THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Ecologie Fonctionnelle et Sciences Agronomiques

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau
Portée par l'Université de Montpellier

Unité de recherche ITAP – Technologies & méthodes pour les agricultures de demain

Development of a crowdsourcing approach adapted to agriculture specificities: the case of monitoring vine water status with the shoot growth approach

Présentée par Léo PICHON

Le 13 Juillet 2021

Sous la direction de Véronique BELLON-MAUREL
et Bruno TISSEYRE

Devant le jury composé de

Marianne CERF, Directrice de recherche, INRAE, France	**Examinatrice**
Romain JULLIARD, Directeur de recherche, CNRS-MNHN, France	**Rapporteur**
Cornelis VAN LEEUWEN, Professeur, Bordeaux Sciences Agro, France	**Rapporteur**
Véronique BELLON-MAUREL, IPEF, INRAE, France	**Directrice de thèse**
Bruno TISSEYRE, Professeur, Institut Agro – Montpellier SupAgro, France	**Invité**

UNIVERSITÉ DE MONTPELLIER

Montpellier SupAgro

Remerciements

Dans ce manuscrit, j'étudie comment un groupe d'individus peut créer de la connaissance en mettant en commun ses observations. Pour moi, le collectif est un moyen de s'élever, de voir plus loin, de servir une cause plus grande que celle de son propre intérêt. C'est aussi un moyen de réaliser des choses qu'on ne pourrait pas faire tout seul. De les faire mieux et de les rendre plus belles parce qu'elles sont partagées. Ce manuscrit a pu être réalisé car j'appartiens à des groupes qui m'ont aidé, qui m'ont inspiré et qui ont alimenté directement ou indirectement ce travail. Je suis très fier de leur consacrer les premières pages de ce document.

Je pense bien sûr d'abord à l'équipe AgroTIC parce que c'est celle avec laquelle je vis au quotidien. Le travail en groupe n'est pas toujours facile, en particulier en contexte covid, mais j'ai la chance d'appartenir à un collectif vivant et soudé. Quel que soit le problème que je rencontre, je pense toujours à au moins trois ou quatre personnes que je peux solliciter à l'improviste et qui feront tout ce qu'elles peuvent pour m'aider. Je n'oublie pas que si j'ai pu dégager du temps pour mes activités de recherche et pour mener à bien cette thèse c'est parce que le groupe a pris en charge une partie de mes responsabilités et qu'il a absorbé les sollicitations que je pouvais avoir. En un mot, qu'il m'a protégé. Merci à tous pour ça. Merci à Yoann, Nina, Guilhem, Sarah, Thomas, Bruno, Laure, Simon, Laetitia et Arnaud. J'adresse un clin d'œil à Philippe, Hazaël et Nicolas qui contribuent eux aussi à la vie de ce groupe. Je remercie particulièrement Laure pour toutes ses petites attentions et Thomas pour toutes les discussions qu'on a eues et qui m'ont fait prendre du recul sur le boulot et le reste. Je tiens à remercier chaleureusement Simon pour tout le travail d'ingénierie pédagogique qu'il a fait et qui m'a dégagé un temps précieux. Je lui adresse pour cela un bisou-cœur (cordial). Je remercie également Christèle qui est la première à m'avoir dit « *Tu veux faire une thèse ? Bah, vas-y, fonce !* ».

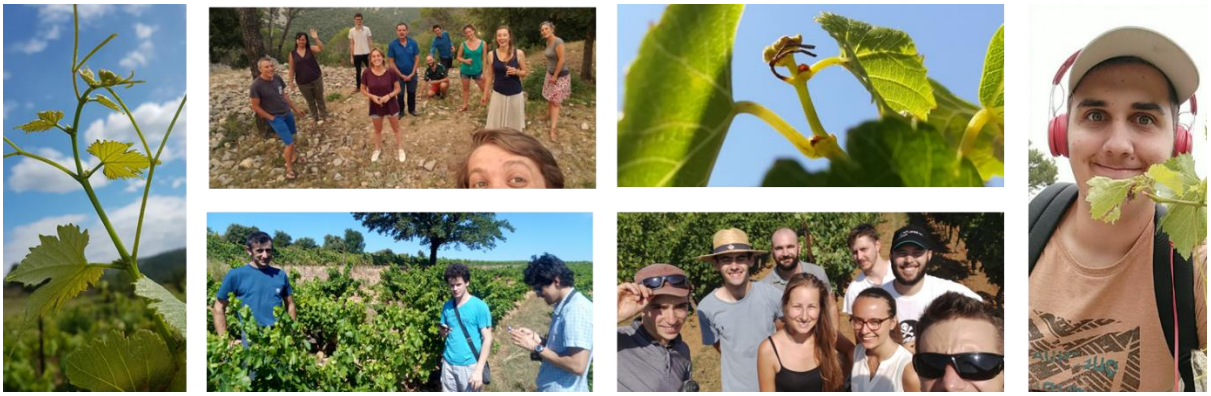
Je pense aussi à l'équipe Démo et en particulier aux doctorants. Merci à Cécile pour avoir su faire émerger ces JeudisDesDoctorants, un espace où l'on parle de science avec bienveillance. Dans l'ordre des générations, je remercie Julien et Corentin dont l'exigence et les questions perpétuelles m'ont d'abord déstabilisé puis inspiré ; Baptiste, Cécile et Anice pour nos longs échanges toujours bienveillants et constructifs qui m'ont qui fait grandir en tant qu'aspirant chercheur ; et dernièrement Eva puis Florian et Daniel avec qui j'espère pouvoir travailler encore dans les mois et les années à venir. Je tiens à remercier également James, notre animateur d'équipe Démo et Véro, ma directrice de thèse. Tous les deux, vous avez fait preuve à mon égard de beaucoup de bienveillance et de pédagogie. Ce sont des valeurs qui comptent pour moi. Que des scientifiques comme vous puissent les partager m'a donné confiance en moi et en la recherche.

Je remercie évidemment l'équipe du projet ApeX-Vigne : Jean Christophe et Guilhem. Merci pour vos compétences, bien sûr, mais merci aussi pour votre investissement. Le projet ApeX-Vigne vous doit beaucoup. Guilhem d'abord qui a développé l'application. Merci pour ta souplesse, ta rigueur, tes idées. Avec toi tout parait simple et un projet comme celui-ci est presque un long fleuve tranquille. C'est juste un plaisir de travailler avec toi. Jean Christophe ensuite qui a largement contribué à la diffusion d'ApeX-Vigne. Je pense sincèrement que si tu n'avais pas testé avec acharnement l'application et que tu n'en avais pas parlé sur tout l'arc méditerranéen, la profession ne se la serait tout simplement pas appropriée. Ton infatigable bonne humeur et ta volonté d'accompagner des projets collectifs au service de la profession donnent du sens à nos projets.

Enfin, merci à tous les participants de l'application ApeX-Vigne dont le plus assidu, Octave, et les 800 autres, sans qui tout cela n'aurait pas beaucoup de sens.

Bruno. Tu es celui qui, sans conteste, mérite le plus de remerciements pour cette thèse et tout ce qu'elle englobe. Merci pour la confiance que tu m'as accordée y'a longtemps quand pas grand monde (y compris moi) ne croyait en mes capacités. Merci de m'avoir fait entrevoir la beauté de la recherche en m'emmenant (presque contre mon gré) à mon premier congrès scientifique. C'est là-bas que j'ai compris que je ne voulais pas seulement faire de l'enseignement. Je voulais aussi faire de la recherche. C'est aussi là-bas que j'ai compris ce qu'était une communauté scientifique et le plaisir qu'on pouvait ressentir à en faire partie et à y contribuer. Une fois que j'avais compris ça, la suite ce n'était qu'une question de temps. J'te remercie de me l'avoir laissé, ce temps. De ne pas t'être découragé. Merci pour ton exigence et ta bienveillance. Merci aussi pour tes infatigables relectures, en particulier ces dernières semaines où tu as donné de ta personne pour être là quand j'en avais besoin. Enfin, même si cela dépasse le cadre de cette thèse, merci car je viens au boulot depuis bientôt 10 ans avec le sourire aux lèvres. Y'a pas grand-chose qui vaut autant que ça.

Et puis parce qu'une thèse, ça ne se passe pas tout à fait qu'au travail, je tiens à remercier tous mes proches qui m'ont accompagné et aidé. Je remercie tout d'abord mes parents. Merci à tous les deux de nous avoir aidés, Clo et moi, à gérer ces derniers mois à la logistique disons...inédite. Merci P'pa. Je crois que sans le goût du défi que tu m'as transmis, je n'aurais pas réussi à aller jusqu'au bout de la rédaction de cette thèse. Merci M'man. C'est toi la première qui a perçu l'intérêt que je pourrais avoir pour la recherche, 10 ans (!!) avant que je m'en rende compte moi-même. J'ai aussi une pensée émue pour Papy Louis qui est arrivé au bout de son chemin pendant la rédaction de ce manuscrit. Ton sens du collectif d'ouvrier et ton amour du travail bien fait d'artisan ont été, et sont pour longtemps encore, une source d'inspiration quotidienne dans mon travail. Une pensée non moins émue pour Olivia qui elle, a débuté son chemin il y a tout juste quelques jours. J'accompagne cette pensée d'un gros bisou à ses parents, Sarah et Quentin. Pour conclure, je tiens à remercier mes armes secrètes, mes sources d'énergie, celles qui me donnent la force d'avancer et de me relever : Clo et Iris.



Merci à tous et à chacun. Je vous dois beaucoup.

Table of content

Remerciements 4

Table of content 6

List of figures 10

List of tables 15

Abstract 16

Résumé 17

Chapter 1: Introduction 18

1.1 Collective approaches for answering complex challenges in agriculture 18

1.2 From crowdsourcing to farmsourcing of spatial observations: definitions and concepts 19

1.2.1 Definition of crowdsourcing 19

1.2.2 Success factors of crowdsourcing projects 20

1.2.3 Crowdsourcing projects of spatial observations 23

1.2.4 Farmsourcing projects of Spatial Observations in agriculture 25

1.3 Do Farmsoucing projects of spatial observations have specific success factors? 27

1.3.1 Specificities of Farmsourcing projects of Spatial Observations 27

1.3.2 Impact on success factors 31

1.4 Scientific question and organisation of the manuscript 33

1.4.1 General approach 33

1.4.2 ApeX-Vigne use case 33

1.4.3 Formulated hypothesis 34

1.4.4 Organisation of the manuscript 37

Chapter 2: Monitoring of vine water status in the south of France: an issue at different spatial scales 40

2.1 Vine water status influences vineyard performances in many ways 41

2.2 Vine water status has to be monitored in time and space 42

2.2.1 Vine water status varies over time and space, even at within field scale 42

2.2.2 Approaches for monitoring vine water status are resulting from a trade-off between reliability, spatial and temporal resolutions 44

2.3 Many actors are concerned by monitoring vine water status both at vineyard and regional scales 47

2.3.1 Monitoring vine water status at field scale for managing vineyard 47

2.3.2 Monitoring vine water status at regional scale for setting up collective strategies 47

2.4 French Mediterranean arc is a relevant study zone for the issue of monitoring vine water status 48

2.4.1 The most relevant scale for defining the study zone is a few hundreds of kilometres 48

2.4.2 The French Mediterranean arc has the characteristics of a relevant case study 49

Chapter 3: Observation of shoot growth: a simple and operational decision-making tool for monitoring vine water status in the vineyard 51

3.1 Details about the corresponding paper *52*
3.1.1 Title of the paper 52
3.1.2 Authors: 52
3.1.3 Abstract 52
3.1.4 Keywords 52

3.2 Introduction *53*

3.3 Material and methods *54*
3.3.1 Collected datasets 54
3.3.2 *Vineyard_dataset*: Sampling design 54
3.3.3 Method for collecting observations on shoot growth 55
3.3.4 Vineyard_dataset: Method for collecting reference observations 56
3.3.5 Metrology_dataset: Sampling design 56
3.3.6 Data analysis: 56
3.3.7 Graphs 57

3.4 Results and discussion *57*
3.4.1 Relationship between iG-Apex and ψPD 57
3.4.2 Validity range of iG-Apex 58
3.4.3 Specificity of iG-Apex to vine water status 59
3.4.4 Study of the factors influencing iG-Apex 60
3.4.5 Considerations for using iG-Apex to estimate vine water restriction in a commercial context 62

3.5 Conclusion *64*

Chapter 4: Characterising within-field variability of vine water status with simple visual observations of shoot growth 65

4.1 Details about the corresponding paper *66*
4.1.1 Title of the paper 66
4.1.2 Authors 66
4.1.3 Abstract 66
4.1.4 Key words 66

4.2 Introduction *66*

4.3 Material and methods *67*
4.3.1 Method for collecting observations 67
4.3.2 Study fields 67
4.3.3 Sampling 68
4.3.4 Analysis of variance and spatial structure 69

4.4 Results and Discussion *69*
4.4.1 Temporal evolution of iG-Apex 69
4.4.2 Temporal evolution of iG-Apex's total and spatially structured variances 70
4.4.3 Mapping of iG-Apex's within field variance 72
4.4.4 Study of operator effect 72

4.5 Conclusion *73*

Chapter 5: ApeX-Vigne: a platform designed for collecting farmsourcing spatial observations 74

5.1 Preliminary comment on the development of ApeX-Vigne application *75*

5.2 Details about the corresponding paper *76*
5.2.1 Title of the paper 76
5.2.2 Authors 76
5.2.3 Abstract 76
5.2.4 Keywords: 76
5.3 Introduction *77*
5.4 Materials and methods *77*
5.5 Results *80*
5.5.1 Application interface 80
5.5.2 Spatial distribution of observations 81
5.5.3 Temporal distribution of observations 81
5.6 Discussion *82*
5.7 Conclusion *83*
Chapter 6: Evaluating the quantity of contributions in a farmsourcing project of spatial observations 84
6.1 Details about the corresponding paper *85*
6.1.1 Title of the paper 85
6.1.2 Authors 85
6.1.3 Abstract 85
6.1.4 Keywords: 86
6.2 Introduction *86*
6.3 Materials and method *88*
6.3.1 The index of Growing Apex (iG-Apex) for monitoring shoot growth 88
6.3.2 General technical characteristics of the ApeX-Vigne application 88
6.3.3 Application interface and publication 88
6.3.4 Study zone: 89
6.3.5 Data Access 91
6.3.6 Considered data 91
6.3.7 Approach for evaluating the adoption of ApeX-Vigne 92
6.3.8 Approach for illustrating the uses of ApeX-Vigne 92
6.3.9 Analyses and graphs 93
6.4 Results and discussion *93*
6.4.1 ApeX-Vigne application adoption - downloads and surveys 93
6.4.2 ApeX-Vigne application adoption – Location of surveys 95
6.4.3 Opportunities for using the application at the within-field, field and study zone scales 96
6.5 Conclusion *102*
Chapter 7: Evaluating the quality of contributions in a farmsourcing project of spatial observations 103
7.1 Details about the corresponding paper *104*
7.1.1 Title of the paper 104
7.1.2 Authors: 104
7.1.3 Abstract 104
7.1.4 Keywords 104

7.2 Introduction *105*

7.3 Material and methods *107*
7.3.1 Problem statement 107
7.3.2 Formalization of the approach 109
7.3.3 Implementation of the approach 110
7.3.4 Use case 112
7.3.5 Validation of the approach 114
7.3.6 Graphs and maps: 116

7.4 Results *116*
7.4.1 Classification of outliers : 116
7.4.2 Classification of surprising observations: 121

7.5 Discussion *123*

7.6 Conclusion *124*

Chapter 8: Conclusion and perspectives 126

8.1 Looking back to the proposed approach *127*

8.2 Looking forward to future research work *128*

8.3 Opening up new perspectives towards on-farm experimentation *129*

References 131

Chapter 9: Appendices 146

9.1 Screenshots of the latest version of ApeX-Vigne mobile application (release of 2021 May 7th) *146*

9.2 Press release sent to the wine press for the official launch of the ApeX-Vigne application *147*

9.3 Other peer-reviewed articles published during the thesis *148*

Résumé étendu 155

List of figures

Figure 1: Sequence diagram of the different components involved in a crowdsourcing project (adapted from Zhao and Zhu, 2014).....20

Figure 2: Factors influencing the success of a crowdsourcing project (adapted from Rechenberger et al. 2015).....20

Figure 3: Number of scientific papers available on the Scopus platform with on the one hand the keyword "Agriculture" and and the other hand one of the keywords "Crowdsourcing", "Citizen Science", "Mashup" and "User-generated content"......26

Figure 4: From crowdsourcing to Farmsourcing of Spatial Observations (FSO): synthesis of the main linkages and areas of interest for FSO projects......27

Figure 5: Map of precipitation on Champagne vineyard estimated by radar for August 14th 2014 (week 33) with 1-kilometre-wide pixels......28

Figure 6: Temporal evolution of cumulative weekly rainfall during spring and summer 2014 in the Champagne vineyards. 28

Figure 7: Monthly observations submitted to eButterfly platform for butterfly observations in North America from 2012 to 2016 (from Prudic et al., 2017)......29

Figure 8: Specific success factors of Farmsourcing projects of Spatial Observations tested by hypothesis 1 and 2 and their relation with generic success factors. 36

Figure 9: Checking specific success factors and testing hypotheses along the steps of the use case in the proposed systemic approach......37

Figure 10: Organisation of the manuscript: implementation of the systemic approach along chapters 2 to 7 based on the main steps identified in Figure 9.. 38

Figure 11: Summary of the main vineyard performances impacted by vine water status......41

Figure 12: Schematic representation of possible vine water status trajectories according to objectives of production (redrawn after Ojeda and Saurin 2014). 42

Figure 13: Examples of spatial and spatio-temporal variations of vine water status: (a) within-field and (b) inter-field variability of vine water status measured by predawn leaf water potential at the end of summer near Gruissan, Languedoc, France (from Acevedo et al. 2008) and by $\delta^{13}C$ near Bordeaux, France (from Van Leeuwen et al., 2018), respectively. (c) Temporal evolution of within-field variability of vine water status measured by stem water potential near Simonsberg, South-Africa (from Jasse et al., 2021). 43

Figure 14: The different approaches for assessing vine water status (Rienth and Scholasch, 2019) are resulting from a trade-off between reliability, spatial and temporal resolutions. For each approach, spatial and temporal resolutions are represented in X and Y axis (according to Acevedo-Opazo et al., 2008) and reliability is represented with different colours (according to Rienth and Scholasch, 2019)......46

Figure 15: Extract from a weekly technical report published by the "Côtes du Rhône" syndicate on the monitoring of vine water status at the whole appellation scale. Polygons represent

homogeneous pedoclimatic zones and points represent reference fields on which observations are made weekly (from Bécart et al. 2020). 48

Figure 16: The study zone encompasses several wine-growing regions of importance in southern France with a large variability in soil and climate conditions. (a) The location of the study zone within France, (b) the location of vineyards within the study zone that have a controlled denomination of origin (INAO, 2020). Two principal pedo-climatic characteristics of this zone, (c) soil water holding capacity (IGN, 2020) and (d) mean annual precipitation over the last past 30 years (Météo France, 2020). 50

Figure 17: Location and cultivar of fields on which reference and shoot growth observations were collected for the Vineyard_dataset. 54

Figure 18: Average monthly temperatures and cumulative monthly rainfall for the 2008 (a), 2009 (b), 2010 (c), 2011 (d) and 2012 (e) vintages compared to the respective averages over the period 1976 – 2005. 55

Figure 19: Examples of shoot classified in the categories "full growth" (a), "moderate growth" (b) and "stopped growth" (c). 56

Figure 20: Mean of Predawn Leaf Water Potential (ψPD) depending on iG-Apex values for the 474 observations of Vineyard_dataset. 58

Figure 21: Scatter plots of Predawn Leaf Water Potential (ψPD) vs iG-Apex for: (a) observations with strong water restriction (ψPD< -0.8) and (b) observations with active shoot growth (iG-Apex > 0.9). 59

Figure 22: Standard deviation of Predawn Leaf Water Potential (ψPD) depending on iG-Apex values for the 474 observations of Vineyard_dataset. 60

Figure 23: Mean and standard deviation of Predawn Leaf Water Potential (ψPD) depending on iG-Apex values for the 236 observations made on cv. Cinsault and cv Grenache out of the 474 observations of Vineyard_dataset. 61

Figure 24: (a) Location of the fields in southern France and location of points within the fields, (b) Syrah with n =101 samples, and (c) Grenache with n = 98 samples. (d) The timings of the weekly iG-Apex observations from early June to late August are indicated by points in the calendar, along with the timing of canopy management and rainfall events. 68

Figure 25: The two fields planted with Syrah (101 observations) and Grenache (98 observations) have similar iG-Apex temporal dynamics throughout the season with a gradual cessation of growth between weeks 24 and 34. 70

Figure 26 : The total variance (C1+C0) of iG-Apex observed at the intra-field scale increased and then decreased through the season for the two fields planted respectively with Syrah (green line; n = 101) and Grenache (grey line; n = 98) (a). The proportion of this variance that is spatially organised (C1 / (C1 + C0)) also increased and then decreased (b). 71

Figure 27: Experimental and model semi-variograms of the iG-Apex for week 28 for both study fields planted respectively to Grenache (a) and to Syrah (b) highlight a significant spatial organisation of iG-Apex values . The corresponding maps (c and d) confirm this spatial structure and illustrate the related spatial patterns. 72

Figure 28: Development cycles of ApeX-Vigne mobile application.....75

Figure 29: This diagram shows the iterative design process to discover, define, design and develop a digital solution that meets user needs......78

Figure 30: Technologies used for the development of ApeX......79

Figure 31: Screenshot of both main interfaces of ApeX......80

Figure 32: Observations collected with Apex during 2018 in southern France, a) example at the within-field level, b) spatial distribution of measurements at the regional level......81

Figure 33: Use of ApeX over the season and examples of results (June – October) (a) the count of sessions made each fortnight and (b) the average recorded STI at approximately weekly intervals in a field......82

Figure 34: Screenshot of the two main interfaces of the ApeX-Vigne application: the input (a) and output (b) screens.....89

Figure 35: The study zone encompasses several wine-growing regions of importance in southern France with a large variability in soil and climate conditions. (a) The location of the study zone within France, (b) the location of vineyards within the study zone that have a controlled denomination of origin (INAO, 2020). Two principal pedo-climatic characteristics of this zone, (c) soil water holding capacity (IGN, 2020) and (d) mean annual precipitation over the last past 30 years (Météo France, 2020)......90

Figure 36: (a) Average weekly temperatures and (b) cumulative weekly rainfall measured during the 2019 and 2020 vintages at a meteorological station located in Montpellier, in the centre of the study zone......91

Figure 37: Basic statistics of the use of the ApeX-Vigne application. (a) Evolution of the number of installations and uninstallations during the 2019 and 2020 seasons (launched in week 21, 2019) and (b) number of surveys carried out per week over the key usage period (Weeks 22 – 35) of these two seasons.....94

Figure 38: (a) Location of surveys made in the study zone using the ApeX-Vigne mobile application during weeks 22 to 36 of the 2019 and 2020 growing season. Local examples of (b) surveys made for a field scale monitoring and (c) for a within-field scale monitoring in week 28 of the 2020 growing season......96

Figure 39: The ApeX-Vigne application was used by farmers and advisers for monitoring vines at the within-field scale: (a) Mean and standard deviation evolutions for the index of Growing Apex (iG-Apex) collected at the within-field scale in 2020 by a user of the ApeX-Vigne application, (b) example of an iG-Apex kriged map obtained from data collected in week 28, 2020 and, (c) the corresponding iG-Apex semi-variogram at this stage......97

Figure 40: The ApeX-Vigne application as used by farmers and advisers for monitoring the evolution of vine water status at the field scale. (a) Examples from three individual fields in different wine-growing regions of the study zone (b) showing the evolution of the index of Growing Apex (iG-Apex) score from week 22 to week 36 in 2020......98

Figure 41: iG-Apex mean and standard deviation per week over the study zone during the 2019 and 2020 seasons......99

Figure 42: Fitted variograms of weekly iG-Apex data illustrating the evolution over time of the spatial structure in these data for the 2019 and 2020 seasons. Arrows indicate the shift associated with an apparent two-week time-lag in response between the two seasons. 100

Figure 43: Main steps of existing approaches for classifying outliers and surprising observations in spatialized crowdsourcing datasets. 108

Figure 44: Main steps of the implemented approach for identifying outliers and surprising observations in spatialized crowdsourcing datasets. 112

Figure 45: The study zone is a wine-growing region with a large variability in soil and climate conditions and in which the ApeX-Vigne application is relevant and widely used. (a) Location of the study zone within France, (b) location of vineyards within the study zone that have a controlled denomination of origin (INAO, 2020). Two principal pedo-climatic characteristics of this zone: (c) soil water holding capacity (IGN, 2020) and (d) mean annual precipitation over the last past 30 years (Météo France, 2020). 113

Figure 46: (a) Evolution of the shoot growth index (iG-Apex) during the season in an early and late developing vineyard and (b) location within the study zone of the "early" (9 observations) and "late" (10 observations) reference fields. 116

Figure 47: Location of the 4381 iG-Apex observations having $si \in \mathcal{AI}$ *and of the 23 observations having* $si \notin \mathcal{AI}$*. Points are overlapping. Therefore, at this wide spatial scale, very close observations may appear as a single point.* 117

Figure 48: (a) Representation of the ApeX-Vigne dataset in 2-dimensional space $iG\text{-}Apex\ vs\ Time$ *and estimation of parameters* $\epsilon 1$ *and* $MinPts1$ *for classification of outliers from: (b) density of observed distance between two observations and (c) density of number of neighbours observed for this distance. Green and yellow lines represent the values chosen for parameters* $\epsilon 1$ *and* $MinPts1$ *respectively.* 118

Figure 49: Location of observations correctly and wrongly classified as outliers by the proposed approach according to expert reference classification plotted in the 2-dimensional space $iG\text{-}Apex\ vs\ Time$*.* 119

Figure 50: (a) Representation of the non-outliers observations of the ApeX-Vigne dataset in the 3 dimensional space ($Fweekj-1, Fweekj, Fweekj+1$*), and estimation of parameters* $\epsilon 2$ *and* $MinPts2$ *for classification of surprising observations from: (b) density of observed distance between two observations and (c) density of number of neighbours observed for the most frequent distance. Green and yellow lines represent the values chosen for parameters* $\epsilon 2$ *and* $MinPts2$ *respectively.* 121

Figure 51: Classification of surprising observations plotted in the 2-dimensional space $iG\text{-}Apex\ vs\ Time$*.* 122

Figure 52: (a) Interpolated map of iG-Apex during week 30 at the study zone scale made from the non-surprising observations. (b) Detail of this map representing the surprising observations on the same iG-Apex scale. 122

Figure 53: Du crowdsourcing au farmsourcing d'observations spatialisées : synthèse de l'imbrication des principaux concepts. 155

Figure 54: Facteurs de succès spécifiques des projets de FOS testés par les hypothèses 1 et 2 et relation avec les facteurs de succès génériques. 156

Figure 55: Exemples d'apex de vigne classés dans les catégories (a) « pleine croissance », (b) « croissance ralentie » et (c) « croissance arrêtée ». 157

Figure 56 : Potentiel hydrique foliaire de base moyen (ψPD) en fonction des valeurs d'iG-Apex. 157

Figure 57: Les semi-variogrammes expérimentaux et modélisés des valeurs d'iG-Apex pour la semaine 28 sur les deux parcelles d'étude Grenache (a) et Syrah (b) mettent en évidence une organisation spatiale des valeurs d'iG-Apex. Les cartes correspondantes (c et d) confirment cette structure spatiale et illustrent les motifs spatiaux associés. 158

Figure 58: Principaux écrans de l'application mobile ApeX-Vigne. 159

Figure 59 : Zone d'étude le long de l'arc méditerranéen français. 159

Figure 60 : Statistiques d'utilisation de l'application ApeX-Vigne. (a) Évolution du nombre d'installations et de désinstallations au cours des saisons 2019 et 2020 (lancement en semaine 21, 2019) et (b) nombre d'observations réalisées par semaine sur la période d'utilisation (semaines 22 - 35) de ces deux saisons. 160

Figure 61: Représentation dans l'espace en 2 dimensions $iG\text{-}Apex\ vs\ Temps$ des observations labélisées comme surprenantes par l'approche proposée. 161

List of tables

Table 1: Different participants' type of motivation (adapted from Batson et al., 2002) 22

Table 2: Participants' type of contribution for each type of crowdsourcing (adapted from Minet et al., 2017).....23

Table 3: Characteristics of the different project elements for unstructured, semistructured and structured crowdsourcing project of spatial observations (adapted from Kelling et al. (2019)).25

Table 4: Potential impact of specificities of Farmsourcing projects of Spatial Observations on the success factors of crowdsourcing projects......32

Table 5: Six main steps of a Farmsourcing project of Spatial Observations for the general case and for the ApeX-Vigne use case.....34

Table 6: Results of the ANOVA performed on 474 iG-Apex observations of Vineyard_ dataset with three factors (cultivar, classes of ψPD and field). 61

Table 7: Results of the ANOVA performed on 54 iG-Apex observations with two factors (operator and repetition). 62

Table 8: Main notations and definitions.....109

Table 9: Description of the experts and their characteristics (1 = low expertise; 2 = medium expertise and 3 = high expertise) 114

Table 10: Outliers' classification confusion matrix for the proposed approach compared to expert reference classification on the ApeX-Vigne original dataset. 119

Table 11: Outliers' classification confusion matrix for the proposed approach compared to expert reference classification over 1000 bootstrap's iterations on the original ApeX-Vigne dataset. Bold values indicate the median value from the bootstrapping while values in parentheses are confidence intervals for $\alpha = 0.1$ 120

Abstract

Crowdsourcing is an approach consisting in answering a question defined by an organisation (research laboratory, company, etc.) by relying on the collective intelligence of a community of contributors. To date, crowdsourcing is not widely spread in agriculture, but it has great potential for collecting georeferenced observations to monitor phenomena at regional scale (e.g. diseases, pests or abiotic stresses monitoring). These crowdsourcing projects in agriculture have specificities in terms of participants (professional contributors, importance of the role of advisors), studied phenomena (with strong spatial and temporal covariances) and datasets collected (asynchronous and heterotopic) that have led some authors to coin the concept of farmsourcing to describe them. These specificities of farmsourcing projects influence the design of the projects and the involvement of the different stakeholders. They also influence the criteria and indicators for evaluating the success of such projects. Finally, they influence the methods for identifying outliers and surprising observations in corresponding datasets. To date, there is no existing approach taking into account the specificities of farmsourcing projects. The objective of this thesis is to propose tools and methods to develop a farmsourcing approach in both the design and the evaluation of the project (How to foster the contribution of participants? How to evaluate the success of a project?) and then in the characterisation of the quality of the resulting observations (How to identify outliers and surprising observations? How can these approaches be automated?) The thesis is based on a systemic approach with the implementation of a case study. This case study is the monitoring of the vine water status at regional scale using i) an indicator (iG-Apex) based on observations of vine shoot growth and ii) the development of a dedicated farmsourcing application (ApeX-Vigne).

Firstly, the work demonstrated the value of a simple but noisy approach, such as the one based iG-Apex, for characterising an agronomic variable of interest (in this case, the vine water status) at the field and intra-field levels in a decision support context. They demonstrated how an approach like this could be used to promote participation in farmsourcing projects. The work carried out explored the technological and methodological choices for designing and deploying on a large scale a mobile application promoting the gathering of georeferenced farmsourcing observations. It also proposed a simple approach based on the study of spatial structure to assess the capacity of these projects to provide relevant information at the regional scale. Finally, the work carried out explored an approach for automatically identifying outliers and surprising observations in farmsourcing datasets. This approach is based on density-based clustering methods taking into account spatial, temporal and attribute characteristics of observations.

In the coming years, this work should enable the development of farmsourcing tools and projects giving access to new sources of information for decision support at different spatial scales.

Résumé

Le crowdsourcing est une approche qui consiste à répondre à une question définie par une organisation (laboratoire de recherche, entreprise, etc.) en s'appuyant sur l'intelligence collective d'une communauté de contributeurs. En agriculture, le crowdsourcing est pour l'instant peu développé mais il présente un intérêt potentiel fort pour la collecte d'observations géolocalisées permettant un suivi de phénomènes à l'échelle régionale (ex : suivi des maladies, des ravageurs ou des stress abiotiques). Ces projets de crowdsourcing en agriculture possèdent de spécificités au niveau des participants (contributeurs professionnels, importance du rôle des conseillers), des phénomènes étudiés (fortes covariances spatiale et temporelle) et des jeux de données collectées (asynchrones et hétérotopes) qui ont poussé certains auteurs à définir le concept de farmsourcing pour les décrire. Ces spécificités des projets de farmsourcing influent sur la conception des projets et sur l'implication des différents acteurs. Elles influent également sur les critères et les indicateurs qui permettent d'évaluer la réussite de tels projets. Enfin, elles influent sur les méthodes qui peuvent être mises en œuvre pour identifier les observations aberrantes et surprenantes dans les données issues de ces projets. Aujourd'hui, il n'existe pas d'approches qui prennent en compte ces différents aspects en considérant les spécificités des projets de farmsourcing. L'objectif de cette thèse est de proposer des outils et des méthodes pour développer une approche de farmsourcing à la fois dans la conception et l'évaluation du projet (Comment favoriser la contribution des participants ? Comment évaluer la réussite d'un projet ?) puis dans la caractérisation de la qualité des observations qui en sont issues (Comment identifier les données aberrantes et surprenants ? Comment automatiser ces approches ?). La thèse s'appuie sur une démarche systémique avec la mise en place d'un cas d'étude. Ce cas d'étude est le suivi de l'état hydrique de la vigne à l'échelle régionale à l'aide i) d'un indicateur (iG-Apex) basé sur des observations de la croissance végétative et ii) du développement d'une application de farmsourcing dédiée (ApeX-Vigne).

Dans un premier temps, les travaux ont permis de démontrer l'intérêt d'une approche simple mais bruitée, comme celle basée sur iG-Apex, pour caractériser une grandeur agronomique d'intérêt (ici l'état hydrique de la vigne) à l'échelle parcellaire et intra-parcellaire dans un contexte d'aide à la décision. Ils ont permis de démontrer comment une approche comme celle-ci pouvait être utilisée comme un levier pour favoriser la participation à des projets de farmsourcing. Les travaux menés ont ensuite exploré les choix technologiques et méthodologiques permettant de concevoir et de déployer à large échelle une application mobile favorisant la collecte d'observation de farmsourcing géolocalisées. Ils ont également permis de proposer une approche simple basée sur l'étude de la structure spatiale pour évaluer la capacité de ces projets à fournir une information pertinente à l'échelle régionale. Enfin, les travaux menés ont permis d'explorer une approche pour identifier de manière automatique les observations aberrantes et surprenantes dans des jeux de données de farmsourcing. Cette approche repose sur la prise en compte des caractéristiques spatiales, temporelles et attributaires des observations à l'aide d'approches automatiques basées sur leur densité.

Ces travaux méthodologiques devraient permettre le développement dans les années à venir de projets et d'outils de farmsourcing donnant accès à de nouvelles sources d'information pour l'aide à la décision à différentes échelles spatiales.

