distribution can be estimated. A priori distributions (priors) will be chosen for each parameter and these distributions will be updated by taking into account the field observations. Consequently, the yield performance will also be modeled as a probability distribution. The most probable yield values will be considered as the expected yield performances and the dispersion of the distribution will indicate the uncertainty associated with this forecast. The yield performance is required to have a positive probability distribution with a single mode. Therefore, a Gamma distribution is chosen to model yield with a shape parameter superior to 1, noted a, and a positive scale parameter, noted b. This choice offers the advantage of easily recovering the yield expectation and the variance, respectively equal to $a * b$ and $a * b^2$. a and b are modeled as the exponential of a linear combination involving several candidate covariates. These covariates are taken among the essential variables : blocks characteristics, yield components, historical yield data and also include site-specific weather indicators as they were identified by the BLiSS method as proposed in this PhD work. Two random effects will also be added, one corresponding to a year effect and the other to a plot effect.

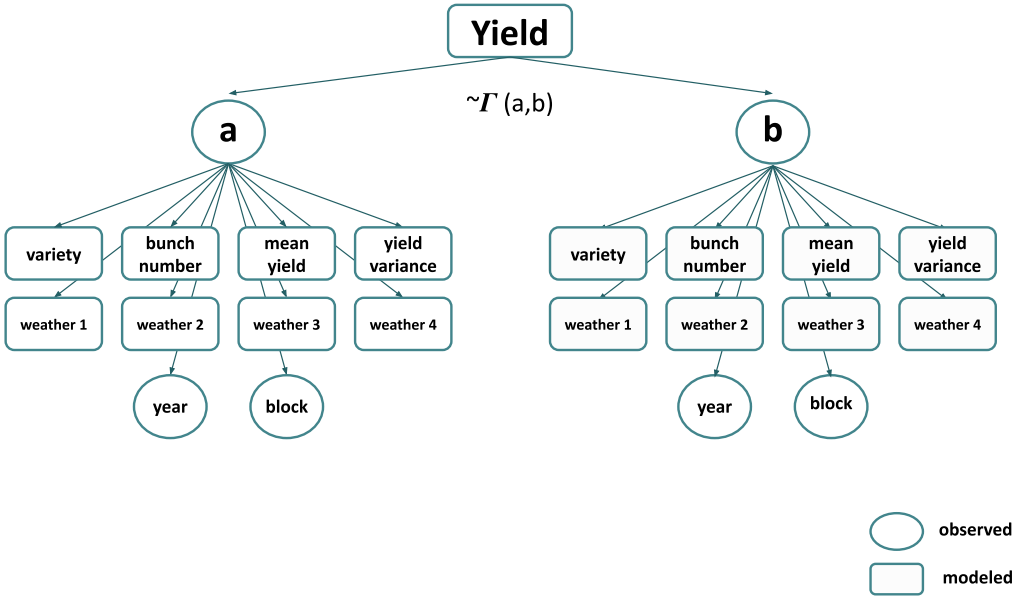

Figure 6.3 – Preliminary representation of a hierarchical Bayesian model for grape yield forecasting. Yield is considered to follow a Gamma law of parameter a and b. These a and b parameters are modelled according to covariates referring to the variety, the mean number of bunches per vine, the mean and variance of previous yield performances for the considered block, weather indicators and two random effects (year and block effects).

The priors of covariates parameters will all correspond to normal distributions centered at 0. This corresponds to the prior assumption that the covariates effect on the yield distribution can be positive or negative. The variance parameters of each covariate will be calculated in relation with the magnitude of the covariate so that all covariates have the same magnitude in their a priori effect on the yield response.
A model selection will be conducted using the Bayes Factor and a criterion such as Bayesian information criterion (Schwarz 1978) or Widely Applicable Bayesian Information Criterion (Watanabe 2013). This will lead to the selection of the most influential covariates. The numerical estimation of the model parameters will be performed with a Gibbs Sampler.

This modelling process will be used each time a yield forecast will be released. The forecast dates will be positioned at the end of a period of weather influence, as detected during the PhD project, or when a new annual indicator e.g. a yield component, will be collected.

6.5 How could the PhD proposal be further assessed ?

The most immediate work perspective is to develop and implement a Bayesian model for grape yield forecasting. Once this is done, the challenge/issue will be to evaluate properly the proposed method for grape yield forecasting in an operational context. This is not trivial because it has to be compared to a large variety of empirical methods used in the field (Clingeleffer et al. 2001; De la Fuente et al. 2015) but it should also be evaluated in cases where no method has effectively been implemented. **Moreover, the form in which the yield forecast is delivered, i.e. a probability distribution of expected yield performances, is different from the single value classically announced. Finally, the forecasting precision should be the first criterion of evaluation but there are others, especially in relation with the adoption of the method in the operational routine of commercial vineyards/wineries.**

To address this issue, the evaluation should be performed in two stages. **The first stage would evaluate the enabled improvement of grape yield forecasting in terms of precision.** The dispersion of the yield forecast (i.e. of the expected yield distribution) should then be compared to the dispersion of the

observed (and recorded) yield performances for each vineyard. Then the yield forecast should be compared to the result of an empirical method typically used in the field as described by De la Fuente et al. 2015. Such a model could correspond to a linear regression of the mean number of bunches per block against the yield performance. The mean number of bunches per block could be chosen because it corresponds to an essential indicator unlike other yield components such as bunch weight. The most probable value taken from the yield forecast should be compared to the empirical yield prediction. The dispersion of the yield forecast should also be compared to a forecast interval surrounding the empirical prediction. However, it should be noted that such an interval is never calculated in the field. Another evaluation of the forecast could consist in establishing a recall proportion i.e. the proportion of cases when the real yield performance is within an interval at 30, 50 or 75% of the forecast distribution.

The second stage of the method evaluation should address its operational relevance. In principle, the method implementation should not require more time for the measurement and sampling of indicators since these are analyzed as they are usually collected. Therefore, the second evaluation would be more focused on users satisfaction with the forecast precision in relation with the effort put into the data collection. However, the data quality is a strong determinant of the method results and this contribution cannot really be distinguished from the method itself for a given data set. Would the user continue to use the yield forecasting method and even improve the way he/she collects data if the results are not satisfying ? This second evaluation is indeed difficult to perform without testing in real conditions by professionals of the wine industry. However, customer satisfaction can also come from the rationalization of the decision-making of certain operations, in particular in the vineyard. In fact, connecting a yield indicator with other agronomic variables such as the vine water status can allow the formulation of more advanced viticultural recommendations in relation with the production objectives, which is Fruition Sciences' expertise.

6.6 How far can the PhD case study be generalized to the agronomic approach for operational purposes using farm data?

6.6.1 How far is generalizable to other variables than grape yield ?

Forecasting another grapevine variable implies working with the same data sets which were studied in this PhD project. The essential variables and their corresponding indicators should remain unchanged e.g. blocks characteristics, time series of weather data or yield performances. Therefore, the conceptual model of any other grape variable should be designed in the same way, the weather time series should be analyzed in the same way and a Bayesian model could also be implemented. However, the relevance of the new method should depend on the correlation between the essential variables, in particular the weather variables, and the variable to be modelled. Yet, most of the agronomic variables to be forecast are somehow correlated to the essential variables because these represent input variable (Lamanda et al. 2012).

A subsequent question is: would these additional variables be sufficient to forecast the variability of response with a relevant precision ? In particular, a better integration of the effect of cultural practices in the forecasting model might be requested. This an area for further improvement for forecasting grape yield or any other grape variable. However, the traceability of cultural practices often remains insufficient. Taking the example of bunch thinning, a quantitative indicator corresponding to the number of left bunches could be built. Still, the indicator must often be downgraded by setting only two categories: thinned and unthinned, which is not likely to improve the yield forecast.

6.6.2 How far is generalizable to other crops than grapevine ?

The general method proposed to perform agronomic research from farm data, starting from the transcription of the operational needs, constraints and data sets, should be applicable to other crops. However, it would be more relevant for perennial crops, whose yield determinants are less understood compared to annual crops and whose physiology is characterized by numerous trajectory and memory effects to be addressed by time series analyses.

In addition, the conceptual model designed for grape yield development corresponds to a vineyard-specific system, which is estimated as a black blox and which is modulated by block-specific effects. In this PhD, the block specific effect included the plant material and its interaction with the environment. In the case of grapevine or other perennial crops, the environment can be considered as partly constant because it corresponds to the same blocks year after year. However, this is not true for crops in rotation. Thus, it is still

possible to apply the eGDD or the BLiSS method for all years of a given crop but it should be considered that the results would incorporate the differences between blocks. This would be better managed in the forecasting model since several "variety" or "crop" and "block" effects could be introduced. However, the management of the interactions between these effects should be designed or voluntarily left aside.

Chapter 7

Résumé étendu de la thèse

7.1 Chapitre I : Introduction, problématique et structure du document

L'agronomie vise à améliorer la compréhension des mécanismes des agrosystèmes et à soutenir le processus de prise de décision sur le terrain, en fournissant des informations et des règles de décision pertinentes à différentes échelles spatiales et temporelles (Damour, Navas, and Garnier 2018). Dans le but de générer des connaissances sur les agrosystèmes, de nombreux modèles ont été construits en intégrant les résultats des expériences de recherche (Lamanda et al. 2012; Doré et al. 2006).**Cependant, peu de ces modèles ont été conçus pour être fonctionnels dans un contexte opérationnel c'est-à-dire pour répondre aux besoins et contraintes de terrain et fonctionner à partir de données agricoles. Aussi y a-t-il un fort enjeu à aborder cette problématique pour soutenir le processus de prise de décision sur le terrain.**

Premièrement, la diversité des agrosystèmes réels dépasse celle étudiée par la recherche, ce qui peut remettre en cause la validité des modèles proposés. **Or, cette diversité est susceptible de générer des connaissances sur l'évolution du rendement dans une grande variété de situations.** Deuxièmement, les données disponibles dans les exploitations agricoles (données agricoles) ne sont pas nécessairement adaptées aux paramètres couramment utilisés dans les modèles agronomiques. Les données agricoles sont généralement caractérisées par une qualité de données inférieure à celle des données issues des expériences de recherche. Cela empêche une compréhension très précise des phénomènes agronomiques mais correspond à un degré d'incertitude avec lequel la prise de décision opérationnelle doit composer. De plus, les données agricoles sont souvent caractérisées par un nombre parcimonieux de variables. En effet, les cycles longs (un ou deux ans) et la charge de travail élevée qui caractérisent le travail agricole créent un potentiel de changement des pratiques relativement contraint et lent. Par ailleurs, le développement actuel de nombreuses technologies de mesure concerne souvent des données auxiliaires ou celles-ci ne sont pas encore opérationnelles sur le terrain (Weiss, Jacob, and Duveiller 2020; Laurent, Oger, et al. 2021; Tardaguila et al. 2021). Le nombre de variables susceptibles d'être utilisées comme paramètres de modèles agronomiques évolue donc peu. **Dans le même temps, les jeux de données agricoles contiennent souvent de longs historiques de données dont le potentiel informationnel est encore peu exploité par les travaux de recherche.** Troisièmement, les contraintes et les attentes opérationnelles ne sont souvent pas prises en compte dans la manière dont l'information agronomique est construite et délivrée sur le terrain. **Le contexte opérationnel appelle donc à repenser les modèles conceptuels et les méthodes statistiques lorsque la recherche agronomique vise à soutenir la prise de décision sur le terrain plus qu'à générer des connaissances.**

Pour répondre à cette problématique, un pan de recherche agronomique davantage orienté vers une application opérationnelle a été développé, souvent en collaboration avec des entreprises privées. Fruition Sciences est l'une de ces entreprises. Elle est basée en Californie et en France. Elle aide les vignobles à comprendre et à réagir aux conditions du millésime en fonction de leur terroir grâce à l'analyse de leurs propres données. L'expertise de Fruition Sciences consiste principalement en une lecture de la saison en fonction de l'état hydrique de la plante. Cet indicateur est ensuite confronté à d'autres variables agronomiques telles que l'expression végétative ou le rendement pour fournir des recommandations viticoles. Les domaines viticoles clients de Fruition Sciences sont situés dans le monde entier et correspondent donc à des

situations opérationnelles très différentes en termes d'environnement, de système de culture, de contexte socio-économique et d'objectifs de production. **Le présent projet de doctorat a été initié par Fruition Sciences** via un financement CIFRE et s'est donc développé sur l'exemple international de ces vignobles internationaux. Cette thèse fait l'hypothèse que la recherche agronomique peut fournir des informations locales et actionnables pour la prise de décision sur le terrain si elle se base sur l'analyse de données agricoles et se conforme aux contraintes et attentes opérationnelles. La filière viticole est intéressante à considérer dans le cadre de cette hypothèse car il s'agit d'un secteur relativement conservateur, pour lequel de nombreuses décisions sont encore prises de manière experte. En revanche, d'autres cultures comme le blé, le maïs ou le soja bénéficient d'un grand nombre de modèles pour appuyer leurs décisions de gestion (Dury et al. 2012; Silva and Giller 2020; Young, Ros, and de Vries 2021). De plus, le caractère pérenne de la vigne encourage la constitution de jeux de données sur plusieurs années pour des parcelles fixes. On suppose donc que les effets de l'année et de la parcelle dans les processus agronomiques sont plus facilement identifiables dans les jeux de données viticoles que dans des jeux de données incluant des parcelles en rotation.

Le projet de thèse s'est concentré sur la prévision du rendement en viticulture comme étude de cas. La prévision du rendement est en effet l'un des indicateurs de performance utilisés pour décider de la plupart des opérations dans le vignoble, dans le chai ainsi que dans la gestion commerciale et économique de diverses organisations viticoles, vinicoles et de négoce. C'est également une pierre angulaire de la recherche agronomique et écophysiologique en viticulture. Il existe de nombreuses publications récentes sur l'estimation, la prédiction et la prévision des rendements (Nuske, Wilshusen, et al. 2014; Nogueira Júnior et al. 2018; Sirsat et al. 2019; Zhu, Fraysse, et al. 2020; Arab et al. 2021) mais peu d'entre elles abordent réellement le contexte opérationnel, c'est-à-dire les besoins, les contraintes et les données des domaines en production. **Par conséquent, les questions de recherche du projet de doctorat se sont concentrées sur l'utilisation des données agricoles pour fournir des prévisions de rendement pertinentes pour la prise de décision opérationnelle. Il s'agit d'identifier et de développer un modèle conceptuel d'élaboration du rendement, de concevoir une méthode de prévision du rendement, d'employer des méthodes statistiques appropriées et de proposer une mise en œuvre pouvant être adaptée à la diversité des situations opérationnelles en viticulture.**

Pour cela, une revue de la littérature scientifique et des connaissances de terrain a d'abord été réalisée afin d'identifier les enjeux, les méthodes et les défis de la prévision du rendement viticole dans un contexte opérationnel (chapitre 2). Cet état de l'art a permis de déterminer les questions de recherche et la démarche scientifique du projet de thèse, qui sont également présentés dans le chapitre 2. Le chapitre 3 présente les premiers résultats du projet de thèse : i) les jeux de données collectés et leur caractérisation, ii) un modèle conceptuel de l'élaboration du rendement dans un contexte opérationnel, iii) la conception de la méthode de prévision du rendement correspondante et iv) la présentation des verrous scientifiques traités dans la thèse pour mettre en œuvre la méthode de prévision du rendement à partir de jeux de données agricoles. Les travaux relatifs à ces verrous scientifiques font l'objet des chapitres 4 et 5. Enfin, le chapitre 6 discute de la contribution et des perspectives de ce travail de thèse à la prévision du rendement en viticulture, mais aussi à la valorisation des données agricoles pour la recherche agronomique en général.

7.2 Chapitre II : Bibliographie et présentation de la démarche de recherche

7.2.1 Etat de l'art des questions, méthodes et perspectives pour l'estimation, prédiction et prévision de rendement en viticulture (Laurent et al., 2021)

Le rendement viticole correspond le plus souvent à la quantité de vendange exprimée en masse de raisins ou volume de vin par unité de surface pour un cycle cultural. L'élaboration du rendement est un processus dynamique et complexe, en interaction avec de nombreux facteurs dont le climat, le sol, les agressions biotiques ou les opérations culturales. Ainsi, pour une qualité de vendange, il existe non pas un seul rendement fixé par défaut mais une gamme de rendement possibles. C'est pourquoi la prévision de rendement permet, au cours de la saison, de décider des opérations au vignoble adéquates afin d'optimiser le rendement final et ce, pour une qualité fixée par les objectifs de production. Ces opérations incluent par exemple la fertilisation, l'irrigation, la manipulation de la canopée et l'éclaircissage. La prévision de rendement est aussi utile à d'autres acteurs de la filière viti-vinicole : elle permet d'anticiper la logistique humaine et matérielle des vendanges et des vinifications, elle constitue une information clef pour les services de commer-

cialisation et de gestion comptable des domaines et structures de négoce. A l'échelle du territoire, la prévision de rendement permet également de gérer les volumes produits sur l'appellation et de fixer les réglementations en conséquence. Les divers acteurs de la filière ont des besoins différents en termes de prévision, notamment en termes de date et d'échelle spatiale auxquelles la prévision est formulée mais aussi d'unité d'expression du rendement. Ces différents acteurs ont aussi des contraintes opérationnelles, notamment liées aux moyens humains et financiers pour déployer la méthode de prévision et aux données disponibles sur le terrain. Ces contraintes rendent difficiles l'utilisation de la plupart des méthodes proposées dans la littérature scientifique en contexte opérationnel. **Par ailleurs, les propositions de la littérature ne concernent souvent qu'une étape du processus de prévision de rendement à savoir i) l'estimation de variables par des approches de mesure et d'échantillonnage et ii) la modélisation du rendement final grâce aux variables explicatives préalablement estimées.**

En conséquence, les propositions de la littérature scientifique ne permettent actuellement pas de satisfaire les attentes de la filière en termes de prévision de rendement. **Dans le cadre de la thèse, il est considéré que l'intégralité du processus de prévision de rendement doit être étudié pour pouvoir être amélioré.** Ainsi, en confrontant les méthodes proposées pour chaque étape de la prévision de rendement au processus d'élaboration du rendement, aux besoins et aux contraintes en contexte opérationnel, trois pistes encore peu explorées ont été identifiées pour leur potentiel d'amélioration. **Il s'agit de développer une méthode i) pertinente opérationnellement c'est-à-dire adaptée aux besoins de la filière en termes de date, d'échelle et d'unité de prévision et fonctionnant à partir des données opérationnelles, ii) qui prenne davantage en compte la temporalité de l'élaboration du rendement pour mieux exploiter les séries temporelles de données souvent présentes dans les données opérationnelles et mieux gérer l'incertitude liée à l'évolution future du rendement pendant le reste de la saison et iii) qui considère un schéma local d'élaboration du rendement pour une plus grande précision du modèle de prévision, valable pour le domaine uniquement.**

7.2.2 Présentation des questions de recherche abordées dans la thèse

A travers l'étude de cas de la prévision du rendement en viticulture, l'article précédent a permis de formaliser la démarche de recherche agronomique en trois étapes : la mesure et l'estimation d'indicateurs agronomiques et la modélisation d'une variable réponse basée sur ces indicateurs. Il a révélé que la plupart des travaux de recherche visant une application opérationnelle se positionnent sur les questions de mesure et d'échantillonnage et peu sur celles de modélisation. Il existe donc une opportunité importante de travailler sur des approches de modélisation pour améliorer la prévision du rendement dans un contexte opérationnel. En outre, la revue de la littérature a mis en évidence le potentiel de l'exploitation des données agricoles à des fins opérationnelles. Il existe deux types de données agricoles : celles qui ont déjà été collectées, dont la mesure, l'échantillonnage et même la traçabilité sont imposées et ne peuvent être changées, et celles qui seront collectées dans le futur, pour lesquelles il est possible de décider des processus de mesure et d'échantillonnage. **Le projet de thèse se concentre sur les données agricoles déjà collectées.**

Les défis concernant le processus de modélisation basé sur les données de l'exploitation et à des fins opérationnelles ont été repérés par l'état-de-l'art. Ces défis ont été classés en trois catégories : i) la recherche de la pertinence opérationnelle, ii) la prise en compte de la temporalité du développement du rendement et iii) le développement d'une approche spécifique au site du développement du rendement. Les questions scientifiques du projet de thèse portent donc sur la mise en œuvre d'une recherche agronomique répondant à ces trois exigences avec l'exemple de la prévision du rendement en viticulture : **i) quel modèle conceptuel d'élaboration du rendement de la vigne pourrait répondre à une telle approche opérationnelle, temporelle et locale de l'élaboration du rendement ? ii) quelle méthode de prévision du rendement pourrait être mise en œuvre pour tenir compte d'un tel modèle ? iii) quelles approches de modélisation statistique pourraient être utilisées pour exploiter les données agricoles ? iv) comment mobiliser de telles approches statistiques dans la méthode de prévision du rendement viticole ?**

Une mise en correspondance des défis identifiés dans l'état-de-l'art avec les questions scientifiques de la thèse est proposée dans la Fig. 7.2.