Chapter 1: Introduction

1.1 Collective approaches for answering complex challenges in agriculture

Agriculture is in transition. It is currently facing complex and sometimes conflicting challenges. How to produce healthy and abundant food at affordable prices while paying farmers with dignity? How to farm in a climate where average temperatures are 1, even 2 or 3 degrees higher than they were at the end of the last century and where extreme weather events are more frequent? How to limit the consumption of fossil energy and the production of greenhouse gases? How to take advantage of the opportunities offered by digital tools to better observe and better understand the living world? To answer these complex questions, agriculture has undertaken agro-ecological, climate, energy and digital transitions. Every agricultural stakeholder is involved in these transitions, from farmers and their advisors to researchers and policymakers.

Farmers are facing these challenges in their daily work. They experiment, alone or in groups, with solutions to address them. They share information with peers (e.g. http://www.agricool.net/forum/) by organising themselves collectively to share their results and compare them with those obtained in similar contexts (Slimi et al. 2021). Public policies support these collective projects, which are seen as a way for agriculture to succeed in its transition (e.g. https://collectifs-agroecologie.fr/). Tools and services are being developed to help farmers to objectify their observations, to analyse their results and to share their conclusions (e.g. https://landfiles.com/). This transition also raises fundamental research questions requiring the production of knowledge. The scientific community is studying this transition in all its dimensions, whether agronomic (Garcia et al. 2018), technological (Bellon-Maurel and Huyghe 2017) or socio-economic (Hu 2020), etc. The challenge for the scientific community is to produce knowledge that can answer both the global challenge of transition and the very local problems of its implementation. In some cases, the scientific community produces fundamental knowledge that is appropriated by agriculture stakeholders (Montes de Oca Munguia et al. 2021). In other cases, it studies innovations created by these stakeholders and tries to understand and formalize the underlying mechanisms (Salembier et al. 2020). Research projects bringing together researchers and stakeholders with different points of view in the same group seem particularly well suited to the emergence of new knowledge that can support agriculture in its transition (Moneyron et al. 2017).

The project led by Van Etten, Beza, et al. (2019) is emblematic of this approach of bringing together a collective of farmers and researchers to produce knowledge on complex issues. These authors have proposed a research approach for comparing several seed varieties by involving many farmers. To do so, farmers are given a random subset of three varieties to grow under their own conditions. They make observations and rank the varieties according to their performance. The observations are shared into a large dataset. This dataset is used to generate knowledge about the adaptation of the tested varieties to different soil and climatic conditions. Finally, results are presented and discussed between researchers and farmers during feedback workshops. This approach helps researchers to collect a large number of observations on experiments carried out in real conditions that they would not have been able to collect by other means. It also helps

farmers to access knowledge by getting involved in the experimentation and by exchanging information with peers and scientists. This approach helps agriculture in its transition by generating useful knowledge for dealing with the complex issues it faces (Van Etten, de Sousa, et al. 2019). The work carried out by Van Etten, Beza, et al. (2019) focused on seeds, but these authors have set up a platform (https://climmob.net/blog/) offering software and methodological tools to duplicate this methodology to other use cases. Even if there have been no reports of this methodology being tested in a systematic study of technical innovations or being applied to complete farming systems, it offers exciting opportunities to collectively generate knowledge and support agriculture in its transition. This approach, based on information collected by the crowd, is called crowdsourcing. This thesis falls within this approach and is adopting crowdsourcing methods to the perspective of crop monitoring for decision support.

The aim of this chapter is to define crowdsourcing and to describe concepts and issues associated with it in order to position the work of this thesis in relation to existing work and to identify the scientific issues and their originality. This chapter first describes the general concepts of crowdsourcing and the specificities of the projects that will be studied (1.2). It presents how these specificities can influence the success of these projects (1.3). Finally, it describes the scientific questions arising, the approaches used to answer them and the corresponding organisation of the manuscript (1.4).

1.2 From crowdsourcing to farmsourcing of spatial observations: definitions and concepts

1.2.1 Definition of crowdsourcing

Crowdsourcing emerged in the 2000s, facilitated by the expansion of internet access. It is based on the concept of the wisdom of crowds defined by Surowiecki (2005), which states that a crowd of individuals is capable of making better decisions and predictions than the individuals who compose it. A few years later, Brabham (2008) offered a use case analysis based on early projects and an introduction to the concept of crowdsourcing. According to Brabham (2008), crowdsourcing is an approach consisting in answering a question by relying on the collective intelligence of a community of contributors. Zhao and Zhu (2014) identified three components in these projects: i) the organisation (a research laboratory, a company, etc.) that defines the question to be addressed, ii) the participants (neophyte, interested amateur, expert amateur/professional, expert authorities or unaware volunteer (Bordogna et al. 2014) who contribute by providing part of the answer to the investigated problem, and iii) the platform that provides the infrastructure (often using digital tools) to help the organisation to collect contributions. Depending on the project, the platform may be managed by the organisation itself or by a third party.

In crowdsourcing projects, the organisation first formalises the question to be addressed and then defines unitary tasks to be carried out by the participants to answer this question (Figure 1). The role of the platform is to create the conditions (technical, logistical, financial, etc.) for participants to want to contribute and to be able to perform these tasks. The participants then contribute to the platform, which sends the collected data to the organisation. Finally, the organisation sets up

a data validation procedure and rewards the participants for their contribution. Depending on the project, the reward can be financial but it can also take many other forms (see section 1.2.2).

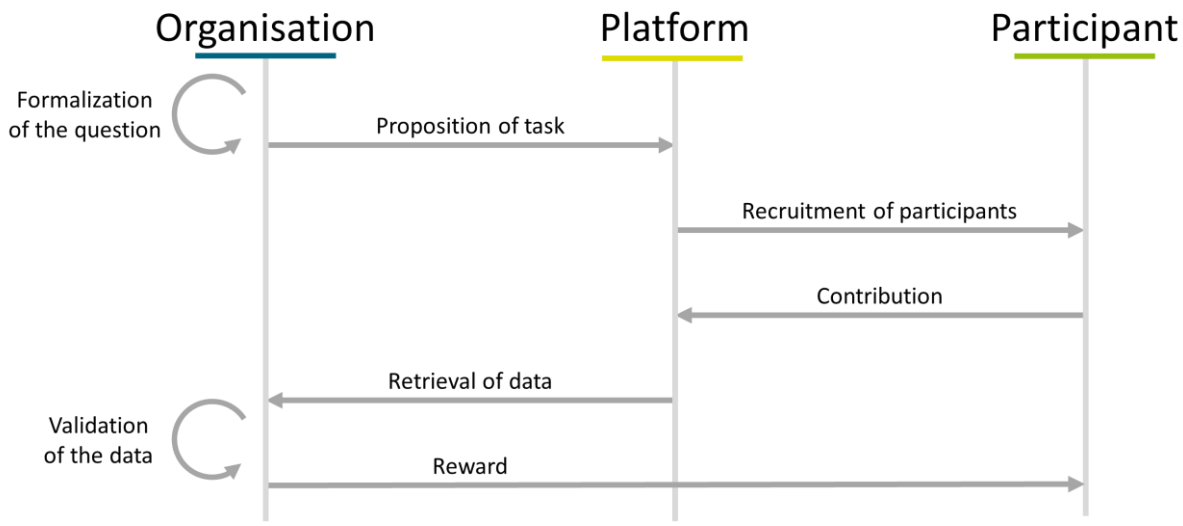


Figure 1: Sequence diagram of the different components involved in a crowdsourcing project (adapted from Zhao and Zhu, 2014)

1.2.2 Success factors of crowdsourcing projects

A crowdsourcing project is considered successful if participants' contributions help to answer the question asked by the organisation. This success depends on many factors. Rechenberger et al. (2015) identified 41 of them and described the positive or negative interactions between them. Figure 2 presents a summary version of this research work in which the main factors are represented. A successful crowdsourcing project must i) be carried out in a secure environment, ii) generate a large number of contributions and iii) set up strategies to ensure that contributions are of good quality. The following sections explain and describe the corresponding success factors.

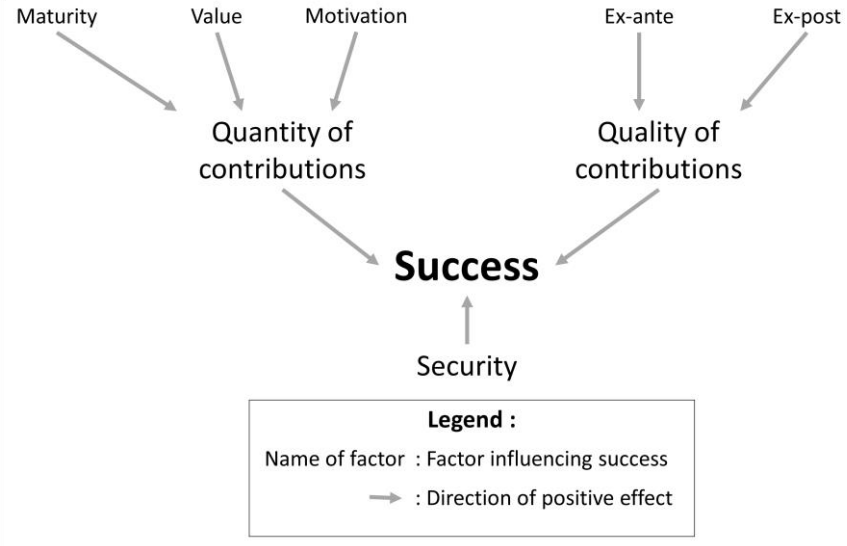


Figure 2: Factors influencing the success of a crowdsourcing project (adapted from Rechenberger et al. 2015).

1.2.2.1 *Security*

A crowdsourcing project must be carried out in a climate of security (Rechenberger et al. 2015). This encompasses both i) financial security ensuring that the project has the necessary financial resources to operate (Ebitu et al. 2021), ii) contractual security ensuring that ownership and access to data is clearly defined and in accordance with applicable laws (Ryan et al. 2018) and iii) digital security ensuring that hardware and software infrastructures are correctly constructed and ready to face possible external attacks (Yan et al. 2017).

1.2.2.2 *Quantity of contributions*

A crowdsourcing project must be able to collect a significant number of contributions (Rechenberger et al. 2015). This number of contributions depends on four major success factors:

Success factor 1: Platform maturity (MATURITY)

The older the platform that supports participation, the more likely participants are to contribute (Walter and Back 2011). Participants prefer to contribute on platforms and projects that they are familiar with, i.e. that have been around for some time and that are mature.

Success factor 2: Value and ease of use (VALUE)

The easier the platform is to use, the more likely participants are to contribute (Rechenberger et al. 2015). This ease of use depends on the one hand on how ergonomic the platform is and on the other hand on the functions it offers to participants. Asingizwe et al. (2019) showed that designing a crowdsourcing platform using co-design methods greatly enhances its ergonomic nature. According to these authors, co-design methods allow for the involvement of participants at a very early stage in the design of the platform. This offers them the opportunity to give their opinion on how ergonomic the platform is, which afterwards favours their appropriation and ownership of it. As a consequence, they contribute more readily (Nguyen et al. 2017).

Success factor 3: Motivation (MOTIVATION)

- Participants must be motivated to ensure a high number of contributions are made. The motivation of participants is certainly the participation factor that has been most studied in the literature. Batson et al. (2002) have proposed a framework that identifies four types of motivations for contributors. These are summarised in table 1 where they are presented in no order of importance.

- Egoism: Participants get involved to serve their own interest. The collective project is an unintended consequence or a means to serve this interest. Within this type of motivation, it is possible to distinguish between intrinsic motivations that are satisfied simply by the fact of contributing (e.g. having fun) and extrinsic motivations for which the accomplishment of the task is only an indirect means to achieve the real objective (e.g. obtaining money, information, recognition, etc.) (Kaufmann et al. 2011).

- Altruism: Participants get involved to answer to the interest of others for whom they feel empathy. The collective project is an unintended consequence or a means to serve this interest.

- Collectivism: Participants get involved to answer to the interest of a group or collective. In this case, the collective project is at the centre and it is the desire to create the common good that guides the involvement of participants.
- Principlism: Participants get involved in the name of moral principles and higher interests, such as freedom or justice. The collective project is an unintended consequence or a means to serve this interest.

Table 1: Different participants' type of motivation (adapted from Batson et al., 2002)

Type of motivation	Goal of the contribution	Example
Egoism (intrinsic)	Serving directly his own interest	Having fun
Egoism (extrinsic)	Serving indirectly his own interest	Financial benefit
Altruism	Serving someone else's interest	Helping researchers
Collectivism	Serving a group's interest	Contributing to common good
Principlism	Acting on behalf of a moral principle	Acting for justice

In every crowdsourcing project and for every participant, these four types of motivation are present, but their respective importance depends on the type of project and on the way that the objectives are presented (Land-Zandstra et al. 2016). Moreover, these proportions may change over time. Egoism is often the main type of motivation when first contributing and then altruism and collectivism may play a more important role in the choice of participants to continue to contribute (Rotman et al. 2012). Finally, regardless of the stage of the project, participants are more motivated to participate if they are confident in the organisation driving the project (Kosonen et al. 2013) and if they identify meaning in the tasks they are given (Chandler and Kapelner 2013).

1.2.2.3 *Quality of contributions*

Good quality contributions must be collected to properly answer the question asked by the organisation (Wiggins et al. 2011). In the literature, this quality is characterised by many metrics (Senaratne et al. 2016) and there are two strategies to improve it: *ex ante* and *ex post* strategies (Bordogna et al. 2014) .

Success factor 4: Ex ante strategy (EX ANTE)

This strategy is also called by some authors "quality assurance" (Jonietz et al. 2017). It consists of setting up mechanisms before the collection of observations for ensuring that collected observations are of good quality. For example, these mechanisms can consist in the implementation of training for participants, the definition of a common vocabulary or strict collection protocols (Criscuolo et al. 2016). The use of sensors to collect data is a good way to implement this strategy as it promotes standardisation and minimises the risk of input errors. This sensor-based approach has been widely developed for the collection of spatial information, especially with the development of smartphones and Global Navigation Satellite System (GNSS)

receivers (Crall et al. 2010). Finally, Kelling et al. (2019) stressed the importance of collecting information not only about the phenomenon under study but also about the collection process itself. According to them, the collection of metadata associated with the participants' contributions (who contributes, where, when, according to which protocol, etc.) does not directly improve the quality of the collected data but it facilitates understanding and analysis *a posteriori*.

Success factor 5: Ex post strategy (EX POST)

This strategy is also called by some authors "quality assessment" (Jonietz et al. 2017). It consists of assessing the quality of the data once it has been collected so that poor quality data can be eliminated. Criscuolo et al. (2016) distinguished three approaches to achieve this:

- The organisation assesses the quality of the collected data expertly.
- The community assesses this quality either by comparing the data collected by several participants having performed the same task (Raykar and Yu 2012) or by assessing the reputation of the participants (Jabeur et al. 2018). This reputation corresponds to the community's trust in a participant. It is generally assessed by comparing his or her contributions to the opinion of experts (Saoud et al. 2020) or by studying the modifications made to his or her contributions by other participants (Yijiang Zhao et al. 2016).
- The quality of contributions is assessed automatically either by comparing them to reference data (Massad and Dalyot 2018) or by studying the characteristics of the resulting dataset (Touya et al. 2017). The latter approach is particularly used in the crowdsourcing of spatial observations by relying on the spatial coherence of phenomena (Goodchild and Li 2012). In the literature, these three approaches are frequently combined to assess the quality of the collected data (Fonte et al. 2017).

1.2.3 Crowdsourcing projects of spatial observations

1.2.3.1 *Types of crowdsourcing projects*

Although crowdsourcing projects all share the same principles and success factors, they are very diverse in the nature of the contributions that are made by participants. Minet et al. (2017) defined three types of crowdsourcing projects according to the nature of these contributions (Table2):

Table 2: Participants' type of contribution for each type of crowdsourcing (adapted from Minet et al., 2017)

Type of crowdsourcing	Type of contribution	Example
Crowdsourcing of knowledge	Sharing knowledge	Treude et al. 2011
Crowdsourcing of tasks	Doing unitary tasks	Altwaijry et al. 2016
Crowdsourcing of spatial observations	Collecting spatial information	Sullivan et al. 2009

- The crowdsourcing of knowledge in which participants contribute by sharing their knowledge is not always seen as crowdsourcing, although this type of project is well known to the general public. Examples of this include the online encyclopaedia Wikipedia (https://www.wikipedia.org/) (Westerman 2009) or Q&A forums, such as Stackoverflow (https://stackoverflow.com/), that is widely used by the computer development community (Treude et al. 2011).
- The crowdsourcing of tasks in which participants do unitary tasks of all kinds from their computers is also called microworking. This has been widely enabled by Amazon's Mechanical Turk platform (https://www.mturk.com/). It is widely used to perform tedious tasks that can only be done by humans. One of the most common and iconic examples is the labelling of images by participants to train deep learning algorithms (Altwaijry et al. 2016).
- Crowdsourcing of spatial observations in which participants collect spatialized observations involve projects that are mainly described by the keywords "crowdsourcing", "Mashup", "User-generated content" and "citizen sciences" (See et al. 2016). Minet et al. (2017) distinguished between visual observations and observations collected with sensors, but in all cases these observations are associated with a geolocation system. The increased availability of GNSS receivers and then smartphones has largely contributed to the growth of this type of crowdsourcing (Massad and Dalyot 2018).

In agriculture, crowdsourcing projects of spatial observations are of particular interest because the collection of spatialized and dated observations is at the heart of the profession of both farmers and agronomists (Ebitu et al. 2021). This type of project is the one that is particularly studied in the rest of this document.

1.2.3.2 *Characteristics of crowdsourcing projects of spatial observations*

Kelling et al. (2019) proposed to classify crowdsourcing projects of spatial observations according to their level of structuration (Table 3). According to them, unstructured projects correspond to a totally opportunistic observation gathering strategy, whereas structured projects correspond to a gathering strategy where everything is defined in advance by the organisation. According to Kelling et al. (2019), the majority of crowdsourcing projects of spatial observations choose an intermediate strategy called semi-structured, with clear objectives and an established protocol but with participation that remains open and an opportunistic gathering strategy compensated by the collection of metadata.

Table 3: Characteristics of the different project elements for unstructured, semistructured and structured crowdsourcing project of spatial observations (adapted from Kelling et al. (2019))

		Structuration of crowdsourcing project of spatial observations		
		Unstructured	*Semistructured*	*Structured*
Project elements	*Survey design*	Opportunistic	Opportunistic	Totally designed
	Protocol	Not defined	Partially defined	Totally defined
	Recruitment	Open and flexible	Open and flexible	Limited
	Metadata	Not collected	Collected	Not collected

These projects can address both static and dynamic phenomena. For example, the OpenStreetMap platform (https://www.openstreetmap.org), which aims to build a free geographic database of the world, is certainly the most studied case for observation of a static phenomenon (i.e. the location of buildings, roads, etc.) (e.g. Haklay (2010); Senaratne et al. (2016) among others). This crowdsourcing platform claims more than seven million contributors and more than 8 billion geo-referenced points uploaded to the platform (OpenStreetMap, 2021). For the monitoring of dynamic phenomena, many research studies have been carried out in the field of environmental monitoring, for example to produce maps for monitoring the spatial distribution of different species (Sullivan et al. 2009). Some projects are also emerging in agriculture, for example for monitoring crop pests (Malek et al. 2018).

1.2.4 Farmsourcing projects of Spatial Observations in agriculture

In recent years, several articles have proposed reviews of crowdsourcing projects in agriculture (Ebitu et al. 2021; Minet et al. 2017). Despite the particular interest in crowdsourcing of spatial observations in agriculture, existing review articles have only focused on crowdsourcing projects as a whole, without specifically addressing agriculture-based projects. A Boolean search of articles associating the keyword "agriculture" with each of the four keywords describing crowdsourcing of spatial observations indicates up a first inventory of the existing research in this topic. For this purpose, the Scopus platform (https://www.scopus.com/) was used and the searches were carried out on 25 March 2021 only considering the title, abstract and keywords of articles within the database.

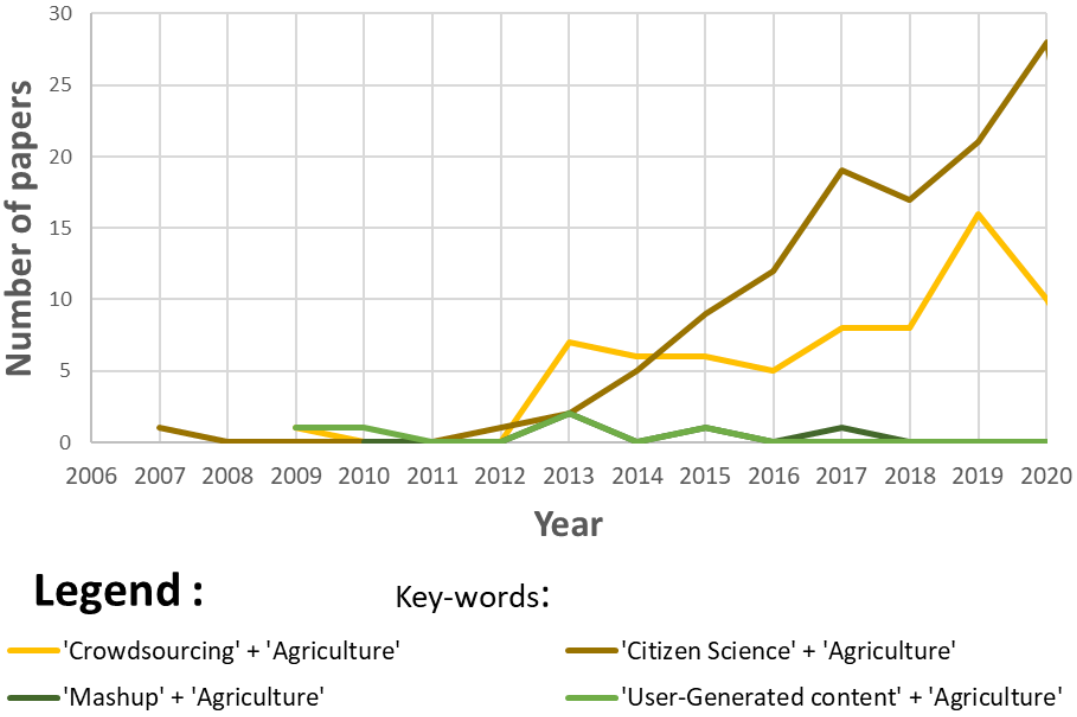


Figure 3: Number of scientific papers available on the Scopus platform with on the one hand the keyword "Agriculture" and and the other hand one of the keywords "Crowdsourcing", "Citizen Science", "Mashup" and "User-generated content".

Published research on crowdsourcing of spatial observations in agriculture first appeared in the late 2000s and really started to emerge in 2013 (Figure 3). From 2013, there have been more and more articles associating the keywords "crowdsourcing" and "citizen science" with agriculture. Nevertheless, only six out of 187 articles (3%) associated the keywords "Crowdsourcing" and "Citizen science" with agriculture simultaneously. This keyword analysis indicated that there are two sub-communities that study agriculture from different perspectives. The "Crowdsourcing" sub-community focuses on agriculture more from the perspective of the farmer and decision making (Weddagala et al. 2020). It is related to other keywords like "Remote sensing", "Crops", "Mapping" and "Decision making". The "Citizen Science" sub-community studies agriculture as part of a larger system, such as the ecosystem, biodiversity or climate change (Billaud et al. 2021). This sub-community associates other keywords like "Biodiversity", "Agricultural land", "Land use", "Ecosystem" and "Climate change". As a result, participants are more likely to be amateurs or the general public in the case of projects associated with the keyword "Citizen science", while they are more likely to be agricultural professionals (e.g. farmers, advisors) in the case of projects associated with the keyword "Crowdsourcing" (Ebitu et al. 2021). Minet et al. (2017) proposed the term "Farmsourcing" to describe crowdsourcing projects that involve professionals from the agricultural sector. The sub-community of crowdsourcing projects of spatial observations involving agricultural professionals and interested in decision support can be described by the term "Farmsourcing of Spatial Observations".

For ease of understanding, Figure 4 summarises the different concepts presented and their interrelationships. The work carried out in this thesis focuses on Farmsourcing of Spatial Observations (FSO) projects.

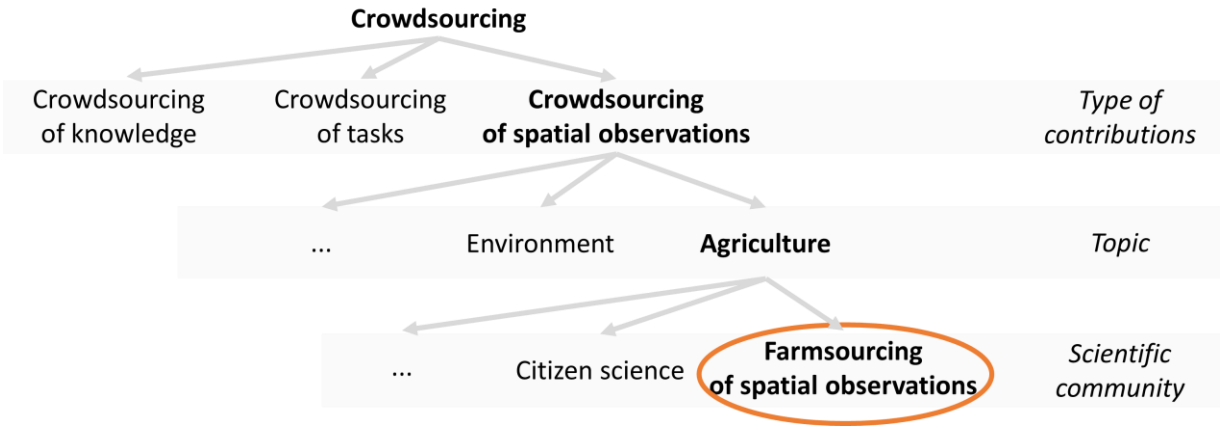


Figure 4: From crowdsourcing to Farmsourcing of Spatial Observations (FSO): synthesis of the main linkages and areas of interest for FSO projects.

1.3 Do Farmsoucing projects of spatial observations have specific success factors?

The aim of this section is to understand the influence of FSO projects' specificities to their own success. It first details these specificities (1.3.1) and then outlines the influence they may have on the success factors (1.3.2).

1.3.1 Specificities of Farmsourcing projects of Spatial Observations

FSO projects have six main specificities compared to other crowdsourcing projects. These are due to both the type of contributions (spatial observations) and the agricultural context in which these contributions are gathered (farmsourcing).

Specificity 1: Phenomenon are structured in time and space (STRUCTURED)

The studied phenomena concern agricultural issues, such as the presence of diseases, pests (Awuor et al. 2019) or abiotic stresses (Dell'Acqua et al. 2018). These phenomena are often dependent on environmental factors, such as soil (e.g. pedology, topometry) or climate (e.g. temperature, rainfall). These environmental factors are often spatially organised (Colaço et al. 2019). Rainfall is a good example of this. As an illustration, Figure 5 shows the precipitation estimated by radar over the entire Champagne viticulture region for August 14th 2014 (week 33). The rainfall is spatially organised from east to west. On this day, some areas of the region received < 5 mm of precipitation while other areas received up to 40 mm. In this case, the spatial organisation of the precipitation was particularly strong.

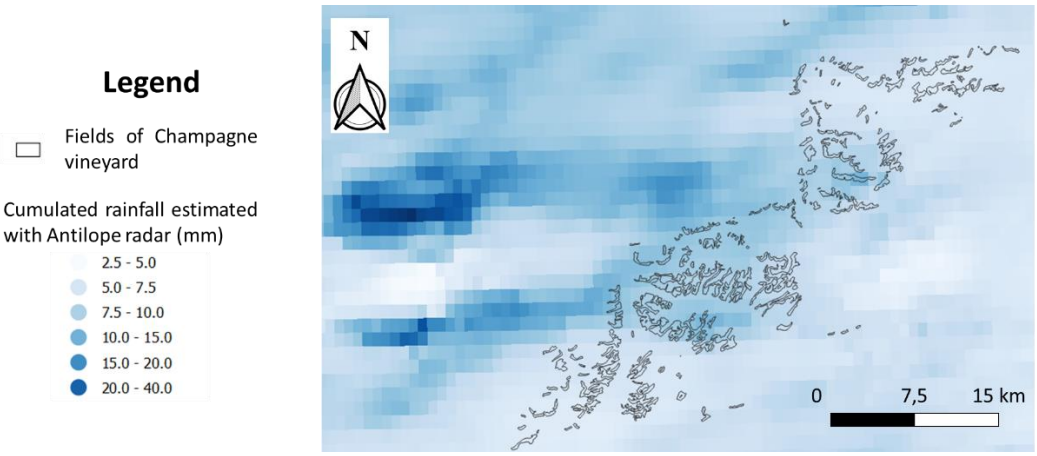


Figure 5: Map of precipitation on Champagne vineyard estimated by radar for August 14th 2014 (week 33) with 1-kilometre-wide pixels.

Environmental factors can also vary over time. Rainfall is also a good example of it. As an illustration, Figure 6 shows the cumulative weekly rainfall over the same Champagne region during the 2014 spring and summer. Two particularly rainy periods appear at the beginning of spring (weeks 17 to 21) and in mid-summer (weeks 27 to 33) interspersed with drier periods (weeks 22 to 25 and 35 to 37). During this year 2014, the temporal organisation of the rainfall was also strong.

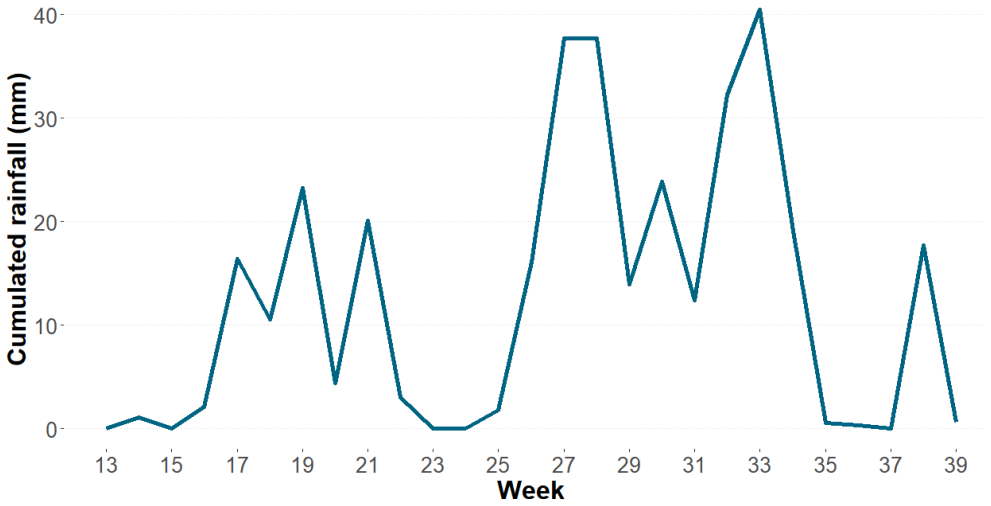


Figure 6: Temporal evolution of cumulative weekly rainfall during spring and summer 2014 in the Champagne vineyards.

Agricultural phenomena, such as biotic or abiotic stresses, that are studied in FSO projects depend on environmental factors. They are therefore influenced by the spatial and temporal structures of the environmental factors. As a result, the agricultural phenomena will also be structured in space and time (Jasse et al. 2021). This means that, for the variable under study, two observations collected at close dates and sites have a greater probability of presenting similar attribute values than two observations that are distant in time and space. Consequently, variables collected by FSO tend to show strong temporal and spatial covariances.

Specificity 2: Phenomenon are seasonal and dynamic (SEASONAL)

Studied phenomena are generally seasonal and dynamic. This means that the phenomenon appears each year at the same period and evolves during the season. This is, for example, the case for the monitoring of crop pests that appear every year in spring and whose population evolves during the summer period (Maistrello et al. 2018). Naturally, participants are trying to observe this phenomenon and tend to gather their observations during this period. As a result, the number of observations over time varies very strongly in this type of study. Observations collected on the eButterfly platform are a good illustration of this (Prudic et al. 2017). The number of observations is very high during the summer (June to August in the northern hemisphere), which corresponds to the period when butterflies are present (Figure 7). On the contrary, it is very low during winter (December to February), which corresponds to a period when the butterflies are in larva stages and not observable. In this representative example of FSO projects, the monthly number of observations thus evolves in a dynamic and seasonal manner.

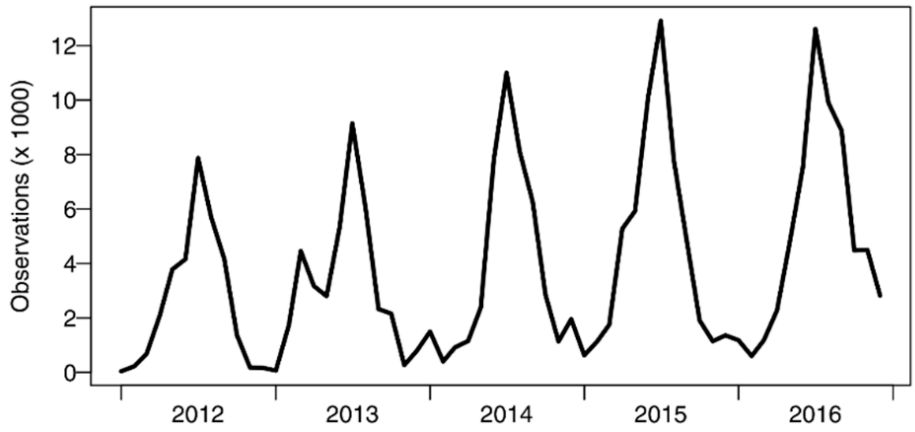


Figure 7: Monthly observations submitted to eButterfly platform for butterfly observations in North America from 2012 to 2016 (from Prudic et al., 2017).

Specificity 3: Datasets are asynchronous and heterotopic (ASYNCHRONOUS)

FSO projects are generally semi-structured (Kelling et al. 2019) with an open and flexible recruitment of contributors and opportunistic gathering strategy. The major consequence is that observations are collected at sites and dates that are only determined by the participants according to their individual constraints or motivations. These sites and dates are not known in advance by the organisation and are not necessarily optimal for the observation of the phenomenon of interest. The resulting datasets are therefore:

- Asynchronous, i.e. observations made at two different sites are usually made on two different dates.
- Heterotopic, i.e. observations made on two different dates are usually located at two different sites.

These characteristics are practically non-existent in conventional experimental approaches in agriculture. Experimentation is usually based on randomised replicate designs where measurements are systematically made on the same date and repeated on the same plots, which greatly simplifies data analysis. Therefore, the asynchronous and heterotopic nature of FSO data is a fundamental feature of these data and will condition the data processing methods that allow knowledge to be

inferred. In this context, the gathering of metadata on the observations is a major issue because *a posteriori* knowledge of the sites and dates of collection is needed in order to manage this asynchronous and heterotopic character.

Specificity 4: Participants are professionals (PROFESSIONALS)

FSO project participants contribute to these projects as part of their professional activity (Minet et al. 2017). Therefore, the participants are mostly stakeholders in the agricultural sector (farmers, advisors, etc.) and not enthusiastic amateurs or the general public, as is the case in many crowdsourcing projects (Prudic et al. 2017). This characteristic of the participants influences their motivations, which are rather egoist and extrinsic, i.e. by contributing, they try to answer a need related to their professional activity (e.g. to know the performance of a seed variety (Van Etten, Beza, et al. 2019)). This characteristic of the participants also influences their knowledge of the phenomena studied. As professionals in the agricultural sector, they contribute to the observation of phenomena that they know well. Therefore, it can be assumed that their level of expertise is higher than in other crowdsourcing projects, although no published literature was found that has studied this assumption in detail. It is likely that this characteristic also influences the sites at which participants will carry out observations. Farmers tend to make observations on the fields they work on and advisors on those they support. It is quite rare that several different participants collect observations on the same field or site. Finally, extension services are one of the key actors in FSO projects. As for the adoption of any innovation in agriculture (De Souza Filho et al. 2021), extension services play an important role in the participation of farmers to crowdsourcing projects because i) they benefit from their trust, ii) they are usually well established in the territory, and iii) they provide a link between the participants and the organisation (Ryan et al. 2018).

Specificity 5: Observations are only collected once (ONCE)

The repetition of the same task by several participants is a frequent strategy for assessing the quality of contributions in different types of crowdsourcing projects (Rodrigues et al. 2013), 2013). In the case of FSO projects, the observed phenomena are dynamic (Specificity SEASONAL). The repetition of the same observation by several participants implies that these participants carried out the same observation on the same site and on the same date. Furthermore, as agricultural professionals, participants tend to make observations only in their own areas of interest (Specificity PROFESSIONALS). From a practical point of view, the repetition of the same observation by several participants is very unlikely and almost never happens. Therefore, it can be considered that the observations of the phenomena studied are collected only once at a given site and on a given date.

Specificity 6: Reference data is not available (REFERENCE)

The reference data that are usually used in different types of crowdsourcing projects are data from an authoritative structure (e.g. the State) (Arsanjani et al. 2015). These reference data are often expensive and time-consuming to acquire. They are well suited to characterise a static phenomenon (e.g. the location of buildings, roads, etc.) (Brovelli et al. 2017). In the case of monitoring dynamic phenomena, reference data are rarely acquired or available because they remain very expensive and the duration of the acquisition protocols is often incompatible with the temporal dynamic of the phenomenon, which is often rapid. FSO projects rarely have reference data or, if they are available, they are only available very locally. As a result, the observations

made by the participants are in the vast majority of cases the only data available to describe the phenomena studied (Sullivan et al. 2009).

1.3.2 Impact on success factors

The specificities of FSO projects have an influence on the success factors of these projects (Table 4). The following section describes this influence in order to identify the most impacted success factors.

Impact on security:

FSO projects need financial, contractual and digital security to be successful. Security is an important factor but it is not influenced by any of the specificities of these projects (Table 4). Therefore, security is a success factor for FSO projects as it is a success factor for any crowdsourcing project or indeed for all projects regardless of the field of activity. This issue, although strong, is not specific to FSO projects and does not raise specific questions and will not be directly addressed in this thesis.

Impact on quantity of contributions:

The specificities of FSO projects do not influence the importance of having a mature and easy-to-use platform to get many contributions (Table 4). However, they do have a strong influence on the motivation of participants. Indeed, although agricultural stakeholders are rather enthusiastic about participating in crowdsourcing projects (Dehnen-schmutz et al. 2016), they tend to contribute only if they perceive a direct benefit for their agricultural production or business (Ebitu et al. 2021). The desire to be part of scientific research or to contribute to a collective project is less strong than in other sectors (Beza et al. 2017). Participant motivation is more egoist and extrinsic (Specificity PROFESSIONALS). Moreover, as the observed phenomena are seasonal (Specificity SEASONAL), it is likely that their contribution and motivation also changes seasonally. There are few published works that have studied FSO projects (Ebitu et al. 2021) and the impacts of their specificity on participant motivation remains an open question.

Impact on quality of contributions:

Specificities of FSO projects do not influence the *ex ante* strategy for managing the quality of collected observations (Table 4). However, these specificities can have a strong impact on the *ex post* strategy. Indeed, in these projects, the difficulties in acquiring reference data (Specificity REFERENCE) or in comparing the collected data with those collected by other participants (Specificity ONCE) limit the possible *ex post* strategies. The only possible approaches are to study the characteristics of the collected dataset (Goodchild and Li 2012) or to assess the reputation of the participants (Jabeur et al. 2018). In the case of reputation assessment, the approaches classically proposed in the literature are difficult to transpose to FSO projects. Most approaches have assessed the reputation of a participant by studying the modifications of their contributions by other participants (number of modifications, time to first modification, etc.) (Fogliaroni et al. 2018). These methods require that the same object or phenomenon is observed by different operators. However, in the case of FSO projects, observations are usually collected only once at a given site and date (Specificity ONCE). Moreover, they are considered to be made by the best "local expert", so that modification by other participants can hardly be justified. Therefore, existing *ex post* strategies in the literature are faced with their limitations in adapting to FSO

projects. Nevertheless, the phenomena studied have interesting characteristics to consider original *ex post* strategies. Spatial and temporal structures of the studied phenomena (Specificity STRUCTURED), as well as their seasonality (Specificity SEASONAL), make it possible to estimate expected values. This offers the opportunity for identifying observations that deviate from these expected values, either because the phenomenon studied effectively presents a local anomaly or because the observation considered is erroneous. Finally, in order to be implemented on large amounts of data, tools and methods dedicated to the evaluation of the quality of contributions must be automatable and adaptable to asynchronous and heterotopic datasets (ASYNCHRONOUS specificity). *Ex post* strategies adapted to FSO projects are currently little studied. However, they represent both a major challenge for the development of these projects and raise original questions for the scientific community.

Table 4: Potential impact of specificities of Farmsourcing projects of Spatial Observations on the success factors of crowdsourcing projects.

		Specificities of Farmsourcing projects of Spatial Observations					
		STRUCTURED	*SEASONAL*	*ASYNCHRONOUS*	*PROFESSIONALS*	*ONCE*	*REFERENCE*
Success factors of crowdsourcing project	*Security*						
	Security						
	Quantity of contributions						
	Platform maturity						
	Value and ease of use						
	Motivation						
	Quality of contributions						
	Ex-ante strategy						
	Ex-post strategy						

Potential impact

Legend:

	No impact
	Low impact
	Strong impact

1.4 Scientific question and organisation of the manuscript

1.4.1 General approach

The specificities of FSO projects have an impact on both the quantity and quality of contributions. In particular, they influence the motivation of participants and the *ex post* quality management strategy. In the literature, a few studies have mentioned these success factors (Van Etten, Beza, et al. 2019), but the analyses are often focused on specific, studied projects. To the best of our knowledge, no study has investigated systemically the specific success factors of FSO projects. Therefore, a study of these factors is not possible by conducting a literature review, such as the one proposed by Rechenberger et al. (2015) for crowdsourcing projects in general. However, it is possible to make assumptions about these specific factors based on the generic success factors of crowdsourcing projects (1.2.2) and on the influence that the specificities of FSO projects may have on these factors (1.3.2).

The work carried out in this thesis aims to formulate these hypotheses and then to test them through the implementation of a systemic approach. This approach consists in conducting a real FSO project through its main stages, including its definition, implementation and evaluation (Table 5). This approach then consists of testing the hypotheses formulated during the different stages of the project and generalising the results in order to validate (or not) these hypotheses. In order to ensure this generalisation, the project chosen must have characteristics that are representative of all FSO projects.

1.4.2 ApeX-Vigne use case

The ApeX-Vigne project was chosen as a use case to implement the systemic approach (Table 5). This FSO project aims to map vine water status at a regional scale. It was mainly developed in the south of France. Participants contributed by gathering observations on the vine shoot growth during summer using the iG-Apex method. They used the mobile application ApeX-Vigne, which was designed and developed during this thesis. In this project, the *ex post* quality strategy was ensured by methods that automatically identify surprising observations and outliers. This project is presented in detail in the following chapters. The ApeX-Vigne was officially launched in 2019. Its evaluation of the quality and quantity of contributions focuses on observations collected during the 2019 and 2020 seasons.

Table 5: Six main steps of a Farmsourcing project of Spatial Observations for the general case and for the ApeX-Vigne use case

		General case	ApeX-Vigne use case
Defining the project	1	Choosing the agricultural issue	Choosing vine water status issue
	2	Defining the contribution for participants	Defining the iG-Apex method
Building the project	3	Designing the platform	Designing the ApeX-Vigne application
	4	Designing methods for *ex post* strategy	Designing automatic methods for outlier and surprising observation detection
Evaluating the project	5	Evaluating the quantity of contributions	Evaluating the quantity of ApeX-Vigne contributions in 2019 and 2020 seasons
	6	Evaluating the quality of contributions	Evaluating the quantity of ApeX-Vigne contributions in 2020 season

1.4.3 Formulated hypothesis

Two main hypotheses on the specific success factors of FSO projects will be tested in this thesis. They focus on the success factors that are most impacted by the specificities of FSO projects. The first hypothesis concerns the specific success factors that would allow for the collection of a significant amount of observations i.e. the quantity of contributions collected. The second hypothesis focuses on the specific success factors that would favour the quality of these contributions (Figure 8).

Hypothesis 1: FSO projects must address multiscale agricultural issue, propose useful and simple tasks to participants and rely on an accessible platform to generate a significant amount of observations.

- *Specific success factor studied 1: The project addresses an agricultural issue both at the field or farm scale and at the regional scale (MULTISCALE)*

  An organisation sets up an FSO project to understand a phenomenon at the scale that they are interested in, which is the regional scale. Participants contribute to this project to understand a phenomenon at the spatial scale they are concerned with, i.e. the field or farm scale (specificity PROFESSIONALS). It is hypothesised that the agricultural issue addressed in an FSO project must meet this double expectation. It must address both a regional issue to interest the organisation and a local issue to interest participants.

- *Specific success factor studied 2: The task proposed to the participants is simple and offers them a counterpart directly useful in their professional activity (USEFUL)*

  Participants in FSO projects contribute as part of their professional activity (specificity PROFESSIONALS). They often have little time to contribute and expect a direct feedback. Therefore, it is hypothesised that the contribution requested from participants must be simple and must give them access to information that is useful for their professional activity for the FSO to be successful.

- *Specific success factor studied 3:* The platform is accessible to a high number of potential participants (ACCESSIBLE).

  A FSO platform must first of all be easy to find by potential participants. It is hypothesised that it should also follow the same rules as any other crowdsourcing platform, in particular regarding its maturity (factor MATURITY) and ease of use (factor VALUE). It must enable the centralization of collected observations so that the organisation can use them. Finally, as participants contribute in the context of their professional activity (specificity PROFESSIONALS), they are likely to be particularly sensitive to the trust they give to the organisation (factor MOTIVATION).

Hypothesis 2: FSO projects must rely on a standardized gathering protocol, the recording of metadata and ex post quality methods based on attribute, spatial and temporal characteristics of observations to generate contributions of good quality.

- *Specific success factor studied 4: The platform enables the gathering of contributions according to a standardised protocol. (STANDARDIZED)*

  In order for observations collected by different operators and under different conditions to be studied together and eventually compared, it is necessary to limit the variation factors affecting the measurement (factor EX-ANTE). It is hypothesised that collecting observations according to a standardised protocol will permit the collection of good quality observations.

- *Specific success factor studied 5: The platform enables the gathering of metadata about the conditions under which contributions were acquired. (METADATA)*

  Time and space strongly structure the studied phenomena (specificities STRUCTURED and SEASONAL). Therefore, gathering metadata about the site and the date of collection of observations is as important as the actual observation in order to be able to properly interpret the collected data (factor EX-ANTE). It is hypothesised that recording the site and date of each observation collected will facilitate the analysis of the data and thus would ameliorate indirectly data quality.

- *Specific success factor studied 6: The ex post quality methods are based on attribute, spatial and temporal characteristics of the observations. (MULTI-CHARACTERISTIC)*

  Quality improvement strategies after the collection of observations are difficult considering the unique characteristics of FSO datasets (specificities ASYNCHRONOUS, ONCE and REFERENCE). Nevertheless, time and space strongly structure the phenomena studied

(specificities STRUCTURED and SEASONAL) and offer opportunities for quality improvement. It is hypothesised that the implementation of analytic methods based on the attribute, spatial and temporal characteristics of the observations could improve the quality of the observations gathered, in particular by identifying outliers and surprising observations.

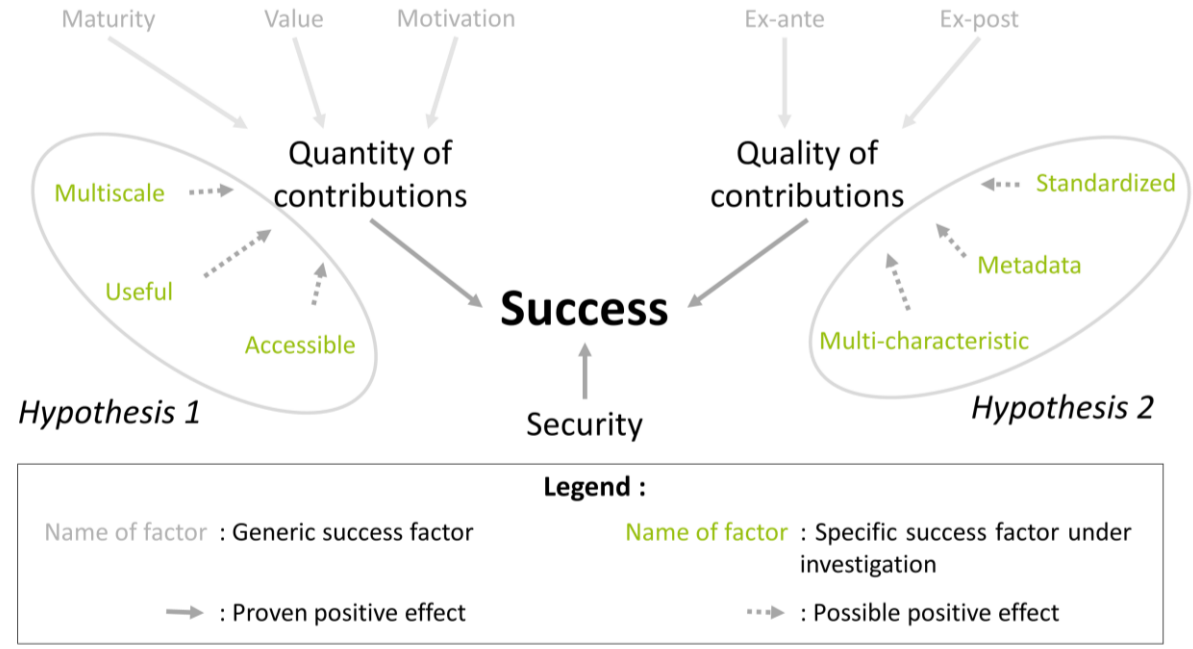


Figure 8: Specific success factors of Farmsourcing projects of Spatial Observations tested by hypothesis 1 and 2 and their relation with generic success factors.

According to the systemic approach implemented in this thesis, the two hypotheses were tested throughout the main stages of the studied use case (i.e. ApeX-Vine) (Figure 9). First, it was demonstrated that the use case did effectively present the specific success factors under study in order to be able to test hypothesis 1 (i.e. MULTISCALE, USEFUL and ACCESSIBLE) and hypothesis 2 (i.e. STANDARDIZED, METADATA, MULTI-CHARACTERISTIC). Secondly, these hypotheses were tested and the results were generalised.

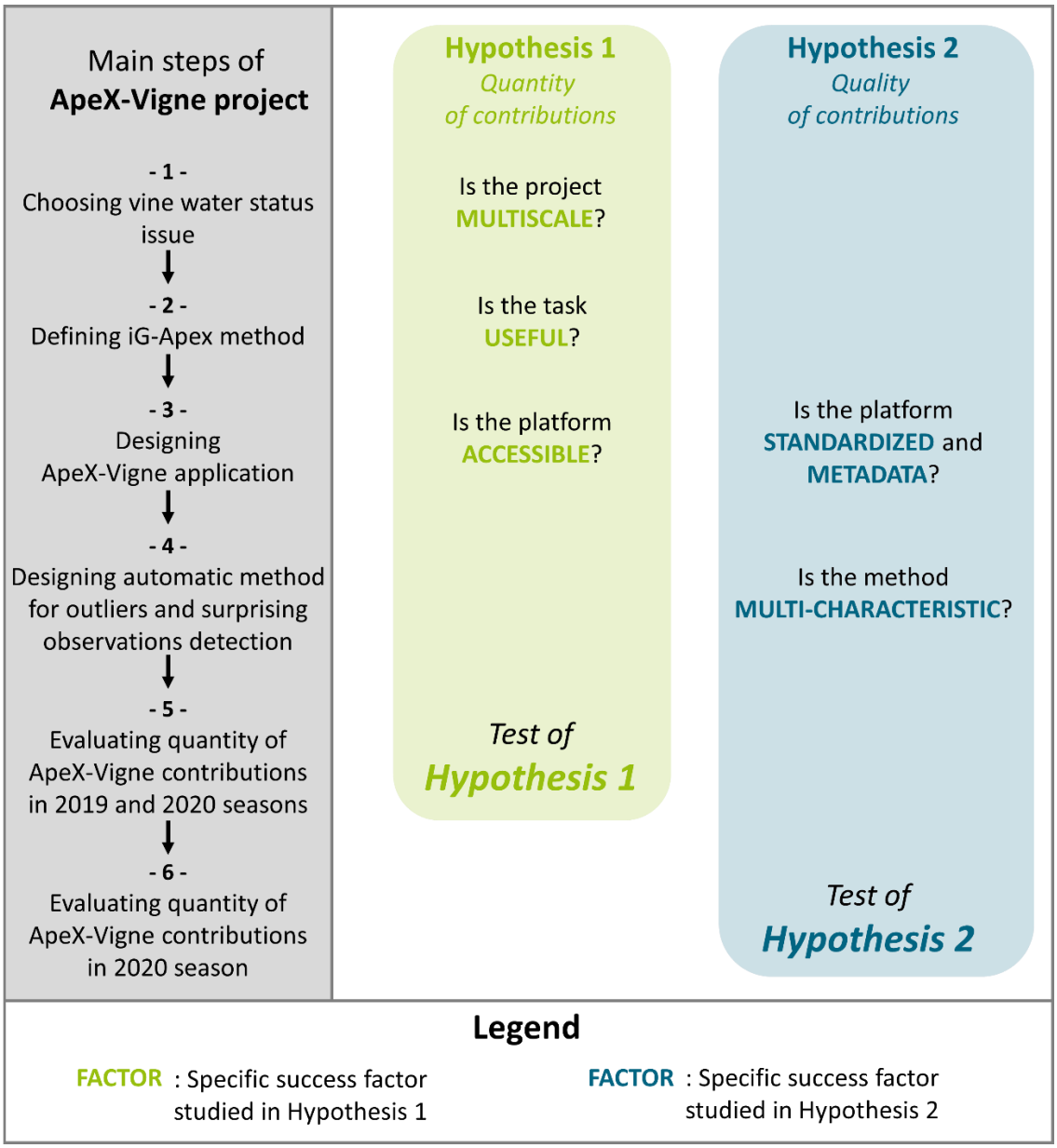


Figure 9: Checking specific success factors and testing hypotheses along the steps of the use case in the proposed systemic approach.

1.4.4 Organisation of the manuscript

After this first introductory chapter, the manuscript is structured in 6 chapters (numbered from 2 to 7) following the demonstration and the main steps of the use case (Figure 10). Chapter 8 concludes the manuscript and opens up perspectives for future research.

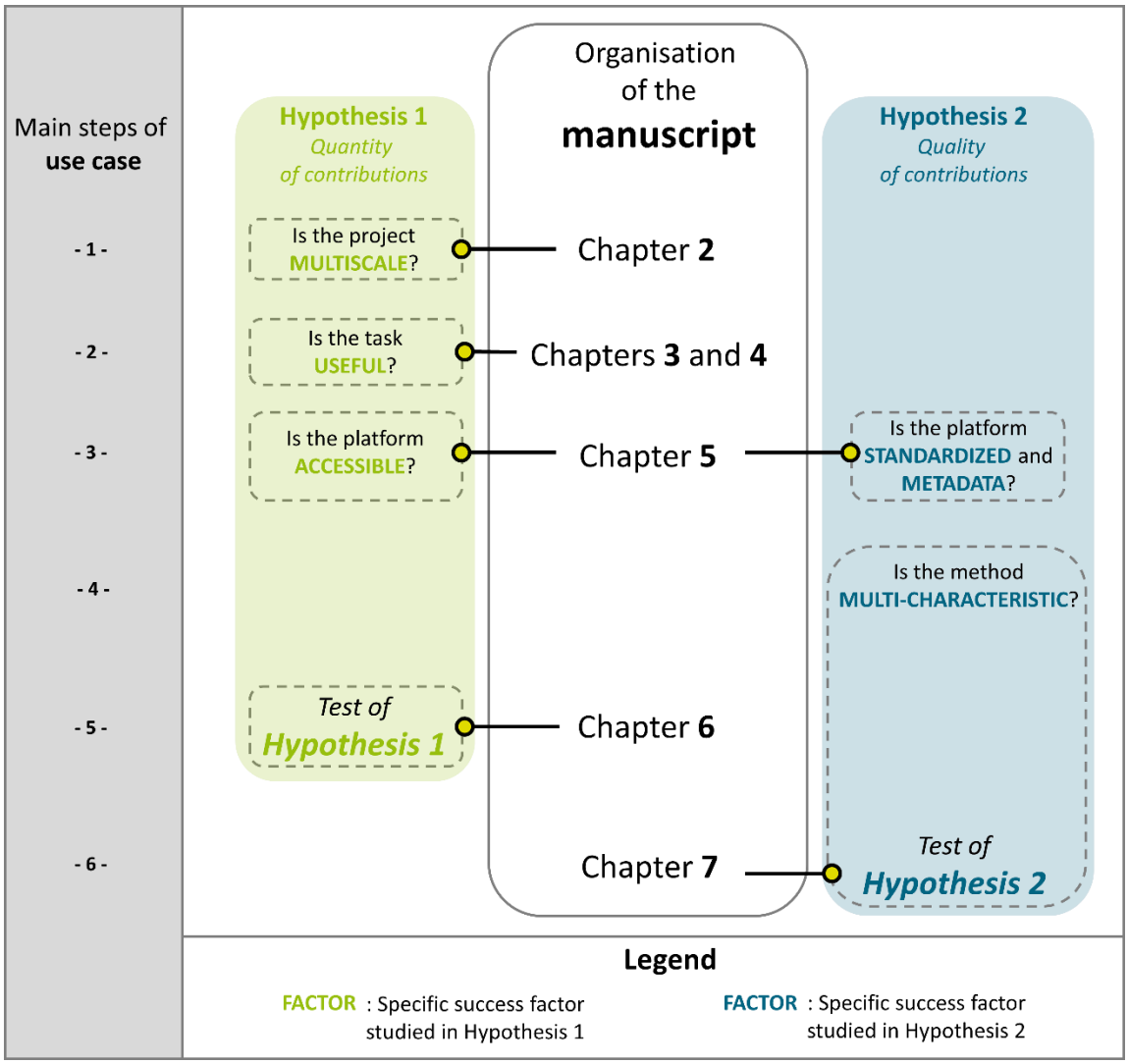


Figure 10: Organisation of the manuscript: implementation of the systemic approach along chapters 2 to 7 based on the main steps identified in Figure 9..

Chapter 2: Monitoring of vine water status in the south of France: an issue at different spatial scales

The objective of chapter 2 is to present the use case of vine water status monitoring and to detail its issues at different spatial scales (specific factor MULTISCALE). This chapter does not aim to produce knowledge. It simply provides a quick literature review on the vine water status monitoring issue in order to demonstrate how important it is at different spatial scales and for different stakeholders.

Chapter 3: Observation of shoot growth: a simple and operational method to monitor vine water status in the vineyard

The objective of chapter 3 is to demonstrate that the iG-Apex method based on the measurement of vine shoot growth is a relevant indicator from an operational and scientific point of view for monitoring vine water status in the vineyard (specific factor USEFUL). The questions explored in this chapter are: Is it possible to establish a relationship between iG-Apex and a reference method

for monitoring vine water status? Under what conditions is this relationship valid? What recommendations for use can be drawn from it?

Chapter 4: Characterising within-field variability of vine water status with simple visual observations of shoot growth

The objective of chapter 4 is to demonstrate that the iG-Apex method can also be used to characterise the within-field variability of vine water status (USEFUL specific factor). The question explored in this chapter is: Is the uncertainty associated with the iG-Apex method low enough to allow participants to characterise vine water status at the within-field level and to identify the optimal period for using this method as an operational decision support tool?

Chapter 5: ApeX-Vigne: a platform designed for collecting farmsourcing spatial observations

The objective of chapter 5 is to demonstrate that the ApeX-Vigne mobile application is easily accessible and usable by potential participants (specific factor ACCESIBLE), that it enables data to be collected according to a standardised protocol (specific factor STANDARDIZED) and that it records contextual information on the observations gathered (specific factor METADATA). The question explored in this chapter is: What methodological and technological choices should be made to provide an easily accessible and ergonomic mobile application that enables the collection and centralisation of standardised, dated and geo-referenced vineyard observations?

Chapter 6: Evaluating the quantity of contributions in a farmsourcing project of spatial observations

The objective of chapter 6 is to test hypothesis 1 and to generalise the results on specific success factors regarding the quantity of contributions. The question explored in this chapter is: How can the adoption and uses of an FSO application be evaluated? Did the ApeX-Vine application allow the collection of a large quantity of observations?

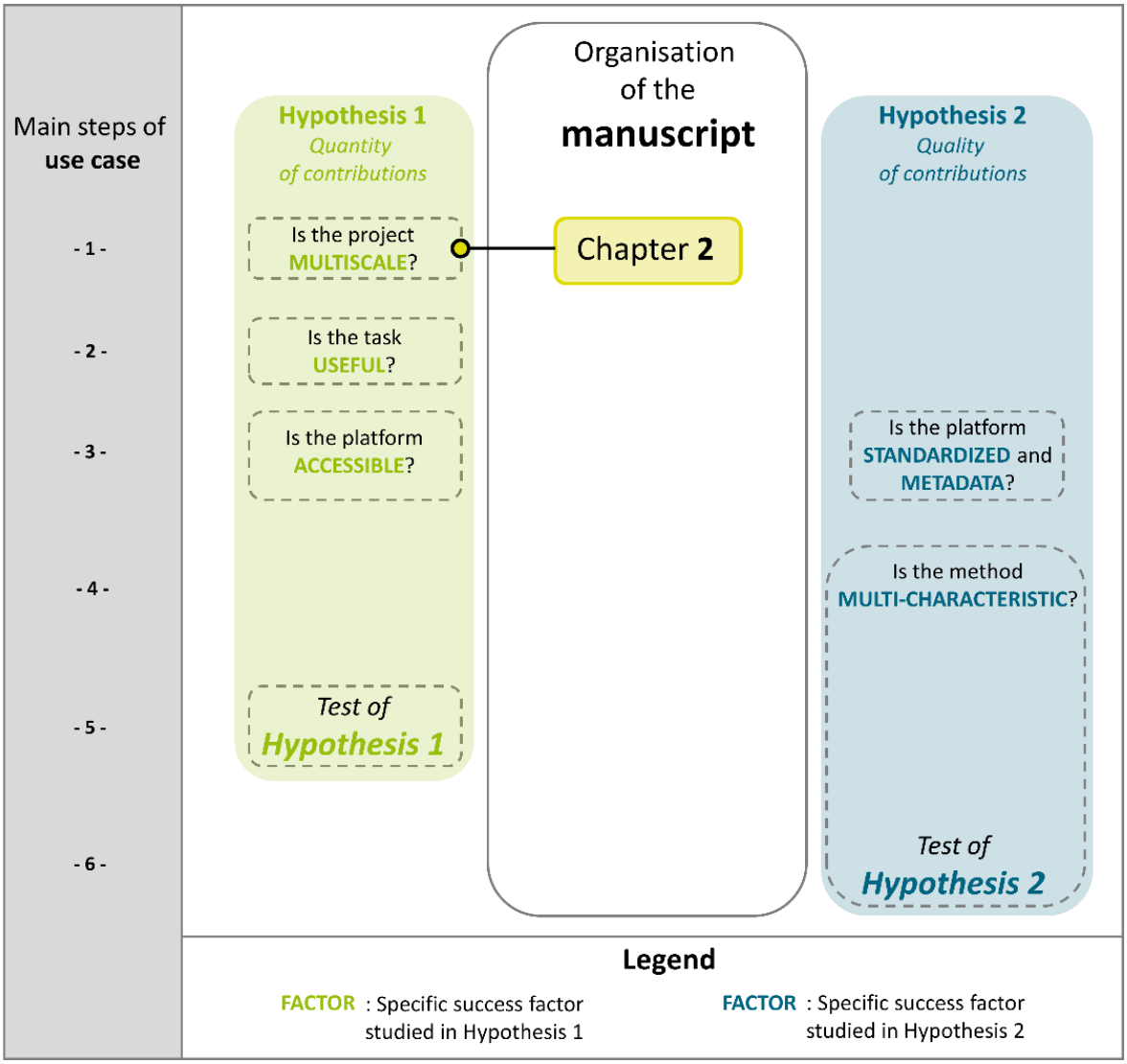
Chapter 7: Evaluating the quality of contributions in a farmsourcing project of spatial observations

The first objective of chapter 7 is to demonstrate that the *ex post* quality management strategy implemented in the ApeX-Vigne project is based on the attribute, spatial and temporal characteristics of observations (MULTI-CHARACTERISTIC). The second objective is to test hypothesis 2 and generalise the conclusions about specific success factors concerning the quality of contributions. The questions explored in this chapter are: How to propose a method based on the attribute, spatial and temporal characteristics of observations to automatically identify outliers and surprising observations in a FSO dataset? What are the performances of this method?

Chapter 8: Conclusion and perspectives

The objective of chapter 8 is to conclude on the validation of the two formulated hypotheses and to give directions for future research work. Hypotheses 1 and 2 can be considered as validated for the ApeX-Vigne case study and for all FSO projects. The study of time series of observations and the consideration of participants' reputation are ways to improve the proposed approach. The use of observations from FSO projects jointly with other sources of information is a direction that will be explored in future research in particular on the inversion of mechanistic models.

Chapter 2: Monitoring of vine water status in the south of France: an issue at different spatial scales



The objective of chapter 2 is to present the use case of vine water status monitoring and to detail its issues at different spatial scales (specific factor MULTISCALE). This chapter does not aim to produce knowledge. It simply provides a quick literature review on the vine water status monitoring issue in order to demonstrate how important it is both at field and regional scales and for different stakeholders.

2.1 Vine water status influences vineyard performances in many ways

Vine water status is a physiological quantity describing the state of the plant regarding water resources. When this resource is limiting, vine water status corresponds to what is called a water constraint. This water constraint plays a particularly important role in vineyard management because it influences vegetative growth (Lebon et al. 2006), grape quality (Van Leeuwen et al. 2009) and yield (Medrano et al. 2003) (Figure 11).

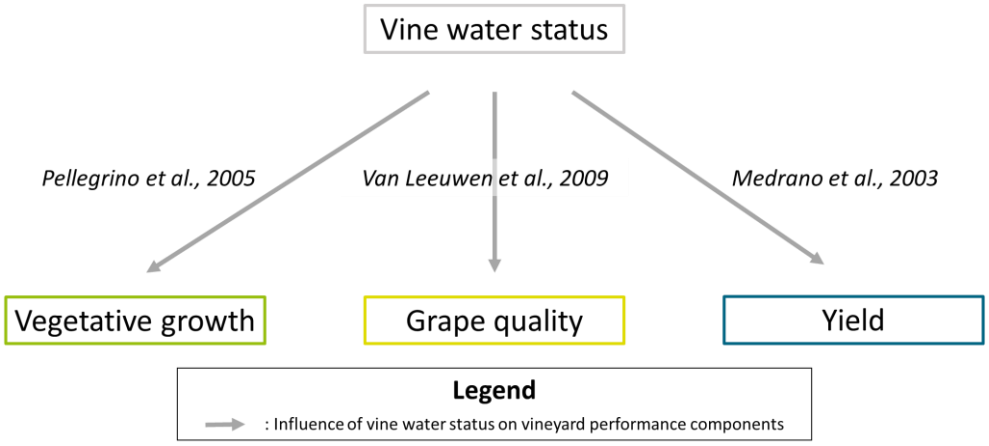


Figure 11: Summary of the main vineyard performances impacted by vine water status.

The impact of water constraint on vineyard performance depends on its intensity, its duration and its period of occurrence. Many works have addressed these issues and the literature on the subject is abundant (Mirás-Avalos and Silva Araujo 2021; Scholasch and Rienth 2019). The objective of this section is simply to expose the issues of water constraint for vineyard management purposes by highlighting the main phenomena involved:

- The period between budburst and flowering corresponds to the development of the vine's vegetative system. If a water constraint, even low, occurs during this period, the vegetative growth is slowed down or even stopped (Lebon et al. 2006). The main consequence is a decrease in the total leaf area per vine (Pellegrino et al. 2005) and thus a reduction in the plant's capacity to perform photosynthesis (Hsiao 1973). It has also been demonstrated that water constraint around flowering in year n partly determine the maximum yield that can be reached in year n+1 (Guilpart et al. 2014).
- The period between setting and veraison corresponds to the herbaceous growth of the berry. A water constraint occurring during this period affects the average berry size and consequently reduces final yield. This effect can be targeted because it also increases final berry quality through concentration of quality determining compounds in the berry (Ojeda et al. 2001).
- The period between veraison and harvest corresponds to berry ripening. A water constraint occurring during this period mainly influences berry quality. This quality is complex to assess but is often described by four main characteristics (Acevedo-Opazo et al. 2010; Zufferey et al. 2017): berry size, sugar content, total acidity and total polyphenol content. All of these characteristics are affected by water constraint (Mirás-avalos and Intrigliolo 2017). A water constraint occurring during this period can also reduce yield, but to a lesser extent than if it occurs before veraison (Scholasch and Rienth 2019).

The influence of vine water status on vineyard performance is therefore complex. It evolves along the vegetative cycle and may require compromises, for example between quality and yield. Typical trajectories corresponding to different production objectives have been proposed to assist winegrowers in their vine water status management strategy (Figure 12) (Ojeda and Saurin 2014). These trajectories illustrate the importance of monitoring vine water status and therefore measuring or estimating it throughout the season.