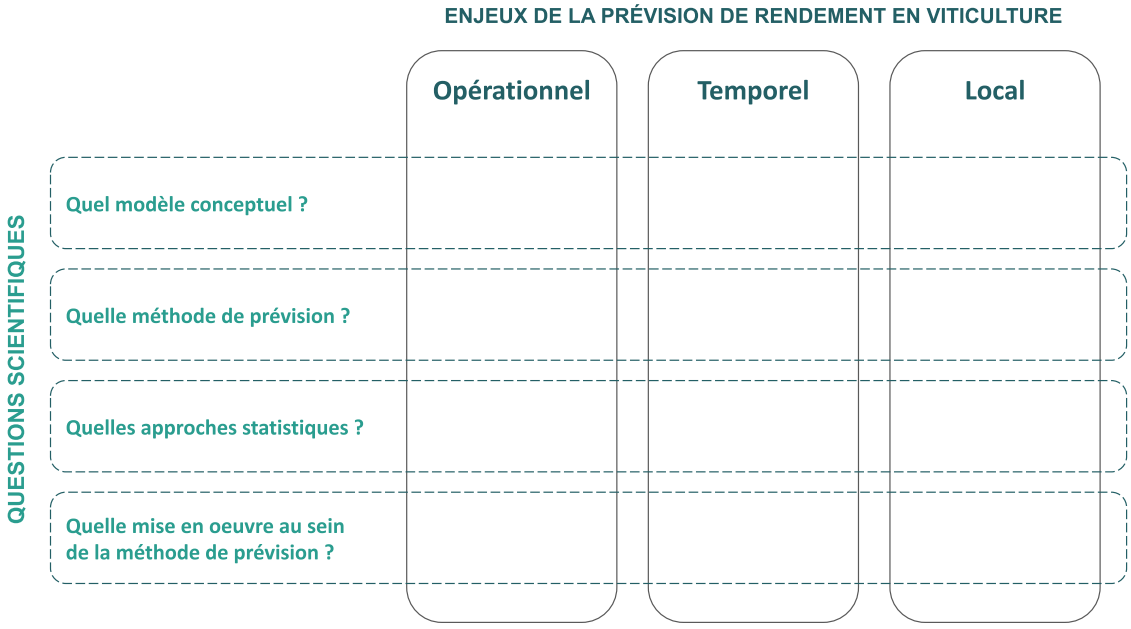

Figure 7.1 – Positionnement des axes de recherche en fonction de la méthode de prévision du rendement et organisation du manuscrit correspondante.

Pour répondre à ces quatre questions scientifiques, la démarche de recherche a été organisée en 5 étapes :

1. **Collecte de jeux de données agricoles et consolidation des retours d'expérience sur le terrain concernant la prévision du rendement viticole**
2. **Sélection des jeux de données pour les études de cas**
3. **Définition d'un modèle conceptuel pour l'élaboration du rendement**
4. **Conception d'une méthode de prévision du rendement viticole**
5. **Choix de l'approche statistique et mise en oeuvre des recherches associées à chaque étape de la méthode de prévision du rendement**

7.3 Chapitre III : Conception d'une méthode de prévision du rendement à partir de la caractérisation de jeux de données agricoles

7.3.1 Caractérisation des jeux de données agricoles et sélection de trois cas d'étude

Table 7.1 – Liste non-exhaustive des variables et indicateurs présents dans les jeux de données agricoles

Variable type	Indicateurs essentiels	Indicateurs optionnels	Echelle temporelle	Echelle spatiale
Caractéristiques parcellaires	Topographie	Exposition	constant	parcelle
	Surface	Caractéristiques du sol	constant	parcelle
	Cépage, porte-greffe		constant	parcelle
	Densité de plantation	Nombre de manquants	constant	parcelle
	Date de plantation		constant	parcelle
	Système de conduite		constant	parcelle
Rendement	Rendement final	Nombre de baies par grappe	annuel	parcelle
	Nombre de grappes par cep	Masse d'une grappe	annuel	parcelle
		Masse d'une baie	annuel	parcelle
Météo	Température	Rayonnement	journalier	vgoble
	Précipitations	Humidité relative	journalière	vignoble
Statut hydrique		Tension du sol	journalier ou infra-annuel	vignoble
		Potentiel hydrique	journalier ou infra-annuel	vignoble
	Flux de sève	journalier ou infra-annuel	vignoble	
Nutrition		Analyses de sol	annuel	parcelle
		Analyses pétiolaires	annual	parcelle
Phénologie	Débourrement, Floraison	Autre stade phénologique	annuel	parcelle
	et Véraison	Croissance des rameaux	annuel	parcelle
Pratiques culturales		Taille(date et intensité)	annuel	parcelle
		Eclaircissage (date et intensité)	annuel	parcelle
		Enherbement (surface, date de destruction)	annuel	parcelle
		Fertilisation (dates et doses)	infra-annuel	parcelle
		Irrigation(dates et volumes)	infra-annuel	parcelle

7.3.2 Développement d'un modèle conceptuel pour l'élaboration du rendement adapté au contexte opérationnel

On fait l'hypothèse que les données présentes dans tous les jeux de données (données dites essentielles) pourraient être suffisantes pour alimenter une méthode de prévision du rendement pertinente sur le plan opérationnel. Cette hypothèse implique que tout vignoble possédant des données essentielles pourrait avoir accès à une prévision de rendement, dont la précision dépendrait alors du volume et de la qualité de son jeu de données. Un modèle conceptuel d'élaboration du rendement a donc été proposé sur la base du jeu de variables essentielles préalablement défini. Il est illustré en Fig. 7.2

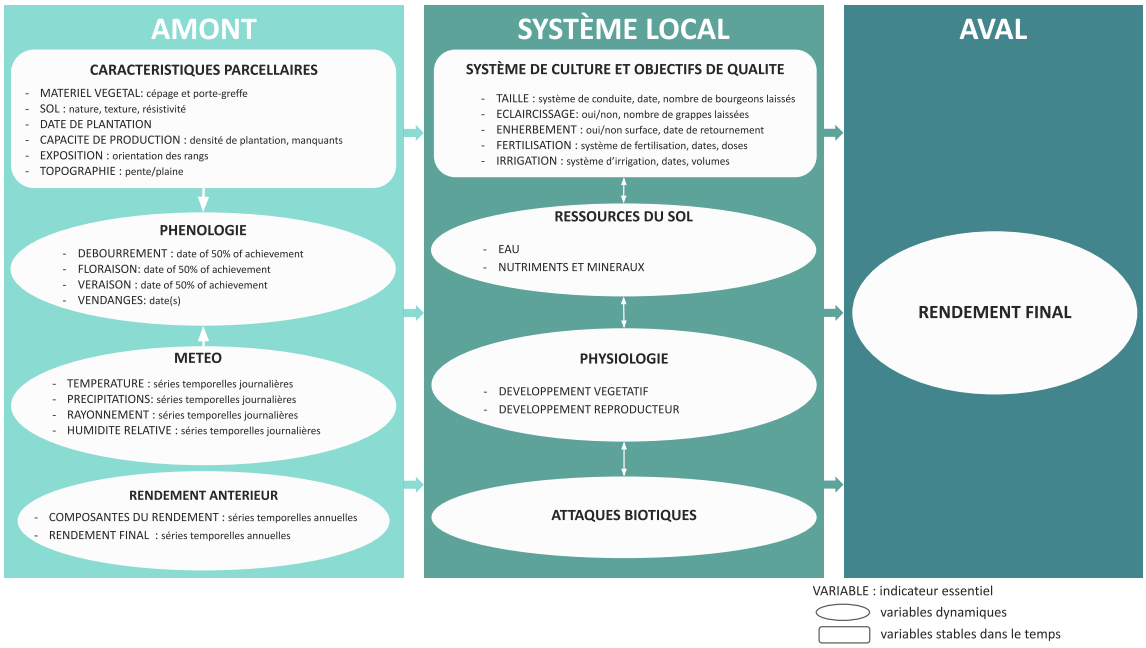

Figure 7.2 – Modèle conceptuel de l'élaboration du rendement d'après la bibliographie. En contexte opérationnel, le système local est considéré comme une boîte noire.

Le modèle conceptuel du rendement est établi à l'échelle parcellaire. Cependant, une partie du fonctionnement de la parcelle est commune à toutes les parcelles du vignoble de par i) une certaine cohérence des influences environnementales à l'échelle du vignoble et ii) la plupart des déterminants du système de culture sont généralement raisonnées à l'échelle du vignoble ou de groupes de parcelles. **On peut donc considérer qu'une partie de l'élaboration du rendement suit un schéma spécifique au vignoble et que ce dernier est modulé par des variables parcellaires.** La plupart des variables essentielles sont des variables dites amont c'est-à-dire qu'elles influencent l'élaboration du rendement sans subir rétro-influence (Lamanda et al., 2012). Ce sont aussi des variables qui caractérisent l'échelle parcellaire. **On peut donc estimer l'effet vignoble dans l'élaboration du rendement comme une boîte noire spécifique au vignoble, alimentée par des variables parcellaires et étalonnée par apprentissage de données.**
Les variables impliquées dans le modèle peuvent être différentes d'un vignoble à l'autre en fonction des données disponibles. Les indicateurs qui informent chaque variable peuvent aussi différer dans leur protocole de mesure et d'échantillonnage. Ils peuvent donc varier dans leur définition et unité en fonction du vignoble. Par exemple, le rendement peut être mesuré en tonnes par acre avant ou après le transport de la vendange ou elle peut être mesurée en hectolitres par hectare après pressurage ou après soutirage. Les données météorologiques peuvent également être mesurées par plusieurs stations situées à l'intérieur du vignoble ou par une seule station située à l'extérieur du vignoble.

7.3.3 Développement d'une méthode de prévision du rendement viticole en contexte opérationnel

Pour prendre la temporalité de l'élaboration du rendement en compte, la méthode envisagée doit permettre de produire **plusieurs prévisions de rendement judicieusement positionnées dans la saison.** Chacune d'entre elles ferait intervenir un modèle différent mais toujours spécifique au vignoble, basé sur les données disponibles c'est-à-dire celles qui ont déjà été collectées à ce moment de la saison.
Pour que cette prévision du rendement soit une aide pertinente pour la décision opérationnelle sur le terrain, elle doit être accompagnée d'une caractérisation de sa fiabilité. En d'autres termes, cette prévision doit bien être une prévision, pas une prédiction ou une estimation du rendement. **Celle-ci est exprimée par une distribution des rendements probables au lieu d'une valeur prédite unique.**
Une illustration du rendement attendu pour une parcelle donnée est présentée dans la Fig. 7.3.

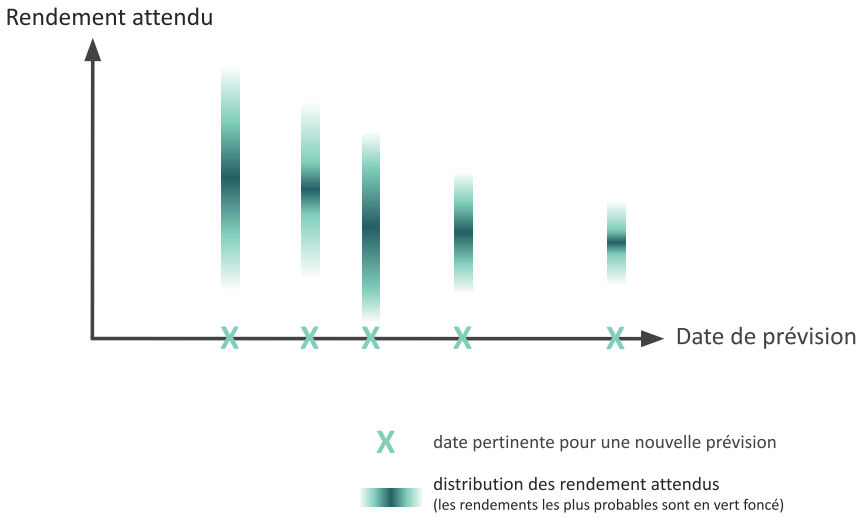

Figure 7.3 – Cahier des charges de la méthode de prévision de rendement attendue : elle doit fournir une distribution des rendements probables en fonction de la date de prévision

7.3.4 Présentation des verrous scientifiques explorés dans le projet de doctorat

Les données essentielles doivent être préalablement traitées avant d'être intégrées comme variables explicatives dans un modèle de prévision du rendement. En particulier, une caractéristique commune des jeux de données agricoles étudiés est la présence de séries temporelles de données. Ces séries temporelles peuvent être classées en deux catégories : i) les séries de données avec un pas de temps annuel i.e. qui ont une observation par an, par exemple les séries temporelles de rendement ou d'observation des composantes du rendement et ii) les séries avec un pas de temps infra-annuel i.e. qui ont de nombreuses observations tout au long de la saison, par exemple les séries temporelles de données météorologiques (cf. Tableau 7.1). Les séries de données météorologiques sont intéressantes car elles fournissent des informations continues et permettent donc de mettre en œuvre plus de modèles de prévision de rendement que d'autres indicateurs, qui ne sont récoltés qu'une fois dans la saison. Cependant, elles ne sont pas triviales à analyser pour plusieurs raisons. La première raison est que ces séries sont collectées simultanément à l'élaboration continue du rendement. Leur comparaison dans le cadre de toute analyse implique donc que le rythme d'élaboration du rendement soit le même d'une série à l'autre, c'est-à-dire d'une parcelle à l'autre ou d'une année à l'autre. On sait que ce n'est pas le cas lorsque le temps est exprimé selon le calendrier Grégorien (Wang 1960; Cross and Zuber 1972; Grigorieva, Matzarakis, and Freitas 2010). Les séries temporelles de données météorologiques doivent donc être synchronisées en fonction du rythme d'élaboration du rendement pendant ou après leur collecte. En outre, il est supposé que l'élaboration du rendement suit un schéma local, spécifique au domaine voire à la parcelle. La méthode de synchronisation doit donc tenir compte de cet effet de site. **Un premier axe de recherche a donc porté sur la synchronisation locale de séries de données météorologiques. Il est abordé dans le chapitre 4.**

La deuxième raison pour laquelle les données de séries temporelles de données météorologiques ne sont pas faciles à analyser est qu'elles contiennent beaucoup d'informations, plus ou moins pertinentes par rapport à l'élaboration du rendement et aussi beaucoup de bruit. Or, l'information extraite de la série temporelle est destinée à être intégrée dans un modèle de prévision du rendement avec d'autres variables. Elle doit donc être sélectionnée avec parcimonie afin de ne pas systématiquement sur-ajuster le modèle de prévision du rendement. Enfin, une troisième raison est liée au fait que les séries temporelles ne peuvent pas être analysées avec les méthodes classiques d'analyse statistique en raison de leur autocorrélation. Ainsi, **un deuxième axe de recherche porte sur la valorisation des données de séries temporelles de données météorologiques une fois qu'elles ont été synchronisées en fonction du rythme d'élaboration du rendement. Cet axe de recherche est traité dans le chapitre 5 et aboutit à la**

construction d'indicateurs météorologiques.

Les deux axes de recherche abordés au cours de la thèse sont présentés en lien avec la mise en œuvre de la méthode de prévision du rendement dans la Fig. 7.4. L'organisation du manuscrit correspondant est également mentionnée.