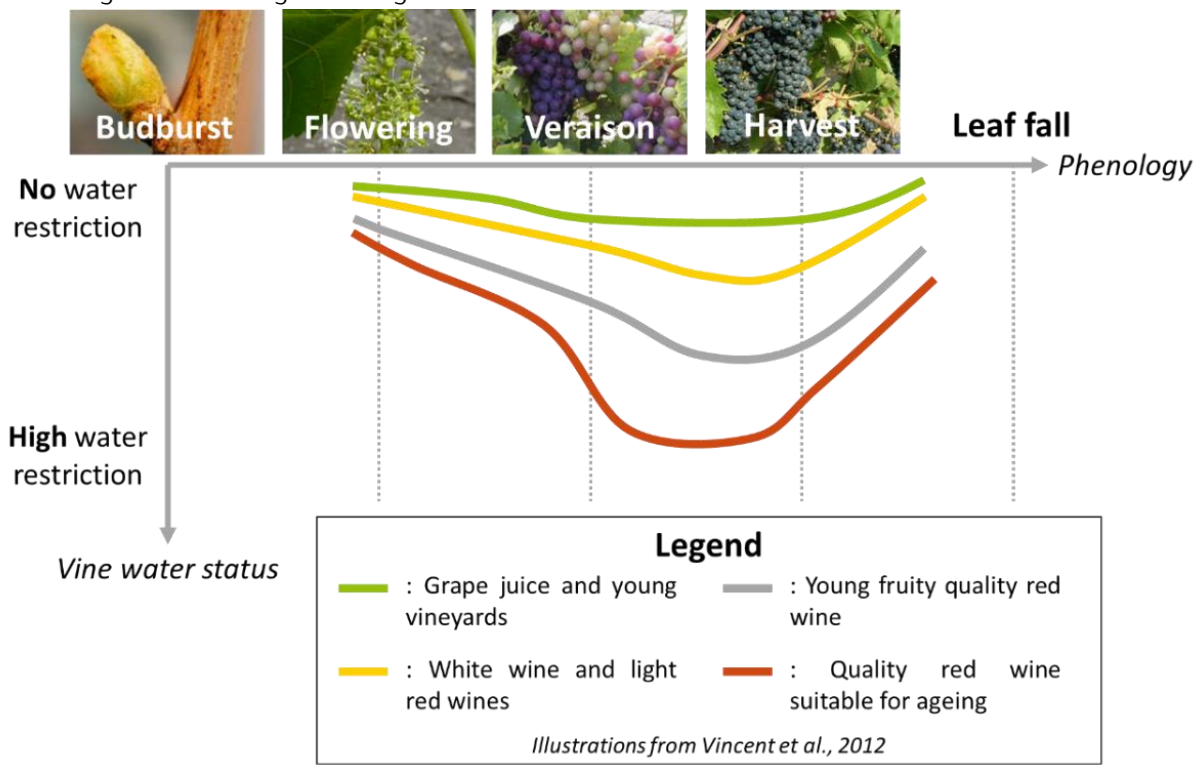


Figure 12: Schematic representation of possible vine water status trajectories according to objectives of production (redrawn after Ojeda and Saurin 2014).

2.2 Vine water status has to be monitored in time and space

2.2.1 Vine water status varies over time and space, even at within field scale

Vine water status depends mainly on soil and climatic conditions. The climate determines both the amount of water the plant receives through precipitation (Mirás-Avalos and Silva Araujo 2021) and the amount of water consumed through evapotranspiration (Lebon et al. 2003). The soil, in particular its texture, determines the amount of water that can be stored and then gradually mobilised by the plant (Van Leeuwen et al. 2018). However, climate is variable both in time and space (Neethling et al. 2019) and soil characteristics vary in space (Román Dobarco et al. 2017). As a result, the vine water status also varies greatly in time and space. Figure 13 illustrates this phenomenon with three maps showing, on the one hand, the spatial variability of the vine water status at within-field (Figure 13a) and inter-field (Figure 13b) scales and, on the other hand, the temporal evolution of this spatial variability at the within-field scale (Figure 13c). These maps are derived from works published in the literature. They are based on three types of measurements of vine water status that will be detailed in the following section: Predawn Leaf Water Potential, $\delta^{13}C$ and Stem Water Potential, respectively. These measurements provide information on the

vine water status at the time of the measurement (water potentials) or throughout the season ($\delta^{13}C$) (see next section).

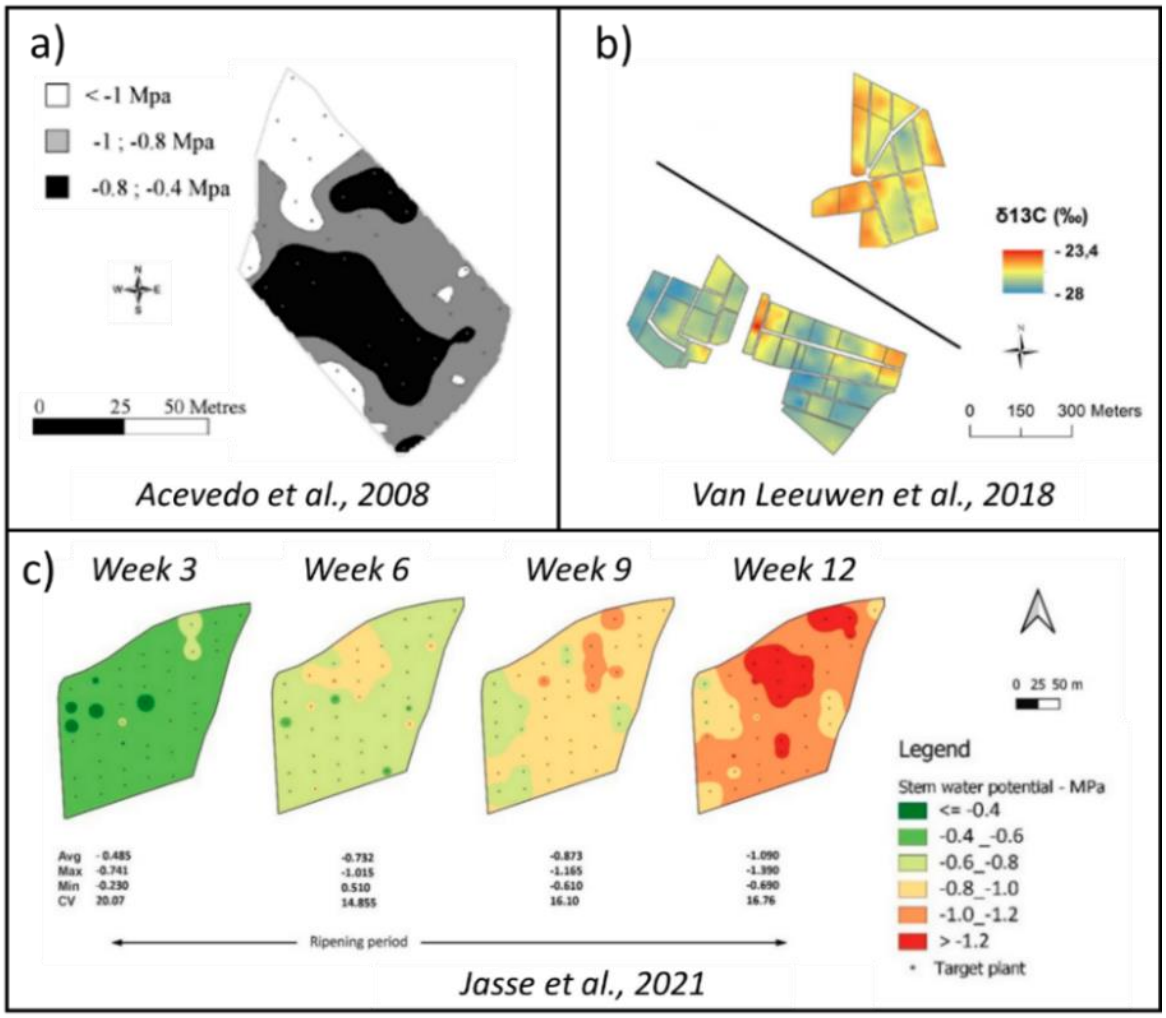


Figure 13: Examples of spatial and spatio-temporal variations of vine water status: (a) within-field and (b) inter-field variability of vine water status measured by predawn leaf water potential at the end of summer near Gruissan, Languedoc, France (from Acevedo et al. 2008) and by $\delta^{13}C$ near Bordeaux, France (from Van Leeuwen et al., 2018), respectively. (c) Temporal evolution of within-field variability of vine water status measured by stem water potential near Simonsberg, South-Africa (from Jasse et al., 2021).

Given the agronomic importance of vine water status monitoring and its variability in time and space, the objective for vineyard management is therefore to be able to estimate it as accurately as possible, specifically, regularly and if possible exhaustively. The following section details the existing tools and methods for carrying out these measurements.

2.2.2 Approaches for monitoring vine water status are resulting from a trade-off between reliability, spatial and temporal resolutions

Many approaches have been developed to estimate vine water status. They are based on measurements carried out either i) directly on the plant (Choné et al. 2001), ii) or on factors influencing vine water status such as soil (Gardner et al. 2001) or climate (Lebon et al. 2003), iii) or indirectly, on factors that are influenced by vine water status such as vegetation growth (Martinez-De-Toda et al. 2010) or berry composition (Gaudillère et al. 2002). Approaches based on plant measurements are often more reliable but also more complex to implement. The operational constraints of these approaches limit the frequency and the number of observations that can be made in a given time. In their state of the art, Rienth and Scholasch (2019) identified eight main approaches. Depending on their reliability, these approaches can be considered as reference methods, approaches with high reliability or approaches with some limitations. Acevedo-Opazo et al. (2008) estimated for each of them the area that could be measured in one day and the temporal resolution that could be achieved given the operational constraints of their implementation. The section below briefly describes all of these approaches, focusing on their operational limitations in the context of a spatio-temporal monitoring of vine water status.

- *Pressure chamber (PC):* This approach is considered as a reference measurement in many research works (James A. Taylor et al. 2012). It is used in research studies, for technical experimentation and in some cases in commercial vineyards. The principle of the measurement is to estimate the water potential of the vine (Ψ) by measuring the suction pressure required to extract a drop of sap from a leaf. When the measurement is carried out between 13.00h and 16.30h by forcing transpiration to stop, it provides information on the plant's potential (Stem Water Potential) (Choné et al. 2001). When it is taken before sunrise, it provides information on the soil's potential (Predawn Leaf Water Potential) (Améglio et al. 1999). In both cases, operational constraints are heavy and limit the number of observations that can be made within a day and during a whole season (Acevedo-Opazo et al. 2008).
- *Stomatal conductance and leaf gas exchange measurements (Sto)*: This approach is considered as a reference measurement in many research works (Pellegrino et al. 2005). It is mainly used in research studies. This approach consists of estimating the closure of stomata, which is one of the first physiological responses of the vine to, even very low, water restriction. It is based on measuring the flow of water vapour and/or CO2 between the leaves and the atmosphere using a porometer (Jones et al. 2002) or an infrared gas analyser (Caravia et al. 2016). In both cases, this equipment is specific and expensive which limit the use of this approach. A few dozen observations can be made in one day (Acevedo-Opazo et al. 2008).
- *Sap flow based measurement (SF)*: This approach consists of estimating the transpiration of the plant, by measuring the sap flow between roots and leaves. It is used in some commercial vineyards and it is considered reliable. Several methods are available to carry out this measurement (Rienth and Scholasch 2019) but in all cases they require instrumenting a vine stock with specific and expensive equipment. This limits the number of vines monitored and therefore the characterisation of spatial variability (Acevedo-Opazo

et al. 2008). With this approach, sensors can be connected and observations can be collected automatically at very short time intervals.

- *Soil sensors (SS)*: This approach consists of estimating the soil moisture which is one of the main factors influencing vine water status. It is quite frequently used in commercial vineyards and in technical experimentation. The measurement is carried out by estimating the percentage of water in a given volume of soil or by measuring energy with which water is withheld by the soil (Gardner et al. 2001). With this approach, sensors can be connected and observations can be collected automatically at very short time intervals. However, the measurement is carried out on a single site, which limits the characterisation of spatial variability. Moreover, this approach is often imprecise because the depth of measurement is rarely sufficient to take into account the entire volume of soil explored by vine's roots. In addition, vine water status is not directly related to soil water content, which may limit the reliability of the approach in some cases (Rienth and Scholasch 2019).
- *Water balance methods (WB):* This approach consists of estimating the total amount of transpirable soil water (TTSW) which influences vine water status (Lebon et al. 2003). Estimates are made using a water balance which accounts for water inputs (irrigation plus precipitation) and outputs (runoff plus drainage plus evapotranspiration) to the soil. This approach is used in some research work (Gaudin et al. 2014) and in some commercial vineyards. The interest of this approach is to estimate TTSW at very short time intervals and at many sites. However, it is dependent on the quality of the initial TTSW estimate and the risk of overestimating water constraint is high (Rienth and Scholasch 2019).
- *Atmosphere based methods (Atm)*: This approach consists of estimating evapotranspiration of a vineyard, which is correlated to vine water status, by carrying out an energy balance (S. Li et al. 2008). It is mainly used in research studies since it is complex and requires expensive sensors not adapted to a commercial vineyard. Simplifying hypotheses make it possible to implement it, but the estimation of evapotranspiration solely due to the vine is still under investigation (Rienth and Scholasch 2019).The interest of this approach is to estimate evapotranspiration at very short time intervals and at many sites but the many simplifying assumptions limit the reliability of the estimation (Rienth and Scholasch 2019).
- *Visual observations (VO)*: This approach consists of characterising the stop of shoot growth, which is one of the first physiological functions affected by water constraint (Martinez-De-Toda et al. 2010). It is commonly used in production vineyards and in technical experimentation. The measurement is carried out by observing the growth of 30 to 50 apexes. With this approach, observations can be made quickly and without specific equipment. About 100 observations can be made in one day. During a season, the same site is generally sampled a dozen times. However, this approach has its limits, as vegetative growth can also be affected by other phenomena than water restriction (see chapter 3).
- *Carbon isotope discrimination (*$\delta^{13}C$*)*: This approach consists in measuring the ratio (called $\delta^{13}C$) between ^{13}C and ^{12}C isotopes present in berries at maturity. This ratio is considered as an integrative indicator of the water constraint experienced by the vine during grape ripening (Gaudillère et al. 2002). This approach is quite commonly used in research work (Bota et al. 2016) and in commercial vineyards. It is reliable but can only be carried out at the end of the season. About 100 observations can be made in one day (Acevedo-Opazo et al. 2008).

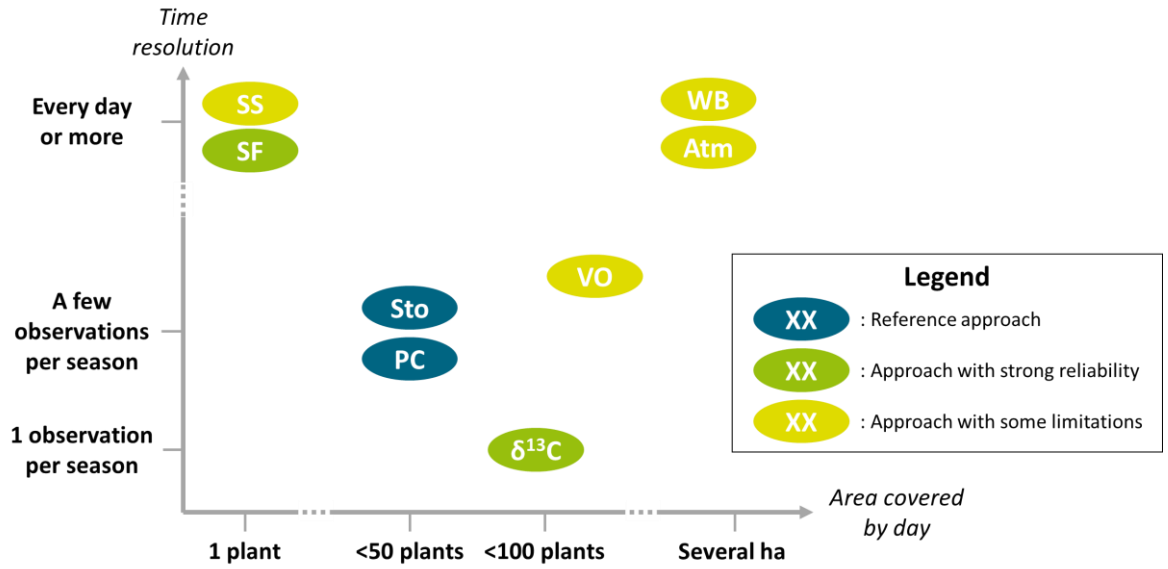


Figure 14: The different approaches for assessing vine water status (Rienth and Scholasch, 2019) are resulting from a trade-off between reliability, spatial and temporal resolutions. For each approach, spatial and temporal resolutions are represented in X and Y axis (according to Acevedo-Opazo et al., 2008) and reliability is represented with different colours (according to Rienth and Scholasch, 2019).

Each of these methods corresponds to a compromise between reliability, and spatial and temporal resolutions (Figure 14). The most reliable methods considered as a reference (i.e. Sto and PC) have operational constraints that limit their spatial and temporal resolutions. Vine water status monitoring with these methods can therefore only be carried out on a small number of sites and at a small number of dates during a season. Other methods, which are still very reliable, make it possible to increase either the spatial resolution (i.e. δ13C) or the temporal resolution (i.e. SF) of the measurement. However, their operational constraints do not allow to improve both dimensions at the same time. Methods allowing to monitor the vine water status at many sites and at many dates (i.e. Atm, WB and to a lesser extent VO) often require the estimation of parameters that are difficult to measure properly (i.e. the estimation of the initial TTSW). They therefore require strong assumptions before implementation.

This brief review of methods for estimating vine water status shows the diversity of existing approaches and the specificities of each of them. It appears that none of these methods can simultaneously meet all the operational expectations for monitoring vine water status: i) precise and specific measurements, ii) tools easy to deploy and iii) an approach that can simultaneously cover large areas and ensure monitoring over time. In operational conditions, the choice of a method for monitoring the vine water status will therefore necessarily be a compromise between all these characteristics.

2.3 Many actors are concerned by monitoring vine water status both at vineyard and regional scales

2.3.1 Monitoring vine water status at field scale for managing vineyard

The vine water status has a complex influence on the quantity and quality of the harvest. For growers, the issue is to be able to monitor vine water status at sufficiently small spatial and temporal resolutions to be able to support their decision-making. On the one hand, this monitoring must help them to take tactical decisions, i.e. decisions concerning farming practices to be implemented to meet the quality and quantity objectives they have set for the current season. These decisions may, for example, concern irrigation, cover crop destruction or reduction of leaf area by shoot trimming. On the other hand, monitoring of vine water status should also support winegrowers to make strategic decisions, i.e. decisions concerning farming practices to be implemented to reach the objectives they have set over the long term (i.e. several years to several decades). These decisions may, for example, concern the installation of shading or the planting of a new vineyard including the choice of cultivar, rootstock, planting density, row orientation, training system, etc. Winegrowers are sometimes supported by advisors providing them with individualised advice. These advisors therefore share the same issues as the winegrowers and have the same expectations regarding monitoring of vine water status. For advisors, the number of fields monitored is nevertheless much greater than for the winegrowers because each advisor generally assists several dozen winegrowers. This greater number of fields to be monitored can influence the vine water status monitoring approach the advisors may choose and promote.

2.3.2 Monitoring vine water status at regional scale for setting up collective strategies

Monitoring of vine water status is also an issue for organisations supporting farmers of a given territory, such as cooperative wineries, denomination of origin producers' union, chambers of agriculture, etc. These structures work on territories that are a few kilometres to a few dozen kilometres wide. Their mission is to animate collective dynamics between growers, to improve production overall quality or to collectively manage water resources. Monitoring vine water status at their territory scale help them to understand and anticipate the main trends at work regarding water resources. These structures can therefore identify periods of occurrence of strong water constraint or zones regularly impacted by water stress. This information can be used, for example, to define a collective strategy, to anticipate harvest logistic, to apply for irrigation authorisation or to define sectors where the installation of irrigation equipment should be prioritised. These structures also use this information to provide generic and often free advice to the winegrowers of their territory. This is the case, for example, at the scale of a denomination of origin zone (Figure 15) (Bécart et al. 2020) or at a French department scale (Michelot 2020).

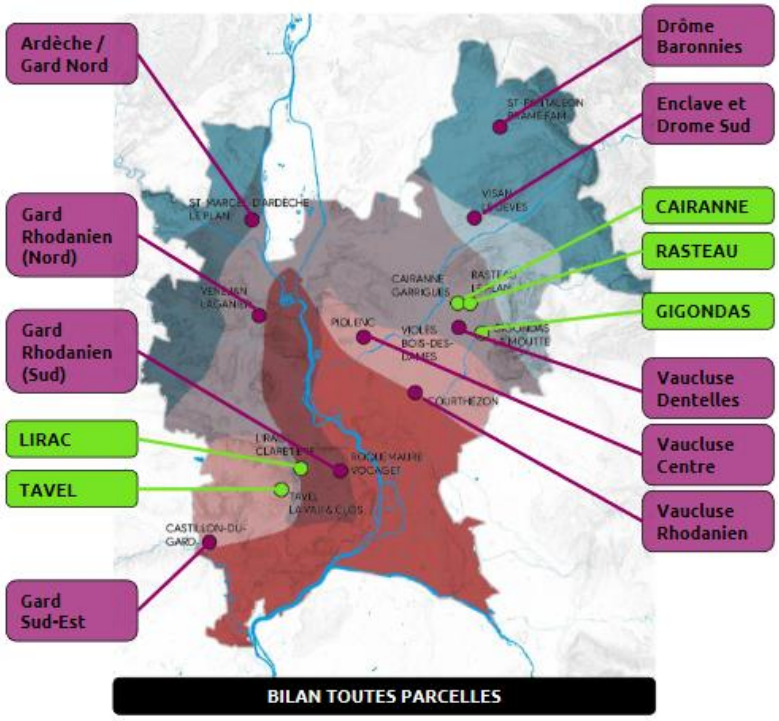


Figure 15: Extract from a weekly technical report published by the "Côtes du Rhône" syndicate on the monitoring of vine water status at the whole appellation scale. Polygons represent homogeneous pedoclimatic zones and points represent reference fields on which observations are made weekly (from Bécart et al. 2020).

In addition, monitoring vine water status on larger scales, such as regional scale, can also be a good indicator of climate. This information can, for example, be used to characterise the effects of climate change in order to better understand the dynamics of the phenomenon. The scientific community has identified vine water status monitoring as a challenge for characterising the consequences of climate change in the coming years (Van Leeuwen et al. 2019).

2.4 French Mediterranean arc is a relevant study zone for the issue of monitoring vine water status

2.4.1 The most relevant scale for defining the study zone is a few hundreds of kilometres

In the systemic approach implemented in this thesis, the choice of the study zone to carry out the farmsourcing project of spatial observations is essential because this zone must offer the conditions allowing hypotheses 1 and 2 to be tested. This study zone must therefore include numerous vineyard in order to have a high number of potential participants. It also has to be regularly confronted to water restriction phenomena so that monitoring of vine water status is an issue for all the stakeholders in the study zone. These stakeholders must be confronted to diverse situations in terms of soil, climate and socio-economic conditions so that the case study can be generalised as much as possible. These stakeholders must also be as organised and structured as

possible at the territorial and sectoral level so that information about the project circulates well and a significant number of observations can be collected quickly. Finally, the study zone must belong to a unique macroclimate so that the study of the vine water status at this scale makes sense. A scale of a few hundred kilometres seemed to be a good compromise between a zone encompassing a large number of territories with different situations and a zone belonging to a unique macroclimate. It is therefore this spatial scale that was chosen to define the study zone.

2.4.2 The French Mediterranean arc has the characteristics of a relevant case study

The chosen study zone was situated in the south of France in an area encompassing the vineyards of the Languedoc, Provence and Côtes du Rhone viticulture regions (Figure 16a). The study zone covers an area of ~49 500 km^2 along the Mediterranean Sea and encompasses 57 different controlled denominations of origin with > 300 000 ha of vines and > 26 000 growers (Agreste, 2010) (Figure 16b). It includes a wide range of pedo-climatic conditions. The soils mostly have low soil water holding capacity (< 50 mm) (Figure 16c) and the mean annual precipitation over the last 30 years varies from relatively low values (500 – 700 mm yr^{-1}) in coastal areas to relatively high values (> 1 300 mm yr^{-1}) in the hinterland (Figure 16d).

This study zone is of interest because i) it is one of the largest contiguous wine producing areas in France, ii) the majority of the vineyards are non-irrigated and the climate is Mediterranean, i.e. characterized by hot and dry summers, and iii) the pedo-climatic conditions are diverse and broadly representative of conditions that can be found in many vineyards around the world that are located in regions with a Mediterranean-type climate (e.g. west coast of the USA, southern Australia, western coast of South America and the northern and southern extremities of Africa).

In this study zone, the monitoring of vine water status is therefore a major issue both at the vineyard and regional scales. Every of the more than 26,000 winegrowers and numerous collective structures are concerned. It is impossible to mention all these structures but the main ones that monitor vine water status at the scale of their territory are chambers of agriculture (e.g. “Chambre d’agriculture de l’Hérault”), denomination of origin producer’s union (e.g. “Syndicat des côtes du Rhône”), large wineries (e.g. Advini company), advice companies (e.g. ICV group), cooperative wineries (e.g. Cantalric cooperative), regional development companies (e.g. BRL) or irrigators' associations (e.g. “ASA de Gignac”). For all these structures, the vine water status is a major issue and they are seeking to monitor it at the scale of their territory.

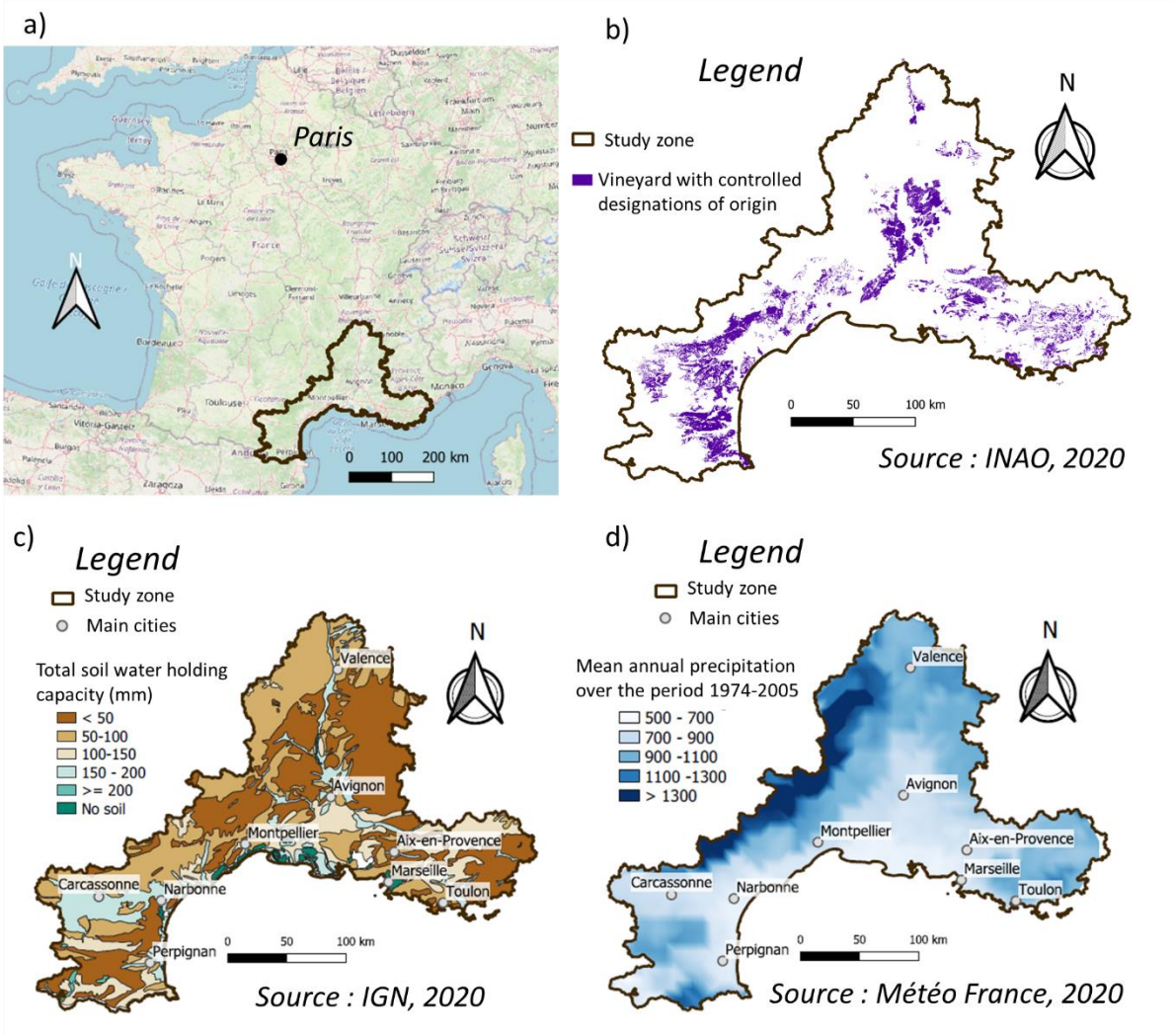
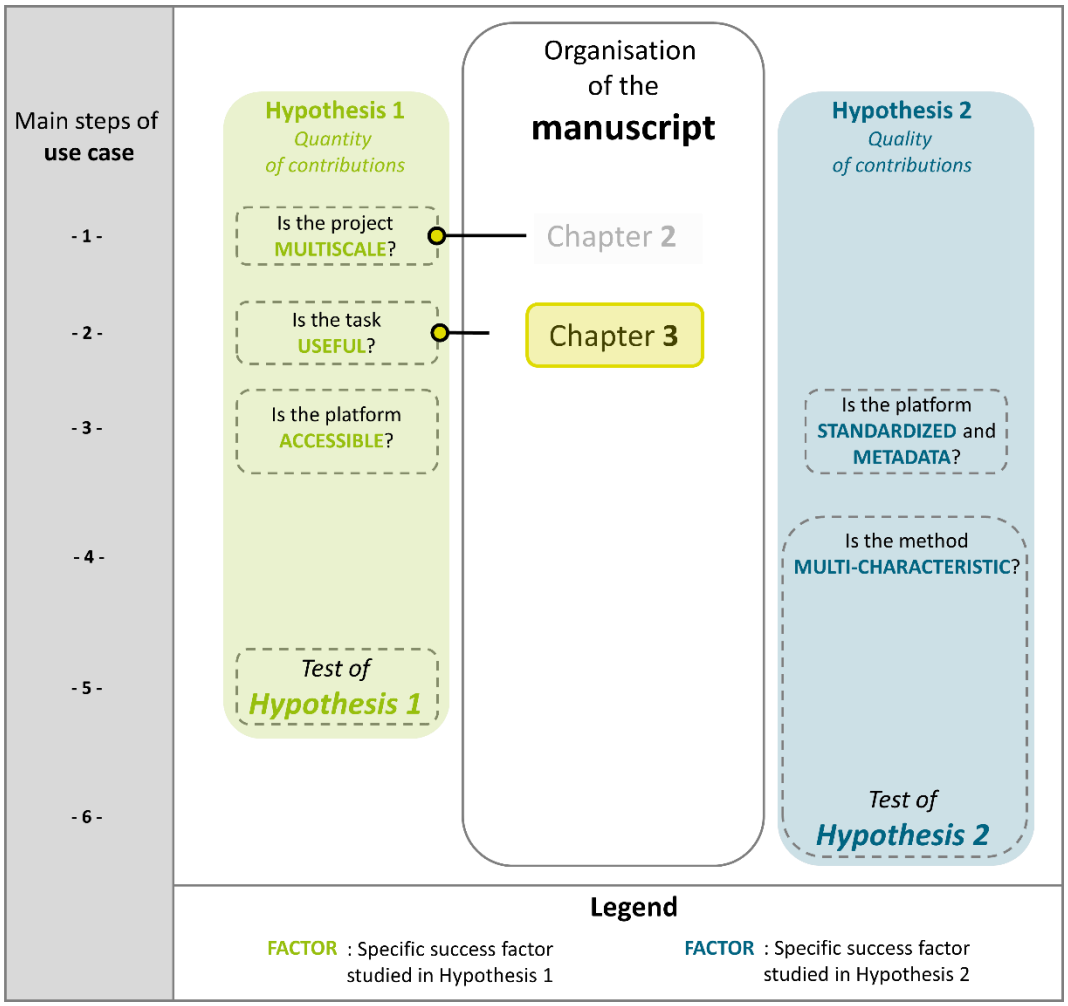


Figure 16: The study zone encompasses several wine-growing regions of importance in southern France with a large variability in soil and climate conditions. (a) The location of the study zone within France, (b) the location of vineyards within the study zone that have a controlled denomination of origin (INAO, 2020). Two principal pedo-climatic characteristics of this zone, (c) total soil water holding capacity (calculated from soil depth and texture) (IGN, 2020) and (d) mean annual precipitation over the period 1974-2005 (Météo France, 2020).

Chapter 3: Observation of shoot growth: a simple and operational decision-making tool for monitoring vine water status in the vineyard



The objective of chapter 3 is to demonstrate that observation of vine shoot growth (i.e. iG-Apex) is a relevant indicator from an operational and scientific point of view for monitoring vine water status in the vineyard (specific factor USEFUL). The scientific questions explored in this chapter are: "Is it possible to establish a relationship between the vine shoot growth approach and a reference method for monitoring vine water status? Under which conditions is this relationship valid? What recommendations for use can be drawn from it? "

3.1 Details about the corresponding paper

3.1.1 Title of the paper

This chapter will be submitted as a short communication in the journal *Oeno One* under the title "Observation of shoot growth: a simple and operational decision-making tool for monitoring vine water status in the vineyard".

3.1.2 Authors:

Léo Pichon[1], Cécile Laurent[123], Jean Christophe Payan[4], Bruno Tisseyre[1]

[1]*ITAP, Institut Agro Montpellier, INRAE, Univ. of Montpellier, Montpellier, France*

[2]*ABSYS, Institut Agro Montpellier, INRAE, Univ. of Montpellier, Montpellier, France*

[3]*Fruition Sciences, Montpellier, France*

[4]*Institut Français de la Vigne et du Vin, Le Grau du Roi, France*

3.1.3 Abstract

Monitoring vine water status is a major issue for vineyard management because it influences both the quality of the harvest and the yield. Numerous tools and methods have been proposed to monitor vine water status, but they are often costly and complex to implement. Methods based on the observation of vine shoot growth seem to be interesting alternatives because they are simple to carry out and therefore potentially more adapted for use in production vineyards. However, these methods have never been evaluated and compared to reference measurements on several cultivars and during several vintages. The objective of this article is to study their characteristics (validity range, specificity, sensitivity) in order to propose recommendations for their rigorous implementation in an experimental or operational context. The study was carried out using the iG-Apex method to monitor vine shoot growth and predawn leaf water potential as a reference measure on 55 fields located in the Tavel vineyard (Occitanie, France) during 2008 to 2012 vintages. The results showed that iG-Apex can be used as an operational tool, for monitoring vine water status at field scale for water restrictions ranging from -0.2 MPa to -0.8 MPa. Nevertheless, precautions must be taken when interpreting the results as the method is not specific to water restriction and is also sensitive to other phenomena. The use of this method to carry out collective monitoring of vine shoot growth over large spatial areas in addition to more precise and more localised monitoring carried out with reference measurements also seems relevant.