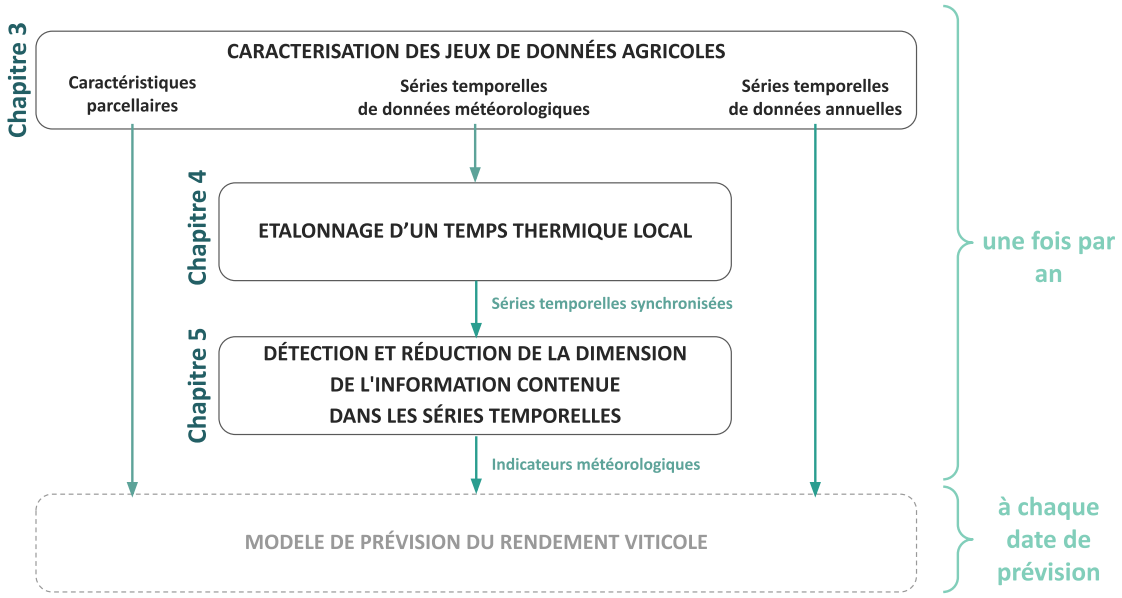

Figure 7.4 – Positionnement des axes de recherche traités dans le cadre la thèse en fonction de la méthode de prévision des rendements et organisation du manuscrit correspondante. Le modèle en gris clair correspond à une perspective de travail.

7.4 Chapitre IV : Etalonnage d'un temps local exprimé en fonction d'un indice thermique pour effectuer des analyses des données en cohérence avec la phénologie du vignoble

7.4.1 Objectif du chapitre

La température correspond à l'un des principaux facteurs de terroir influençant la physiologie de la vigne et a fortiori sa phénologie. Des indices thermiques ont donc été développés pour prédire la date de réalisation des stades phénologiques clés (Zapata, Salazar-Gutierrez, et al. 2015; Zapata, Salazar-Gutierrez, et al. 2017; Parker, De Cortázar-Atauri, et al. 2011; Fila et al. 2014). Ils ont ensuite été utilisés pour synchroniser des séries de données temporelles afin d'étudier l'effet d'une variable environnementale sur une réponse physiologique de la vigne (Guilpart, Metay, and Gary 2014). En outre, plusieurs études ont suggéré l'intérêt de calibrer ces indices thermiques de manière locale afin de prendre en compte l'interaction spécifique au site des nombreuses variables qui déterminent la réponse de la vigne à la température en termes de phénologie.

On fait donc l'hypothèse que l'utilisation d'un indice thermique local pour exprimer le temps améliore la synchronisation des séries temporelles de données et réduit ainsi le bruit dû à leur décalage phénologique pour des analyses ultérieures. Dans le cadre de cette thèse, on fait aussi l'hypothèse de la faisabilité de la construction d'un tel indice thermique à partir de données agricoles. Enfin, dans la littérature, un indice thermique donné peut implicitement être utilisé à des fins de Prédiction ou de Synchronisation sans distinction. Ainsi, une troisième hypothèse de travail considère que **l'étalonnage local doit être guidé par l'utilisation prévue de l'indice thermique pour une meilleure performance.** Cette hypothèse a conduit à clarifier et à formaliser la différence entre une utilisation de Prédiction ou de Synchronisation d'un indice thermique donné.

Le travail de recherche présenté dans ce chapitre a été réalisé en deux étapes. Premièrement, les trois hypothèses de travail ont été explorées pour preuve de concept à travers une étude rapide impliquant l'étalonnage d'indices thermiques construits empiriquement (Laurent, Oger, et al. 2021). Deuxièmement, une méthode appelée eGDD pour Extended Growing Degree Days a été conçue et développée pour généraliser la construction d'un indice thermique local. Cette méthode est testée sur trois jeux de données agricoles. Elle fera l'objet d'un article en deux parties qui sera soumis au journal Agriculture and Forest Meteorology.

7.4.2 Présentation de la méthode Extended Growing Degree Days (eGDD)

La méthode eGDD propose quatre améliorations par rapport à l'approche classique des Degrés Jours (GDD, Zapata, Salazar-Gutierrez, et al. 2017): i) elle considère des seuils de température supplémentaires pour définir la relation entre la phénologie de la vigne et la température, ii) elle utilise une approche d'optimisation sous contraintes pour étalonner ces seuils de manière locale, iii) elle oriente le processus d'étalonnage en fonction de l'objectif d'utilisation de l'indice thermique (Prédiction ou Synchronisation) et iv) elle propose une intégration sur la saison de la fonction reliant lé phénologie de la vigne à la température (fonction PAST) sur la saison en y associant la photopériode. Un exemple de sorties de la méthode eGDD est donné en Fig. 7.5.

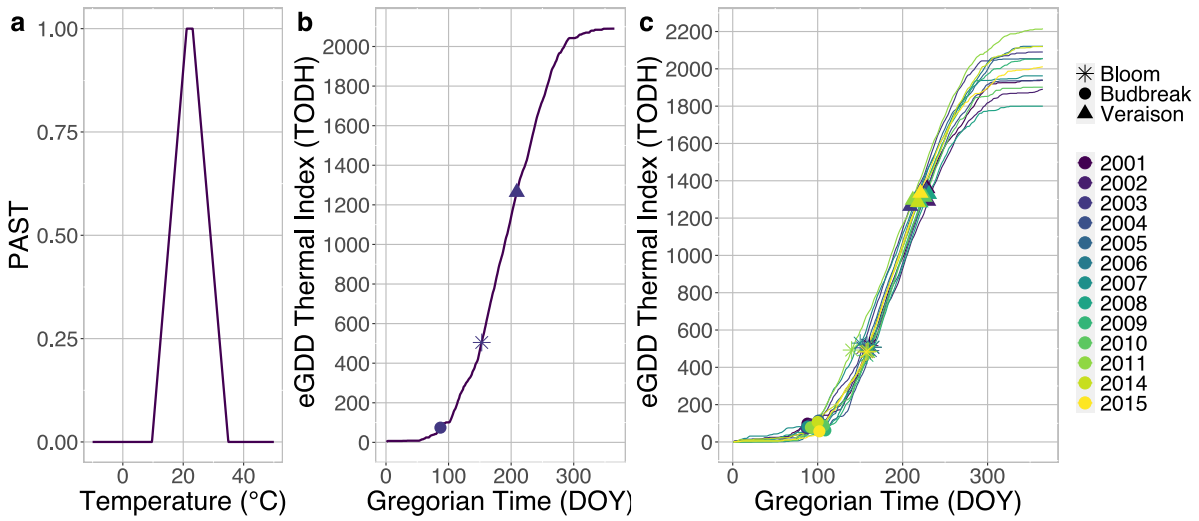

Figure 7.5 – Exemple des sorties de la méthode eGDD pour une parcelle avec (a) la fonction PAST optimisée, (b) l'indice thermique eGDD obtenu par intégration de la fonction PAST sur l'historique des températures de 2003 et (c) les historiques de température de toutes les années du jeu de données. Les dates de réalisation du débourrement, de la floraison et de la véraison exprimées dans l'indice thermique eGDD et dans le calendrier grégorien sont indiquées comme références.

7.4.3 Validation de la méthode eGDD

La méthode eGDD a été validée sur trois jeux de données agricoles, correspondant à trois domaines en production respectivement situés en Califonie , en Israël et en France (vignobles A, B et C). Les deux utilisations de la méthode eGDD (Prédiction ou Synchronisation) sont étudiées. Les résultats de la méthode eGDD sont comparés au calendrier Grégorien et à l'approche GDD classiquement utilisée dans la littérature.
Pour l'utilisation de la prédiction, la comparaison entre les différents calendriers a été effectuée selon une erreur quadratique moyenne croisée (RMSE) entre les dates prédites et observées de débourrement, de floraison et de véraison respectivement. La précision de la synchronisation a été évaluée en fonction du plus grand nombre de périodes pouvant être discrétisées dans chaque série temporelle tout en garantissant que le débourrement, la floraison et la véraison se produisent respectivement dans la même période pour toutes les années d'une même parcelle. Un grand nombre de périodes discrétisées signifie que les séries temporelles sont bien synchronisées.

La méthode eGDD a permis d'obtenir de meilleurs résultats de prédiction que le calendrier Grégorien ou que l'approche GDD classique. On peut en effet voir sur la Fig. 7.6 que la méthode eGDD permet d'obtenir les erreurs de prédiction parmi les plus faibles pour le débourrement, la floraison et la véraison et ce, pour les trois vignobles étudiées. Si un seul indice thermique doit être utilisé pour prédire les dates d'atteinte des trois stades phénologiques à la fois, la méthode eGDD apparaît ainsi comme le meilleur compromis pour les trois vignobles étudiés.

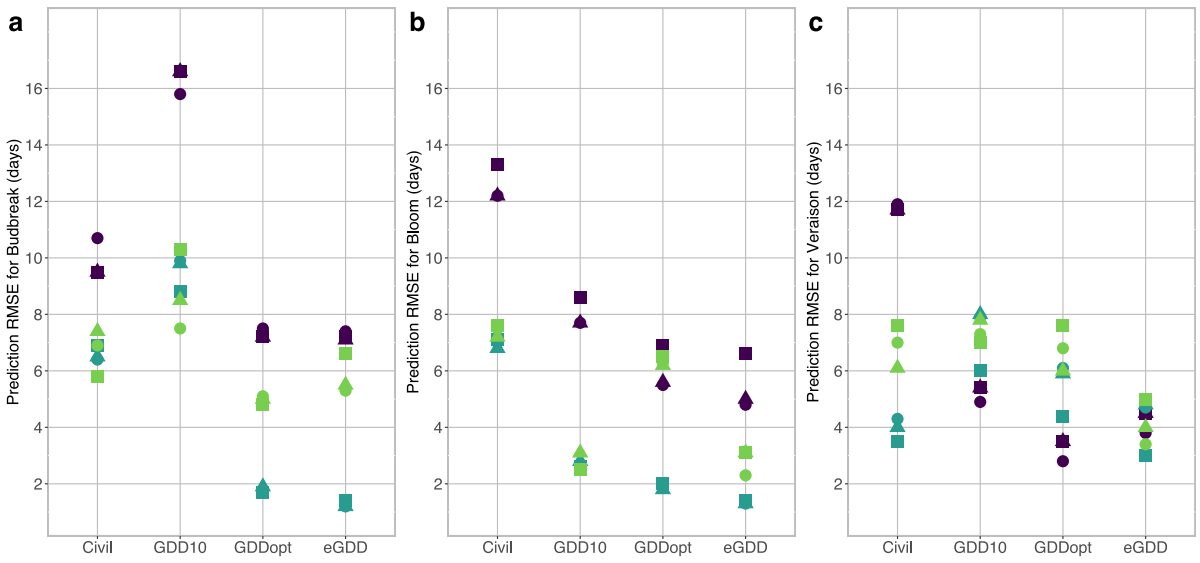

Figure 7.6 – Erreur quadratique moyenne (RMSE) de prédiction des stades phénologiques obtenue par validation croisée et exprimée en jours pour le débourrement (a), la floraison (b) et la véraison (c). Les RMSE obtenues avec le calendrier grégorien (Civil), l'approche des Degrés Jours avec une température de base de 10°C (GDD_{10}), l'approche Degrés Jours avec une température de base optimisée (GDD_{opt}) et l'approche eGDD sont comparées pour chaque parcelle des trois vignobles étudiés.

La méthode eGDD a aussi permis d'obtenir de meilleurs résultats de synchronisation que le calendrier Grégorien ou que l'approche GDD classique. On peut en effet voir sur la Fig. 7.7 que la méthode eGDD a permis de discrétiser le plus grand nombre de périodes pour les vignobles B et C. Ce résultat correspond à une amélioration significative par rapport à la meilleure des trois autres approches : le nombre de périodes a été augmenté de 11 en moyenne pour le vignoble B et de 6 en moyenne pour le vignoble C. Dans le cas du vignoble A, l'approche GDD_{opt} a donné un nombre de périodes supérieur de 1 ou 2 à celui de la méthode eGDD pour les blocs 1 et 2. La méthode eGDD a réalisé la meilleure synchronisation des séries temporelles pour le bloc 3.

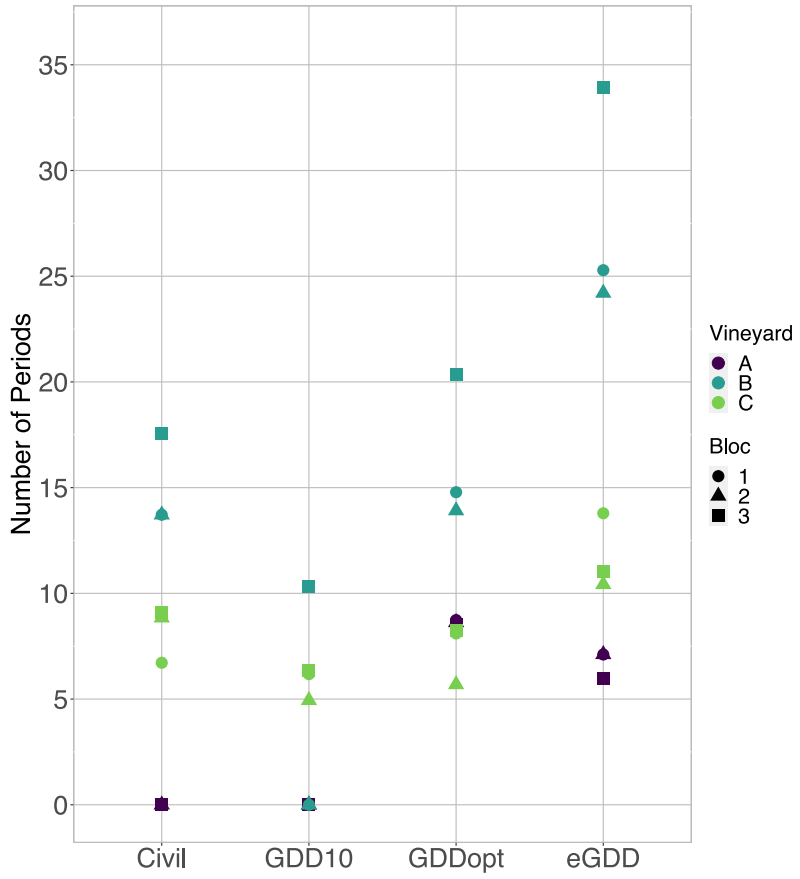

Figure 7.7 – Nombre maximal de périodes pouvant être discrétisées dans les séries temporelles de chaque parcelle en respectant la contrainte selon laquelle le débourrement, la floraison et la véraison de toutes les années doivent avoir lieu pendant la même période pour le calendrier grégorien (Civil), la méthode des Degrés Jours avec une température de base de 10°C (GDD_{10}), la méthode des Degrés Jours avec une température de base optimisée (GDD_{opt}) et la méthode eGDD avec utilisation de la synchronisation (eGDD) pour chaque parcelles des trois vignobles étudiés.

7.4.4 Définition d'une échelle spatiale d'importance primaire pour la mise en oeuvre de la méthode eGDD

Les indices thermiques obtenus avec la méthode eGDD étaient cohérents entre les parcelles d'un même vignoble mais différents entre les vignobles (Fig. 7.8). L'échelle du vignoble semble donc permettre la plus grande plue-value lors de la mise en œuvre de la méthode eGDD. Toutefois, ce constat doit être modulé lorsque des cépages différents sont étudiés.

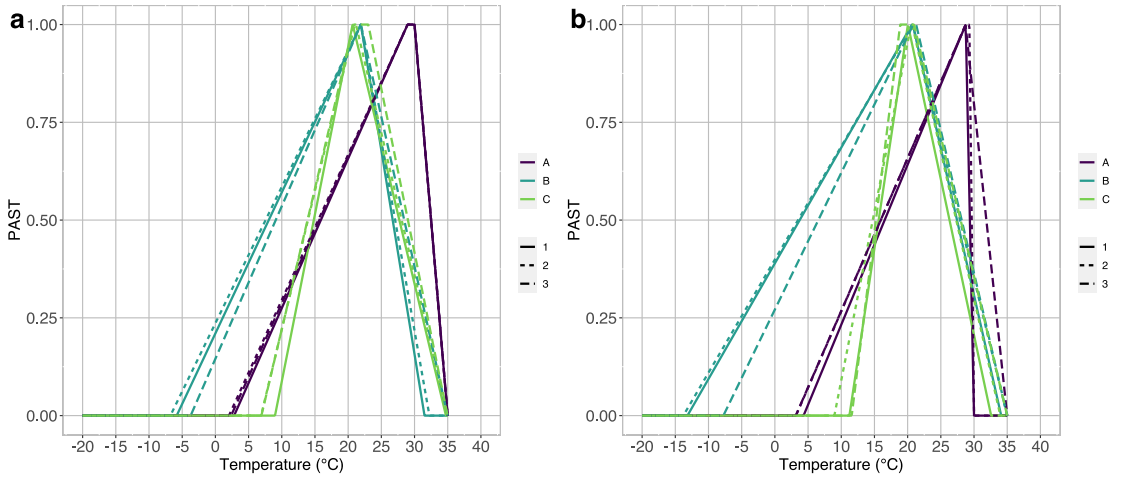

Figure 7.8 – Profils de la fonction PAST incluant les valeurs optimisées de la température de base (T_b), des températures optimales (T_o^1 et T_o^2) et de la température critique (T_c) obtenues avec la méthode eGDD pour chacun des jeux de données pour les utilisations Prédiction (a) et Synchronisation (b).