3.1.4 Keywords

Grapevine, Vegetative indicator, Vitis vinifera, Water potential, Water stress,

3.2 Introduction

Many authors have shown that water stress decreases photosynthesis and transpiration (Hsiao 1973), thus limiting vegetative growth (Lebon et al. 2006) and influencing berry ripening (Van Leeuwen and Seguin 1994). In the vineyard, water stress has been shown to limit yield (Ojeda et al. 2001), influence grape quality (Van Leeuwen et al. 2009) and impact the following year's production (Guilpart et al. 2014). Measuring and monitoring water status is therefore a major issue for vineyard management. Many tools and methods have been proposed to estimate the vine water status (Rienth and Scholasch 2019). They all rely on the measurement of a variable varying with or explaining water status and can be measured either on the soil (Gardner et al. 2001), on the plant (Pons et al. 2008) or on climate parameters (Gaudin et al. 2014). Most of these approaches have been designed to meet the requirements of research and experimentation where the objective is to obtain the most reliable proxy of the plant's water status. Sensitivity (i.e. the ability to vary with water status) and specificity (i.e. the ability to vary only with water status) of the proxy has therefore been a priority in the design of these tools. As a result, they often require the use of expensive equipment and the implementation of complex protocols. These operational constraints limit the use of these tools at a large scale in the vineyard. The Predawn Leaf Water Potential (ψ_{PD}) may be a good example of this. It is considered as a reference method (Améglio et al. 1999) but remains relatively little used in commercial vineyards because it requires to carry heavy and voluminous equipment and to make observations under restrictive conditions with qualified operators.

Methods based on shoot growth observation seem to be interesting alternatives to estimate the vine water status (Pellegrino et al. 2005). They are simple to implement and therefore potentially more suitable for use in commercial situations. Schultz and Matthews (1988) have shown that shoot extension rate is more rapidly affected by water stress than leaf extension rate. As a result, comparing these two extension rates by simply folding the last unfolded leaves on the shoot is a relevant indicator for estimating how water restriction affects the growth. Based on this principle, Rodriguez Lovelle et al. (2009) proposed a protocol for observing 50 shoots spread over 10 vines. Martinez-De-Toda et al. (2010) proposed an operational indicator based on the observation of these shoots' apexes and their classification into three categories: i) apex in full growth (FG) when organogenesis is active, ii) apex in moderate growth (MG) when organogenesis is reduced and iii) apexes with stopped growth (SG) when the apex has fallen or has dried out. These authors proposed to summarize these observations by an S indicator (*Equation 1*).

$$S = w_{FG}S_{FG} + w_{MG}S_{MG} + w_{SG}S_{SG} \qquad \textit{Equation 1}$$

where w_{FG}, w_{MG}, w_{SG} stand respectively for the proportions of full growth, moderate growth and stopped growth apexes. S_{FG}, S_{MG}, S_{SG} are the coefficients associated with each type of apex (Equation 1). Despite the simplicity of this shoot growth approach (SGA), the few papers that have studied its interest for decision support have been limited to a single grape variety and two vintages (Martinez-De-Toda et al. 2010). As a result, there are very few scientific references to determine the potential and limitations of SGA. The objective of this paper is to compare this approach with a reference method through a study on several cultivars and several vintages in order to provide references on i) the relationship between SGA and a reference measurement, ii) the range of validity of SGA, iii) its specificity to the vine water status and iv) its sensitivity to

other factors. The objective is to propose recommendations for the rigorous implementation of this approach in an experimental or operational context.

3.3 Material and methods

3.3.1 Collected datasets

Two datasets were collected. The objective of the first dataset, entitled *Vineyard_dataset*, was to compare shoot growth observations with reference observations under production conditions and over a large number of fields, cultivars and vintages. The objective of the second dataset, entitled *Metrology_dataset*, was to evaluate the metrological qualities of shoot growth observations by repeating measurements with several operators.

3.3.2 *Vineyard_dataset*: Sampling design

Observations (shoot growth and reference) were collected in the vineyard of Tavel, in the southern Rhône Valley, France (WGS84; X= 4.682064; Y= 44.009484). The climate in this region is Mediterranean with low annual rainfall and high summer temperatures. The soils are relatively fertile and the main factor limiting the vine shoot growth is access to water resources (Martinez-Vergara et al. 2014).

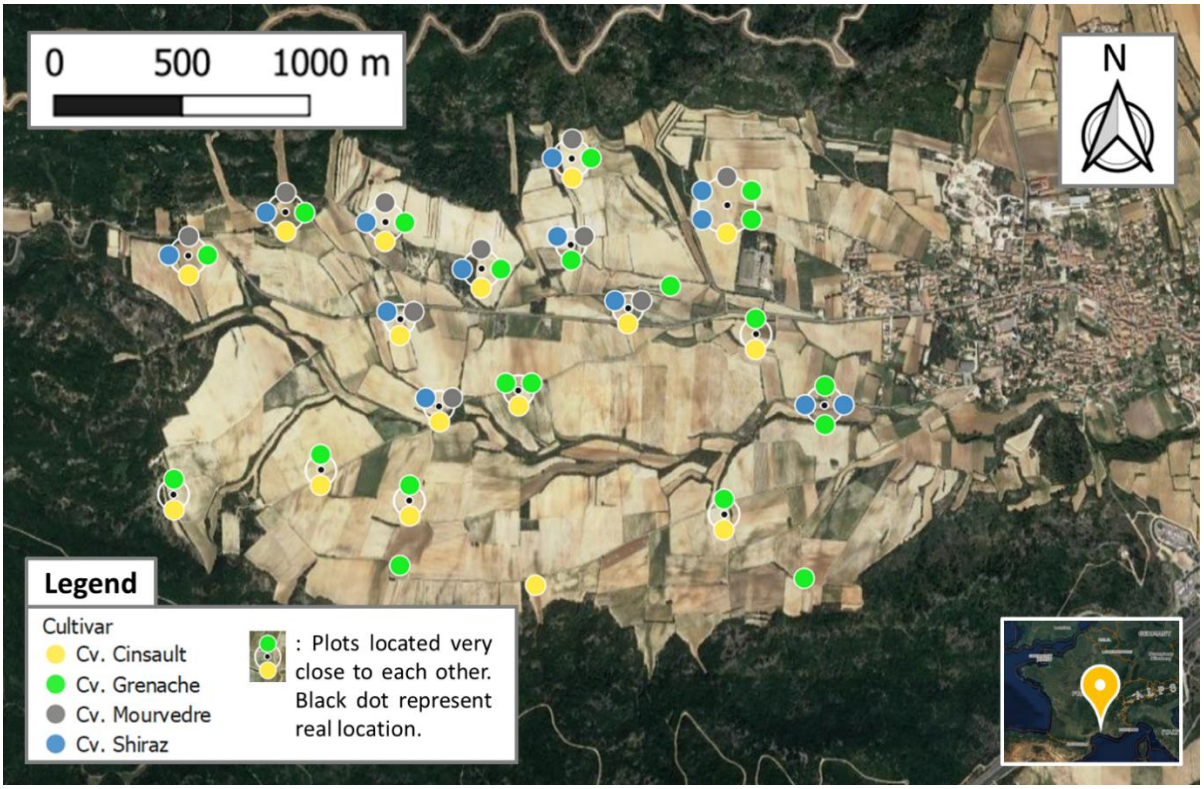


Figure 17: Location and cultivar of fields on which reference and shoot growth observations were collected for the Vineyard_dataset.

55 measurement sites corresponding to 55 different fields were defined on the vineyard. The fields were situated on the same soil unit. The soil was sandy-clay with a small amount of calcareous stones on the surface. Each measurement site consisted of 10 consecutive vines along the row within each field. The fields were chosen in order to i) take into account the diversity of cultivars in the region (cv. Cinsault, cv. Grenache, cv. Mourvèdre and cv. Syrah) and to cover a wide variety of situations in the study zone (Figure 17). The study was conducted over 5 consecutive years from 2008 to 2012. These vintages presented a variety of temperature and rainfall profiles that allowed a wide range of climatic conditions to be explored (Figure 18). For example, 2008 was characterized by an exceptionally wet spring and early summer, while in contrast, the early summer of 2011 was particularly dry. The year 2009 was warm compared to the temperatures observed over the last 30 years but with rather average rainfall. In contrast, 2010 summer was dry with no precipitation.

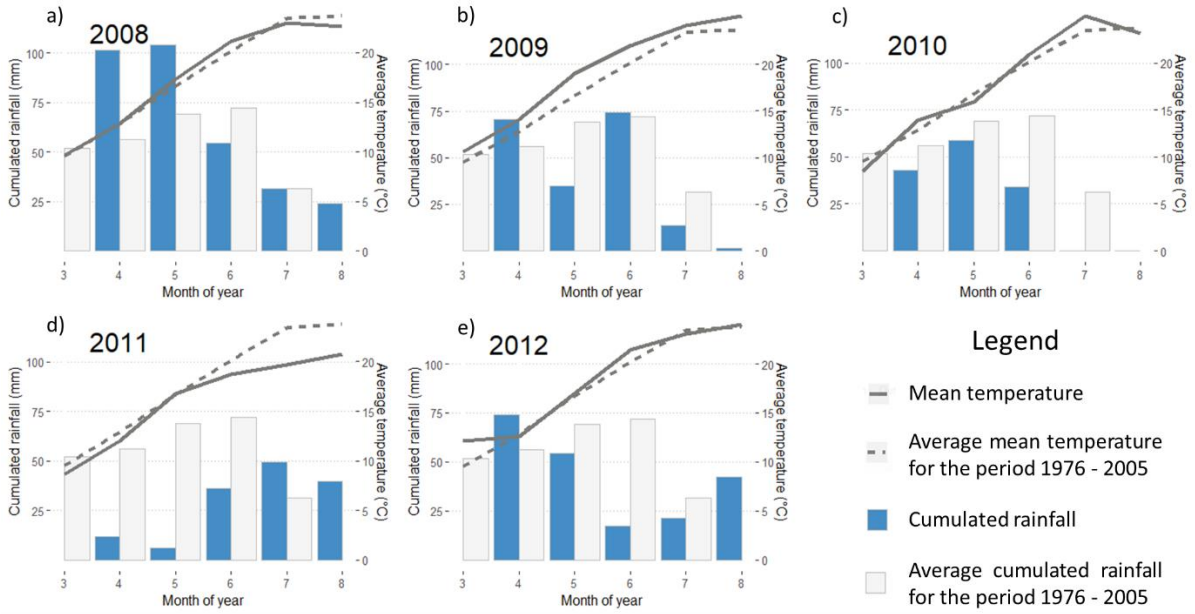


Figure 18: Average monthly temperatures and cumulative monthly rainfall for the 2008 (a), 2009 (b), 2010 (c), 2011 (d) and 2012 (e) vintages compared to the respective averages over the period 1976 – 2005.

At each site, observations (shoot growth and reference) were collected on a weekly basis. Observations were all collected between May 30th and September 1st, but each year the date of the first observation was dependant on the weather and the resulting precocity or delay of the water restriction. At the end of the protocol, the *Vineyard_dataset* consisted in 474 pairs of shoot growth and reference observations.

3.3.3 Method for collecting observations on shoot growth

At each studied site, 50 apexes were observed and classified into the FG, MG, SG categories (Figure 19). For the calculation of the S indicator (Equation 1), the recommendations of the Institut Français de la Vigne et du Vin (IFV) were followed (Payan 2020). The values 1, 0.5 and

0 were defined for the parameters S_{FG}, S_{MG} and S_{SG} respectively, defining the iG-Apex indicator. The term "iG-Apex" will be used in the rest of the paper to describe the shoot growth index.



Figure 19: Examples of shoot classified in the categories "full growth" (a), "moderate growth" (b) and "stopped growth" (c).

3.3.4 Vineyard_dataset: Method for collecting reference observations

The ψ_{PD} measured with a pressure chamber (Scholander et al. 1965) was chosen as the reference measurement. For each of the 10-vine sites, measurements were taken on the five central vines (vines 4 to 8). On each vine, measurements were taken on a primary leaf located in the middle zone of the vegetation. The average of these 5 measurements was considered as the reference ψ_{PD} value for the site. The ψ_{PD} measurements were performed according to the method described by Améglio et al. (1999) by experienced operators who regularly calibrated each other to limit operator effects.

3.3.5 Metrology_dataset: Sampling design

iG-Apex observations were collected in conditions allowing the widest possible range of shoot growth to be observed at the same date. These observations were collected on 3 fields planted with 3 cultivars of varying vigour and precocity (cv. Chenin, cv. Mourvèdre and cv. Syrah). These fields were located in Montpellier (WGS84; Lat: 43.617592; Long: 3.855987) in a Mediterranean climate. Each field consisted in a site of 10 consecutive vines along the row. Observations were collected on July 25th 2019 when water restriction was likely to impact shoot growth. Six inexperienced operators repeated the iG-Apex observations three times at each of the three sites with 30-minute intervals. At the end of the protocol, the *Metrology_dataset* consisted in 54 iG-Apex observations.

3.3.6 Data analysis:

3.3.6.1 *Relationship between iG-Apex and ψ_{PD}*

The objective was to study the relationship that could be established between iG-Apex and the ψ_{PD}. For this purpose, all 474 observations of the Vineyard_dataset were considered. The average of the ψ_{PD} values was plotted for every 0.1 iG-Apex interval. A linear regression was performed between iG-Apex and the corresponding average ψ_{PD}. The corresponding r squared was calculated to evaluate the quality of the regression.

3.3.6.2 *Validity range of iG-Apex*

The objective was to identify the range of vine water status in which this relationship was relevant. Only the extreme ranges of *Vineyard_ dataset* iG-Apex values were considered in order to identify the situations where this indicator appeared saturated or not sensitive to ψ_{PD}. Considered observations were plotted and studied in a two-dimensional graph $\psi_{PD}\ vs\ iG - Apex$. Linear regressions were performed and corresponding r squared were calculated.

3.3.6.3 *Specificity of iG-Apex to vine water status*

The objective was to study the uncertainty associated with the prediction of ψ_{PD} from iG-Apex. As a first approach, it was considered that the standard deviation may be a good proxy for the uncertainty of this prediction. The standard deviation of ψ_{PD} values was plotted for each 0.1 iG-Apex interval.

3.3.6.4 *Study of the factors influencing iG-Apex*

The sensitivity of iG-Apex to different factors was studied using two ANOVAs. The first ANOVA was performed on all 474 *Vineyard_ dataset* observations. The objective was to test the influence of factors that could potentially impact the vine shoot growth under commercial conditions. The first factor studied was ψ_{PD}. It was grouped into four classes (ψ_{PD}>-0.2MPa; -0.2MPa>ψ_{PD}>-0.4MPa; -0.4MPa>ψ_{PD}>-0.6MPa; -0.6MPa> ψ_{PD}) corresponding to vine water stress (i) mild or absent, (ii) mild to moderate, (iii) moderate severe and (iv) severe, respectively (Deloire et al. 2004). The second factor studied was the cultivar as it has been shown that the genetics of the plant influence its shoot growth (Prieto et al. 2010). Finally, shoot growth can be influenced by other factors such as nitrogen access (Gaudillère et al. 2002), farming practices (Meissner et al. 2019) or fruit load (Hardie and Martin 2000). As a first approach, it was considered that the influence of all these factors could be summarized in a field effect that encompassed all the farming practices. This analysis simply evaluated the overall effect of all these factors without being able to differentiate their respective contributions. The second ANOVA was performed on the *Metrology_ dataset.* The objective was to test the influence of factors that could potentially impact the iG-Apex measurement process. The reproducibility and repeatability of the measurement were evaluated by studying the operator effect and the effect of the repetition of the same measurement by the same operator, respectively.

3.3.7 Graphs

All analyses and graphs were performed with R 3.6.1 (R Core Team, 2020).

3.4 Results and discussion

3.4.1 Relationship between iG-Apex and ψ_{PD}

Results obtained with the 474 measurement of the Vineyard_dataset highlight a clear linear relationship (r^2 = 0.97) between classes of iG-Apex and the mean ψ_{PD} observed for each considered class (Figure 20). The lower the iG-Apex, the lower the average ψ_{PD}. Low iG-Apex values corresponded on average to vines experiencing water restriction. This result clearly shows

that iG-Apex is related to vine water status and that it can be used as a relevant surrogate in commercial situations. However, for extreme class of iG-Apex (iG-Apex < 0.1 and iG-Apex > 0.9), the general trend highlighted by figure 20 is less straightforward. For example, for iG-Apex class > 0.9 a plateau is clearly observed showing that mean ψ_{PD} for this iG-Apex class is not that different from mean ψ_{PD} observed for the next class (iG-Apex = [0.8; 0.9]). The next section provides an insight to better understand the relationship between iG-Apex and ψ_{PD} for these extreme values in order to identify the range of validity of this relationship.

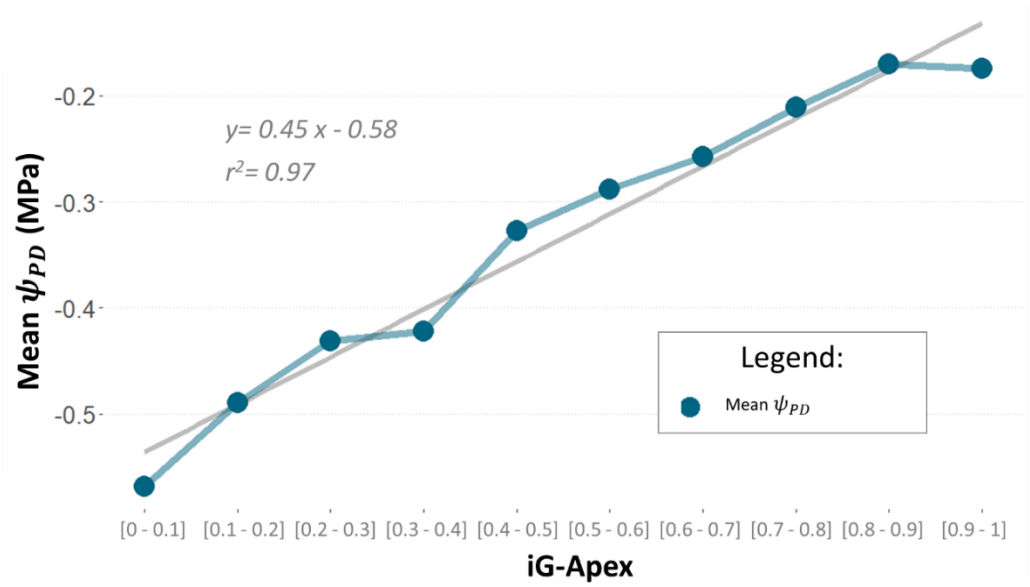


Figure 20: Mean of Predawn Leaf Water Potential (ψ_{PD}) depending on iG-Apex values for the 474 observations of Vineyard_ dataset.

3.4.2 Validity range of iG-Apex

Figure 21 shows relationship between classes of iG-Apex and the mean ψ_{PD} observed for extreme class of iG-Apex values (iG-Apex < 0.1 and iG-Apex > 0.9).

For observations having a high shoot growth (iG-Apex >0.9), ψ_{PD} vary over the whole range from -0.3 MPa to 0 MPa (Figure 21b). In this range, the linear model (ψ_{PD} vs iG-Apex) explains only a very small part of the variability (r^2=6.8x10^{-3}). This result can be interpreted by the fact that shoot growth is not impacted by differences in water restriction when this latter is very low. Under these conditions, all the apexes were in full growth and the iG-Apex values remained very close to 1. The upper threshold of validity for the relationship between ψ_{PD} and iG-Apex is close to -0.2 MPa. These results are consistent with the scientific literature indicating that between 0 and -0.2 MPa, no real water restriction can be considered (Deloire et al. 2004).

When water restriction is high (ψ_{PD} < -0.8MPa), almost all observations showed very low shoot growth with iG-Apex of zero or very close to 0 (Figure 21a). Again, in this case, the linear model (ψ_{PD} vs iG-Apex) explains only a very small part of the variability in this range (r^2=1.6x10^{-2}). This result shows that beyond a given water restriction threshold (around -0.8 MPa), the shoot growth

of all the apexes has stopped. When the vine water status kept on decreasing, the apexes remained in the same state and the iG-Apex indicator is no longer sensitive to variations in vine water status. This result is consistent with the work carried out on potted vines by Schultz and Matthews (1988) who demonstrated that leaf and internode growth are very low for ψ_{PD} around -0.8 MPa and completely stopped for ψ_{PD} of -1 MPa.

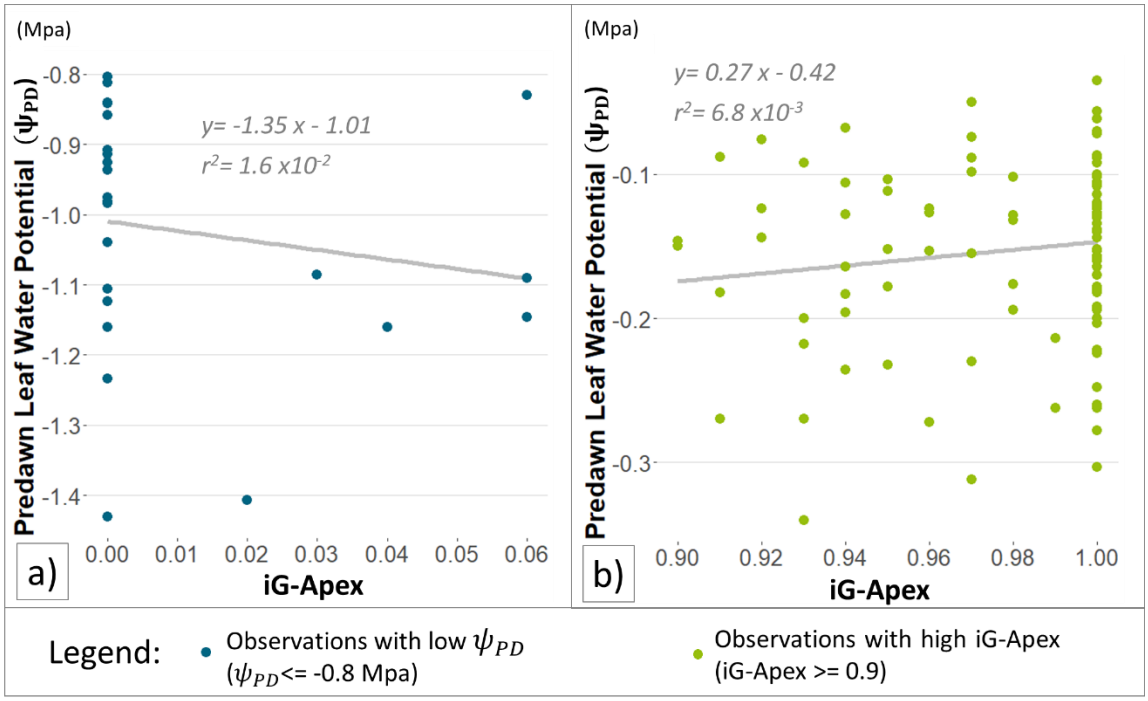


Figure 21: Scatter plots of Predawn Leaf Water Potential (ψ_{PD}) vs iG-Apex for: (a) observations with strong water restriction (ψ_{PD}< -0.8) and (b) observations with active shoot growth (iG-Apex > 0.9).

For vine water status monitoring purposes, the range of validity of iG-Apex to be considered is therefore approximately between -0.8 MPa and -0.2 MPa. Outside this range, the observation of shoot growth cannot be used as a surrogate to estimate vine water status.

3.4.3 Specificity of iG-Apex to vine water status

The uncertainty associated with the estimation of ψ_{PD} from iG-Apex can be estimated by the standard deviation of ψ_{PD} values observed for the same iG-Apex value. This standard deviation is relatively small (around 0.06 MPa) for high iG-Apex and higher (around 0.20 MPa) for low iG-Apex (Figure 22). The lower the iG-Apex, the higher the uncertainty in the relationship between ψ_{PD} and iG-Apex.

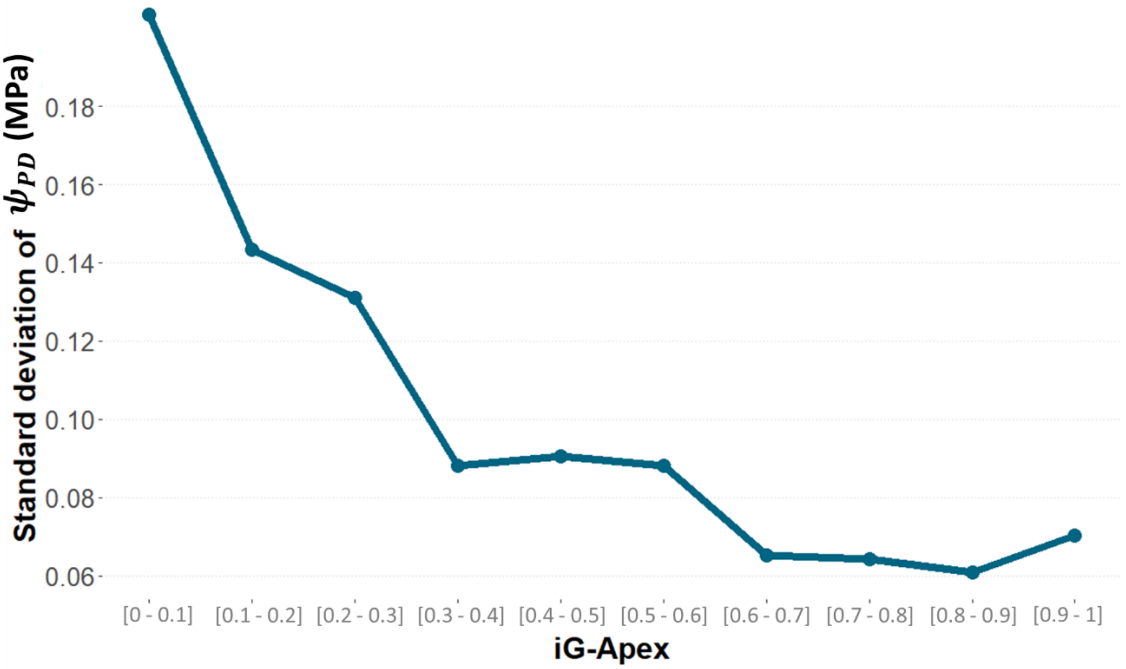


Figure 22: Standard deviation of Predawn Leaf Water Potential (ψ_{PD}) depending on iG-Apex values for the 474 observations of Vineyard_ dataset.

This result may show that in our conditions, the vine water status is not the only factor influencing iG-Apex. For example, a low value of iG-Apex may correspond to a high water restriction but may also correspond to a moderate water restriction with another factor limiting shoot growth (e.g. access to nitrogen). For low values, iG-Apex may therefore be less specific to the vine water status and may be influenced by other factors. The objective of the following section is to study the relative influence of these other factors in the context of the experiment.

3.4.4 Study of the factors influencing iG-Apex

As expected, the class of ψ_{PD} is the main factor explaining the variability of iG-Apex (Table 6). This factor explains 96.4% (Mean square = 12.728) of this variability. This result is in accordance with the relationship presented in Figure 20). Compared to similar observations made either in controlled conditions (Pellegrino et al. 2005) or on a small number of fields (Martinez-De-Toda et al. 2010), this result confirms the relevance of iG-Apex under commercial conditions and on a large diversity of fields to monitor vine water status in commercial conditions. Farming practices, estimated through the field effect, have a less significant effect (p-value < 0.01) on iG-Apex values explaining only 0.005% of variability (Mean square = 0.065). In this experiment, field effect includes all the factors related to vineyard management (fertilisation, weed control, trellising system, etc.) that are likely to affect iG-Apex. Two hypotheses may be formulated for interpreting this relatively low field effect. Either this effect is very low compared to the vine water status one, or the vineyard management is relatively similar in the zone under consideration. Cultivar had a significant effect (p-value < 0.001) although the proportion of variability explained by this factor was only 2.8% (Mean square = 0.371).

Table 6: Results of the ANOVA performed on 474 iG-Apex observations of Vineyard_dataset with three factors (cultivar, classes of ψ_{PD} and field).

	Degrees of freedom	Sum of Squares	Mean Squares	F-value	Probability	
Class of ψ_{PD}	3	$3.82*10^{1}$	$1.27*10^{1}$	$3.22*10^{2}$	$<2*10^{-16}$	***
Cultivar	3	1.11	$3.71*10^{-1}$	9.39	$5.16*10^{-6}$	***
Field	54	3.50	$6.48*10^{-2}$	1.64	$4.36*10^{-3}$	**
Residuals	414	$1.63*10^{1}$	$3.93*10^{-2}$			

This cultivar effect is illustrated by the curve of average ψ_{PD} versus iG-Apex for the two cultivars : cv. Cinsault and cv. Grenache (Figure 23). For the same range of iG-Apex, lower average ψ_{PD} are observed for cv. Grenache than for cv. Cinsault. When considering variety, the standard deviation of ψ_{PD} is lower and therefore the prediction uncertainty is reduced if only one cultivar is considered. This result is consistent with the work carried out in ecophysiology (Prieto et al. 2010). Although the cultivar effect on iG-Apex variability is small compared to the water restriction effect, this result shows that taking into account the cultivar is an interesting way to improve the accuracy of the estimation especially for moderate to high water restriction levels (< -0.5 MPa). Our experiment does not bring sufficient material to conclude whether a specific model would be necessary for each cultivar or if models for a few groups of cultivars with similar behaviour would be sufficient in a commercial context.

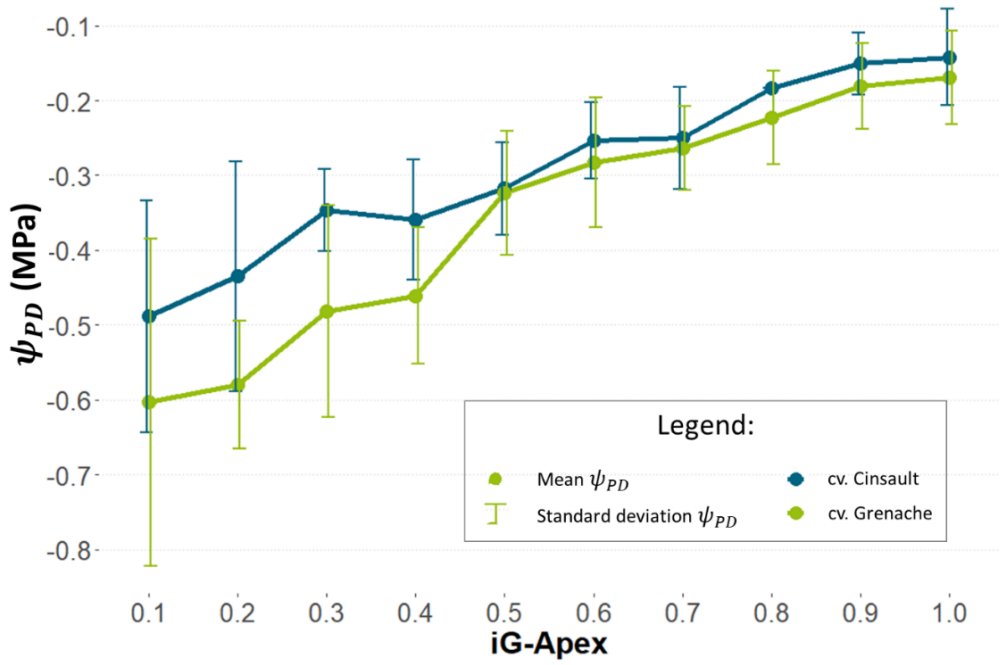


Figure 23: Mean and standard deviation of Predawn Leaf Water Potential (ψ_{PD}) depending on iG-Apex values for the 236 observations made on cv. Cinsault and cv Grenache out of the 474 observations of Vineyard_dataset.

Considering the measurement process of iG-Apex and the factors that may influence it, the ANOVA shows that neither the operator nor the repetitions by the same operator have a significant effect on the iG-Apex values (Table 7). This result indicates that the method is repeatable and reproducible. Remember that this experiment was conducted on purpose with non-expert users. Under these conditions, the results observed reinforce the repeatability and the reproducibility of the method.

Table 7: Results of the ANOVA performed on 54 iG-Apex observations with two factors (operator and repetition).

	Degrees of freedom	Sum of Squares	Mean Squares	F-value	Probability	
Operator	5	$1.73*10^{-1}$	$3.46*10^{-2}$	1.22	$3.17*10^{-1}$	
Repetition	2	$3.0*10^{-3}$	$1.51*10^{-3}$	$5.30*10^{-2}$	$9.48*10^{-1}$	
Residuals	46	1.31	$2.85*10^{-2}$			

3.4.5 Considerations for using iG-Apex to estimate vine water restriction in a commercial context

iG-Apex is a relevant indicator for the operational estimation of vine water status. This study demonstrates that it can be used for low to moderate water stress (from -0.3 MPa to -0.8 MPa) and in particular for dynamic monitoring of vine water status at field level. However, for operational use, it is important to note that iG-Apex is not able to make the distinction between a field with moderate to strong water restriction and a field with very strong water restriction. In both cases, iG-Apex values will be close to 0 but the first field will produce good quality grapes while yield and grape quality of the second field will be reduced. iG-Apex is therefore an approach that can be used at the beginning of the season to identify the onset of water restriction. However, this approach is less relevant for in-season monitoring of fields undergoing moderate to severe water restriction.

iG-Apex is based on the observation of the vine's response to water restriction. It provides information on the water restriction experienced by the plant during the few days before the measurement but not on vine water status at the exact moment of the measurement. This method is therefore particularly suitable for regions where summer rainfall is rare and onset of water restriction is rather linear. On the other hand, the relationship between iG-Apex and PHFB may not be valid in regions where summer rainfall is more frequent or in irrigated fields. iG-Apex therefore seems particularly suitable for use in regions with a Mediterranean climate and on non-irrigated vineyards.

According to the measurements characteristics (repeatability and reproducibility) of iG-Apex, it is possible to compare observations acquired by different operators. This approach can then be used to compare vine water status dynamics of two different fields planted with the same cultivar in a neighbourhood of a few kilometres. Nevertheless, interpretations must be tempered taking into account other biotic or abiotic factors that are likely to influence shoot growth for moderate to high water restriction. Comparing dynamics of two fields with different cultivars is also possible since cultivar effect remains small compared to water restriction effect. However, iG-Apex values

from two different cultivars must be interpreted with caution especially for moderate to high water restriction.

The study highlights the potential complementarity between iG-Apex and other reference methods such as ψ_{PD} as decision support. The low material and human cost of iG-Apex measurements offers the possibility to carry out numerous observations with a regular frequency and at many sites at the beginning of the season, i.e. before high or moderate water restriction occurs. In these conditions iG-Apex values are relevant and accurate even when performed by several operators on different cultivars. With an increasing water restriction, if vine water status monitoring requires accurate estimation, iG-Apex may be less appropriate since it is less specific to water restriction. In these latter conditions, reference values like ψ_{PD} may be more appropriate. The uncertainty (standard deviation) as presented in Figure 22 may constitute a decision support to decide when it is necessary to switch from one method to the other. In commercial conditions, such a strategy is relevant since it benefits from advantage of both approaches either to minimise operational constraints or to maximise accuracy. Indeed, the uncertainty threshold to decide the change from iG-Apex to a reference method has to be defined by the users according to expected accuracy in vine water status monitoring.

For the practitioner, this strategy can also be guided on the basis of empirical knowledge of spatial variability. Indeed, many studies have shown a significant spatial variability of vine water status at different spatial scales (Acevedo-Opazo et al. 2008; Taylor et al. 2009). This variability is stable over time because it depends on stable environmental parameters, in particular the soil (Kazmierski et al. 2011). Starting the monitoring of vine water status can then be carried out preferentially in zones where the shoot growth stop is the most important and therefore where the vine is the most likely to experience water restriction early in the season. This strategy, implemented at field or farm scale, can also be implemented at a larger scale. For example, at a small territory scale such as a cooperative, iG-Apex observations can be collected each week by winegrowers and shared with their advisors. The latter can thus identify areas of early onset of water restriction on which to focus their monitoring with ψ_{PD}.

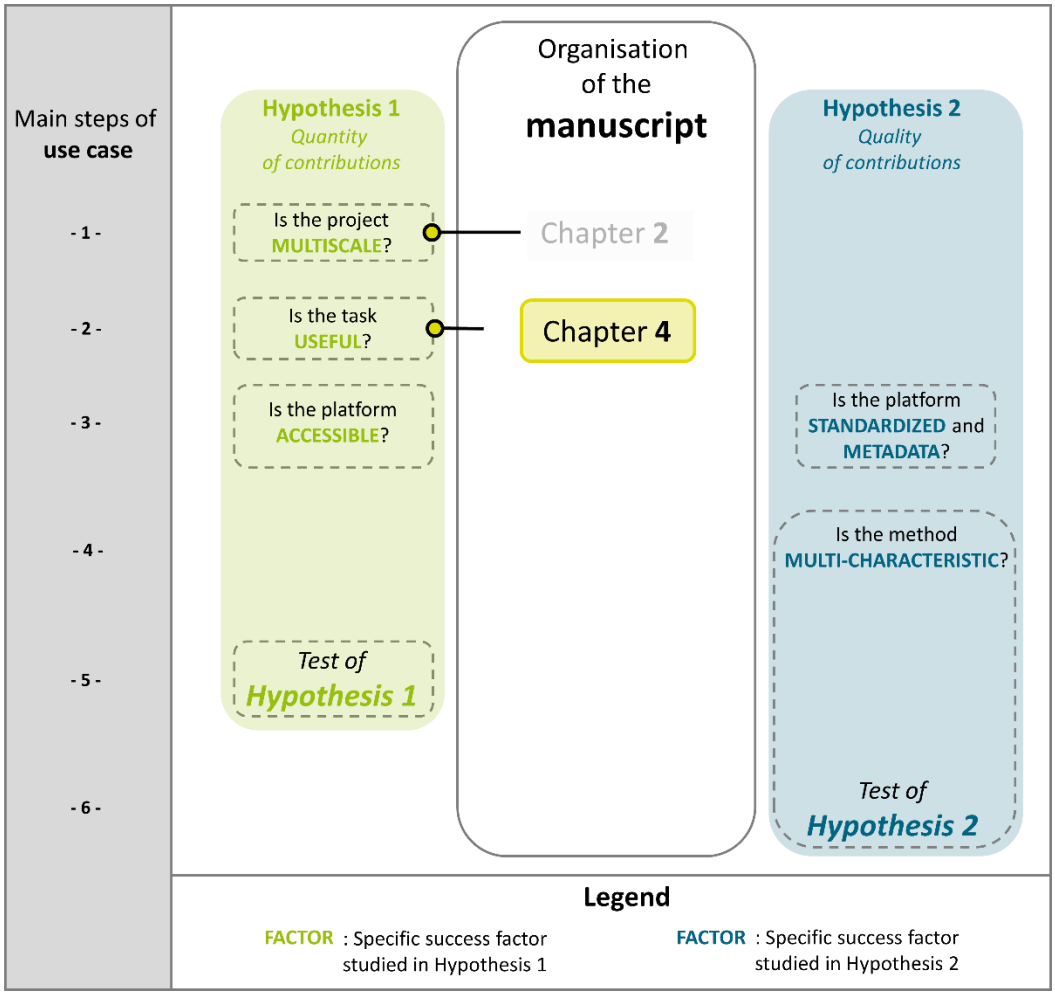
Although the iG-Apex approach is relatively simple, it does require a number of tedious and time-consuming operations, i.e. counting, classifying apex, calculating iG-Apex values and recording date and location for each measurement. To date, these practicalities have limited the widespread use of this approach. The development of mobile applications simplifying all of these operations (See Chapter 5:) should encourage the adoption of this approach by winegrowers and their advisors. This study should help operators in using it as rigorously as possible.

Finally, future research is needed to better characterise the relationship between iG-Apex approach and reference method, particularly as a function of cultivar or group of cultivars. It is also important to remember that this study is considering only the shoot growth of first order branches. However, it is widely accepted that second order branches are not influenced in the same way by water restriction (Pellegrino et al. 2005). In some wine-growing regions, particularly in the south of France, shoot trimmings are frequent, which favours the growth of second order branches. There is currently no reference that would allow this different response to be taken into account in order to adapt the iG-Apex calculation. This issue could be the subject of future investigations if the method is developed in the wine industry.

3.5 Conclusion

This study showed that a relationship could be established between a method of vine shoot growth, i.e. iG-Apex, and a reference method for monitoring the vine water status, i.e. ψ_{PD}. These results obtained on measurements carried out on several cultivars and several vintages demonstrated that an approach based on the observation of vine shoot growth can be used as an operational tool for monitoring vine water status in vineyard conditions. The validity range of the approach goes from -0.2 MPa to -0.8 MPa. Although vine water status is the main factor explaining variations of shoot growth, the approach is not specific to it. This approach is sensitive to other factors such as cultivar or farming practices that may affect the measurement, especially for high water restriction. It can be used alone or jointly with reference methods to increase the number and frequency of observations. Thanks to the good metrological qualities of this approach, it is possible to share observations made by several operators to do monitoring over large spatial areas. Nevertheless, for decision support purposes, other biotic or abiotic factors that may influence shoot growth must be taken into account in the interpretation of iG-Apex observations. The development of dedicated tools, such as a mobile application, should encourage the use of this approach and the sharing of observations between different operators of the same area.

Chapter 4: Characterising within-field variability of vine water status with simple visual observations of shoot growth



The objective of chapter 4 is to demonstrate that iG-Apex method can also be used to characterise the within-field variability of vine water status (specific factor USEFUL). The question explored in this chapter is: "Is the uncertainty associated with iG-Apex method low enough to allow participants to characterise vine water status at within-field level and to identify the optimal period for using this method as an operational decision support tool?"

4.1 Details about the corresponding paper

4.1.1 Title of the paper

This chapter has been accepted as an oral presentation in the 13th European Congress on Precision Agriculture (ECPA – Budapest – July 19th to 22th 2021) under the title "Characterising within-field variability of vine water status with simple visual observations of shoot growth". The proceedings will be published in the *Precision Agriculture* journal.

4.1.2 Authors

Leo Pichon[1], Octave Bopp[1], Bruno Tisseyre[1]

[1]*ITAP, Institut Agro Montpellier, INRAE, Univ. of Montpellier, Montpellier, France*

4.1.3 Abstract

Existing vine water status monitoring tools are often complex and expensive. Shoot growth observations have been demonstrated to be an interesting alternative for field-level management but not yet at the within-field scale. The study reported in this paper was conducted to determine whether a simple shoot observation method could be satisfactorily used to characterize the spatial structure of shoot growth related to water stress at within-field scale. Based on a geostatistical approach, this study showed that shoot growth observations can be used to delineate within field management zones of vine water status if observations are made by a single operator and during periods when vine water stress is increasing.

4.1.4 Key words

Decision support, precision viticulture, spatial analysis, *Vitis vinifera*

4.2 Introduction

Monitoring vine water status is a major issue for vineyard management as it influences both the quantity and quality of the harvest (Zufferey et al. 2017). Several authors have demonstrated the importance of monitoring vine water status from the regional scale (Baralon et al. 2012) to the field scale (Taylor et al. 2009) and down to the within-field scale (Tisseyre et al. 2005). At the within-field scale, this information helps winegrowers and their advisors to take both tactical (pruning, canopy management, irrigation) and strategic (planting resistant grape varieties, setting up shadings) decisions. Many tools have been developed to monitor vine water status (Rienth and Scholasch 2019) and estimate its spatial variability. Among them, reference methods, such as the pressure chamber (Scholander et al. 1965), often require specialised equipment and specific skills. In a commercial context, operational constraints usually limit the use of these reference methods and they are rarely used to characterize the vine water status at the within-field scale. Methods giving access to high spatial resolution information (Tisseyre et al. 2007) have also been proposed such as remote sensing images or soil electrical resistivity maps. However, these sources of

information remain costly and difficult to use for precise temporal monitoring. Moreover, they often require a calibration step that can be difficult to implement in a commercial context.

In order to tackle these issues, Martinez-De-Toda et al. (2010) proposed a method to estimate the vine water status from shoot growth observations. This method hypothesises that water availability is the main factor affecting vine growth. It is based on the classification of apexes (tips of the vine shoots) into three categories: i) apex in full growth (FG) when organo-genesis is active, ii) apex in moderate growth (MG) when organo-genesis is reduced, and iii) apex with stopped growth (SG) when the apex has fallen off or dried out. Rodriguez Lovelle et al. (2009) proposed to implement this method by doing 50 observations across 10 vines and summarizing them in an S index (Equation 2).

$$S = w_{FG}S_{FG} + w_{MG}S_{MG} + w_{SG}S_{SG} \qquad \textit{Equation 2}$$

where w_{FG}, w_{MG}, w_{SG} are the proportions of full growth, moderate growth and stopped growth apexes, respectively, and S_{FG}, S_{MG}, S_{SG} are the coefficients associated with each type of apex.

Although based on visual observations, Rodriguez Lovelle et al. (2009) demonstrated that the uncertainty associated with this protocol was acceptable to characterize the vine water status at the field level. However, this method has never been tested at the within-field scale. At this scale, it may be of interest to delineate water status zones at a low cost; however, it raises issues related to the observations' uncertainty that may be too high to highlight short scale variability in vine water status. Therefore, the objectives of this study were to determine if the uncertainty associated with this method is sufficiently low to enable growers to characterize the within-field vine water status spatial structure and to identify the optimum period for using this method as an operational decision support tool.

4.3 Material and methods

4.3.1 Method for collecting observations

The French Vine and Wine Institute (IFV) recommends using the equation proposed by Martinez-De-Toda et al. (2010) (Equation 2) with the coefficients S_{FG}, S_{MG}, S_{SG} set to 1, 0.5 and 0, respectively, to suit production conditions in the south of France (Payan, 2020). These settings define an index called the Index of Growing Apex (iG-Apex), which was considered to be the most appropriate index based on Equation 2 in the context of this study. The iG-Apex index varies between 0 and 1, typically with values generally to 1 around flowering and then decreases, with increasing water stress, to values close to 0 around veraison.

4.3.2 Study fields

Observations were collected in 2020 on two non-irrigated grapevine (*Vitis vinifera*) fields located near Corbières in southern France (Figure 24a) (WGS 84: latitude=43.1692; longitude=2.5629). One field was planted with cv. Grenache in 1989 and the other with cv. Syrah in 1995, with respectively an area of 1.3 and 0.8 ha and an elevation of 190 and 95 m. Plantation density was

4000 vines per ha for both fields. Training and management practices were typical for this viticulture region.

During the season, a single rainfall event of 38 mm was observed on July 23rd (week 30). The Grenache field was uniformly trimmed on July 9th (week 28). No canopy management operations affecting shoot growth were carried out on the Syrah field.

4.3.3 Sampling

There were 101 and 98 sampling sites on the Syrah (Figure 24b) and Grenache (Figure 24c) fields, respectively. The sites were defined on a regular grid to which a random distance was added to each point to introduce variance in distances between sites.

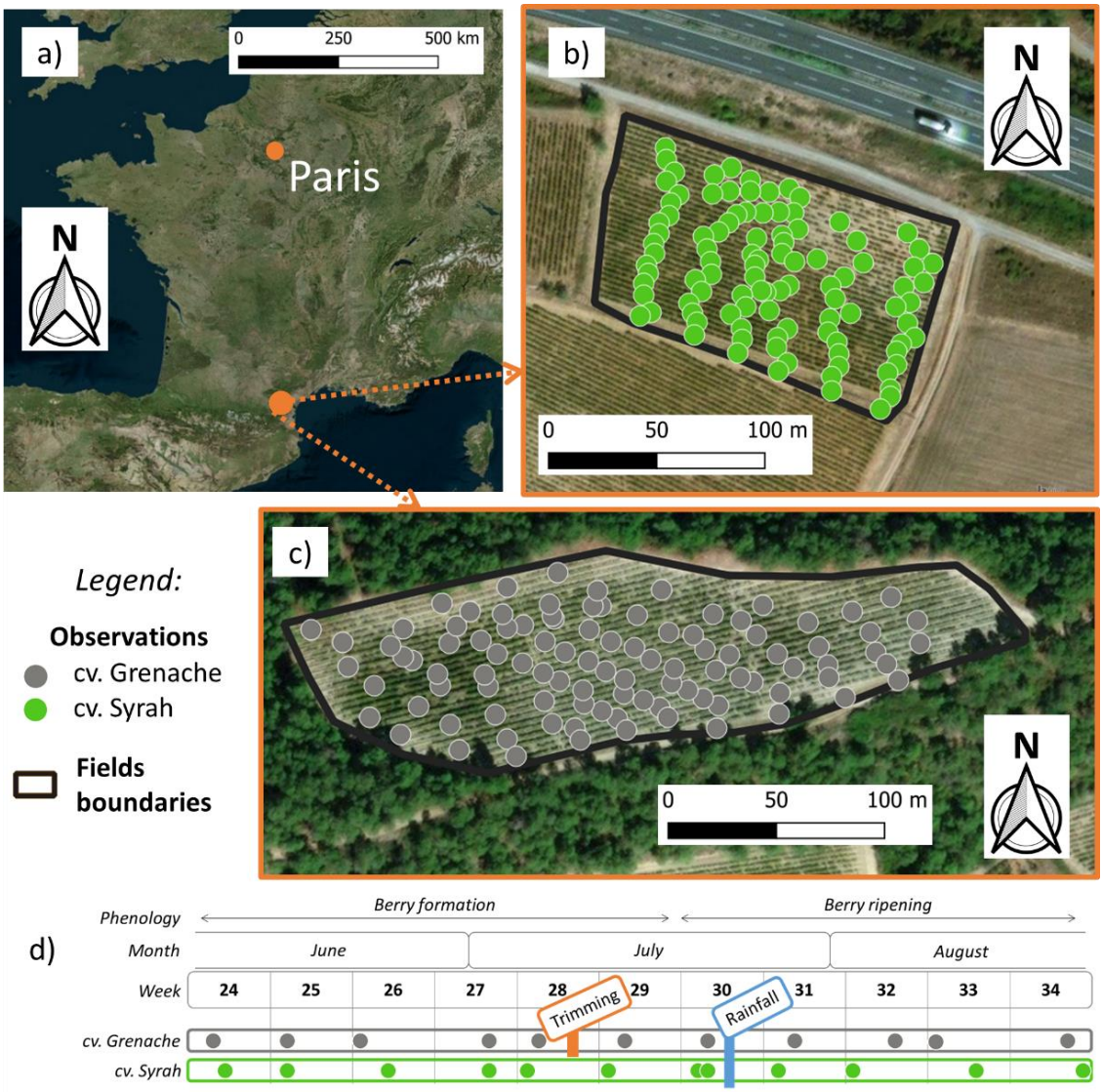


Figure 24: (a) Location of the fields in southern France and location of points within the fields, (b) Syrah with n =101 samples, and (c) Grenache with n = 98 samples. (d) The timings of the weekly iG-Apex observations from early June to late August are indicated by points in the calendar, along with the timing of canopy management and rainfall events.

At each sampling site, 50 apex measurements over 10 consecutive vines were collected by a single operator using the ApeX-Vigne smartphone application (See Chapter 5:). Each apex was classified as one of three categories (FG, MG or SG) and iG-Apex calculated by the Apex-Vigne app. For each sample site, the position of the fifth vine was geo-referenced using a GNSS receiver (R1, Trimble, Sunnyvale, USA) with a SBAS/EGNOS correction service. The iG-Apex observations were carried out weekly at every sampling site in both fields from week 24 to week 34, i.e. from flowering to veraison (Figure 24d).

In addition, specific iG-Apex observations were performed on July 25th on the Syrah field in order to i) assess the variance due to repeatability and reproducibility of the measurement and ii) to compare it to the within-field variance. These iG-Apex observations were performed by six different operators replicated three times at the same site and analysed using ANOVA.

4.3.4 Analysis of variance and spatial structure

As a first approach, iG-Apex dynamics were studied for each field by following the evolution of their mean and variance during the season. Then, the spatial structure of each field was studied and illustrated by variography and mapping of iG-Apex. Finally, the operational feasibility of implementing this approach was assessed by comparing the spatially structured variance to the variance due to operator and repetition effects.

The analysis of the iG-Apex within-field spatial structure was carried out using semi-variograms at every date. Semi-variogram models were fitted with the gstat package using REML (Pebesma, 2004). When the model was second order stationary, the estimated stochastic variance (nugget effect - C0) and spatially structured variance (C1) were extracted. When the model was non-stationary, stochastic variance (C0) was extracted and the total variance of the field was considered as an estimate of C1 + C0. These characteristics were extracted for each week of the study period for both fields.

The evolution of total variance and the spatially structured variance were studied using the indices C1 + C0 and C1/(C1+C0) respectively. Interpolated maps were generated by ordinary kriging. Operator and repetition effects were evaluated with ANOVA after testing homoscedasticity with the Levene test. Maps were created with Qgis 3.4.8-Madeira (QGIS Development Team, 2020) and geostatistical analyses, graphs and ANOVA with R 3.6.1 (R Core Team, 2020).

4.4 Results and Discussion

4.4.1 Temporal evolution of iG-Apex

Both fields showed similar iG-Apex temporal dynamics (Figure 25). The mean iG-Apex of these two fields was close to 1 in week 24 and then gradually decreased and approached or equalled zero in week 34. This progressive slowdown in shoot growth can be explained by the high evaporative demand during summer with no rainfall leading to the gradual occurrence of water stress. The rainfall event observed in week 30 did not significantly influence this trend. These results are consistent with the literature (Martinez-De-Toda et al. 2010) and confirmed the value

of iG-Apex to monitor water status at field level when water availability is the main driving factor of vine growth.

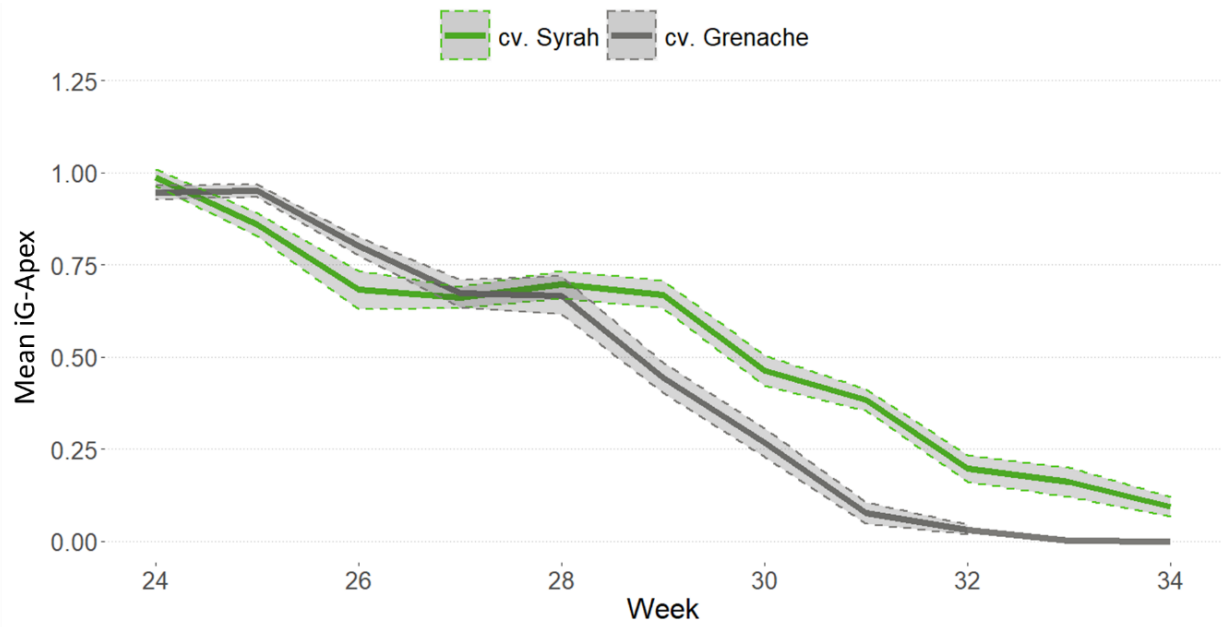


Figure 25: The two fields planted with Syrah (101 observations) and Grenache (98 observations) have similar iG-Apex temporal dynamics throughout the season with a gradual cessation of growth between weeks 24 and 34.

4.4.2 Temporal evolution of iG-Apex's total and spatially structured variances

For both study fields, total variance was low at the beginning and at the end of the season and relatively high between weeks 26 and 30 (Figure 26a). Over the same period, the spatially structured variance followed the same trend, reaching values >60 % of total variance between weeks 26 and 30 (Figure 26b). The stochastic (random) variance, although not negligible, remained lower during this period.

Evolution of variance over the season was mainly explained by changes in the spatially structured variance. This can be related to within-field variability of environmental factors that affect vine available water (e.g. soil composition, topography). Vine shoot growth in zones with low water availability was affected (slowed and stopped) earlier and was reflected by an earlier decrease in iG-Apex. Conversely, vine shoot growth in zones with high water availability was affected later leading to a delay in the decline in iG-Apex. The heterogeneity in vine available water in both fields resulted in an increase in the total variance observed at the within-field level when water restriction was increasing (from week 24 to week 30). Later in the season (from week 30 to week 34), total variance (Figure 26a) and spatially structured variance (Figure 26b) both decreased as the majority of apexes had stopped growing, generating observed iG-Apex values close to zero.

However, even if the dynamics were similar for the two fields, the total variance and spatial structure increased earlier and decreased later for the Syrah field compared to the Grenache field. These differences can be explained by the characteristics of the fields (cultivar, soil type, etc.) but this short study does not permit further investigations here into their respective influence.

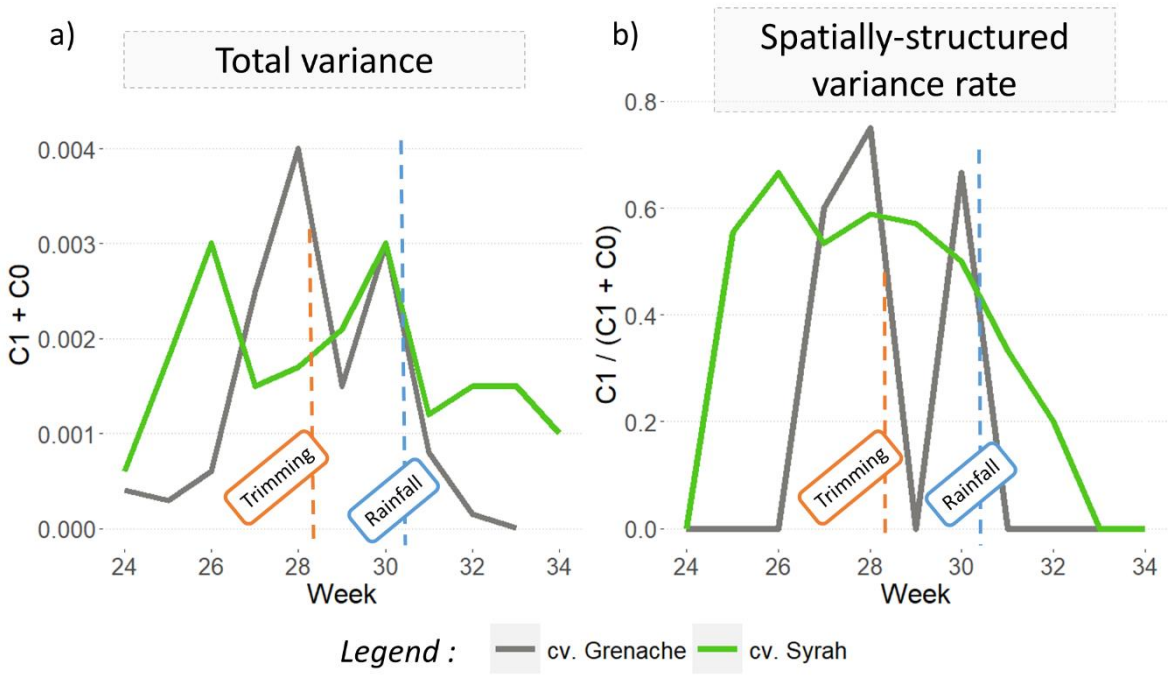


Figure 26 : The total variance (C1+C0) of iG-Apex observed at the intra-field scale increased and then decreased through the season for the two fields planted respectively with Syrah (green line; n = 101) and Grenache (grey line; n = 98) (a). The proportion of this variance that is spatially organised (C1 / (C1 + C0)) also increased and then decreased (b).

Another point of note is the response in week 29 for the Grenache field that showed a sharp decrease in total variance (Figure 26a) and spatially structured variance (Figure 26b) immediately after a canopy trimming operation in the field. By trimming shoots uniformly over the field, this canopy management operation homogenized the observed shoot growth a few days later, reducing the total variance and completely masking the spatial structure. However, by 10 days after trimming, shoot growth had restarted; subject to water availability, the total variance had increased and the spatial structure was present again. While trimming had a marked effect on variance, the rainfall event in week 30 did not appear to influence the iG-Apex response.

From an operational point of view, these results showed that the period around weeks 28-30 was the best period for characterising the within-field spatial structure of iG-Apex. If invasive canopy management is planned, then observations should preferably be made before canopy operations. If this is not possible, then a delay of approximately ten days appeared to be sufficient to allow the spatial structure in iG-Apex to be re-expressed at the within-field scale. In the conditions of this study, week 28 showed the highest spatial variability and this date was chosen for variography analysis and to map iG-Apex.

4.4.3 Mapping of iG-Apex's within field variance

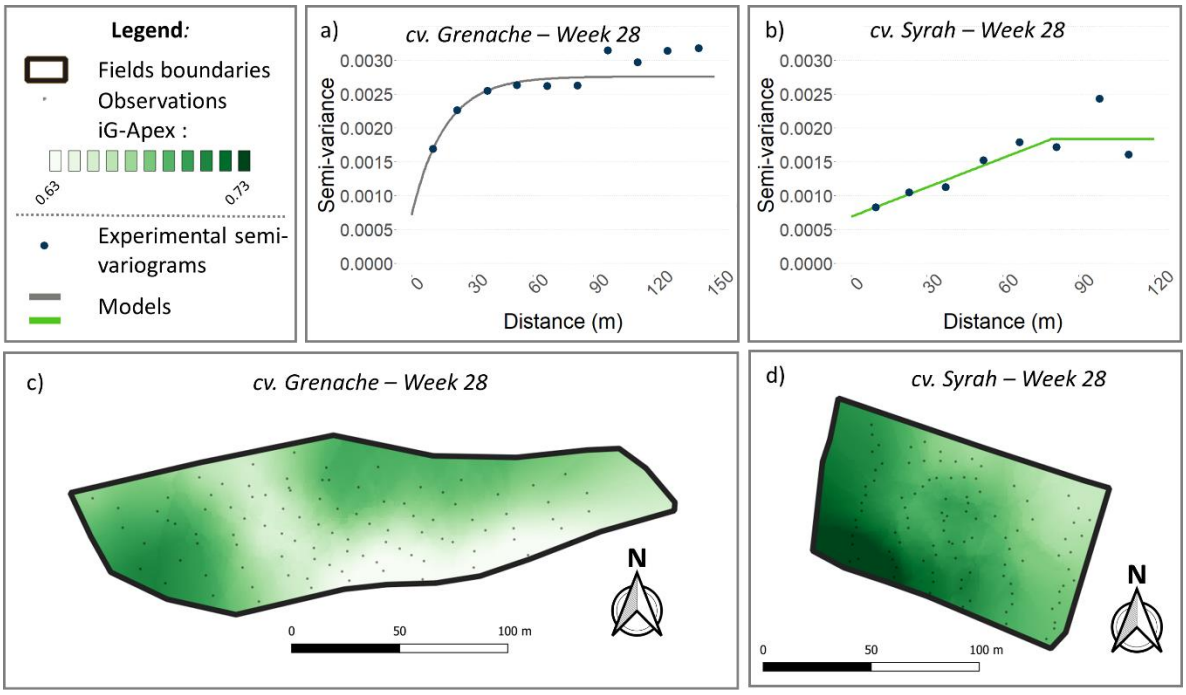


Figure 27: Experimental and model semi-variograms of the iG-Apex for week 28 for both study fields planted respectively to Grenache (a) and to Syrah (b) highlight a significant spatial organisation of iG-Apex values . The corresponding maps (c and d) confirm this spatial structure and illustrate the related spatial patterns.

The semi-variograms for the two fields exhibited a second order stationary phenomena with ranges of ~60 m. The C1/(C1+C0) ratio was respectively 0.74 and 0.61 for Grenache (Figure 27a) and Syrah (Figure 27b) indicating a strong spatial structure in both cases. These spatial structures are illustrated by the corresponding kriged maps. The Grenache field map (Figure 27c) shows a zone with weaker shoot growth in the south and centre of the field, while the Syrah field map (Figure 27d) shows a south-west to north-east gradient. In both cases, these spatial structures of iG-Apex values are driven by spatial patterns of environmental factors i.e. topography, variability of soil characteristics and their resulting influence on vine water availability.

4.4.4 Study of operator effect

In order to study the possibility of implementing within-field iG-Apex mapping in an operational context, it is important to compare the characteristics of the studied phenomenon (C1, C0) to the metrological characteristics of the method. The ANOVA indicated that, for iG-Apex observations, the variance due to the measurement effect was 2.8e-04 (results not shown). By way of comparison, the nugget effect (stochastic variance) was respectively 7.1e-04 and 6.9e-04 for the Grenache (Figure 27a) and Syrah (Figure 27b) fields in week 28. Therefore, approximately 40% of the observed random variance was estimated to be due to measurement uncertainty. The rest of the stochastic variability can be explained by the phenomenon of shoot growth, which can vary over very short distances (inter-shoot or inter-plant variability). Nevertheless, this stochastic variance was very small compared to the spatially structured variance, 2.0e-03 and 1.1e-03

respectively for the Grenache and Syrah fields (Figure 27a-b). Results of observations collected by six different operators indicated that the operator had a significant effect on the mean iG-Apex (p-value = 5.6e-07) as variance due to the operator effect was 1.0e-02.

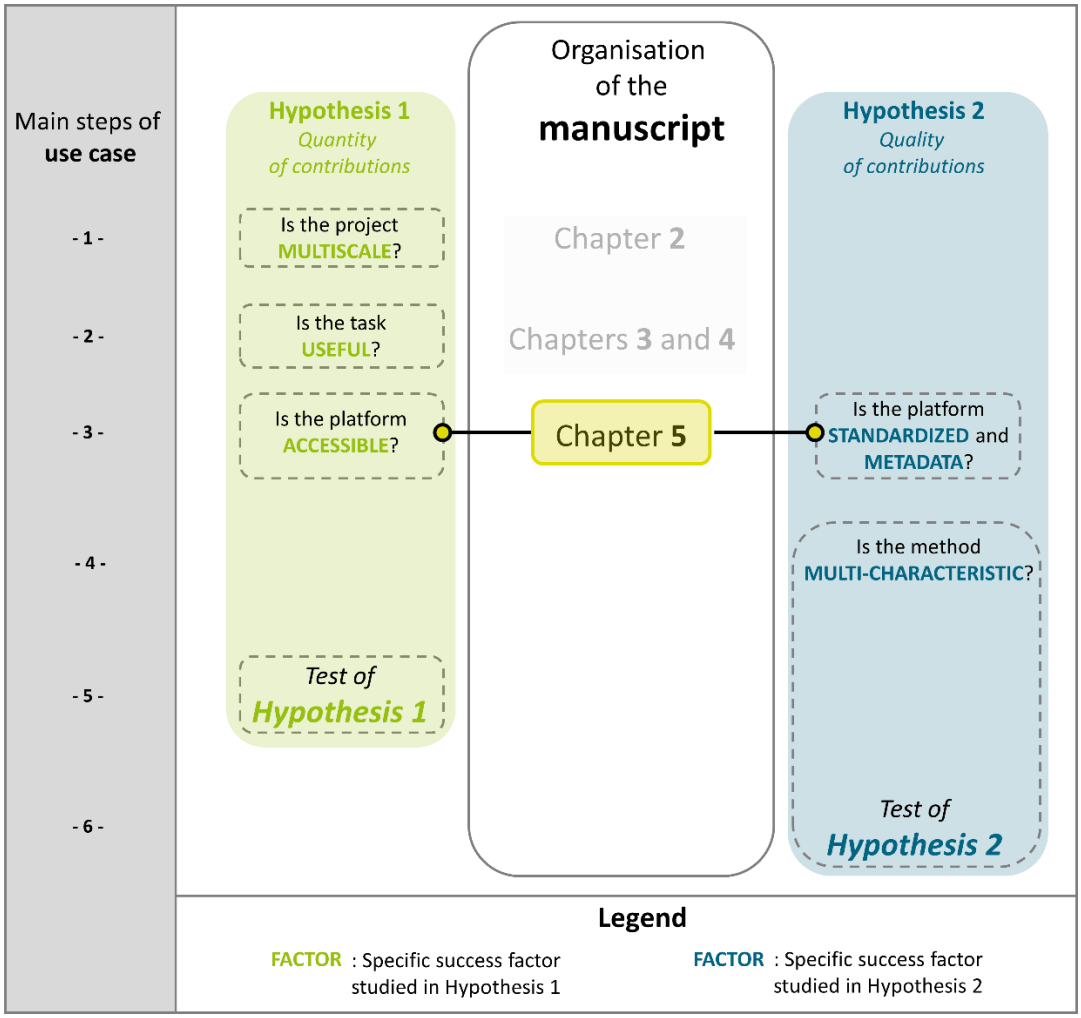
These results demonstrated the relevance of using iG-Apex to produce maps of shoot growth and to identify spatial water restriction patterns at the within-field level, when observations are collected by the same operator. Although based on a subjective human observation, this approach is still relevant to be used as a decision support tool by winegrowers and their advisors to identify and manage within field vine water restriction zones. However, as it is subjective, it is prone to operator error. The factor of 10 difference that exists between the variance due to the operator effect and the spatially structured variance clearly indicated that this type of map should only be produced from observations collected by a single operator. Future work may explore the possibility of reducing the operator effect, either before measurement by setting up a calibration procedure between operators or after measurement by using homogenisation methods, such as those developed by Pichon et al. (2019).

For a practical implementation, the identification of the optimal observation period will be a major issue. One option could be to carry out, at the beginning of the season, a temporal monitoring of a few zones based on targeted sampling from ancillary data e.g. free remote sensed images. When the variance between the different zones exceeds a given threshold, the acquisition of spatially dense iG-Apex observations would be triggered to produce a map. Future research may be set up to develop and test this method in real conditions and to define relevant variance thresholds to implement it.

4.5 Conclusion

Although the measurement is based on human observations, the Index of Growing Apex (iG-Apex) can be used at the within-field scale to highlight zones with different water status of non-irrigated vineyards. The iG-Apex maps have the advantage of being produced at a low cost and represent an interesting decision tool for small wine producers in the south of France who do not have the capacity to invest in expensive methods. However, this work showed that care must be taken with data collection as the operator effect remains significant. Ideally, observations should be collected by the same operator. In the case where several operators are involved, iG-Apex will require a preliminary phase of calibration between operators and/or observations would need to be homogenised. The best time to characterise within field variability using iG-Apex is the middle of summer (July), before veraison, when the spatial structure is the strongest. Finally, this work helped to identify future research issues for the joint use of iG-Apex maps with other easily accessible sources of information.