7.5 Chapitre V : Prise en compte de la temporalité de l'élaboration du rendement par l'analyse de séries temporelles de données météorologiques

7.5.1 Objectif du chapitre

Les séries temporelles sont définies comme un ensemble d'observations organisées séquentiellement dans le temps comme la réalisation d'un processus stochastique (Brockwell et David, 2009). Par conséquent, les observations ne peuvent pas être considérées comme des variables explicatives indépendantes pour des analyses classiques telles que les régressions linéaires multivariées. Pour contourner ce problème, la plupart des études bibliographiques se sont concentrées sur l'utilisation de variables météorologiques se résumant à quelques stades phénologiques clés connus (Buttrose 1974; Pouget 1981) ou à quelques périodes régulières mais grossières (Guilpart, Metay, and Gary 2014; Molitor and Keller 2017), qui peuvent donc être considérés comme indépendants. Cependant, ces approches classiques présentent des limites : i) elles dépendent de choix subjectif des périodes, ii) il est souvent nécessaire de supprimer des données ou de n'analyser que certaines parties d'une série temporelle et iii) l'autocorrélation des séries temporelles n'est pas prise en compte. De plus, l'information extraite de la série temporelle est destinée à être intégrée dans un modèle de prévision du rendement avec d'autres variables. Elle doit donc être sélectionnée avec parcimonie afin de ne pas systématiquement sur-ajuster le modèle de prévision du rendement. Enfin, l'influence de la météo sur le développement du rendement est supposée inclure à la fois des influences générales et spécifiques au site résultant d'interactions uniques entre de nombreux facteurs spécifiques au site dans chaque vignoble. L'objectif du chapitre 5 est donc de développer une démarche analytique à mettre en œuvre localement pour valoriser les séries temporelles de données météorologiques agricoles. Cette démarche doit permettre d'identifier avec parcimonie les périodes pendant lesquelles les variables météorologiques ont le plus d'influence sur le schéma d'élaboration local du rendement.

7.5.2 Présentation de la démarche analytique incluant les méthodes eGDD et BLiSS

La méthode Bayesian functional Linear regression with Sparse Step functions (BLiSS, Grollemund et al. 2019) vise à fournir une estimation bayésienne du support de la fonction de coefficient d'une régression linéaire fonctionnelle sur une réponse scalaire. En d'autres termes, elle vise à détecter les périodes où une variable fonctionnelle, telle qu'une série temporelle météorologique prise en fonction du temps, influence une réponse scalaire telle que le rendement viticole. Une telle méthode pourrait donc être utilisée pour analyser les séries temporelles de données météorologiques de manière appropriée par rapport à leurs caractéristiques

et extraire des indicateurs météorologiques pertinents (variables météorologiques sur des périodes données) et parcimonieux de l'élaboration du rendement. Elle est mise en œuvre sur les séries temporelles de données météorologiques de trois vignobles, préalablement synchronisées grâce à la méthode eGDD.

La Fig. 7. 9 présente un exemple de détection de périodes pendant lesquelles la température moyenne journalière influence l'élaboration du rendement. Ces périodes correspondent aux dates pour lesquelles l'estimateur de β est différent de 0. La gradient de couleur représente la fiabilité avec laquelle la période est détectée (le rouge représente une fiabilité de 100%). Tout d'abord, on peut remarquer que les périodes d'influence sont plus fortement détectées pour les vignobles B et C que pour le vignoble A, sûrement en relation avec le nombre d'individus analysés. Pour le vignoble A, une seule période d'influence peut être identifiée de manière fiable pour les sections temporelles 12 à 14. Deux autres périodes peuvent être présumées pour les périodes 1 à 5 et 9 à 11. En ce qui concerne la période allant de la section temporelle 12 à 14, la valeur de l'estimateur BLiSS est positive, c'est-à-dire que la température moyenne quotidienne observée pendant cette période est positivement corrélée avec le rendement (plus la température est élevée, plus le rendement est élevé). Pour le vignoble B, 4 périodes d'influence peuvent être identifiées. Une cinquième période peut même être détectée dans la section temporelle 16 bien qu'elle n'ait pas été sélectionnée par l'étape éparse de l'estimateur BLiSS. La 2ème et la 4ème période sont positivement corrélées avec le rendement, tandis que la 1ère, la 3ème et la 5ème période le sont négativement. Pour le vignoble C, 4 périodes d'influence sont détectées. Les 2ème et 3ème périodes sont positivement corrélées avec le rendement, tandis que les 1ère et 4ème périodes sont négativement corrélées avec celui-ci.

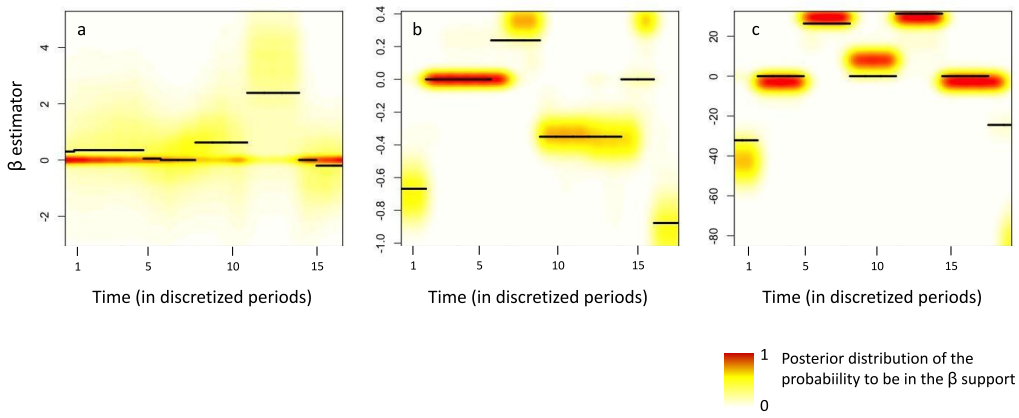

Figure 7.9 – Estimation BLiSS pour la série temporelle synchronisée des données de température moyenne quotidienne pour les vignobles A, B et C (respectivement a, b et c)

7.5.3 La démarche analytique permet d'identifier des indicateurs météorologiques locaux de l'élaboration du rendement

Plusieurs périodes de sensibilité de l'élaboration du rendement à la température ont été détectées avec la méthode BLiSS (Tableau 7.2). Certaines d'entre elles ont déjà été mises en évidence par des travaux de recherche antérieurs et d'autres sont pour l'instant inconnues. Les périodes mises en évidence diffèrent d'un vignoble à l'autre en termes de variable de température influente (température moyenne, maximale ou minimale journalière, précipitations), de date, de durée et de direction de la corrélation avec le rendement (Tableau 7.2).

Table 7.2 – Date (rang des périodes dans la série temporelle), durée (nombre de périodes) et sens de la corrélation avec le rendement (- : négatif, + : positif) des influences sur l'élaboration du rendement détectées avec la méthode BLiSS pour les séries temporelles des données de température moyenne, maximale et minimale journalières de chaque vignoble. Le gradient de couleurs vertes représente respectivement les périodes de débourrement, de floraison et de véraison de l'année n-1 et de l'année n pour chaque vignoble. La couleur grise représente des périodes non considérées pour le vignoble en question.

		Periods number over the year n-1										*Periods number over year n*								
Vineyard	*Temperature indicator*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*
A	*Tmean*												+	+	+					
A	*Tmin*	-	-	-					-	+	+	+								
A	*Tmax*		+	+	+	+														
A	*Rain*			+	+	+														
B	*Tmean*	-	-						+	+	-	-	-	-			+	-	-	
B	*Tmin*				+	-	-	-	+		-							-	-	
B	*Tmax*	-												-	-	+	+		-	
B	*Rain*	+	+	+					-	-										
C	*Tmean*	-	-				+	+	+				+	+	+				-	-
C	*Tmin*			+	+	+	+	+	+		-	-					+	+	+	-
C	*Tmax*						+	+	+		+	+	+				+	+	+	-
C	*Rain*		+	+	+	+	+						-	+	+					

7.6 Chapitre VI : Discussion et perspectives

La prévision du rendement est un élément important pour tous les maillons de la chaîne viticole. Pour l'instant, les attentes de la filière en termes de prévision de rendement ne sont pas satisfaites. Ces trois ans de thèse ont été l'occasion d'en explorer les possibles explications. Premièrement, la plupart des propositions de la recherche se sont concentrées sur la mesure et l'échantillonnage de composantes du rendement et non sur la modélisation du rendement à partir des variables collectées. Deuxièmement, les méthodes de prévision proposées ne prennent pas en compte les besoins et contraintes opérationnelles dans la manière dont la prévision de rendement est formulée et délivrée sur le terrain. Troisièmement, les méthodes proposées ne sont pas adaptées aux données agricoles disponibles.

Sur la base de ce constat, la thèse a proposé une démarche agronomique permettant d'aboutir à une méthode de prévision de rendement qui vise à valoriser les données locales et à répondre aux besoins et contraintes de terrain. Elle est résumée en Fig. 7.10. La dernière partie de cette démarche correspond à la mise en œuvre d'un modèle de prévision du rendement dans un cadre Bayésien.

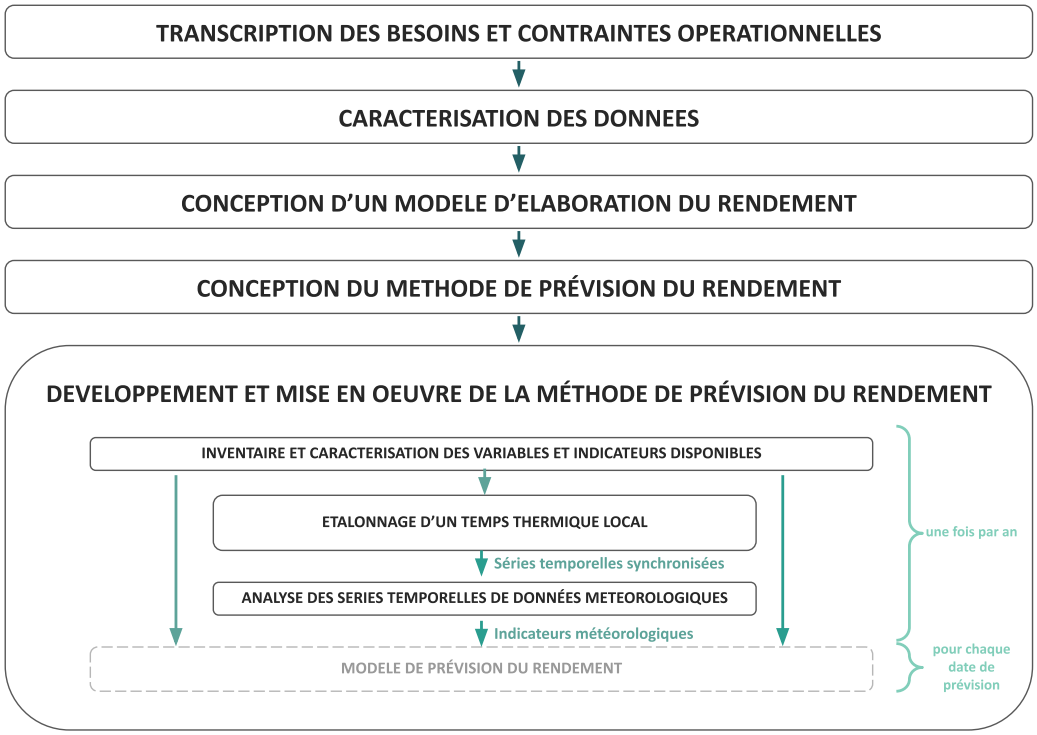

Figure 7.10 – Estimation BLiSS pour la série temporelle synchronisée des données de température moyenne quotidienne pour les vignobles A, B et C (respectivement a, b et c)

La démarche proposée fonctionne à partir des données agricoles. Elle propose de mieux prendre en compte la temporalité de l'élaboration du rendement en formulant plusieurs prévisions de rendement à des dates judicieusement choisies et en quantifiant le niveau de fiabilité à accorder à chacune de ces prévisions. La formulation de plusieurs prévisions de rendement au cours de la saison s'appuie notamment par une analyse plus poussée des séries temporelles de données météo grâce aux méthodes eGDD et BLiSS. Enfin, cette méthode considère un schéma local d'élaboration du rendement, qu'elle approche par apprentissage de données pour chaque vignoble.

Cette thèse a également permis de sélectionner des approches statistiques permettant d'analyser des données agricoles de manière générale selon plusieurs critères. Premièrement, elle a proposé des choix méthodologiques simples et robustes pour gérer la parcimonie des données. Elle a aussi mis en évidence la nécessité de prendre en compte l'interdépendance des données collectées pour les mêmes années et mêmes parcelles. Elle a aussi proposé des adaptations de l'analyse statistique pour prendre en compte un déséquilibre dans le plan d'échantillonnage d'un jeu de données en lien avec des données manquantes. Enfin, elle a proposé de considérer des analyses dans un cadre Bayésien pour mieux gérer l'incertitude due à au volume et à la qualité des jeux de données agricoles.

L'interprétation que l'on peut faire des résultats issus de l'analyse de données agricoles a également été revisitée. En effet, il faut considérer que ces résultats intègrent systématiquement la réponse de la vigne aux variables mais aussi la représentativité des données par rapport aux conditions réellement vécues par les vignes sur le terrain et du bruit dû à la qualité des données. Par exemple, un indice thermique calculé avec la méthode eGDD intègre la réponse de la vigne à la température en termes de phénologie mais aussi les variations spatiales de température entre la station météo et les parcelles étudiées, des erreurs d'estimation des dates phénologiques et le bruit dû à la mesure de la température par une station météorologique donnée. Ainsi, il y a un décalage entre la manière d'interpréter les résultats expérimentaux et les résultats opérationnels issus de l'analyse des jeux de données agricoles. Ce n'est pas un problème si cette différence est identifiée car c'est ainsi que les agriculteurs doivent interpréter leurs données sur le terrain. Cependant, il y a bien un intérêt à formaliser et à traiter cette incertitude qui est habituellement gérée de manière experte par les agriculteurs. Cela permettrait de supporter la prise de décision sur le terrain d'une manière plus reproductible et traçable.

Ceci peut être réalisé en utilisant une approche de prévision dans un cadre bayésien.

Une perspective de travail est donc la mise au point d'un modèle de prévision du rendement viticole dans un cadre Bayésien. Ceci implique que tous les paramètres du modèle sont considérés comme des variables aléatoires pour lesquelles une distribution de probabilité peut être estimée. Des distributions a priori seront choisies pour chaque paramètre et ces distributions seront mises à jour en prenant en compte les observations de terrain. Par conséquent, le rendement sera également modélisé comme une distribution de probabilité. Les valeurs de rendement les plus probables seront considérées comme la gamme de rendements attendus et la dispersion de la distribution indiquera l'incertitude associée à cette prévision.

La méthode de prévision du rendement ainsi proposée doit pouvoir être évaluée de deux manières : en fonction de l'amélioration qu'elle apporte dans la précision de la prévision du rendement par rapport aux méthodes empiriques qui sont actuellement utilisées sur le terrain, par exemple une méthode consistant en une régression linéaire entre le nombre de grappes par cep et le rendement final (Clingeleffer et al. 2001; De la Fuente et al. 2015). Elle doit aussi faire ces preuves en termes de capacité à être adoptée par les professionnels de la filière viticole. Cependant, ce dernier point n'est pas trivial puisque la satisfaction des utilisateurs de la méthode en termes de qualité de la prévision dépend directement des données qui seront mis en entrée (données phénologiques, données de rendement, données météorologiques etc.). Il y a donc un vrai enjeu à faire entrer les utilisateurs de la méthode dans une boucle d'amélioration continue, où les sorties de la méthode de prévision inciteraient les utilisateurs à améliorer leur pratiques de collecte de données.

Enfin, la démarche de cette thèse visant à construire une méthode de prévision du rendement étudiée est vue comme un cas d'étude du développement d'une recherche agronomique pour des fins opérationnelles et à partir de données agricoles. Une perspective de ce travail consisterait à évaluer la possibilité d'extrapoler cette méthode à d'autres variables viticoles et à d'autres cultures. Cette méthode pourrait effectivement être appliquée à d'autres cas d'étude avec deux précautions particulières : i) vérifier la pertinence des données disponibles par rapport pour modéliser la variable d'intérêt, par exemple les variables météorologiques sont-elles parmi les facteurs primordiaux de l'élaboration de la variable d'intérêt ? et ii) adopter l'interprétation des résultats à la réalité du cas d'étude, par exemple, un effet parcelle ne contiendra pas les déterminants propres à la culture pour des cultures en rotation.

Appendix I : Vineyard Interview Guide

1- What would be the estimated parameter (date, unit) ? Ex: cluster number, harvest mass, harvest volume etc.

2- When would you like to receive the yield assessment ?

3- When would you like to receive the yield assessment ? Ex: within-field, block, winery etc.

4-Which precision do you expect ?

5-How would you like to visualize the yield assessment ?

6- Who would be the user of this yield assessment ? What would he/she use it for ?

7- What time and means are you ready to allocate to the yield assessment process ?

8- What would encourage you to use a yield assessment method practically speaking ?

9- What would discourage you from using a yield assessment method practically speaking ?

10- What is the benefit of yield assessment for you? How would you quantify it ?

Bibliography

Abdelghafour, F., B. Keresztes, C. Germain, and J. P. Da Costa (2017). « Potential of On-Board Colour Imaging for in-Field Detection and Counting of Grape Bunches at Early Fruiting Stages ». *Advances in Animal Biosciences* 8.2, pp. 505–509. DOI: 10.1017/S2040470017001030.

Abdelghafour, F., R. Rosu, B. Keresztes, C. Germain, and J. P. Da Costa (2019). « A Bayesian Framework for Joint Structure and Colour Based Pixel-Wise Classification of Grapevine Proximal Images ». *Computers and Electronics in Agriculture* 158, pp. 345–357. DOI: 10.1016/j.compag.2019.02.017.

Adamchuk, V.I., R. A. Viscarra Rossel, D. B. Marx, and A. K. Samal (2011). « Using Targeted Sampling to Process Multivariate Soil Sensing Data ». *Geoderma* 163.1, pp. 63–73. DOI: 10.1016/j.geoderma.2011.04.004.

Allan, Catherine, Thi Phuoc Lai Nguyen, Giovanna Seddaiu, Ben Wilson, and Pier Paolo Roggero (2013). « Integrating Local Knowledge with Experimental Research: Case Studies on Managing Cropping Systems in Italy and Australia ». *Italian Journal of Agronomy* 8.2 (2), e15–e15. DOI: 10.4081/ija.2013.e15.

Aquino, A., I. Barrio, M.P. Diago, B. Millan, and J. Tardaguila (2018). « vitisBerry: An Android-Smartphone Application to Early Evaluate the Number of Grapevine Berries by Means of Image Analysis ». *Computers and Electronics in Agriculture* 148, pp. 19–28. DOI: 10.1016/j.compag.2018.02.021.