Chapter 5: ApeX-Vigne: a platform designed for collecting farmsourcing spatial observations



The objective of chapter 5 is to demonstrate that the ApeX-Vigne mobile application is easily accessible and usable by potential participants (specific factor ACCESIBLE), that it enables data to be collected according to a standardised protocol (specific factor STANDARDIZED) and that it records contextual information on the observations gathered (specific factor METADATA). The question explored in this chapter is: What methodological and technological choices should be made to provide an easily accessible and ergonomic mobile application that enables the collection and centralisation of standardised, dated and geo-referenced vineyard observations?

5.1 Preliminary comment on the development of ApeX-Vigne application

The beta version of ApeX-Vigne mobile application was developed and tested in 2018. The app was officially launched in 2019 and improved in 2020 and 2021. Three types of stakeholders were involved in the development of the application:

- *Development team*: This team was composed of a project manager in charge of leading the project and prioritizing the main decisions and a developer in charge of the design and the implementation of technical solutions.
- *Expert users*: Three expert users were involved in the project. These experts came from advisory and support structures such as the French Technical Institute of Vine and Wine (IFV). They were particularly involved in the project and familiar with shoot growth observations. They were able to synthesize the expectations of many potential users.
- *Beta testers*: These people were farmers and advisers, simple users of the application. Twenty beta testers were involved in the project. They agreed to use the beta version of the application in 2018 (one year before the official launch of the application) and to share their feedbacks.

The ApeX-Vigne application was developed according to the principles of agile methods (Mahmud and Abdullah 2015) which favour short development cycles with frequent discussions between development team and users. The first 3 cycles were carried out in 2018, before the period during which shoot growth observations were gathered (Figure 28). During these cycles, the development team interacted only with expert users. Cycle 4 took place during the 2018 observation period after the feedbacks of beta testers. Cycle 5 took place just before the official launch of the application on June 3rd 2019. Communication was made to the wine industry mainly through the viticulture technical press (See Appendix 9.2). Cycles 6 and 7 were made just before observation periods of years 2020 and 2021, respectively. Their objective was to fix existing bugs and to improve some aspects of the application.

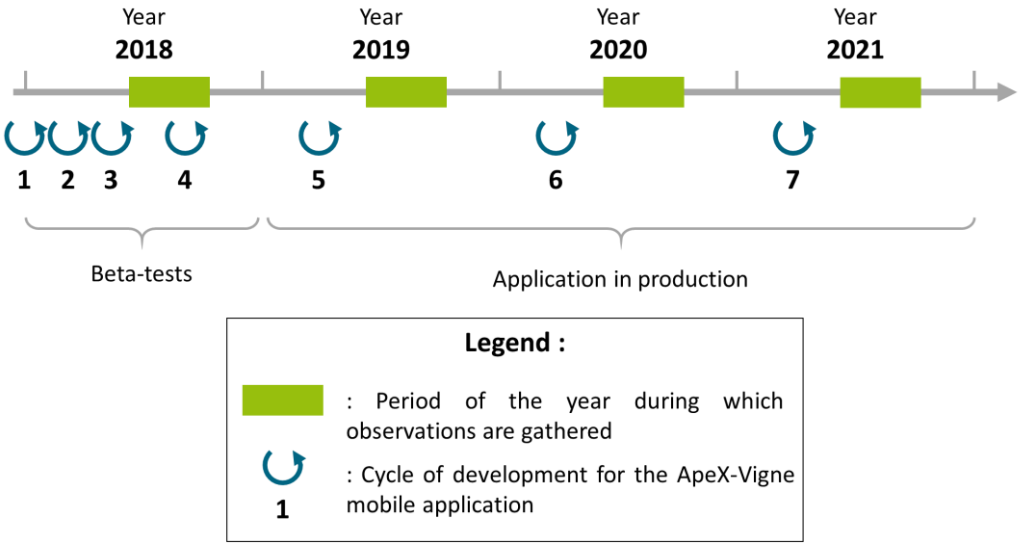


Figure 28: Development cycles of ApeX-Vigne mobile application

This chapter of the manuscript was written at the end of 2018, between development cycles 4 and 5. At this point of the project, several concepts were still being defined and have evolved since then. In this chapter, the application was called "ApeX" instead of "ApeX-Vigne" and the method used to summarize shoot growth observations was the Shoot Tip (STI) Index instead of iG-Apex. The principles remain the same and these terms can be considered synonymous for the purpose of the thesis. The design of the application has also changed and screenshots of the latest version can be found on Appendix 9.1. The main interest of this chapter lies in the material and methods of this article, which presents the technological choices that have been made. The results obtained at the time the article was written were partial. Data collected between development cycles 6 and 7 will be presented and analysed in detail in the next chapter (Chapter 6:).

5.2 Details about the corresponding paper

5.2.1 Title of the paper

This chapter was published in the proceedings of the 12th European Conference on Precision Agriculture under the title "*Easy water stress detection system for vineyard irrigation management*". (https://doi.org/10.3920/978-90-8686-888-9)

Easy water stress detection system for vineyard irrigation management

G. Brunel*, L. Pichon, J. Taylor and B. Tisseyre
ITAP, Irstea, Montpellier SupAgro, Univ. of Montpellier, Montpellier, France; guilhem.brunel@supagro.fr

5.2.2 Authors

G. Brunel[1], L. Pichon[1], J. Taylor[1] and B. Tisseyre[1]

[1]*ITAP, Irstea, Montpellier SupAgro, Univ. of Montpellier, Montpellier, France*

5.2.3 Abstract

The monitoring and management of vine water stress is a major issue for many vineyards. A method based on the observation of shoot tip growth makes it possible to determine the water stress of plants without the drawbacks of standard instrumental methods. However, this "shoot tip" method is rarely used in the field, as observations and index calculations need to be done manually. Therefore, while the data collection can be relatively quick, data interpretation is not well supported. This article presents a mobile application, 'Apex', developed to answer this challenge. The Apex application allows a simple and fast recording of observations and a pooling of spatialized observations. The developed application makes it possible to envisage the collaborative water stress monitoring with crowdsourcing observations.

5.2.4 Keywords:

Vine water status, shoot tip index, smartphone application, *Vitis vinifera*

5.3 Introduction

Many authors have demonstrated the significant role of water status in vine behaviour and berry composition (Cifre et al. 2005). Monitoring vine water status is of importance for yield and quality management purposes. The vine water status naturally changes over time according to the climatic characteristics of the season. In addition, many authors have shown that it presents a significant spatial variability at different scales: at the within field level (Tisseyre et al. 2005), at the vineyard level (Taylor et al. 2009) and at the regional scale (Baralon et al. 2012). Characterizing the temporal and the spatial variability of vine water status is therefore of paramount importance at these different scales to provide the wine industry with relevant decision support tools: either for irrigation scheduling or for other management practices in the case of non-irrigated vineyards.

Several standard instrumental methods have been proposed to measure plant water status (Acevedo-Opazo et al. 2008). However, these are not easy to obtain, since they are manual techniques requiring heavy materials and specific skills. These constraints make systematic spatio-temporal vine water status measurements difficult to carry out and time-consuming. In order to overcome these constraints, simple alternative approaches have been proposed (Acevedo-Opazo et al. 2008). Among them, the shoot tip index (STI) proposed by Trambouze et al. (2009) and Rodriguez Lovelle et al. (2009) is of interest as it is based on simple observations of vine shoots. Although more imprecise than standard methods, it has the advantage of being easily achievable without expensive equipment and by inexperienced operators. This method is therefore interesting to monitor temporal and spatial changes in vine water status when several operators may be involved (workers, growers, advisors) in collecting observations.

However, the STI method has some limitations that limits its dissemination: i) it requires the observers to count the Apex (end of shoots) of approximately 50 shoots and to class them into three different classes (full growth, moderate growth, growth stopped), ii) the STI needs to be calculated based on the Apex observations and then iii) the observed STI value needs to be recorded and transferred to a database with additional information, like the time and the location of observation. Although simple, these different operations can be cumbersome because they require information to be recorded and a calculation to be made in the field. The aim of this article is to present a mobile application, called ApeX, whose purpose is to tackle this issue by providing a simple tool for: i) collecting Apex notations, ii) calculating STI, iii) geo-referencing and timing STI observations and iv) sending STI observation to a server.

5.4 Materials and methods

The specifications of ApeX were defined by involving the end-users in development and the developments followed the precepts of UX Design (Hekkert 2011). This step made it possible to highlight the characteristics that defined the technical choices of the application. The main ones are summarized hereafter:

- Dissemination: The application must be widely deployable to different organisations and stakeholders (growers, advisors, etc.) whose smartphone (model, operating system, etc.) may be different,

- Accessible: Users are not necessarily familiar with professional applications. The use must be simple and user-friendly while maintaining the data integrity,
- Offline mode: Network coverage is variable according to the areas of use, the application must operate in offline mode as well as in connected mode.

The developments followed the precepts of UX Design, which refers to a person's emotions and attitudes about using a product or service. User experience is dynamic over time due to changing circumstances of usage and changes to individual systems, as well as the wider usage context. User needs can be described in three ways: Why, What and How. "Why" corresponds to the motivations, values and vision of the context. "What" describes the features and functionalities necessary to achieve the "Why". And finally, "How" corresponds to the way of handling and aesthetics. As shown in Figure 29, the application design process follows an iterative cycle consisting of four phases based on Design-thinking methods:

- First phase consists of discovering user needs, requirements and constraints,
- Second phase consists of imagining all the scenarios and solutions to meet the needs. At the end of this phase, all participants make a pragmatic choice on the new features to be added,
- Third phase consists of offering the user a visual set that is both operational and functional. The aim is to combine comfort of use without losing efficiency,
- Fourth phase consists of quickly implementing the chosen characteristics and performing tests in real conditions by the end user.

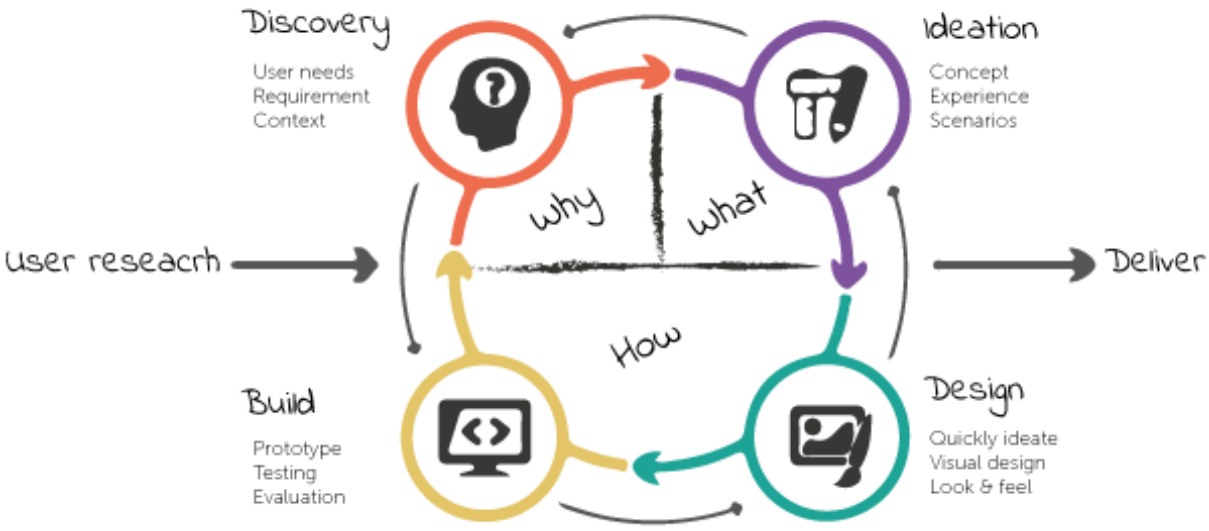


Figure 29: This diagram shows the iterative design process to discover, define, design and develop a digital solution that meets user needs.

Unlike a traditional approach where a finalized solution is produced after the third phase, here the cycle will be renewed after a test period. This is needed because the user experience is dynamic and constantly changing over time. Four experts were integrated into the ApeX development cycles (6 complete cycles), these users were able to test and validate the various updates.

ApeX was developed as a hybrid app. A hybrid app is one that combines elements of both native and web applications. Hybrid applications execute within wrappers that are targeted to each platform, and rely on standards-compliant application programming interface (API) bindings to access each device's capabilities, such as embedded sensors, data, network status, like a native application. A hybrid app approach was chosen to facilitate the dissemination of the application across more than one platform, without having to re-implement it with each platform's language and tool set.

The development of the application was programmed using Ionic framework based on Apache Cordova (Griffith 2017). Apache is an open-source mobile development framework, which allows the use of standard web technologies - HTML5, CSS3, and JavaScript - for cross-platform development. AngularJS is a JavaScript-based open-source front-end web application framework to address many of the challenges encountered in developing applications. The JavaScript components complement Apache Cordova. It aims to simplify both the development and the testing of such applications by providing a framework for client-side model–view–controller (MVC) and model–view–view model (MVVM) architectures, along with components commonly used in rich internet applications. The application uses Typescript, a superset of JavaScript, which allows access to a number of extra features, such as type of declarations and interfaces. JavaScript code runs on a NodeJS server, which is a platform built on Chrome's V8 JavaScript engine. It differs from other platforms in that it uses a non-blocking approach to perform I/O asynchronously. This workspace is coupled with Gulp, which offers a toolkit for automating time-consuming tasks in the development workflow (Figure 30).

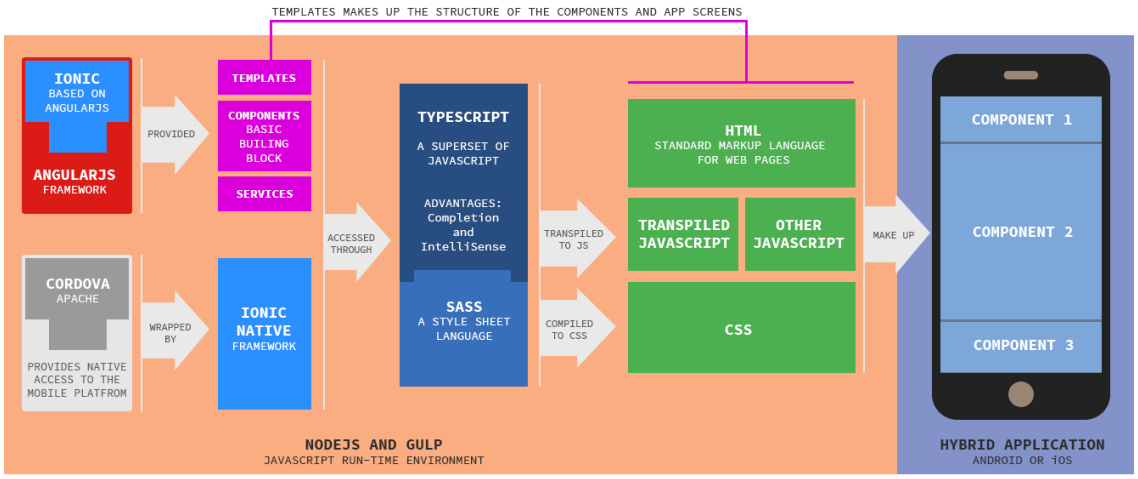


Figure 30: Technologies used for the development of ApeX.

All data created and inserted into the application is synchronized with the ApeX cloud when the smartphone has a 4G or Wi-Fi connection. To operate in areas without a network, all data is stored in a local database supported by SQLite and are automatically synchronized on a MySQL server database when the network allows it. Communication between the smartphone and the server is provided by a JavaScript API. The application uses the smartphone's GNSS to locate observations. The web application was designed and developed to quickly and efficiently visualize the collected data. The web application was developed using HTML5 mark-up language, CSS3 styling and WordPress API for a responsive web design. It automatically adjusts to different screen sizes, allowing it to be easily viewed on any device or at any screen resolution.

The ApeX app was tested by twenty users in the summer of 2018 in the *Occitanie* and *Provence-Alpes-Côte d'Azur* regions of southern France. One session represents fifty observations of shoot tips in a vineyard field, which is necessary for the STI to be calculated to estimate vine water stress. More than 1500 sessions distributed over 270 different vine fields were collected over the season.

5.5 Results

5.5.1 Application interface

The calculation page is divided into three areas (Figure 31a): i) a list to choose the field that the observation is performed in. This list is ordered according to the user's GNSS position, ii) three buttons for shoot tip selection. This area takes up 70% of the screen for comfort of use. Each button has a different vibration to indicate user choice, iii) a dynamic counter at the bottom of the screen to allow the user to know the number of observations made. A long vibration of the phone indicates the achievement of the fifty observations necessary to calculate the STI. The STI button is only clickable after fifty observations to compute STI.

The home page (Figure 31b) navigates to the calculation pages and allow visualization of the results. For each field, the user can access i) the current growth rate of the vine (vigour), ii) the growth dynamics based on past and current growing STI rates and iii) a scale range of the current STI (water stress). To improve the quality of the user experience, other interfaces have been developed (instructions, contact form, field characteristics, etc...).

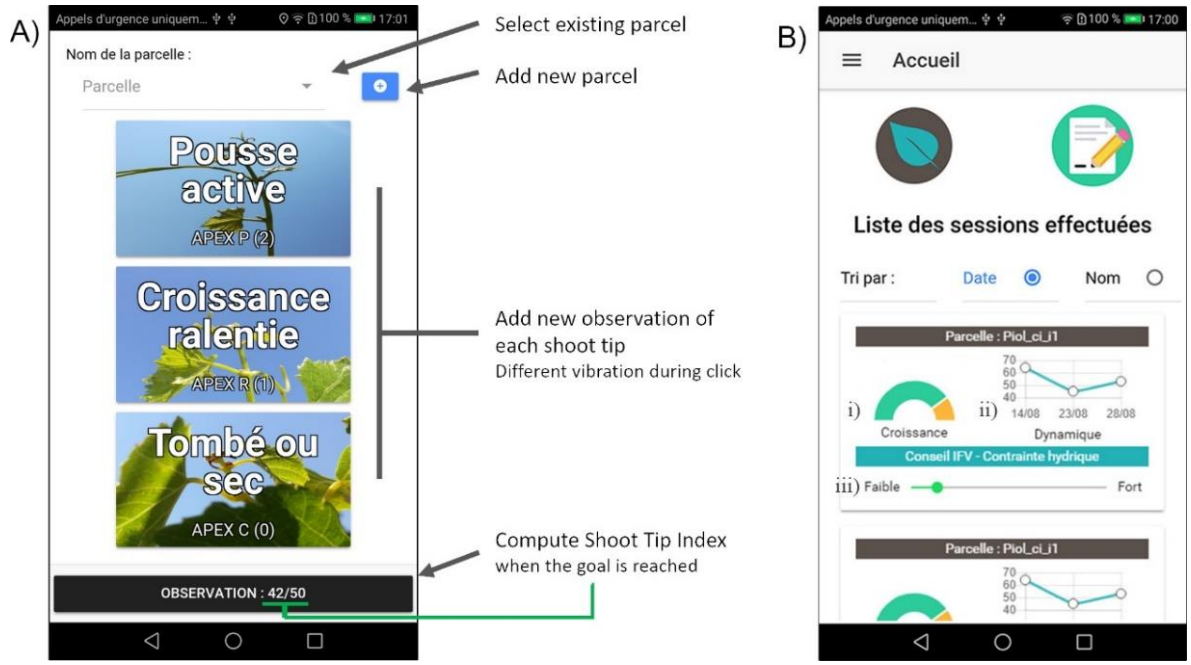


Figure 31: Screenshot of both main interfaces of ApeX.

5.5.2 Spatial distribution of observations

More than 1500 sessions (each of 50 shoot observations), from 22 operators from 7 different companies (extension services, advisers and research institutes) were collected with the ApeX app during 2018. Figure 32 shows the sessions collected at two different scales: the within-field scale (Figure 32a) and the regional scale (Figure 32b).

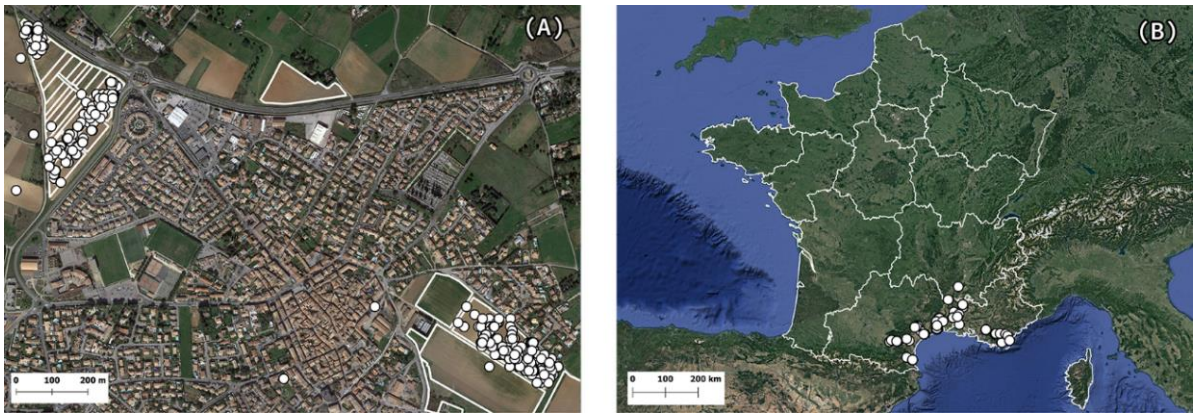


Figure 32: Observations collected with Apex during 2018 in southern France, a) example at the within-field level, b) spatial distribution of measurements at the regional level.

At the within field scale (Figure 32a), the 50 observations were plotted to provide more detailed information on the monitoring that has been carried out. Some observations were positioned outside the fields. This result can be explained by the low quality of the smartphone's geolocation sensor but also by poor operating conditions (fast switching on of the smartphone, locations with poor positioning signal quality etc...). All together, these problems resulted in spatial outliers. It highlighted the need to identify and to correct outliers before the app is made available to professionals.

At the regional scale, the data were collected mainly around the Mediterranean basin (Figure 32b). This spatial distribution of the observations was mostly driven by the location of partners who were interested in using the ApeX application and for whom the monitoring of water stress is a major issue.

5.5.3 Temporal distribution of observations

Figure 33a shows the number of sessions collected during the 2018 season for each two-week interval. The first sessions were made at the beginning of June with about fifty sessions over the first two weeks. The number of sessions then increased sharply to ~400 over the same interval at the beginning of July. It then decreased during August, with only a few sessions during September and October. ApeX was predominantly used to decide the date of the first irrigation during the dry summer period when vine water stress reached critical values. In the Mediterranean region, this decision is usually taken in July. The increased and decreased number of observations corresponded respectively to the start and end of a period of risk for the vine. Most users who started using the application kept on using it during the season. This indicated that it met users' expectations and on the other hand that its usability was satisfactory to them.

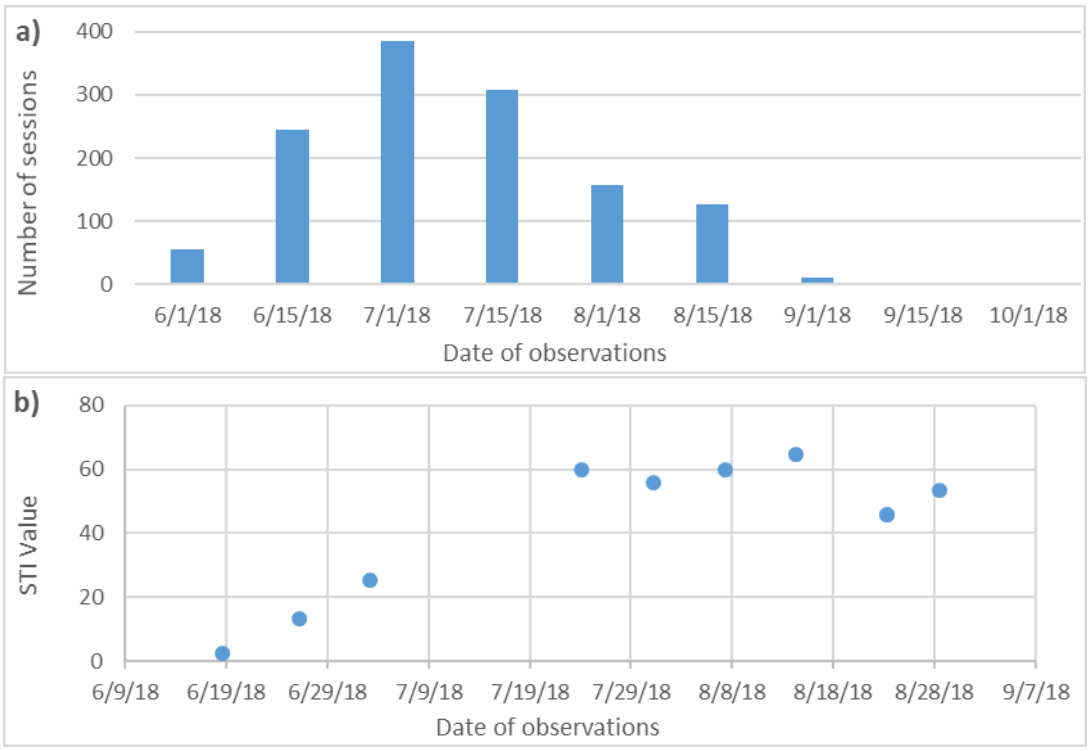


Figure 33: Use of ApeX over the season and examples of results (June – October) (a) the count of sessions made each fortnight and (b) the average recorded STI at approximately weekly intervals in a field.

Figure 33b shows an example of the evolution of the STI indicator over the season in a vineyard. The STI was close to 0 on June 19th, corresponding to significant vegetative growth and no water restriction. The STI increased steadily during June, as vine growth decreased with the occurrence of a significant water restriction. The STI then stabilized at around 60, indicating moderate to weak growth between mid-July and early September. Slight variations in the STI were observed during this period. This information, combined with agronomic expertise, could be used to identify periods of water stress as a decision support vineyard management.

5.6 Discussion

The development of ApeX followed the MVC pattern, in which the user interface (presentation), functional process logic (decision rules code), computer data storage and data access were developed and maintained as independent modules. This application architecture provided a flexible model and a reusable application. By segregating an application into tiers, it was possible to refine the existing model or add new ones based on shoot tip analysis, without having to rebuild the entire application.

The specifications of ApeX were realized by involving the end-users and the developments followed the precepts of UX Design. This reduced development times by directly integrating feedback and requests for improvement. A suggestion form was added in the first version so that the user could feedback directly during use. Consequently, some changes were made, for example, vibrations during clicks were added to allow the user to feel actions without having to check the screen at each observation. This basic feature was widely appreciated, as it considerably increased user

comfort. Users also felt involved in the application, which explains the success and the large amount of data collected during the first test season.

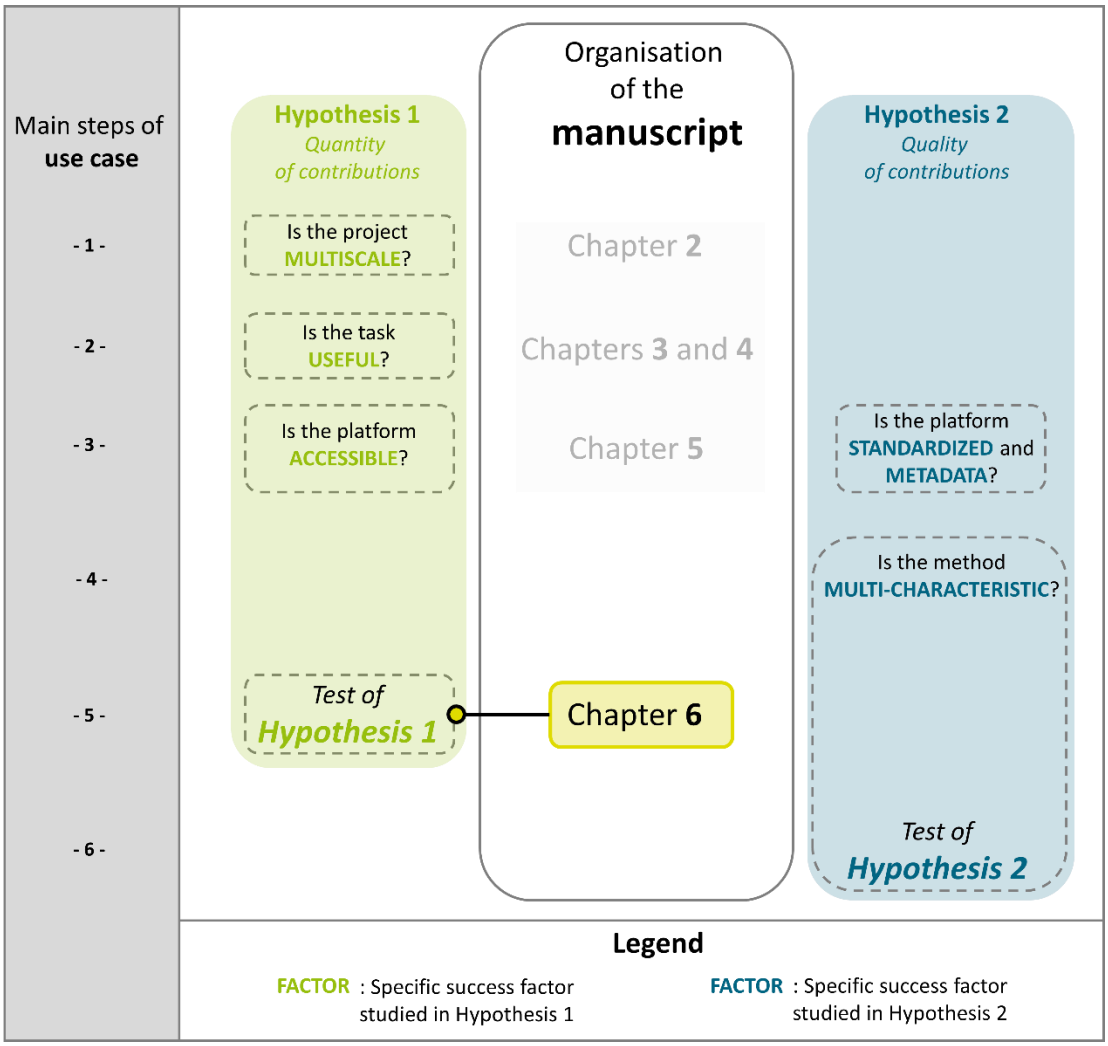
The results of this study showed that the ApeX application allows the collection of observations on vine growth in a simple and rapid way. These observations were georeferenced and shared at different scales. This work assumed that the ApeX method has been validated by comparison with reference methods in previous work (Trambouze et al. 2009). New experiments will have to be conducted to validate whether the use of the ApeX application improves or alters the relationship with reference measurements. In this work, the use of the application was restricted to a limited number of users, but the number of sessions collected was still relatively large. When the ApeX app is freely released to the wider public, it should collect a very large number of sessions at very different scales. This wider use will be an opportunity to develop a tool to help monitor water stress at a large scale. However, it does raise the question of the quality of the data that will be produced by the application. This aspect is indeed a major challenge in the use of crowdsourced data (Minet et al. 2017). It will be necessary to set up mechanisms to characterize and improve the quality of the data collected. Finally, it seems relevant to study how this new data source, that provides data in large quantities of low to medium quality, can complement and enrich the reference data.

5.7 Conclusion

The ApeX application was developed to monitor vine water status using a simple and robust method that can potentially be used by stakeholders with very different skills. The application is free of charge. It is a novel decision support tool that can be an interesting alternative for small growers who cannot afford alternative commercial services. The version presented in this work is a first approach. However, after a first season of use, the results showed that users have adopted the application throughout the monitoring season. This indicated that it meets a need and that the interface is appropriate.

The app can be used at the field or the within-field scale for temporal or spatio-temporal monitoring of vine water status. It can also be used at a regional scale by aggregating data, although one limitation remains the quality and imprecision of the information obtained with this type of approach.

Chapter 6: Evaluating the quantity of contributions in a farmsourcing project of spatial observations

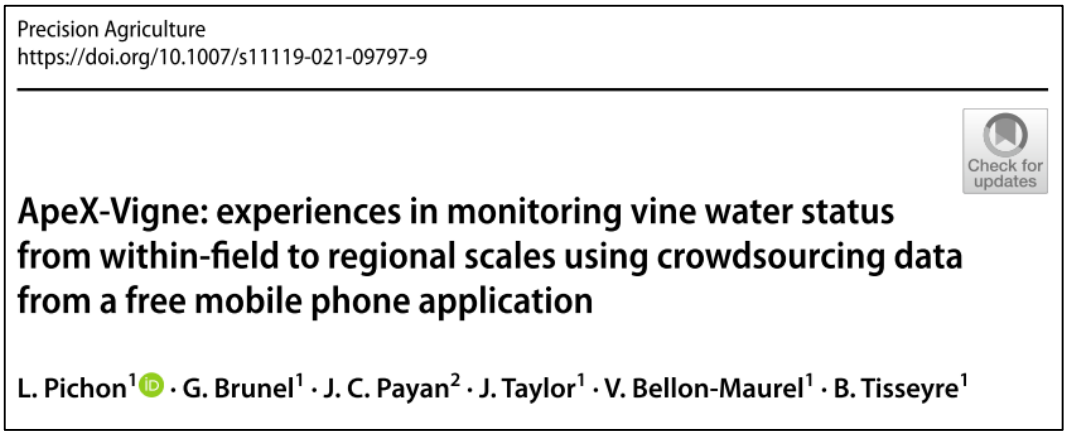


The objective of chapter 6 is to test hypothesis 1 and to generalise the results on specific success factors regarding the quantity of contributions. The question explored in this chapter is: How can the adoption and uses of an FSO application be evaluated? Did the ApeX-Vigne application allow the collection of a large quantity of observations? This chapter also gives an overview of the potential uses of the data collected with ApeX-Vigne at within-field, inter-field and regional scales.

6.1 Details about the corresponding paper

6.1.1 Title of the paper

This chapter was published as a full-paper in the journal "Precision Agriculture" under the title: "*ApeX-Vigne: experiences in monitoring vine water status from within-field to regional scales using crowdsourcing data from a free mobile phone application*". (https://doi.org/10.1007/s11119-021-09797-9)



6.1.2 Authors

L. Pichon[1], G. Brunel[1], J.C. Payan[2], J. Taylor[1], V. Bellon-Maurel[1], and B. Tisseyre[1]

[1]*ITAP, Univ. of Montpellier, Institut Agro Montpellier, INRAE, Montpellier, France*

[2]*Institut Français de la Vigne et du Vin, Le Grau du Roi, France*

6.1.3 Abstract

Monitoring vine water status is a major issue for vineyard management because water constraints impact both the quality and the quantity of the harvest. Existing methods are often costly and complex to implement. ApeX-Vigne is a free mobile application developed to facilitate the collection and geolocation of 50 vine apex observations to characterise vine shoot growth and classify it into 3 growth categories. The application also provides the user with a simple estimate of vine water status based on shoot growth. This paper presents the results obtained over two seasons (2019 and 2020) after the launch of the Apex-Vigne application and its use over a large wine producing region in the south of France. An existing method was adapted for evaluating the interest of the application based on the number of installations and uninstallations. The results showed that the application had more than 1200 downloads and 6000 observations made in the 2020 season. Examples from the commercially collected data showed that ApeX-Vigne can be used as a tool for characterizing water stress at within-field and inter-field scales. Finally, it was also demonstrated that by enabling the massive and centralized collection of spatial field and within-field scale observations of shoot growth, the ApeX-Vigne data was able to characterise the spatial structure of vine water status at the regional scale. Access to this new source of

information offers opportunities for the management of water resources at a regional scale as well as for site- and vineyard-specific management. These results also raised new research questions on the joint use of this new source of spatial data with other sources of high spatial resolution information.