Aquino, A., B. Millan, M.P. Diago, and J. Tardaguila (2018). « Automated Early Yield Prediction in Vineyards from On-the-Go Image Acquisition ». *Computers and Electronics in Agriculture* 144, pp. 26–36. DOI: 10.1016/j.compag.2017.11.026.

Aquino, A., B. Millan, D. Gaston, M.P Diago, and J. Tardaguila (2015). « vitisFlower®: Development and Testing of a Novel Android-Smartphone Application for Assessing the Number of Grapevine Flowers per Inflorescence Using Artificial Vision Techniques ». *Sensors* 15.9, pp. 21204–21218. DOI: 10.3390/s150921204. pmid: 26343664.

Aquino, A., B. Millan, S. Gutiérrez, and J. Tardáguila (2015). « Grapevine Flower Estimation by Applying Artificial Vision Techniques on Images with Uncontrolled Scene and Multi-Model Analysis ». *Computers and Electronics in Agriculture* 119, pp. 92–104. DOI: 10.1016/j.compag.2015.10.009.

Arab, S. T., R. Noguchi, S. Matsushita, and T. Ahamed (2021). « Prediction of Grape Yields from Time-Series Vegetation Indices Using Satellite Remote Sensing and a Machine-Learning Approach ». *Remote Sensing Applications: Society and Environment* 22, p. 100485. DOI: 10.1016/j.rsase.2021.100485.

Araya-Alman, M., C. Acevedo-Opazo, S. Guillaume, H. Valdés-Gómez, N. Verdugo-Vásquez, Y. Moreno, and B. Tisseyre (2017). « Using Ancillary Yield Data to Improve Sampling and Grape Yield Estimation of the Current Season ». *Advances in Animal Biosciences* 8.2, pp. 515–519.

Araya-Alman, M., C. Leroux, C. Acevedo-Opazo, S. Guillaume, H. Valdés-Gómez, N. Verdugo-Vásquez, C. Pañitrur-De la Fuente, and B. Tisseyre (2019). « A New Localized Sampling Method to Improve Grape Yield Estimation of the Current Season Using Yield Historical Data ». *Precision Agriculture* 20.2, pp. 445–459. DOI: 10.1007/s11119-019-09644-y.

Baby, T., M. Gilliham, S.D. Tyerman, and C. Collins (2015). « Differential Fruitset between Grapevine Cultivars Is Related to Differences in Pollen Viability and Amine Concentration in Flowers: Amines and Grapevine Reproduction ». *Australian Journal of Grape and Wine Research* 22.1, pp. 149–158. DOI: 10.1111/ajgw.12191.

Ballesteros, R., D.S Intrigliolo, J.F. Ortega, J. M. Ramírez-Cuesta, I. Buesa, and M.A. Moreno (2020). « Vineyard Yield Estimation by Combining Remote Sensing, Computer Vision and Artificial Neural Network Techniques ». *Precision Agriculture* 21.4, 1_21. DOI: 10.1007/s11119-020-09717-3.

Bates, T.R., R.M. Dunst, and P. Joy (2002). « Seasonal Dry Matter, Starch, and Nutrient Distribution in 'Concord' Grapevine Roots ». *HortScience* 37.2, pp. 313–316. DOI: 10.21273/HORTSCI.37.2.313.

Bates, Terence, Jackie Dresser, Rhiann Eckstrom, Golnaz Badr, Thom Betts, and James Taylor (2018). « Variable-Rate Mechanical Crop Adjustment for Crop Load Balance in 'Concord' Vineyards ». *2018 IoT Vertical and Topical Summit on Agriculture*. IOT Tuscany. Tuscany: IEEE, pp. 1–4. ISBN: 978-1-5386-6930-3. DOI: 10.1109/IOT-TUSCANY.2018.8373046.

Bigard, A., C. Romieu, Y. Sire, M. Veyret, H. Ojéda, and L. Torregrosa (2019). « The Kinetics of Grape Ripening Revisited through Berry Density Sorting ». *OENO One* 53.4. DOI: 10.20870/oeno-one.2019.53.4.2224.

Blom, P. E. and J. M. Tarara (2009). « Trellis Tension Monitoring Improves Yield Estimation in Vineyards ». *HortScience* 44.3, pp. 678–685.

Boss, P.K., E.J. Buckeridge, A. Poole, and M.R. Thomas (2003). « New Insights into Grapevine Flowering. » *Functional Plant Biology* 30.6, pp. 593–606. DOI: 10.1071/fp02112.

Boursiquot, J. M., M. Dessup, and C. Rennes (1995). « Distribution des principaux caractères phénologiques, agronomiques et technologiques chez Vitis vinifera L. » *Vitis* 34.1, pp. 31–35.

Bramley, R.G.V. and S.K. Williams (2001). « A Protocol for Winegrape Yield Maps ». 3rd European Conference on Precision Agriculture. Montpellier, France.

Brisson, N, C Gary, E Justes, R Roche, B Mary, D Ripoche, D Zimmer, J Sierra, P Bertuzzi, P Burger, F Bussière, Y.M Cabidoche, P Cellier, P Debaeke, J.P Gaudillère, C Hénault, F Maraux, B Seguin, and H Sinoquet (2003). « An Overview of the Crop Model Stics ». *European Journal of Agronomy* 18.3-4, pp. 309–332. DOI: 10.1016/S1161-0301(02)00110-7.

Brockwell, Peter J. and Richard A. Davis (2009). *Time Series: Theory and Methods*. Springer Science & Business Media. 591 pp. ISBN: 978-1-4419-0320-4. Google Books: TVIpBgAAQBAJ.

Brunel, G., L. Pichon, J. Taylor, and B. Tisseyre (2019). « Easy Water Stress Detection System for Vineyard Irrigation Management ». *Precision Agriculture '19*. 12th European Conference on Precision Agriculture. Montpellier, France: Wageningen Academic Publishers, pp. 935–942. ISBN: 978-90-8686-337-2 978-90-8686-888-9. DOI: 10.3920/978-90-8686-888-9_115.

Buttrose, M. S. (1969). « Fruitfulness in Grapevines: Effects of Light Intensity and Temperature ». *Botanical gazette* 130.3, pp. 166–173.

— (1970). « Fruitfulness in grapevines: the response of different cultivara to light, temperature and daylength. » *Vitis* 9, pp. 121–5.

— (1974). « Climatic Factors and Fruitfulness in Grapevines. » *Horticultural Abstracts* 44.6, pp. 321–326.

Caffarra, A. and E. Eccel (2010). « Increasing the Robustness of Phenological Models for Vitis Vinifera Cv. Chardonnay ». *International Journal of Biometeorology* 54.3, pp. 255–267. DOI: 10.1007/s00484-009-0277-5.

Camargo-Alvarez, H., M. Salazar-Gutiérrez, M. Keller, and G. Hoogenboom (2020). « Modeling the Effect of Temperature on Bud Dormancy of Grapevines ». *Agricultural and Forest Meteorology* 280, p. 107782. DOI: 10.1016/j.agrformet.2019.107782.

Carmona, M.J., J. Chaïb, J.M. Martínez-Zapater, and M.R. Thomas (2008). « A Molecular Genetic Perspective of Reproductive Development in Grapevine ». *Journal of Experimental Botany* 59.10, pp. 2579–2596. DOI: 10.1093/jxb/ern160.

Carrillo, E., A. Matese, J. Rousseau, and B. Tisseyre (2015). « Use of Multispectral Airborne Imagery to Improve Yield Sampling in Viticulture ». *Precision Agriculture* 17.1, pp. 74–92. DOI: 10.1007/s11119-015-9407-8.

Celette, F. and C. Gary (2013). « Dynamics of Water and Nitrogen Stress along the Grapevine Cycle as Affected by Cover Cropping ». *European Journal of Agronomy* 45, pp. 142–152. DOI: 10.1016/j.eja.2012.10.001.

Cerovic, Z. G., N. B. Ghozlen, C. Milhade, M. Obert, S. Debuisson, and M. Le Moigne (2015). « Nondestructive Diagnostic Test for Nitrogen Nutrition of Grapevine (Vitis Vinifera L.) Based on Dualex Leaf-Clip Measurements in the Field. » *Journal of Agricultural and Food Chemistry* 63.14, pp. 3669–3680. DOI: 10.1021/acs.jafc.5b00304.

Champagnol, F (1984). *Éléments de physiologie de la vigne et de viticulture générale*. Prades-le-Lez: F. Champagnol. ISBN: 978-2-9500614-0-9.

Che, X. and S. Xu (2010). « Bayesian Data Analysis for Agricultural Experiments ». *Canadian Journal of Plant Science* 90.5, pp. 575–603. DOI: 10.4141/CJPS10004.

Cheraïet, A., O. Naud, M. Carra, S. Codis, F. Lebeau, and James Taylor (2020). « An Algorithm to Automate the Filtering and Classifying of 2D LiDAR Data for Site-Specific Estimations of Canopy Height and Width

in Vineyards ». *Biosystems Engineering* 200, pp. 450–465. DOI: 10.1016/j.biosystemseng.2020.10.016.

Chloupek, O., P. Hrstkova, and P. Schweigert (2004). « Yield and Its Stability, Crop Diversity, Adaptability and Response to Climate Change, Weather and Fertilisation over 75 Years in the Czech Republic in Comparison to Some European Countries ». *Field Crops Research* 85, pp. 167–190. DOI: 10.1016/s0378-4290(03)00162-x.

Chuine, I. (2000). « A Unified Model for Budburst of Trees ». *Journal of Theoretical Biology* 207.3, pp. 337–347. DOI: 10.1006/jtbi.2000.2178.

Clingeleffer, P., G. M. Dunn, M. Krstic, and S. Martin (2001). *Crop Development, Crop Estimation and Crop Control to Secure Quality and Production of Major Wine Grape Varieties: A National Approach*. Grape and Wine Resarch & Development Corporation.

Coipel, J, B. R. Lovelle, C. Sipp, and C. Van Leeuwen (2006). « Terroir Effect, as a Result of Environmental Stress, Depends More on Soil Depth than on Soil Type (Vitis Vinifera L. Cv. Grenache Noir, Côtes Du Rhône, 2000). » *Journal International des Sciences de la Vigne et du Vin* 40, pp. 177–185.

Cola, G., L. Mariani, F. Salinari, S. Civardi, F. Bernizzoni, M. Gatti, and S. Poni (2014). « Description and Testing of a Weather-Based Model for Predicting Phenology, Canopy Development and Source–Sink Balance in Vitis Vinifera L. Cv. Barbera ». *Agricultural and Forest Meteorology* 184, pp. 117–136. DOI: 10.1016/j.agrformet.2013.09.008.

Collins, C., R. Coles, J. G. Conran, and B. Rawnsley (2006). « The Progression of Primary Bud Necrosis in the Grapevine Cv. Shiraz (Vitis Vinifera L.): A Histological Analysis ». *Vitis* 45.2, pp. 57–62. DOI: 10.5073/vitis.2006.45.57-62.

Collins, C. and P.r. Dry (2009). « Response of Fruitset and Other Yield Components to Shoot Topping and 2-Chlorethyltrimethyl-Ammonium Chloride Application ». *Australian Journal of Grape and Wine Research* 15.3, pp. 256–267. DOI: 10.1111/j.1755-0238.2009.00063.x.

Coviello, L., M. Cristoforetti, G. Jurman, and C. Furlanello (2020). « GBCNet: In-Field Grape Berries Counting for Yield Estimation by Dilated CNNs ». *Applied Sciences* 10.14, p. 4870. DOI: 10.3390/app10144870.

Cozzolino, D., M. F. Dupont, A. Elbourne, V. K. Truong, A. Power, and James Chapman (2020). « Chapter 9 - Role of Sensors in Fruit Nutrition ». *Fruit Crops*. Ed. by A. K. Srivastava and Chengxiao Hu. Elsevier, pp. 111–119. ISBN: 978-0-12-818732-6. DOI: 10.1016/B978-0-12-818732-6.00009-5.

Crespo-Martínez, S., M. Sobczak, E. Różańska, A. Forneck, and M. Griesser (2019). « The Role of the Secondary Phloem during the Development of the Grapevine Berry Shrivel Ripening Disorder ». *Micron* 116, pp. 36–45. DOI: 10.1016/j.micron.2018.09.012.

Cristofolini, F. and E. Gottardini (2000). « Concentration of Airborne Pollen of Vitisvinifera L. and Yield Forecast: A Case Study at S.Michele All'Adige, Trento, Italy ». *Aerobiologia* 16.1, pp. 125–129. DOI: 10.1023/A:1007690802901.

Cross, H. Z. and M. S. Zuber (1972). « Prediction of Flowering Dates in Maize Based on Different Methods of Estimating Thermal Units. » *Agronomy Journal* 64.3, pp. 351–355. DOI: 10.2134/agronj1972.00021962006400030029x.

Cunha, M., A. R. S. Marçal, and L. Silva (2010). « Very Early Prediction of Wine Yield Based on Satellite Data from Vegetation ». *International Journal of Remote Sensing* 31.12, pp. 3125–3142.

Cunha, M., H. Ribeiro, and I. Abreu (2016). « Pollen-Based Predictive Modelling of Wine Production: Application to an Arid Region ». *European Journal of Agronomy* 73, pp. 42–54. DOI: 10.1016/j.eja.2015.10.008.

Damour, Gaëlle, Marie L. Navas, and Eric Garnier (2018). « A Revised Trait-Based Framework for Agroecosystems Including Decision Rules ». *Journal of Applied Ecology* 55.1, pp. 12–24. DOI: 10.1111/1365-2664.12986.

De la Fuente, M., R. Linares, P. Baeza, C. M. Lissarague, and J. R. Lissarague (2015). « Comparison of Different Methods of Grapevine Yield Prediction in the Time Window between Fruitset and Veraison ». *Journal International Scientifique de la Vigne et du Vin* 49, pp. 27–35.

Demestihas, C., S. Debuisson, and A. Descotes (2018). « Decomposing the Notion of Vine Vigour with a Proxydetection Shoot Sensor: Physiocap® ». *E3S Web of Conferences* 50, p. 03003. DOI: 10.1051/e3sconf/20185003003.

De Rességuier, Laure, Séverine Mary, Renan Le Roux, Théo Petitjean, Hervé Quénol, and Cornelis van Leeuwen (2020). « Temperature Variability at Local Scale in the Bordeaux Area. Relations With Envi-

ronmental Factors and Impact on Vine Phenology ». *Frontiers in Plant Science* 11. DOI: 10.3389/fpls.2020.00515. pmid: 32508854.

Di Gennaro, S. F. and A. Matese (2020). « Evaluation of Novel Precision Viticulture Tool for Canopy Biomass Estimation and Missing Plant Detection Based on 2.5D and 3D Approaches Using RGB Images Acquired by UAV Platform ». *Plant Methods* 16.1, p. 91. DOI: 10.1186/s13007-020-00632-2.

Di Gennaro, Salvatore Filippo, Piero Toscano, Paolo Cinat, Andrea Berton, and Alessandro Matese (2019). « A Low-Cost and Unsupervised Image Recognition Methodology for Yield Estimation in a Vineyard ». *Frontiers in Plant Science* 10. DOI: 10.3389/fpls.2019.00559.

Diago, M. P., C. Correa, B. Millan, P. Barreiro, C. Valero, and J. Tardaguila (2012). « Grapevine Yield and Leaf Area Estimation Using Supervised Classification Methodology on RGB Images Taken under Field Conditions ». *Sensors* 12, pp. 16988–17006.

Diago, M. P., J. Tardaguila, N. Aleixos, B. Millan, J. M. Prats-Montalban, S. Cubero, and J. Blasco (2015). « Assessment of Cluster Yield Components by Image Analysis: Cluster Yield Components by Image Analysis ». *Journal of the Science of Food and Agriculture* 95.6, pp. 1274–1282. DOI: 10.1002/jsfa.6819.

Dobriyal, P., A. Qureshi, R. Badola, and S. A. Hussain (2012). « A Review of the Methods Available for Estimating Soil Moisture and Its Implications for Water Resource Management ». *Journal of Hydrology* 458–459, pp. 110–117. DOI: 10.1016/j.jhydrol.2012.06.021.

Dokoozlian, N. K. (2000). « Grape Berry Growth and Development ». *Raisin Production Manual*. UCANR Publications, pp. 30–37. ISBN: 978-1-879906-44-0. Google Books: rMoAET18u9YC.

Dokoozlian, N. K. and W. M. Kliewer (1996). « Influence of Light on Grape Berry Growth and Composition Varies during Fruit Development ». *Journal of the American Society for Horticultural Science* 121.5, pp. 869–874. DOI: 10.21273/JASHS.121.5.869.

Doré, Thierry, Marianne Le Bail, Philippe Martin, Bertrand Ney, Jean Roger-Estrade, and Michel Sébillotte (2006). *L'agronomie Aujorud'hui*. 1ère édition. Versailles: Editions Quae. 384 pp.

Dry, P.R., M.L. Longbottom, S. Mcloughlin, T.E. Johnson, and C. Collins (2010). « Classification of Reproductive Performance of Ten Winegrape Varieties ». *Australian Journal of Grape and Wine Research* 16, pp. 47–55. DOI: 10.1111/j.1755-0238.2009.00085.x.

Dry, Peter R. (2000). « Canopy Management for Fruitfulness ». *Australian Journal of Grape and Wine Research* 6, pp. 109–115.

Duchêne, E., N. Jaegli, and R. Salber (2003). « Effects of the leaf /fruit ratio during a year N on the bud fertility in the year N+2: results for Pinot noir and Gewurztraminer grapevines (Vitis vinifera L.) in the greenhouse ». *OENO One* 37.4, pp. 257–260. DOI: 10.20870/oeno-one.2003.37.4.958.

Duchêne, E., N. Jaegli, R. Salber, and J. P. Gaudillère (2003). « Effects of Ripening Conditions on the Following Season's Growth and Yield Components for Pinot Noir and Gewurztraminer Grapevines (Vitis Vinifera L.) in a Controlled Environment ». *OENO One* 37.1, pp. 39–49. DOI: 10.20870/oeno-one.2003.37.1.1682.

Dunn, G. M. and S. R. Martin (2000). « Do Temperature Conditions at Budburst Affect Flower Number in Vitis Vinifera L. Cv. Cabernet Sauvignon? » *Australian Journal of Grape and Wine Research* 6.2, pp. 116–124. DOI: 10.1111/j.1755-0238.2000.tb00169.x.

Dunn, Gregory M. and Steve Martin (2004). « Yield Prediction from Digital Image Analysis: A Technique with Potential for Vineyard Assessments Prior to Harvest ». *Australian Journal of Grape and Wine Research* 10.3, pp. 196–198. DOI: 10.1111/j.1755-0238.2004.tb00022.x.

Durquety, P.M., E. Naude, and P. Blanchard (1982). « La Prévision de Récolte Sur Petit Manseng (Vitis Vinifera L.) Basée Sur Les Courbes-Niveaux de Fertilité et Les Températures Durant Une Période Critique. » *Pogrès Agricole et Viticole* 99, pp. 440–469.

Dury, Jérôme, Noémie Schaller, Frédérick Garcia, Arnaud Reynaud, and Jacques Eric Bergez (2012). « Models to Support Cropping Plan and Crop Rotation Decisions. A Review ». *Agronomy for Sustainable Development* 32.2, pp. 567–580. DOI: 10.1007/s13593-011-0037-x.

Ebadi, A., P. May, and B.G. Coombe (1996). « Effect of Short-Term Temperature and Shading on Fruit-Set, Seed and Berry Development in Model Vines of V. Vinifera, Cvs Chardonnay and Shiraz ». *Australian Journal of Grape and Wine Research* 2.1, pp. 1–8. DOI: 10.1111/j.1755-0238.1996.tb00087.x.

Etikan, I. and K. Bala (2017). « Sampling and Sampling Methods ». *Biometrics & Biostatistics International Journal* 5.6, pp. 215–217. DOI: 10.15406/bbij.2017.05.00149.

Fernández-González, M., F. J. Rodríguez-Rajo, V. Jato, O. Escuredo, and M. J. Aira (2011). « Estimation of Yield 'Loureira' Variety with an Aerobiological and Phenological Model ». *Grana* 50.1, pp. 63–72. DOI: 10.1080/00173134.2011.561871.

Fila, G., M. Gardiman, P. Belvini, F. Meggio, and A. Pitacco (2014). « A Comparison of Different Modelling Solutions for Studying Grapevine Phenology under Present and Future Climate Scenarios ». *Agricultural and Forest Meteorology* 195–196, pp. 192–205. DOI: 10.1016/j.agrformet.2014.05.011.

Folwell, R. J., D. E. Santos, S. E. Spayd, L. H. Porter, and D. S. Wells (1994). « Statistical Technique for Forecasting Concord Grape Production ». *American Journal of Enology and Viticulture* 45.1, pp. 63–70.

Font, D., M. Tresanchez, D. Martínez, J. Moreno, E. Clotet, and J. Palacín (2015). « Vineyard Yield Estimation Based on the Analysis of High Resolution Images Obtained with Artificial Illumination at Night ». *Sensors* 15.4, pp. 8284–8301. DOI: 10.3390/s150408284.

Fraga, H., J. A. Santos, J. Moutinho-Pereira, C. Carlos, J. Silvestre, J. Eiras-Dias, T. Mota, and A. C. Malheiro (2016). « Statistical Modelling of Grapevine Phenology in Portuguese Wine Regions: Observed Trends and Climate Change Projections ». *The Journal of Agricultural Science* 154.05, pp. 795–811. DOI: 10.1017/S0021859615000933.

Fraga, Helder, Iñaki García de Cortázar Atauri, Aureliano C. Malheiro, and João A. Santos (2016). « Modelling Climate Change Impacts on Viticultural Yield, Phenology and Stress Conditions in Europe ». *Global Change Biology* 22.11, pp. 3774–3788. DOI: 10.1111/gcb.13382. pmid: 27254813.

Gaiotti, F., C. Pastore, I. Filippetti, L. Lovat, N. Belfiore, and D. Tomasi (2018). « Low Night Temperature at Veraison Enhances the Accumulation of Anthocyanins in Corvina Grapes (Vitis Vinifera L.) » *Scientific Reports* 8.1, p. 8719. DOI: 10.1038/s41598-018-26921-4.

Gambetta, G. A., J. C. Herrera, S. Dayer, Q. Feng, U. Hochberg, and S. D. Castellarin (2020). « The Physiology of Drought Stress in Grapevine: Towards an Integrative Definition of Drought Tolerance ». *Journal of Experimental Botany* 71.16. Ed. by J. Zhang, pp. 4658–4676. DOI: 10.1093/jxb/eraa245.

Garcia, L., F. Celette, C. Gary, A. Ripoche, H. Valdés-Gómez, and A. Metay (2018). « Management of Service Crops for the Provision of Ecosystem Services in Vineyards: A Review ». *Agriculture, Ecosystems & Environment* 251, pp. 158–170. DOI: 10.1016/j.agee.2017.09.030.

Garcia De Cortazar Atauri, Inaki, Eric Duchêne, Agnes Destrac, Gérard Barbeau, Laure De Resseguier, Thierry Lacombe, Amber K. Parker, Nicolas Saurin, and Cornelis Van Leeuwen (2017). « Grapevine Phenology in France : From Past Observations to Future Evolutions in the Context of Climate Change ». *OENO One* 51.2, pp. 115–126. DOI: 10.20870/oeno-one.2016.0.0.1622.

Gautam, Deepak and Vinay Pagay (2020). « A Review of Current and Potential Applications of Remote Sensing to Study the Water Status of Horticultural Crops ». *Agronomy* 10.1 (1), p. 140. DOI: 10.3390/agronomy10010140.

González-Fernández, E., A. Piña-Rey, M. Fernández-González, M. J. Aira, and F. J. Rodríguez-Rajo (2020). « Prediction of Grapevine Yield Based on Reproductive Variables and the Influence of Meteorological Conditions ». *Agronomy* 10.5, p. 714. DOI: 10.3390/agronomy10050714.

González-Flor, Cristina, Lydia Serrano, Gil Gorchs, and Josep M. Pons (2014). « Assessment of Grape Yield and Composition Using Reflectance-Based Indices in Rainfed Vineyards ». *Agronomy Journal* 106.4, p. 1309. DOI: 10.2134/agronj13.0422.

Gouot, J.C, J. P. Smith, B. P. Holzapfel, and C. Barril (2019a). « Impact of Short Temperature Exposure of Vitis Vinifera L. Cv. Shiraz Grapevine Bunches on Berry Development, Primary Metabolism and Tannin Accumulation ». *Environmental and Experimental Botany* 168, p. 103866. DOI: 10.1016/j.envexpbot.2019.103866.

Gouot, Julia C., Jason P Smith, Bruno Holzapfel, and Celia Barril (2019b). « Single and Cumulative Effects of Whole-Vine Heat Events on Shiraz Berry Composition. » 21st GIESCO INternational Meeting. Vol. 2. Thessalonik, Greece: Onoe-One, pp. 171–187.

Gourieroux, A. M., M. E. McCully, B. P. Holzapfel, G. R. Scollary, and S. Y. Rogiers (2016). « Flowers Regulate the Growth and Vascular Development of the Inflorescence Rachis in Vitis Vinifera L. » *Plant Physiology and Biochemistry* 108, pp. 519–529. DOI: 10.1016/j.plaphy.2016.08.016.

Greer, D. H. and C. Weston (2010). « Heat Stress Affects Flowering, Berry Growth, Sugar Accumulation and Photosynthesis of Vitis Vinifera Cv. Semillon Grapevines Grown in a Controlled Environment ». *Functional Plant Biology* 37.3, p. 206. DOI: 10.1071/FP09209.

Grigorieva, E., A. Matzarakis, and C. Freitas (2010). « Analysis of Growing Degree-Days as a Climate Impact Indicator in a Region with Extreme Annual Air Temperature Amplitude ». *Climate Research* 42, pp. 143–154. DOI: 10.3354/CR00888.

Grimm, J., K. Herzog, F. Rist, A. Kicherer, R. Töpfer, and V. Steinhage (2019). « An Adaptable Approach to Automated Visual Detection of Plant Organs with Applications in Grapevine Breeding ». *Biosystems Engineering* 183, pp. 170–183. DOI: 10.1016/j.biosystemseng.2019.04.018.

Grocholsky, B. (2011). « A Camera and Laser System for Automatic Vine Balance Assessment ». American Society of Agricultural and Biological Engineers (ASABE) Annual International Meeting. In collab. with S. Nuske, M. Aasted, S. Achar, and T. Bates. Louisville, Kentucky. DOI: 10.13031/2013.38146).

Grollemund, Paul-Marie, Christophe Abraham, Meïli Baragatti, and Pierre Pudlo (2019). « Bayesian Functional Linear Regression with Sparse Step Functions ». *Bayesian Analysis* 14.1, pp. 111–135. DOI: 10.1214/18-BA1095.

Grossetete, M., Y. Berthoumieu, J. P. Da Costa, C. Germain, O. Lavialle, and G. Grenier (2012). « Early Estimation of Vineyard Yield: Site Specific Counting of Berries by Using a Smartphone ». *International Conference of Agricultural Engineering—CIGR-AgEng*. Infomation Technology, Automation and Precision Farming. International Conference of Agricultural Engineering - CIGR-AgEng 2012: Agriculture and Engineering for a Healthier Life. Valencia, Spain.

Guerra, Bibiana and Kerri Steenwerth (2012). « Influence of Floor Management Technique on Grapevine Growth, Disease Pressure, and Juice and Wine Composition: A Review ». *American Journal of Enology and Viticulture* 63.2, pp. 149–164.

Guilpart, N., A. Metay, and C. Gary (2014). « Grapevine Bud Fertility and Number of Berries per Bunch Are Determined by Water and Nitrogen Stress around Flowering in the Previous Year ». *European Journal of Agronomy* 54, pp. 9–20. DOI: 10.1016/j.eja.2013.11.002.

Guilpart, N., S. Roux, C. Gary, and A. Metay (2017). « The Trade-off between Grape Yield and Grapevine Susceptibility to Powdery Mildew and Grey Mould Depends on Inter-Annual Variations in Water Stress ». *Agricultural and Forest Meteorology* 234–235, pp. 203–211. DOI: 10.1016/j.agrformet.2016.12.023.

Hall, A., D. W. Lamb, B. P. Holzapfel, and J. P. Louis (2011). « Within-Season Temporal Variation in Correlations between Vineyard Canopy and Winegrape Composition and Yield ». *Precision Agriculture* 12.1, pp. 103–117. DOI: 10.1007/s11119-010-9159-4.

Herrero-Langreo, A., B. Tisseyre, J. P. Goutouly, T. Scholasch, and C. Van Leeuwen (2013). « Mapping Grapevine (Vitis Vinifera L.) Water Status during the Season Using Carbon Isotope Ratio (13C) as Ancillary Data ». *American Journal of Enology and Viticulture* 64.3, pp. 307–315. DOI: 10.5344/ajev.2013.12125.

Houel, C., R. Chatbanyong, A. Doligez, M. Rienth, S. Foria, N. Luchaire, C. Roux, A. Adivèze, G. Lopez, M. Farnos, et al. (2015). « Identification of Stable QTLs for Vegetative and Reproductive Traits in the Microvine (Vitis Vinifera L.) Using the 18 K Infinium Chip ». *BMC plant biology* 15.1, p. 205. DOI: 10.1186/s12870-015-0588-0.

Howell, G. S. (2001). « Sustainable Grape Productivity and the Growth-Yield Relationship: A Review ». *American Journal of Enology and Viticulture* 52.3, pp. 165–174.

Ibáñez, S., J. Grimplet, E. Baroja, S. Hernaiz, and J. Ibáñez (2019). « Characterization of the Reproductive Performance of a Collection of Grapevine Cultivars ». *Acta Horticulturae* 1248, pp. 345–352. DOI: 10.17660/ActaHortic.2019.1248.50.

Intrigliolo, D. S. and J. R. Castel (2009). « Response of Grapevine Cv. 'Tempranillo' to Timing and Amount of Irrigation: Water Relations, Vine Growth, Yield and Berry and Wine Composition ». *Irrigation Sciences* 28, pp. 113–125. DOI: 10.1007/s00271-009-0164-1.

Ivorra, E., A. J. Sánchez Salmerón, J. G. Camarasa Baixauli, M. P. Diago, and J. Tardaguila (2015). « Assessment of grape cluster yield components based on 3D descriptors using stereo vision ». *Food Control* 50, pp. 273–282. DOI: 10.1016/j.foodcont.2014.09.004.

JCGM (100:2008). *Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement. Bureau International Des Poids et Mesures.*

Jones, J. E, R. C. Menary, and S. J. Wilson (2009). « Continued Development of V. Vinifera Inflorescence Primordia in Winter Dormant Buds ». *Vitis* 48.3, pp. 103–105.

Kaspar, T. C., T. S. Colvin, D. B. Jaynes, D. L. Karlen, D. E. James, David W. Meek, Daniel Pulido, and Howard Butler (2003). « Relationship Between Six Years of Corn Yields and Terrain Attributes ». *Precision Agriculture* 4.1, pp. 87–101. DOI: 10.1023/A:1021867123125.

Keller, M., L. J. Mills, R. L. Wample, and S. E. Spayd (2005). « Cluster Thinning Effects on Three Deficit-Irrigated Vitis Vinifera Cultivars ». *American Journal of Enology and Viticulture* 56.2, pp. 91–103.

Keller, M., J. P. Smith, and B. R. Bondada (2006). « Ripening Grape Berries Remain Hydraulically Connected to the Shoot ». *Journal of Experimental Botany* 57.11, pp. 2577–2587.

Keller, M., J. M. Tarara, and L. J. Mills (2010). « Spring Temperatures Alter Reproductive Development in Grapevines. Australian Journal of Grape and Wine Research ». *Australian Journal of Grape and Wine Research* 16.3, pp. 445–454. DOI: 10.1111/j.1755-0238.2010.00105.x.

Keller, M., Y. Zhang, P. M. Shrestha, M. Biondi, and B. R. Bondada (2015). « Sugar Demand of Ripening Grape Berries Leads to Recycling of Surplus Phloem Water via the Xylem ». *Plant, Cell & Environment* 38.6, pp. 1048–1059. DOI: 10.1111/pce.12465.

Kruskal, W. and F. Mosteller (1979). « Representative Sampling III: The Current Statistical Literature. » *Revue Internationale de Statistique* 47.3, pp. 245–265.

Kurtser, P., O. Ringdahl, N. Rotstein, R. Berenstein, and Y. Edan (2020). « In-Field Grape Cluster Size Assessment for Vine Yield Estimation Using a Mobile Robot and a Consumer Level RGB-D Camera ». *IEEE Robotics and Automation Letters* 5.2, pp. 2031–2038. DOI: 10.1109/LRA.2020.2970654.

Lamanda, N., S. Roux, S. Delmotte, A. Merot, B. Rapidel, M. Adam, and J. Wery (2012). « A Protocol for the Conceptualisation of an Agro-Ecosystem to Guide Data Acquisition and Analysis and Expert Knowledge Integration ». *European Journal of Agronomy* 38, pp. 104–116. DOI: 10.1016/j.eja.2011.07.004.

Launspach, M., J. A. Taylor, and J. Wilson (2017). « Can Temperatures from an Online Weather Forecast Service Be Suitable for Modelling Growth Stages Using a CERES-Wheat Type Phenology Model? » *Advances in Animal Biosciences* 8.2, pp. 684–688. DOI: 10.1017/S2040470017000383.

Laurent, C., M. Baragatti, J. Taylor, T. Scholasch, A. Metay, and B. Tisseyre (2019). « Evaluation of a Functional Bayesian Method to Analyse Time Series Data in Precision Viticulture ». *Precision Agriculture '19*. 12th European Conference on Precision Agriculture. Montpellier, France: Wageningen Academic Publishers, pp. 67–73. ISBN: 978-90-8686-337-2. DOI: 10.3920/978-90-8686-888-9_7.

Laurent, C., T. Scholasch, B. Tisseyre, and A. Metay (2020). « Building New Temperature Indices for a Local Understanding of Grapevine Physiology. » XIIIth International Terroir Congress. Virtual Congress, Adelaïde, Asutralia, pp. 1–8.

Laurent, Cécile, Baptiste Oger, James Arnold Taylor, Thibaut Scholasch, Aurelie Metay, and Bruno Tisseyre (2021). « A Review of the Issues, Methods and Perspectives for Yield Estimation, Prediction and Forecasting in Viticulture ». *European Journal of Agronomy* 130, p. 126339. DOI: 10.1016/j.eja.2021.126339.

Lavee, S. and P. May (1997). « Dormancy of Grapevine Buds - Facts and Speculation ». *Australian Journal of Grape and Wine Research* 3.1, pp. 31–46. DOI: 10.1111/j.1755-0238.1997.tb00114.x.

Lavoie-Lamoureux, A., D. Sacco, P. A. Risse, and C. Lovisolo (2017). « Factors Influencing Stomatal Conductance in Response to Water Availability in Grapevine: A Meta-Analysis ». *Physiologia Plantarum* 159.4, pp. 468–482. DOI: 10.1111/ppl.12530.

Leolini, Luisa, Sergi Costafreda-Aumedes, João A. Santos, Christoph Menz, Helder Fraga, Daniel Molitor, Paolo Merante, Jürgen Junk, Thomas Kartschall, Agnès Destrac-Irvine, Cornelis van Leeuwen, Aureliano C. Malheiro, José Eiras-Dias, José Silvestre, Camilla Dibari, Marco Bindi, and Marco Moriondo (2020). « Phenological Model Intercomparison for Estimating Grapevine Budbreak Date (Vitis Vinifera L.) in Europe ». *Applied Sciences* 10.11 (11), p. 3800. DOI: 10.3390/app10113800.

Leroy, P., N. Smits, P. Cartolaro, L. Delière, J. P. Goutouly, M. Raynal, and A. Alonso Ugaglia (2013). « A Bioeconomic Model of Downy Mildew Damage on Grapevine for Evaluation of Control Strategies ». *Crop Protection* 53, pp. 58–71. DOI: 10.1016/j.cropro.2013.05.024.

Liu, S., S. Cossell, J. Tang, G. Dunn, and M. Whitty (2017). « A Computer Vision System for Early Stage Grape Yield Estimation Based on Shoot Detection ». *Computers and Electronics in Agriculture* 137, pp. 88–101. DOI: 10.1016/j.compag.2017.03.013.

Liu, S., X. Li, H. Wu, B. Xin, J. Tang, P. R. Petrie, and M. Whitty (2018). « A Robust Automated Flower Estimation System for Grape Vines ». *Biosystems Engineering* 172, pp. 110–123. DOI: 10.1016/j.biosystemseng.2018.05.009.

Liu, S. and M. Whitty (2015). « Automatic Grape Bunch Detection in Vineyards with an SVM Classifier ». *Journal of Applied Logic* 13 (4, Part 3), pp. 643–653. DOI: 10.1016/j.jal.2015.06.001.

Liu, S., X. Zeng, and M. Whitty (2020). « A Vision-Based Robust Grape Berry Counting Algorithm for Fast Calibration-Free Bunch Weight Estimation in the Field ». *Computers and Electronics in Agriculture* 173, p. 105360. DOI: 10.1016/j.compag.2020.105360.

Lobell, David B., Christopher B. Field, Kimberly Nicholas Cahill, and Celine Bonfils (2006). « Impacts of Future Climate Change on California Perennial Crop Yields: Model Projections with Climate and Crop Uncertainties ». *Agricultural and Forest Meteorology* 141.2-4, pp. 208–218.

Lopes, C. M., J. Graça, J. Sastre, M. Reyes, R. Guzmán, R. Braga, A. Monteiro, and P. A. Pinto (2016). « Vineyard Yield Estimation by VINBOT Robot - Preliminary Results with the White Variety Viosinho ». 11th International Terroir Congress. Ashland, USA: Southern Oregon University, pp. 458–463.

Mahmud, Kare P., Bruno P. Holzapfel, Yann Guisard, Jason P. Smith, Sharon Nielsen, and Suzy Y. Rogiers (2018). « Circadian Regulation of Grapevine Root and Shoot Growth and Their Modulation by Photope-

riod and Temperature ». *Journal of Plant Physiology* 222, pp. 86–93. DOI: 10.1016/j.jplph.2018.01.006.

Li-Mallet, A., A. Rabot, and L. Geny (2016). « Factors Controlling Inflorescence Primordia Formation of Grapevine: Their Role in Latent Bud Fruitfulness? A Review ». *Botany* 94.3, pp. 147–163. DOI: 10.1139/cjb-2015-0108.

Marchand, Fleur, Lies Debruyne, Laure Triste, Catherine Gerrard, Susanne Padel, and Ludwig Lauwers (2014). « Key Characteristics for Tool Choice in Indicator-Based Sustainability Assessment at Farm Level ». *Ecology and Society* 19.3. DOI: 10.5751/ES-06876-190346.

Martínez, L., C. Miranda, J. B. Royo, J. Urrestarazu, F. Martínez de Toda, P. Balda, and L. G. Santesteban (2016). « Direct and Indirect Effects of Three Virus Infections on Yield and Berry Composition in Grapevine (Vitis Vinifera L.) Cv. 'Tempranillo' ». *Scientia Horticulturae* 212, pp. 20–28. DOI: 10.1016/j.scienta.2016.09.023.

Martínez-Casasnovas, J. A., A. Martín-Montero, and M. Auxiliadora Casterad (2005). « Mapping Multi-Year Cropping Patterns in Small Irrigation Districts from Time-Series Analysis of Landsat TM Images ». *European Journal of Agronomy* 23.2, pp. 159–169. DOI: 10.1016/j.eja.2004.11.004.

Martínez-Zapater, J.M., P. Carbonell-Bejerano, C. Royo, R. Torres-Pérez, N. Diestro, J. Grimplet, N. Mauri, and J. Ibáñez (2019). « Genetic Variation for Grapevine Reproductive Development ». *Acta Horticulturae* 1248, pp. 319–326. DOI: 10.17660/ActaHortic.2019.1248.46.

Matthews, M.A. and M.M. Anderson (1989). « Reproductive Development in Grape (Vitis Vinifera L.): Responses to Seasonal Water Deficits. » *American Journal of Enology and Viticulture* 40, pp. 52–60.

May, P. (2000). « From Bud to Berry, with Special Reference to Inflorescence and Bunch Morphology in Vitis Vinifera L. » *Australian Journal of Grape and Wine Research* 6.2, pp. 82–98. DOI: 10.1111/j.1755-0238.2000.tb00166.x.

May, Peter (2004). *Flowering and Fruitset in Grapevines*. Phylloxera and Grape Industry Board of South Australia. 132 pp. ISBN: 978-0-9751260-6-6. Google Books: p9tJAAAAYAAJ.

McClymont, L., I. Goodwin, M. Mazza, N. Baker, D. M. Lanyon, A. Zerihun, S. Chandra, and M. O. Downey (2012). « Effect of Site-Specific Irrigation Management on Grapevine Yield and Fruit Quality Attributes ». *Irrigation Science* 30.6, pp. 461–470. DOI: 10.1007/s00271-012-0376-7.

Meneghetti, S., M. Gardiman, and A. Calo (2006). « Flower Biology of Grapevine. A Review. » *Advances in Horticultural Science* 20.4, pp. 317–325. DOI: 10.1400/56453.

Metay, A., J. Magnier, N. Guilpart, and A. Christophe (2015). « Nitrogen Supply Controls Vegetative Growth, Biomass and Nitrogen Allocation for Grapevine (Cv. Shiraz) Grown in Pots ». *Functional Plant Biology* 42.1, p. 105. DOI: 10.1071/FP14062.

Meyers, J. M., N. Dokoozlian, C. Ryan, C. Bioni, and J. E. Vanden Heuvel (2020). « A New, Satellite NDVI-Based Sampling Protocol for Grape Maturation Monitoring ». *Remote Sensing* 12.7 (7), p. 1159. DOI: 10.3390/rs12071159.

Millan, B., S. Velasco-Forero, A. Aquino, and J. Tardaguila (2018). « On-the-Go Grapevine Yield Estimation Using Image Analysis and Boolean Model ». *Journal of Sensors*, p. 9634752. DOI: 10.1155/2018/9634752.

Molitor, D. and M. Keller (2017). « Yield of Müller-Thurgau and Riesling Grapevines Is Altered by Meteorological Conditions in the Current and Previous Growing Seasons ». *OENO One* 50.4, pp. 245–258. DOI: 10.20870/oeno-one.2016.50.4.1071.

Molitor, Daniel, Helder Fraga, and Jürgen Junk (2020). « UniPhen – a Unified High Resolution Model Approach to Simulate the Phenological Development of a Broad Range of Grape Cultivars as Well as a Potential New Bioclimatic Indicator ». *Agricultural and Forest Meteorology* 291, p. 108024. DOI: 10.1016/j.agrformet.2020.108024.

Naor, A., Y. Gal, and B. Bravdo (2002). « Shoot and Cluster Thinning Influence Vegetative Growth, Fruit Yield and Wine Quality of 'Sauvignon Blanc' Grapevines ». *Journal of the American Society for Horticultural Science* 127.4, pp. 628–634. DOI: 10.21273/JASHS.127.4.628.

Naulleau, Audrey, Christian Gary, Laurent Prévot, and Laure Hossard (2020). « Evaluating Strategies for Adaptation to Climate Change in Grapevine Production-A Systematic Review ». *Frontiers in Plant Science* 11, p. 607859. DOI: 10.3389/fpls.2020.607859. pmid: 33519859.

Neethling, E., G. Barbeau, C. Coulon-Leroy, and H. Quénol (2019). « Spatial Complexity and Temporal Dynamics in Viticulture: A Review of Climate-Driven Scales ». *Agricultural and Forest Meteorology* 276–277, p. 107618. DOI: 10.1016/j.agrformet.2019.107618.

Nendel, Claas (2010). « Grapevine Bud Break Prediction for Cool Winter Climates ». *International Journal of Biometeorology* 54.3, pp. 231–241. DOI: 10.1007/s00484-009-0274-8.

Netzer, Y., S. Munitz, I. Shtein, and A. Schwartz (2019). « Structural Memory in Grapevines: Early Season Water Availability Affects Late Season Drought Stress Severity ». *European Journal of Agronomy* 105, pp. 96–103. DOI: 10.1016/j.eja.2019.02.008.

Nogueira Júnior, A. F., L. Amorim, S. Savary, and L. Willocquet (2018). « Modelling the Dynamics of Grapevine Growth over Years ». *Ecological Modelling* 369, pp. 77–87. DOI: 10.1016/j.ecolmodel.2017.12.016.

Nuske, S., S. Achar, T. Bates, S. Narasimhan, and S. Singh (2011). « Yield Estimation in Vineyards by Visual Grape Detection ». *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, CA, USA, pp. 2352–2358. DOI: 10.1109/IROS.2011.6095069.

Nuske, S., K. Wilshusen, S. Achar, L. Yoder, S. Narasimhan, and S. Singh (2014). « Automated Visual Yield Estimation in Vineyards: Automated Visual Yield Estimation ». *Journal of Field Robotics* 31.5, pp. 837–860. DOI: 10.1002/rob.21541.

Oger, B., P. Vismara, and B. Tisseyre (2021). « Combining Target Sampling with within Field Route-Optimization to Optimise on Field Yield Estimation in Viticulture. » *Precision Agriculture* 22, pp. 432–451. DOI: 10.1007/s11119-020-09744-0.

Ojeda, H., A. Deloire, and A. Carbonneau (2001). « Influence of Water Deficits on Grape Berry Growth ». *Vitis* 40.3, pp. 141–145. DOI: 10.5073/vitis.2001.40.141-145.

Oliver, M., T. Bishop, and B. Marchant (2013). *Precision Agriculture for Sustainability and Environmental Protection*. Routledge. 302 pp. ISBN: 978-1-136-46825-4. Google Books: ep4lAgAAQBAJ.

Ollat, N., J. P. Carde, J. P. Gaudillère, F. Barrieu, P. Diakou-Verdin, and A. Moing (2002). « Grape Berry Development : A Review ». *OENO One* 36.3, pp. 109–131. DOI: 10.20870/oeno-one.2002.36.3.970.

Ouadi, L., E. Bruez, S. Bastien, J. Vallance, p. Lecomte, J. C. Domec, and P. Rey (2019). « Ecophysiological Impacts of Esca, a Devastating Grapevine Trunk Disease, on Vitis Vinifera L. » *PLoS ONE* 14.9, e0222586. DOI: 10.1371/journal.pone.0222586.

Pagay, V. and C. Collins (2017). « Effects of Timing and Intensity of Elevated Temperatures on Reproductive Development of Field-Grown Shiraz Grapevines ». *OENO One* 51.4. DOI: 10.20870/oeno-one.2017.51.4.1066.

Palacios, F., G. Bueno, J. Salido, M. P. Diago, I. Hernández, and J. Tardaguila (2020). « Automated Grapevine Flower Detection and Quantification Method Based on Computer Vision and Deep Learning from On-the-Go Imaging Using a Mobile Sensing Platform under Field Conditions ». *Computers and Electronics in Agriculture* 178, p. 105796. DOI: 10.1016/j.compag.2020.105796.

Parker, A. K., I. García de Cortázar-Atauri, L. Gény, J. L. Spring, A. Destrac, Hans Schultz, Daniel Molitor, Thierry Lacombe, Antonio Graça, Christine Monamy, Manfred Stoll, Paolo Storchi, Mike C. T. Trought, Rainer W. Hofmann, and Cornelis van Leeuwen (2020). « Temperature-Based Grapevine Sugar Ripeness Modelling for a Wide Range of Vitis Vinifera L. Cultivars ». *Agricultural and Forest Meteorology* 285–286, p. 107902. DOI: 10.1016/j.agrformet.2020.107902.

Parker, A.K., I.G. De Cortázar-Atauri, C. Van Leeuwen, and I. Chuine (2011). « General Phenological Model to Characterise the Timing of Flowering and Veraison of Vitis Vinifera L.: Grapevine Flowering and Veraison Model ». *Australian Journal of Grape and Wine Research* 17.2, pp. 206–216. DOI: 10.1111/j.1755-0238.2011.00140.x.

Parkes, B., T. P. Higginbottom, K. Hufkens, F. Ceballos, B. Kramer, and T. Foster (2019). « Weather Dataset Choice Introduces Uncertainty to Estimates of Crop Yield Responses to Climate Variability and Change ». 14.12, p. 124089. DOI: 10.1088/1748-9326/ab5ebb.

Petrie, P. R. and P. r. Clingeleffer (2005). « Effects of Temperature and Light (before and after Budburst) on Inflorescence Morphology and Flower Number of Chardonnay Grapevines (Vitis Vinifera L.) » *Australian Journal of Grape and Wine Research* 11.1, pp. 59–65. DOI: 10.1111/j.1755-0238.2005.tb00279.x.

Petrie, Paul R., Michael C. T. Trought, and G. Stanley Howell (2000). « Growth and Dry Matter Partitioning of Pinot Noir (Vitis Vinifera L.) in Relation to Leaf Area and Crop Load ». *Australian Journal of Grape and Wine Research* 6.1, pp. 40–45. DOI: 10.1111/j.1755-0238.2000.tb00160.x.

Pichon, L., G. Brunel, J. C. Payan, J. Taylor, V. Bellon-Maurel, and B. Tisseyre (2021). « ApeX-Vigne: Experiences in Monitoring Vine Water Status from within-Field to Regional Scales Using Crowdsourcing Data from a Free Mobile Phone Application ». *Precision Agriculture* 22, pp. 608–626. DOI: 10.1007/s11119-021-09797-9.

Poni, S. and M. Gatti (2017). « Affecting Yield Components and Grape Composition through Manipulations of the Source-Sink Balance ». *Acta Horticulturae* 1188, pp. 21–34. DOI: 10.17660/ActaHortic.2017.1188.4.

Pothen, Z. and S. Nuske (2016). « Automated Assessment and Mapping of Grape Quality through Image-Based Color Analysis ». *IFAC-PapersOnLine*. 5th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2016 49.16, pp. 72–78. DOI: 10.1016/j.ifacol.2016.10.014.

Pouget, R. (1981). « Action de La Température Sur La Differenciation Des Inflorescences et Des Fleurs Durant Les Phases de Pre-Debourrement et Post-Debourrement Des Bourgeons Latents de La Vigne. » *Journal International des Sciences de la Vigne et du Vin* 15.2, pp. 65–79. DOI: 10.20870/oeno-one.1981.15.2.1791.

Prats-Llinàs, Maria Teresa, Héctor Nieto, Theodore M. DeJong, Joan Girona, and Jordi Marsal (2020). « Using Forced Regrowth to Manipulate Chardonnay Grapevine (Vitis Vinifera L.) Development to Evaluate Phenological Stage Responses to Temperature ». *Scientia Horticulturae* 262, p. 109065. DOI: 10.1016/j.scienta.2019.109065.

Pratt, C. (1971). « Reproductive Anatomy in Cultivated Grapes - A Review ». *American Journal of Enology and Viticulture* 22.2, pp. 92–109.

Primicerio, J., G. Caruso, L. Comba, A. Crisci, P. Gay, S. Guidoni, L. Genesio, D. R. Aimonino, and F. P. Vaccari (2017). « Individual Plant Definition and Missing Plant Characterization in Vineyards from High-Resolution UAV Imagery ». *European Journal of Remote Sensing* 50.1, pp. 179–186. DOI: 10.1080/22797254.2017.1308234.

Ramsey, M. H., S. L. R. Ellison, and P. Rostron (2019). *Eurachem/EUROLAB/ CITAC/Nordtest/AMC Guide: Measurement Uncertainty Arising from Sampling: A Guide to Methods and Approaches.* Eurachem. ISBN: 978-0-948926-35-8.

Ravaz, L. and L. Sicard (1911). « L'effeuillage de La Vigne ». *Annales de l'École nationale d'agriculture de Montpellier* 11.

Rawnsley, B. and C. Collins (2005). *Improving Vineyard Productivity through Assessment of Bud Fruitfulness and Bud Necrosis*. South Australian Research and Development Institute (SARDI).

Reis, M.J.C.S., R. Morais, E. Peres, C. Pereira, O. Contente, S. Soares, A. Valente, J. Baptista, P.J.S.G. Ferreira, and J. Bulas Cruz (2012). « Automatic Detection of Bunches of Grapes in Natural Environment from Color Images ». *Journal of Applied Logic* 10.4, pp. 285–290. DOI: 10.1016/j.jal.2012.07.004.

Reynolds, A. G. and J. E. Vanden Heuvel (2009). « Influence of Grapevine Training Systems on Vine Growth and Fruit Composition: A Review ». *American Journal of Enology and Viticulture* 60, pp. 251–268.

Rienth, M. and T. Scholasch (2019). « State-of-the-Art of Tools and Methods to Assess Vine Water Status ». *OENO One* 53.4 (4). DOI: 10.20870/oeno-one.2019.53.4.2403.

Rienth, Markus, Laurent Torregrosa, Nathalie Luchaire, Ratthaphon Chatbanyong, David Lecourieux, Mary T. Kelly, and Charles Romieu (2014). « Day and Night Heat Stress Trigger Different Transcriptomic Responses in Green and Ripening Grapevine (Vitis Vinifera) Fruit ». *BMC Plant Biology* 14.1, p. 108. DOI: 10.1186/1471-2229-14-108.

Ripoche, A., A. Metay, F. Celette, and C. Gary (2011). « Changing the Soil Surface Management in Vineyards: Immediate and Delayed Effects on the Growth and Yield of Grapevine ». *Plant and Soil* 339, pp. 259–271. DOI: 10.1007/s11104-010-0573-1.

Robbez Masson, J.M. and J.C. Foltete (2005). « Localising Missing Plants in Squared-Grid Patterns of Discontinuous Crops from Remotely Sensed Imagery ». *Computers and Geosciences* 31.7, pp. 900–912. DOI: 10.1016/j.cageo.2005.02.013.

Robert, C. (1992). *L'analyse statistique bayésienne (The Bayesian Statistical Analysis)*. Paris: Economica. 393 pp. ISBN: 978-2-7178-2199-4.

Rodriguez Lovelle, B., W. Trambouze, and O. Jacquet (2009). « Evaluation of vegetative growth stage by the "shoot tip method". » *Progrès Agricole et Viticole* 126.4, pp. 77–88.

Roscher, R., K. Herzog, A. Kunkel, A. Kicherer, R. Töpfer, and W. Förstner (2014). « Automated Image Analysis Framework for the High-Throughput Determination of Grapevine Berry Sizes Using Conditional Random Fields ». *Computers and Electronics in Agriculture* 100, pp. 148–158. DOI: 10.1016/j.compag.2013.11.008. arXiv: 1712.05647.

Rose, J. C., A. Kicherer, M. Wieland, L. Klingbeil, R. Töpfer, and H. Kuhlmann (2016). « Towards Automated Large-Scale 3D Phenotyping of Vineyards under Field Conditions ». *Sensors* 16.2, p. 2136. DOI: 10.3390/s16122136. pmid: 27983669.

Sadras, V., M. Moran, and P. Petrie (2017). « Resilience of Grapevine Yield in Response to Warming ». *OENO One* 51.4. DOI: 10.20870/oeno-one.2017.51.4.1913.

Sadras, V. O. and M. A. Moran (2013). « Nonlinear Effects of Elevated Temperature on Grapevine Phenology ». *Agricultural and Forest Meteorology* 173, pp. 107–115. DOI: 10.1016/j.agrformet.2012.10.003.

Sánchez, Luis A. and Nick K. Dokoozlian (2005). « Bud Microclimate and Fruitfulness in Vitis Vinifera L. » *American Journal of Enology and Viticulture* 56.4, pp. 319–329.

Santesteban, L. G., C. Miranda, I. Barbarin, and J. B. Royo (2015). « Application of the Measurement of the Natural Abundance of Stable Isotopes in Viticulture: A Review ». *Australian Journal of Grape and Wine Research* 21.2, pp. 157–167. DOI: 10.1111/ajgw.12124.

Saporta, G. (2011). *Probabilités, analyse des données et statistique*. Editions TECHNIP. 660 pp. ISBN: 978-2-7108-0980-7. Google Books: VFoGF97GPiwC.

Savary, S., A. D. Nelson, A. Djurle, P. D. Esker, A. Sparks, L. Amorim, A. Bergamin Filho, T. Caffi, N. Castilla, K. Garrett, N. McRoberts, V. Rossi, J. Yuen, and L. Willocquet (2018). « Concepts, Approaches, and Avenues for Modelling Crop Health and Crop Losses ». *European Journal of Agronomy*. Recent Advances in Crop Modelling to Support Sustainable Agricultural Production and Food Security under Global Change 100, pp. 4–18. DOI: 10.1016/j.eja.2018.04.003.

Schwarz, Gideon (1978). « Estimating the Dimension of a Model ». *The Annals of Statistics* 6.2, pp. 461–464. DOI: 10.1214/aos/1176344136.

Serrano, E., S. Roussel, L. Gontier, and T. Dufourcq (2005). « Early Estimation of Vineyard Yield : Correlation Between the Volume of a Vitis Vinifera Bunch During Its Growth and Its Weight at Harvest ». XIV International GESCO Viticulture Congress. Geisenheim, Germany, pp. 311–318.

Serrano, L., C. González-Flor, and G. Gorchs (2012). « Assessment of Grape Yield and Composition Using the Reflectance Based Water Index in Mediterranean Rainfed Vineyards ». *Remote Sensing of Environment* 118, pp. 249–258. DOI: 10.1016/j.rse.2011.11.021.

Silva, João Vasco and Ken E. Giller (2020). « Grand Challenges for the 21st Century: What Crop Models Can and Can't (yet) Do ». *The Journal of Agricultural Science* 158.10, pp. 794–805. DOI: 10.1017/S0021859621000150.

Sirsat, M. S., J. Mendes-Moreira, C. Ferreira, and M. Cunha (2019). « Machine Learning Predictive Model of Grapevine Yield Based on Agroclimatic Patterns ». *Engineering in Agriculture, Environment and Food* 12.4, pp. 443–450. DOI: 10.1016/j.eaef.2019.07.003.

Srinivasan, C. and M. G. Mullins (1981). « Physiology of Flowering in the Grapevine - A Review ». *American Journal of Enology and Viticulture* 32.1, pp. 47–63.

Staudt, G. (1982). « Pollen Germination and Pollen Tube Growth in Vivo and the Dependence on Temperature. » *Vitis* 21, pp. 205–216.

Sun, L., F. Gao, M. Anderson, W. Kustas, M. Alsina, L. Sanchez, B. Sams, L. McKee, W. Dulaney, W. White, J. Alfieri, J. Prueger, F. Melton, and K. Post (2017). « Daily Mapping of 30 m LAI and NDVI for Grape Yield Prediction in California Vineyards ». *Remote Sensing* 9.4, p. 317. DOI: 10.3390/rs9040317.

Suter, Bruno, Roberta Triolo, David Pernet, Zhanwu Dai, and Cornelis Van Leeuwen (2019). « Modeling Stem Water Potential by Separating the Effects of Soil Water Availability and Climatic Conditions on Water Status in Grapevine (Vitis Vinifera L.) » *Frontiers in Plant Science* 10. DOI: 10.3389/fpls.2019.01485.

Tarara, Julie M., Bernardo Chaves, Luis A. Sanchez, and Nick K. Dokoozlian (2014). « Use of Cordon Wire Tension for Static and Dynamic Prediction of Grapevine Yield ». *American Journal of Enology and Viticulture* 65, pp. 443–452. DOI: 10.5344/ajev.2014.14021.

Tardaguila, Javier, Manfred Stoll, Salvador Gutiérrez, Tony Proffitt, and Maria P. Diago (2021). « Smart Applications and Digital Technologies in Viticulture: A Review ». *Smart Agricultural Technology* 1, p. 100005. DOI: 10.1016/j.atech.2021.100005.

Tardáguila, J., M. P. Diago, F. M. Toda, S. Poni, and M. Vilanova (2008). « Effects of Timing of Leaf Removal on Yield, Berry Maturity, Wine Composition and Sensory Properties of Cv. Grenache Grown under Non Irrigated Conditions ». *OENO One* 42.4, pp. 221–229. DOI: 10.20870/oeno-one.2008.42.4.810.

Taylor, J. A., L. Sanchez, B. Sams, L. Haggerty, R. Jakubowski, S. Djafour, and T. R. Bates (2016). « Evaluation of a Commercial Grape Yield Monitor for Use Mid-Season and at-Harvest ». *OENO One* 50.2. DOI: 10.20870/oeno-one.2016.50.2.784.

Taylor, J. A., B. Tisseyre, R. Bramley, and A. Reid (2005). « A Comparison of the Spatial Variability of Vineyard Yield in European and Australian Production Systems ». *Precision Agriculture '05*. 5th European Conference on Precision Agriculture. Uppsala, Sweden., pp. 907–914.

Tilbrook, J. and S. D. Tyerman (2009). « Hydraulic Connection of Grape Berries to the Vine: Varietal Differences in Water Conductance into and out of Berries, and Potential for Backflow ». *Functional Plant Biology* 36.6, pp. 541–550. DOI: 10.1071/FP09019.

Tisseyre, B., C. Mazzoni, and H. Fonta (2008). « Within-Field Temporal Stability of Some Parameters in Viticulture: Potential toward a Site Specific Management ». *Journal International des Sciences de la Vigne et du Vin* 42.1. DOI: 10.20870/oeno-one.2008.42.1.834.

Tonietto, J. and A. Carbonneau (2004). « A Multicriteria Climatic Classification System for Grape-Growing Regions Worldwide ». *Agricultural and Forest Meteorology* 124.1-2, pp. 81–97. DOI: 10.1016/j.agrformet.2003.06.001.

Tourmeau, J. (1976). « Différenciation Des Ébauches Florales Chez La Vigne. » MS Thesis. France: Université de Dijon.

Triolo, R., J. P. Roby, A. Pisciotta, R. Di Lorenzo, and C. Van Leeuwen (2019). « Impact of Vine Water Status on Berry Mass and Berry Tissue Development of Cabernet Franc (Vitis Vinifera L.), Assessed at Berry Level ». *Journal of the Science of Food and Agriculture* 99.13, pp. 5711–5719. DOI: 10.1002/jsfa.9834.

Uribeetxebarria, A., J. A. Martínez-Casasnovas, B. Tisseyre, S. Guillaume, A. Escolà, J. R. Rosell-Polo, and J. Arnó (2019). « Assessing Ranked Set Sampling and Ancillary Data to Improve Fruit Load Estimates in Peach Orchards ». *Computers and Electronics in Agriculture* 164, p. 104931. DOI: 10.1016/j.compag.2019.104931.

Vaillant-Gaveau, N., G. Wojnarowiez, A. N. Petit, L. Jacquens, L. Panigai, C. Clement, and F. Fontaine (2014). « Relationships between Carbohydrates and Reproductive Development in Chardonnay Grapevine: Impact of Defoliation and Fruit Removal Treatments during Four Successive Growing Seasons ». *OENO One* 48.4, pp. 219–229. DOI: 10.20870/oeno-one.2014.48.4.1694.

Valdés-Gómez, H., F. Celette, I. Garcia De Cortazar Atauri, f. Jara-Rojas, S. Ortega-Farías, and C. Gary (2009). « Modelling Soil Water Content and Grapevine Growth and Development with the Stics Crop-Soil Model under Two Different Water Management Strategies ». *OENO One* 43.1, pp. 13–28. DOI: 10.20870/oeno-one.2009.43.1.806.

Van Ittersum, M. K., K. G. Cassman, P. Grassini, J. Wolf, P. Tittonell, and Z. Hochman (2013). « Yield Gap Analysis with Local to Global Relevance—A Review ». *Field Crops Research*. Crop Yield Gap Analysis – Rationale, Methods and Applications 143, pp. 4–17. DOI: 10.1016/j.fcr.2012.09.009.

Van Leeuwen, C., P. Pieri, M. Gowdy, N. Ollat, and J. P. Roby (2019). « Reduced Density Is an Environmental Friendly and Cost Effective Solution to Increase Resilience to Drought in Vineyards in a Context of Climate Change: » *OENO One* 53.2, pp. 129–146. DOI: 10.20870/oeno-one.2019.53.2.2420.

Van Leeuwen, c., J. P. Roby, and L. De Rességuier (2018). « Soil-Related Terroir Factors: A Review ». *OENO One* 52.2, pp. 173–188. DOI: 10.20870/oeno-one.2018.52.2.2208.

Vasconcelos, M. C., M. Greven, C. S. Winefield, M. C. T. Trought, and V. Raw (2009). « The Flowering Process of Vitis Vinifera: A Review ». *American Journal of Enology and Viticulture* 60.4, pp. 411–434.

Verdugo Vásquez, N., C. Acevedo Opazo, H. Valdés Gómez, M. Araya Alman, B. Ingram, I. García de Cortázar, and B. Tisseyre (2015). « Spatial Variability of Phenology in Two Irrigated Grapevine Cultivar Growing under Semi-Arid Conditions ». *Precision Agriculture* 17.2, pp. 218–245. DOI: 10.1007/s11119-015-9418-5.

Verdugo-Vásquez, N., C. Acevedo-Opazo, H. Valdés-Gómez, I. García de Cortázar-Atauri, and B. Tisseyre (2017). « Assessment of an Empirical Spatio-Temporal Model of the Grapevine Phenology at the within-Field Scale ». *Advances in Animal Biosciences* 8.2, pp. 534–539. DOI: 10.1017/S2040470017000097.

Verdugo-Vásquez, N., C. Acevedo-Opazo, H. Valdés-Gómez, B. Ingram, I. García de Cortázar-Atauri, and B. Tisseyre (2020). « Towards an Empirical Model to Estimate the Spatial Variability of Grapevine Phenology at the within Field Scale ». *Precision Agriculture* 21.1, pp. 107–130. DOI: 10.1007/s11119-019-09657-7.

Vigiuer, L., L. Bedoussac, E. P. Journet, and E. Justes (2018). « Yield Gap Analysis Extended to Marketable Grain Reveals the Profitability of Organic Lentil-Spring Wheat Intercrops ». *Agronomy for Sustainable Development* 38.4, p. 39. DOI: 10.1007/s13593-018-0515-5.

Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. Krayer von Krauss (2003). « Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support ». *Integrated Assessment* 4.1, pp. 5–17. DOI: 10.1076/iaij.4.1.5.16466.

Wang, J. Y. (1960). « A Critique of the Heat Unit Approach to Plant Response Studies ». *Ecology* 41.4, pp. 785–790. DOI: 10.2307/1931815. JSTOR: 1931815.

Watanabe, Sumio (2013). « A Widely Applicable Bayesian Information Criterion ». *Journal of Machine Learning Research* 14 (Mar), pp. 867–897.

Webster, R. and M. A. Oliver (1992). « Sample Adequately to Estimate Variograms of Soil Properties ». *Journal of Soil Science* 43.1, pp. 177–192. DOI: 10.1111/j.1365-2389.1992.tb00128.x.

Weiss, M., F. Jacob, and G. Duveiller (2020). « Remote Sensing for Agricultural Applications: A Meta-Review ». *Remote Sensing of Environment* 236, p. 111402. DOI: 10.1016/j.rse.2019.111402.

Wolpert, J. A. and E. P. Vilas (1992). « Estimating Vineyard Yields: Introduction to a Simple, Two-Step Method ». *American Journal of Enology and Viticulture* 43.4, pp. 384–388.

Wonnacott, Thomas H. and Ronald J. Wonnacott (1990). *Introductory Statistics for Business and Economics*. Wiley. 846 pp. ISBN: 978-0-471-61517-0. Google Books: D0K0AAAAIAAJ.

Wulfsohn, D., F. Aravena-Zamora, C. Potin-Téllez, I. Zamora, and M. Garcia-Finana (2012). « Multilevel Systematic Sampling to Estimate Total Fruit Number for Yield Forecasts. » *Precision Agriculture* 13.2, pp. 256–275. DOI: 10.1007/s11119-011-9245-2.

Yan, W. and L.A. Hunt (1999). « An Equation for Modelling the Temperature Response of Plants Using Only the Cardinal Temperatures ». *Annals of Botany* 84.5, pp. 607–614. DOI: 10.1006/anbo.1999.0955.

Young, Madaline D., Gerard H. Ros, and Wim de Vries (2021). « A Decision Support Framework Assessing Management Impacts on Crop Yield, Soil Carbon Changes and Nitrogen Losses to the Environment ». *European Journal of Soil Science* 72.4, pp. 1590–1606. DOI: 10.1111/ejss.13024.

Zabawa, L., A. Kicherer, L. Klingbeil, R. Töpfer, H. Kuhlmann, and R. Roscher (2020). « Counting of Grapevine Berries in Images via Semantic Segmentation Using Convolutional Neural Networks ». *ISPRS Journal of Photogrammetry and Remote Sensing* 164, pp. 73–83. DOI: 10.1016/j.isprsjprs.2020.04.002.

Zapata, C., E. Deléens, S. Chaillou, and C. Magné (2004). « Partitioning and Mobilization of Starch and N Reserves in Grapevine (Vitis Vinifera L.) » *Journal of Plant Physiology* 161.9, pp. 1031–1040. DOI: 10.1016/j.jplph.2003.11.009.

Zapata, D., M. Salazar-Gutierrez, B. Chaves, m. Keller, and G. Hoogenboom (2017). « Predicting Key Phenological Stages for 17 Grapevine Cultivars (Vitis Vinifera L.) » *American Journal of Enology and Viticulture* 68.1, pp. 60–72. DOI: 10.5344/ajev.2016.15077.

Zapata, D., M. Salazar-Gutierrez, B. Chaves, Markus Keller, and G. Hoogenboom (2015). « Estimation of the Base Temperature and Growth Phase Duration in Terms of Thermal Time for Four Grapevine Cultivars ». *International Journal of Biometeorology* 59.12, pp. 1771–1781. DOI: 10.1007/s00484-015-0985-y.

Zhou, Guanglin and Quanjiu Wang (2018). « A New Nonlinear Method for Calculating Growing Degree Days ». *Scientific Reports* 8.1 (1), p. 10149. DOI: 10.1038/s41598-018-28392-z.

Zhu, J., R. Fraysse, M. C. t. Trought, V. Raw, L. Yang, M. Greven, D. Martin, and R. Agnew (2020). « Quantifying the Seasonal Variations in Grapevine Yield Components Based on Pre- and Post-Flowering Weather Conditions ». *OENO One* 54.2 (2). DOI: 10.20870/oeno-one.2020.54.2.2926.

Zhu, J., M. Génard, S. Poni, G. A. Gambetta, P. Vivin, G. Vercambre, M. C. T. Trought, N. Ollat, S. Delrot, and Z. Dai (2018). « Modelling Grape Growth in Relation to Whole-Plant Carbon and Water Fluxes ». *Journal of Experimental Botany* 70.9, pp. 2505–2521. DOI: 10.1093/jxb/ery367.

Abstract

Farm data are data commonly collected in commercial farms in the course of their everyday operation. They offer the advantage of accurately representing the production context of a given farm by providing information on the crop system and management, environment, etc. but they are characterized by lower data quality and a parsimonious number of variables. **This study is based on the hypothesis that valuing farm data with adapted statistical methods can help in performing agronomical research that complies with the operational needs and constraints of a commercial farm and result in the delivery of actionable site-specific information.** This PhD thesis was founded by a precision viticulture company, Fruition Sciences, and therefore aimed at investigating this hypothesis in the viticultural industry with the case study of yield forecasting. The main focus was placed on leveraging time series of weather data because they are commonly encountered in farm data sets and because climate influence on grapevine yield development is prevalent. **All the methods proposed in this PhD project constitute an original framework for valuing farm data with various possible applications in agriculture.**

Keywords

Grapevine, Wine, Estimation, Prediction, Time Series of Weather Data, Yield Development, Bayesian

Résumé

Les données agricoles sont des données couramment collectées dans les exploitations agricoles au cours de leur fonctionnement quotidien. Elles présentent l'avantage de représenter fidèlement le contexte de production d'une exploitation donnée en fournissant des informations sur le système et la conduite des cultures, l'environnement, etc. mais elles sont caractérisées par une moindre qualité et un nombre parcimonieux de variables. **Cette thèse est basée sur l'hypothèse que la valorisation des données agricoles avec des méthodes statistiques adaptées peut aider à réaliser des recherches agronomiques tenant compte des besoins et contraintes opérationnelles d'une exploitation et aboutir à la livraison d'informations locales et pertinentes pour l'aide à la décision sur le terrain.** Ce projet a été initié par une entreprise de viticulture de précision, Fruition Sciences, et visait donc à étudier cette hypothèse au sein de la filière viticole à travers l'étude de cas de la prévision du rendement. L'accent a été mis sur l'exploitation des séries temporelles de données météorologiques parce qu'elles sont couramment rencontrées dans les jeux de données agricoles et parce que l'influence du climat sur le développement du rendement de la vigne est prédominante. **Toutes les méthodes proposées dans ce projet de thèse constituent un cadre original pour la valorisation des données agricoles avec diverses applications possibles en agriculture.**

Mots clés

Vigne, Vin, Estimation, Prédiction, Séries Temporelles de Données Météorologiques,
Elaboration du Rendement, Bayésien