6.1.4 Keywords:

Data sharing, digital agriculture, precision viticulture, smartphone, *Vitis vinifera*,

6.2 Introduction

Many authors have demonstrated the significant role of vine water status on vine physiology and berry composition (Cifre et al. 2005; Van Leeuwen and Seguin 1994; Zufferey et al. 2017). Monitoring vine water status is of importance for yield and quality management purposes. The vine water status naturally changes over time according to the climatic characteristics of the season. In addition, many authors have shown that it presents a significant spatial variability at different scales in Mediterranean climate conditions: at the within field level (Tisseyre et al. 2005), at the vineyard level (Taylor et al. 2009) and at the regional scale (Baralon et al. 2012). Characterizing the temporal and the spatial variability of vine water status is of paramount importance at these different scales to provide the industry with relevant decision support tools for various management options, including canopy and irrigation management, fruit maturity and harvest logistics.

Several standard instrumental methods have been proposed to accurately measure plant water status (Rienth and Scholasch 2019). However, these standard methods are not easy to perform as they are manual techniques that require cumbersome equipment and specific skills. As a result, the use of these methods is typically limited to experiments carried out by research or technical institutes and they are rarely adopted by viticulture practitioners (growers or advisors). In order to provide simpler, non-destructive and cost-effective approaches to estimate vine water status, methods to monitor shoot growth have been proposed. When vine access to water becomes limiting, it has been shown that vine growth is affected (Pellegrino et al. 2005). Martinez-De-Toda et al. (2010) proposed a simple and operational method to relate shoot growth to the water restriction experienced by vines. Their method is based on classifying apexes (vine shoot tips) according to three growth levels: i) full growth (FG) when the apex is undergoing active organogenesis, ii) moderate growth (MG) when organogenesis is reduced and iii) stopped growth (SG) when the apex has fallen off or dried out. The authors showed that the observations of apex growth were correlated with stem water potential, a reference measurement of vine water status. The operational implementation of the approach requires counting 50 apexes (Rodriguez Lovelle et al. 2009). The growth of these 50 apexes can be summarised by an indicator S proposed by Martinez-De-Toda et al. (2010) (Equation 3).

$$S = w_{FG}S_{FG} + w_{MG}S_{MG} + w_{SG}S_{SG} \qquad \textit{Equation 3}$$

where w_{FG}, w_{MG}, w_{SG}, correspond respectively to the proportions of full growth, moderate growth and stopped growth apexes. S_{FG}, S_{MG}, S_{SG} are coefficients assigned to each of these growth levels and are adjusted by experts according to the pedo-climatic context.

Although less precise than reference methods, monitoring shoot growth has the advantage of being easily achievable without expensive equipment and by inexperienced operators. Therefore, this method is of interest to monitor temporal and spatial changes in vine water status when several operators (workers, growers, advisors etc.) may be involved in collecting observations. It is of particular interest in regions where the majority of vineyards are managed by small growers, who have limited investment capacities and still carry out many manual in-field operations. However, the monitoring of shoot growth has some limitations that has restricted its adoption until now: i) it requires observers to count the apex of 50 vine shoots, to classify them into three different classes and to carefully record the results of the counts for each class, ii) the final indicator S needs to be calculated based on the apex observations and iii) the observed S value needs to be recorded and transferred to a database with additional information, such as the time and location of the observations. Although simple, these different operations can be cumbersome because they require information to be recorded and a calculation to be made in the field.

In order to tackle this issue and to simplify the use of this operational method for monitoring shoot growth, a dedicated free mobile application (app), called ApeX-Vigne, was developed and launched in 2019 (See Chapter 5:). The purpose of the app is to provide a simple tool to growers and advisers to: i) facilitate the collection of apex observations, ii) automatically calculate the S indicator, iii) geo-reference and time-stamp the S indicator values and iv) automatically transfer the S indicator values to a database on a server to facilitate future analysis and sharing of these data.

Two seasons after the launch of the app, it is now possible to evaluate its adoption and appropriation by users. In the literature, the approaches described for this evaluation are either very generic (Liu et al. 2018) or, on the contrary, very specific to a use case (Rahmati and Zhong 2013). To the authors' knowledge, there is no specific method for understanding user behaviour of an app in a professional agricultural context.

In addition, the use of this app and the collection of observations by a significant number of farmers and advisers paves the way to use this dataset to monitor vine growth at a regional as well as local scale. Characterising vine growth at large spatial scales may be of interest for cooperatives, public services or research for understanding general trends in making common decisions. Using so called 'crowd-sourced' datasets collected by many operators is common in the monitoring of environment conditions (Picaut et al. 2019), biodiversity (Joly et al. 2016) or natural disasters (Rogstadius et al. 2013), but its potential has not been widely studied in agriculture (Minet et al. 2017).

Therefore, the objective of this paper is i) to evaluate how the ApeX-Vigne application is adopted by users, ii) to study whether the application answers the questions that professionals have about monitoring vine water status at the field or within-field scales, and iii) to verify whether the aggregated vineyard data can be used to help address regional-wide monitoring issues by providing a new source of relevant information.

6.3 Materials and method

6.3.1 The index of Growing Apex (iG-Apex) for monitoring shoot growth

Based on the equation proposed by Martinez-De-Toda et al. (2010) (Equation 3), the French Technical Institute of Vine and Wine (Institut Français de la Vigne et du Vin – IFV) recommends that growers set the S_{FG}, S_{MG}, S_{SG} coefficients to respectively 1, 0.5 and 0 to define the index of Growing Apex (iG-Apex) (Payan, 2020). These coefficients have been determined from experiments carried out in the south of France and were selected to best relate the iG-Apex with the occurrence of moderate water restriction (Payan, 2020). These coefficients, as recommended by IFV, have been used to calculate iG-Apex in this study.

The iG-Apex method is used at the field and sometimes at the within-field scale by growers and advisors who measure 50 apexes spread over 10 different vines selected in representative zones of the field (Rodriguez Lovelle et al. 2009). Around full bloom, the iG-Apex is generally close to 1. It then decreases with the occurrence of increasing water restriction to reach values close to 0, generally around veraison.

6.3.2 General technical characteristics of the ApeX-Vigne application

ApeX-Vigne is a free hybrid app combining elements of both native and web applications in order to facilitate its dissemination over several platforms. This application is currently only available on the Android platform via the Google Play Store (https://play.google.com/store/apps/details?id=ag.GB.apex&hl=fr last viewed 02/12/2021). The architecture was conditioned to meet operational constraints and functionalities, such as the ability to use the application in areas where network coverage is intermittent or even absent and to safely send and store the data collected on a central database server. Regarding this latter functionality, when users download the ApeX-Vigne application, they explicitly give their consent for the observations they collect to be used in research projects.

For the off-line mode of the application, data storage is performed on the smartphone. Data are automatically synchronized with the central database when the smartphone has a 3G (or greater) signal or a Wi-Fi connection. Communication between the smartphone and the server is provided by a JavaScript API. For geolocation, the application accesses the smartphone's GNSS.

6.3.3 Application interface and publication

The ApeX-Vigne app has two main screens, an input screen for entering observations (Figure 34a) and an output screen with summarized results (Figure 34b). The input screen is divided into three areas: i) a list to select the field where the observations were performed, ii) three buttons to select the shoot tip class, and iii) a dynamic counter at the bottom of the screen to inform the user of the number of observations made. The button to calculate the iG-Apex only becomes active once 50 observations have been made. This set of 50 observations is called a survey. The geolocation of each observation is recorded as soon as the accuracy estimated by the smartphone's GNSS receiver is under 10 m. The geolocation of the survey corresponds to the mean coordinates of the 50 observations. Only the survey coordinates are accessible to the users and considered in

this study. The output screen (Figure 34b) provides a summary for each survey of the proportion of each type of apex observed and the corresponding level of water restriction based on the calculated iG-Apex value. A more detailed description of the ApeX-Vigne application's characteristics can be found in Brunel et al. (2019).

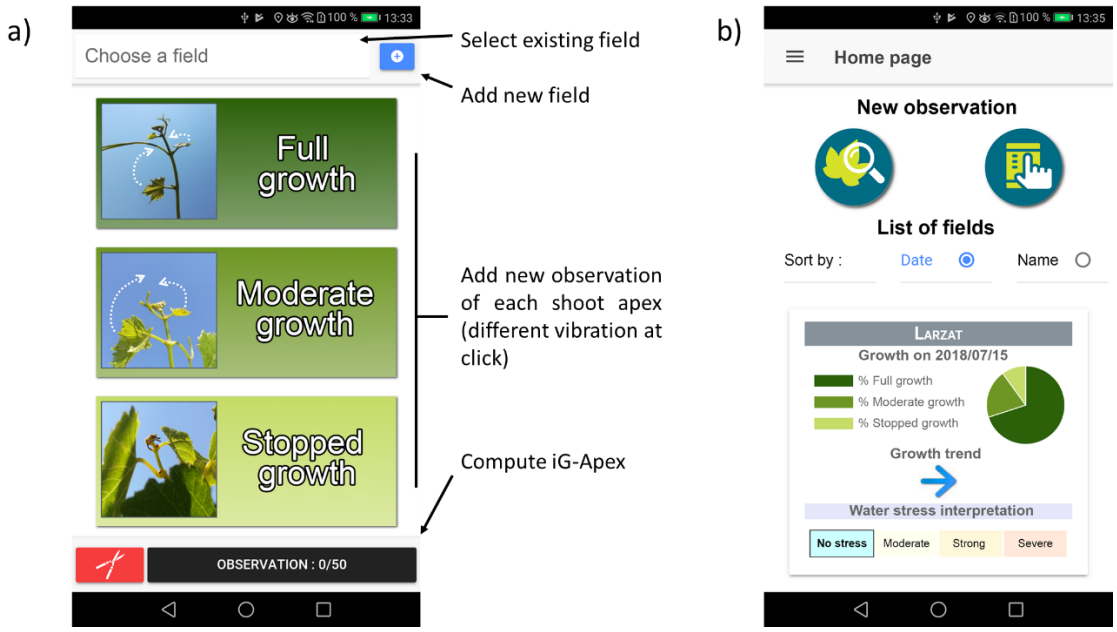


Figure 34: Screenshot of the two main interfaces of the ApeX-Vigne application: the input (a) and output (b) screens.

After a first test campaign in 2018 with approximately 20 beta users, the ApeX-Vigne application was officially launched and widely distributed in late May 2019 (week 21). A new version was released in early June 2020 (week 23). This paper focuses on data obtained from the ApeX-Vigne app over two seasons of use in 2019 and 2020. As the app was designed to assist with management associated with increasing water stress during the growing season, the key period for use in France is considered to be from early June to late August (weeks 23 to 35 in the calendar year)

6.3.4 Study zone:

The approach was tested in the south of France in an area encompassing the vineyards of the Languedoc, Provence and Côtes du Rhone viticulture regions (Figure 35a). The study zone covers an area of ~49 500 km^2 along the Mediterranean Sea (Figure 35a) and encompasses 57 different controlled denominations of origin with > 300 000 ha of vines and > 26 000 growers (Agreste, 2010) (Figure 35b). It includes a wide range of pedo-climatic conditions. The soils mostly have low soil water holding capacity (< 50 mm) (Figure 35c) and the mean annual precipitation over the last 30 years varies from relatively low values (500 – 700 mm yr^{-1}) in coastal areas to relatively high values (> 1 300 mm yr^{-1}) in the hinterland (Figure 35d).

This study zone is of interest because i) it is one of the largest contiguous wine producing areas in France, ii) the majority of the vineyards are non-irrigated and the climate is Mediterranean, i.e. characterized by hot, dry summers, so that the monitoring of vine water status is important for

most growers, which should favour the adoption of the ApeX-Vigne application, and iii) the pedo-climatic conditions are diverse and broadly representative of conditions that can be found in many vineyards around the world that are located in regions with a Mediterranean-type climate (e.g. west coast of the USA, southern Australia, western coast of South America and the northern and southern extremities of Africa).

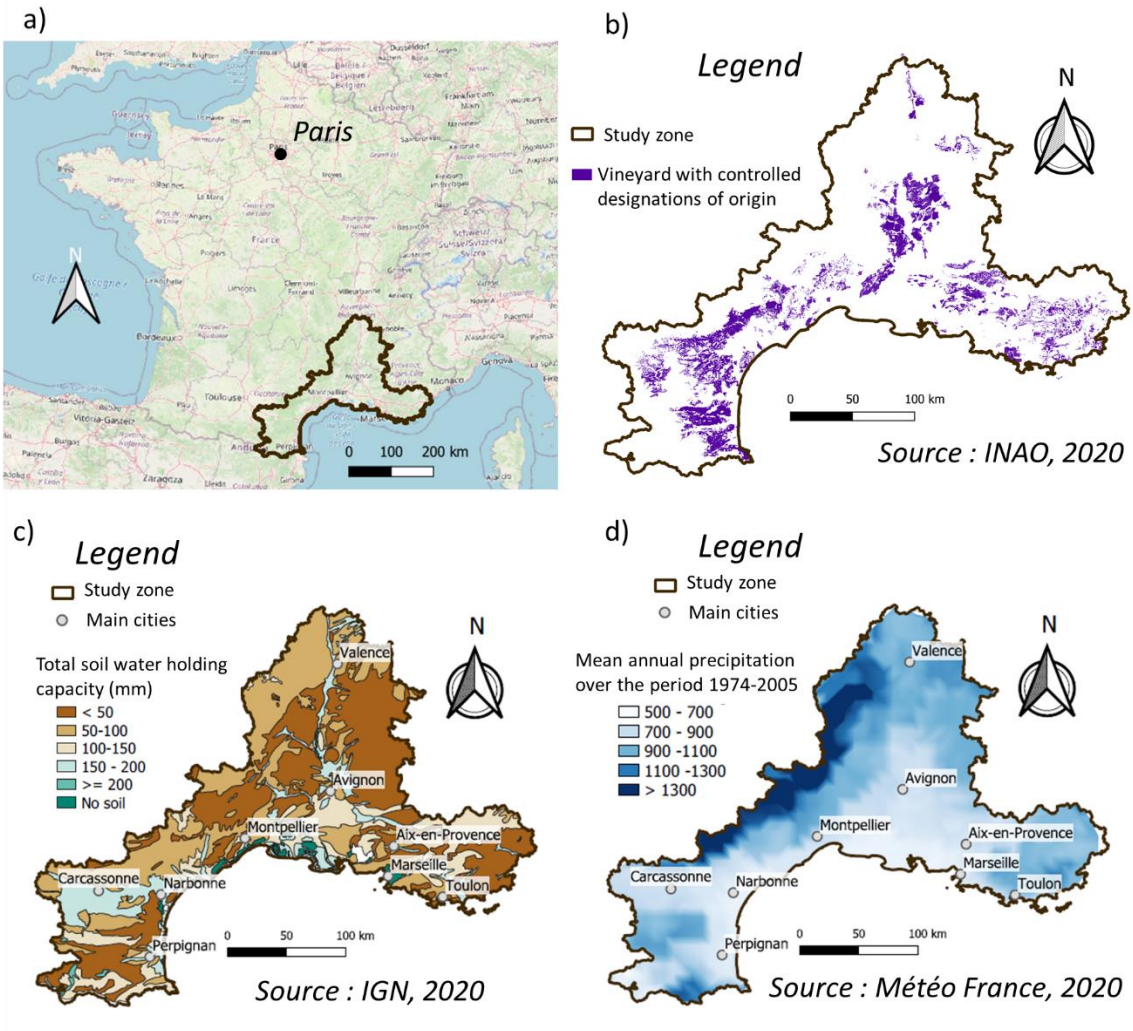


Figure 35: The study zone encompasses several wine-growing regions of importance in southern France with a large variability in soil and climate conditions. (a) The location of the study zone within France, (b) the location of vineyards within the study zone that have a controlled denomination of origin (INAO, 2020). Two principal pedo-climatic characteristics of this zone, (c) soil water holding capacity (IGN, 2020) and (d) mean annual precipitation over the last past 30 years (Météo France, 2020).

Temperatures and rainfall observed during 2019 and 2020 seasons were characteristic of a Mediterranean climate (Météo France, 2020). They are illustrated by the data measured at a meteorological station located in Montpellier in the centre of the study zone (Figure 36). In 2019, the spring was relatively cold and rainy and the summer hot and dry. Unusual heat waves with extreme high temperatures were observed during a few short episodes (around weeks 27-28)

(Figure 36a). In 2020, the spring was warm and relatively wet, while the summer was very dry, particularly during the month of July (weeks 26 to 30) (Figure 36b).

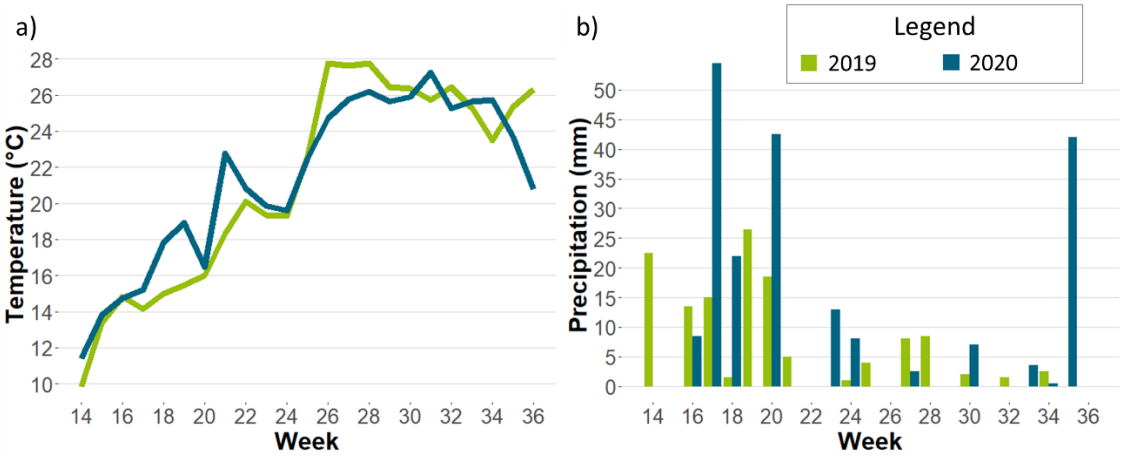


Figure 36: (a) Average weekly temperatures and (b) cumulative weekly rainfall measured during the 2019 and 2020 vintages at a meteorological station located in Montpellier, in the centre of the study zone.

6.3.5 Data Access

The application architecture offers several modes of data access. In its simplest form, each user is able to visualize, on their mobile device, the data collected on their own fields. As well as this individual-level access, the data are collected and centralized on a server database. This also makes it possible for the application administrators to download all the data collected by all users of the application in the form of a CSV file that also includes the user ID, field name, date and geographical coordinates in WGS 84 format.

The two different levels of data access (individual and centralized) permits analyses to be performed at different scales, from the field to the entire region. In both cases, global or local outliers may exist. For these analyses, data were simply filtered to ensure that only surveys collected within the study area and during the relevant time period were considered.

6.3.6 Considered data

The dataset collected during the 2019 and 2020 seasons was considered as a whole. The number and date of installations and uninstallations provided by the Google Play Store platform, in addition to the number and location of surveys provided by the centralized database, were used to characterise the adoption of the ApeX-Vigne application.

In a second step, some examples of data were selected from the dataset to illustrate how growers used the ApeX-Vigne application to answer specific questions that they had about the monitoring of vine water status at the within field and inter-field scale.

- The field used to illustrate monitoring at the within-field scale was chosen as it had a large enough number of performed within-field surveys (n) to model a semi-variogram (n

> 50) (Kerry and Oliver, 2007), and had surveys made on a frequent and regular basis throughout the season.

- The fields used to illustrate monitoring at the inter-field scale were chosen for being located in contrasting soil and climate conditions and for having regular surveys throughout the season.

In a third step, all the surveys carried out across the study zone in 2019 and 2020 were considered to illustrate the potential of the Apex-Vigne application to monitor vine water status at the regional scale.

6.3.7 Approach for evaluating the adoption of ApeX-Vigne

Four experts participated in six development cycles by providing feed-back and improving the application step by step. Previous research work has shown that the ApeX-Vigne application met the users' needs and that the proposed interface was appropriate (Brunel et al. 2019). The objective of this work was to go a step further and to characterise the adoption and uses of the application on the study zone using relevant indicators.

The first indicators were derived from the usage statistics provided by Google for any application uploaded to the Play Store and compared to references established by Liu et al. (2018) based on an analysis of 280 000 applications:

- An application with more than 50 downloads is part of what Liu et al. (2018) called the "popular app set", i.e. the most popular 3500 applications among the 280 000 studied.

- For applications that were part of this "popular app set", Liu et al. (2018) proposed the I/U ratio (Installed/Uninstalled) that represents the number of devices on which the app has been installed divided by the number of devices on which it has been uninstalled. In the Liu et al. (2018) study, the mean I/U ratio was 5.875. Applications with an I/U ratio > 6 were considered to be well adopted.

Although these values were established in a different country (China) and with another application store, Wandoujia (Wandoujia, China), they were assumed to be a reference due to the very large number of applications considered. No equivalent metric was found based on apps delivered via the Google Play or Apple stores.

The second indicator was the evolution of the number of surveys collected per week by all users. This indicator was calculated directly from the database on the ApeX-Vigne server. Its objective was to evaluate the actual use of the application. Finally, spatial distribution of the surveys was mapped to identify whether adoption (and usage) was consistent across the region or possible zones of preferred adoption (and usage).

6.3.8 Approach for illustrating the uses of ApeX-Vigne

Use cases of ApeX-Vigne were illustrated differently at within-field, inter-field and regional scales.

- At the within-field scale, the data were aggregated over time according to the week number in which they were collected. The mean and standard deviation of iG-Apex per

week were calculated and their evolution over time was represented. For the week with the highest iG-Apex standard deviation, the experimental semi-variogram was computed and an exponential model fitted (see Oliver and Webster (2015) for details of modelling and fitting variograms). A map of the iG-Apex interpolated by ordinary kriging was then produced in order to assess the potential interest of ApeX-Vigne app to delineate within-field zones of vine water restriction.

- At the inter-field scale, the evolution of iG-Apex over time for three selected fields were compared. The objective was to assess the potential interest of the ApeX-Vigne app to highlight potential differences in water restriction dynamics within a farm or a region and the potential to use the app data for benchmarking vineyards.
- At the regional scale, the data were also aggregated over time on a per week basis. The mean and standard deviation iG-Apex per week were calculated for both seasons 2019 and 2020. The iG-Apex of each week was compared between the two seasons with the Wilcoxon test (Rey and Neuhäuser, 2011). The spatial structure was again assessed for every week by calculating the experimental semi-variogram and fitting an exponential model. There were ~ 94 – 451 surveys for a given week to compute the semi-variogram. These values were reasonably close or larger than the limit (n > 100) usually reported in the literature to perform this kind of analysis (Kerry and Oliver, 2007). The semi-variogram at a regional scale was assumed to be relevant to highlight meso-scale auto-correlation (tens – hundreds of kilometres). It is commonly used at this spatial scale in soil science (Mishra et al. 2010) and climate studies (Vieira et al. 1997). Depending on the average distance between the surveys' sites, this scale may not highlight any auto-correlation phenomena at short distances, as is commonly used in precision agriculture. However, it remains an interesting approach to study the spatial auto-correlation of data at scales greater than the vineyard (meso- and macro-scale). The objective was to assess the suitability of the iG-Apex to highlight possible zones of different water restriction across the entire study zone.

6.3.9 Analyses and graphs

The different analyses were performed using R 3.6.1 (R Core Team, 2020). The semi-variograms were fitted with the gstat package using REML (Pebesma, 2004) and maps were produced using Qgis 3.4.8-Madeira (QGIS Development Team, 2020).

6.4 Results and discussion

6.4.1 ApeX-Vigne application adoption - downloads and surveys

In 2019, no applications were installed until week 21, corresponding to the official launch date (Figure 37a). A peak in the number of installations was observed in week 24 with more than 100 installations and then a gradual decrease to a value close to 0 around week 35, corresponding to the end of the observation period. The same seasonal dynamic was observed in 2020, with only a few installations per week until week 22, then a strong increase at the beginning of the observation period and finally a decrease to reach values close to 0 around week 35 again. This period

corresponded to mid-late summer when collection of this information is most relevant for decision-making. In total, the ApeX-Vigne app was downloaded 581 and 650 times during the 2019 and 2020 seasons respectively. This total of 1231 downloads (~5% of growers) was well above the threshold of 50 downloads set by Liu et al. (2018), indicating that the app was 'popular'. Note that between the end of the key period for use in 2019 and the beginning of the same period in 2020, the app was still downloaded (around 300 downloads). This result illustrated the curiosity of users for the application outside the target time period of use.

The number of uninstallations was relatively stable over the entire period with ~10 uninstallations per week (Figure 37a). A higher number of uninstallations was observed between weeks 18 to 22 of the 2020 season. This result is certainly explained by the release of a new version of the application that may have triggered uninstallations. The I/U ratio was close to 1 between the two key periods of use (from week 34 of the 2019 season to week 18 of the 2020 season). It was highest at the beginning of the two key period of use (weeks 18 to 26 in 2019 and 2020 seasons) ranging from two to > 10.

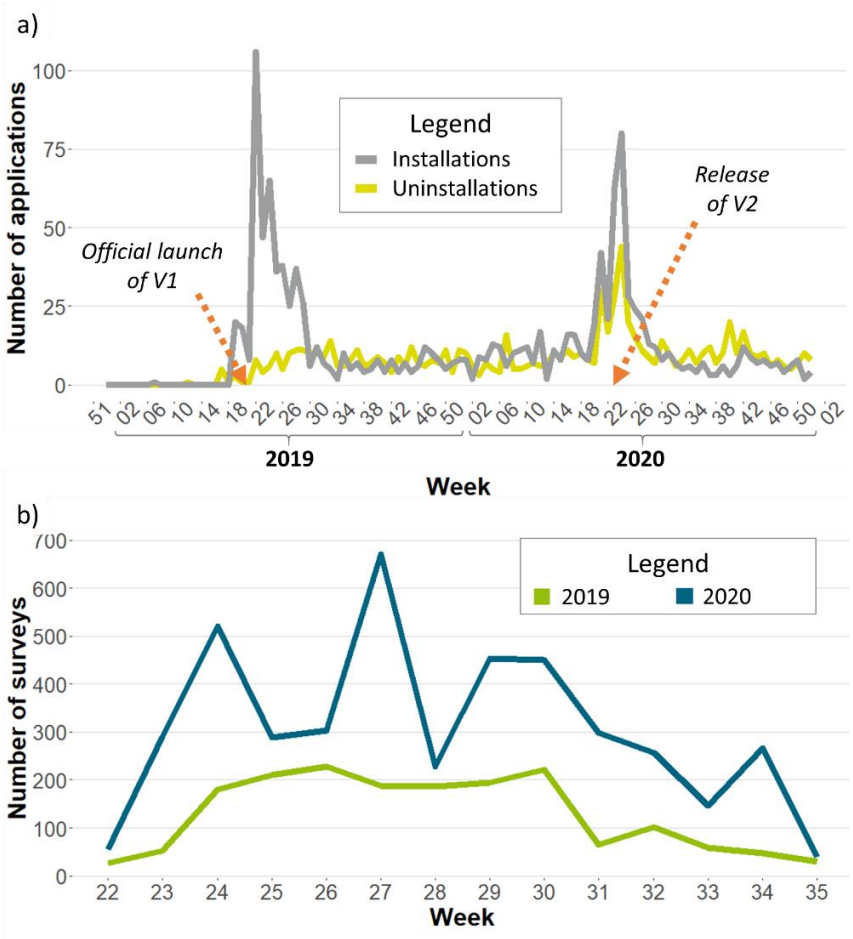


Figure 37: Basic statistics of the use of the ApeX-Vigne application. (a) Evolution of the number of installations and uninstallations during the 2019 and 2020 seasons (launched in week 21, 2019) and (b) number of surveys carried out per week over the key usage period (Weeks 22 – 35) of these two seasons.

The number of surveys per week (Figure 37b) increased as the expected period of water stress began (week 22) and remained fairly high during the key period from weeks 24/25 to weeks 30/31. Moreover, while the trend was similar between the two years, the number of surveys per week increased significantly between 2019 and 2020, with a respective total of ~2000 and 4500 surveys. The total number of fields (608 in 2019 and 902 in 2020) and the average number of surveys per field (3.3 and 4.9 respectively) also increased between the 2019 and 2020 seasons. Comparison of ApeX-Vigne's I/U ratio to standards established by Liu et al. (2018) suggested that the application was not well adopted by users (I/U < 6). On the contrary, the number of surveys collected showed a relatively high and growing interest from users. This contradiction can be explained by two characteristics of ApeX-Vigne that existing metrics do not take into account.

The first is that ApeX-Vigne is an application used in a professional setting. Therefore, potential users are much less numerous than for the mainstream applications studied by Liu et al. (2018). Also, although there are still very few studies on this subject, it seems that the drivers of adoption for mobile applications are not the same in the agricultural sector as in the general population (Michels et al. 2020). Users expect a professional application, such as ApeX-Vigne, to save them time or to improve confidence in their decision making. These expectations are higher than for those of a mainstream application. It is possible that these differences in target audience and expectations may influence installation and uninstallation dynamics and induce I/U ratio values different from the standards set by Liu et al. (2018). The second characteristic of ApeX-Vigne that is not taken into account by existing methods is its periodic nature. The period during which the iG-Apex method is relevant for decision support is limited to two months in the year. Considering a single threshold valid for the whole season does not take this into account.

These two characteristics of ApeX-Vigne may be common for a lot of mobile applications in agriculture. They are guided on the one hand by the stakeholders of the agricultural world who are mainly professionals, and on the other hand by the seasonality of agricultural production, which leads to a periodicity in the use of many agricultural applications. The poor consideration of these characteristics by existing reference metrics illustrates the need for a detailed and specific study to monitor adoption of agricultural applications accounting for specificities like seasonality, spatial distribution, etc.

6.4.2 ApeX-Vigne application adoption – Location of surveys

Figure 38a shows the number and spatial distribution of surveys performed by the users over the study zone in 2019 and 2020. It should be analysed in relation to Figure 35. The surveys were spread across all the winegrowing areas of the region, showing the interest in the app from a large range of viticulture stakeholders, whatever the region or controlled denomination of origin the vineyards belonged to. The density of surveys was higher closer to the Mediterranean Sea and on locations where annual mean precipitation was lower (Figure 35d), showing that the app seems logically more adopted in zones where the risk of having significant water restriction is the higher according to climatic characteristics. The density of surveys was particularly high in the northern part of the study zone and in several locations in the south-eastern part. It appeared that when the use of the app was recommended locally by early adopters, including private or cooperative advisors, the use of the ApeX-Vigne app appeared to be higher. The spatial disparity in adoption may also be explained by the fact that in some areas the iG-Apex method has already been

adopted by local professionals (before the launch of the ApeX-Vigne application), and this may have facilitated adoption of the app.

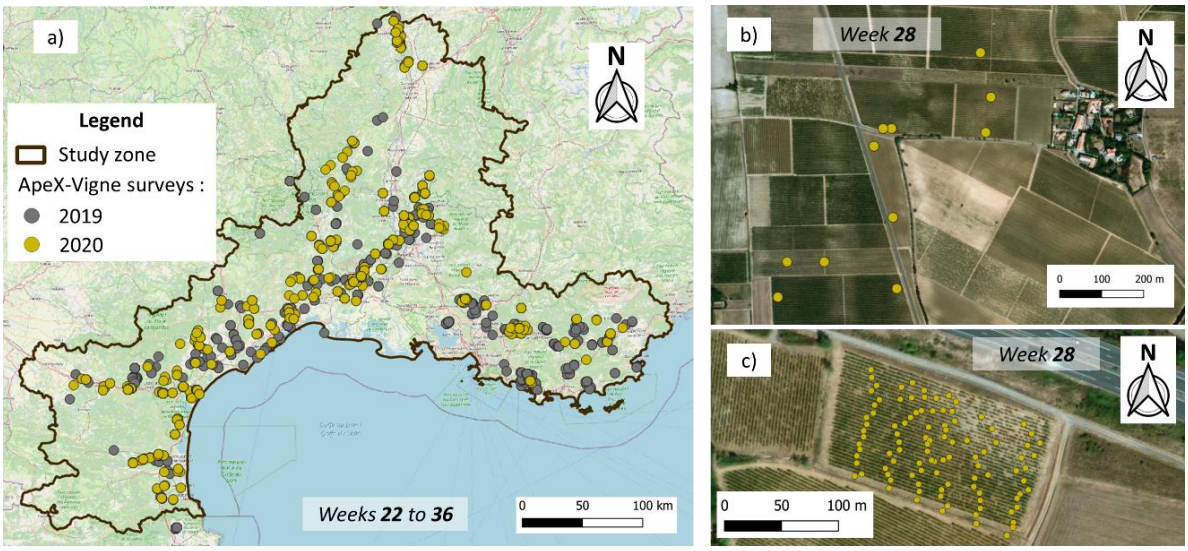


Figure 38: (a) Location of surveys made in the study zone using the ApeX-Vigne mobile application during weeks 22 to 36 of the 2019 and 2020 growing season. Local examples of (b) surveys made for a field scale monitoring and (c) for a within-field scale monitoring in week 28 of the 2020 growing season.

In most cases, users only made one survey (50+ observations) per field and per date (Figure 38b). This common sampling strategy aimed at monitoring the average of the field and characterizing inter-field differences. However, in some cases, several surveys for the same date were carried out within the same field (Figure 38c). The case highlighted here was rare, but indicated an interest among some users to use the application to characterize within field variability and to define potential within-field vine water status zones.

6.4.3 Opportunities for using the application at the within-field, field and study zone scales

Examples taken from ApeX-Vigne's centralized database illustrated how the application was used at different scales to address vineyard management issues.

The first example is the result of within-field scale vineyard monitoring. At this scale, the iG-Apex decreased from a value close to 1 in week 24 to a value approaching 0 in week 33 (Figure 39a). The standard deviation observed was initially very low, then increased between weeks 27 to 30 and become very low again at the end of the observation period. The standard deviation of the within field surveys was the highest around week 28, suggesting that at this time, water availability becomes limiting for the most sensitive zones in the field. The variogram and mapping of iG-Apex at week 28 (Figure 39b-c) showed that shoot growth within the vineyard was spatially structured. Analysis of the variogram indicated that the spatially structured variance was about twice the observed nugget (stochastic) variance (respectively 0.002 and 0.001). According to Martinez-De-Toda et al. (2010), a difference in iG-Apex results in differences in water restriction experienced by the vines. As a result, the characterization of this spatial structure made it possible to

understand the spatial distribution of vine's response to water restriction at the within-field scale and provided information to the grower or their advisor for decision making at this scale.

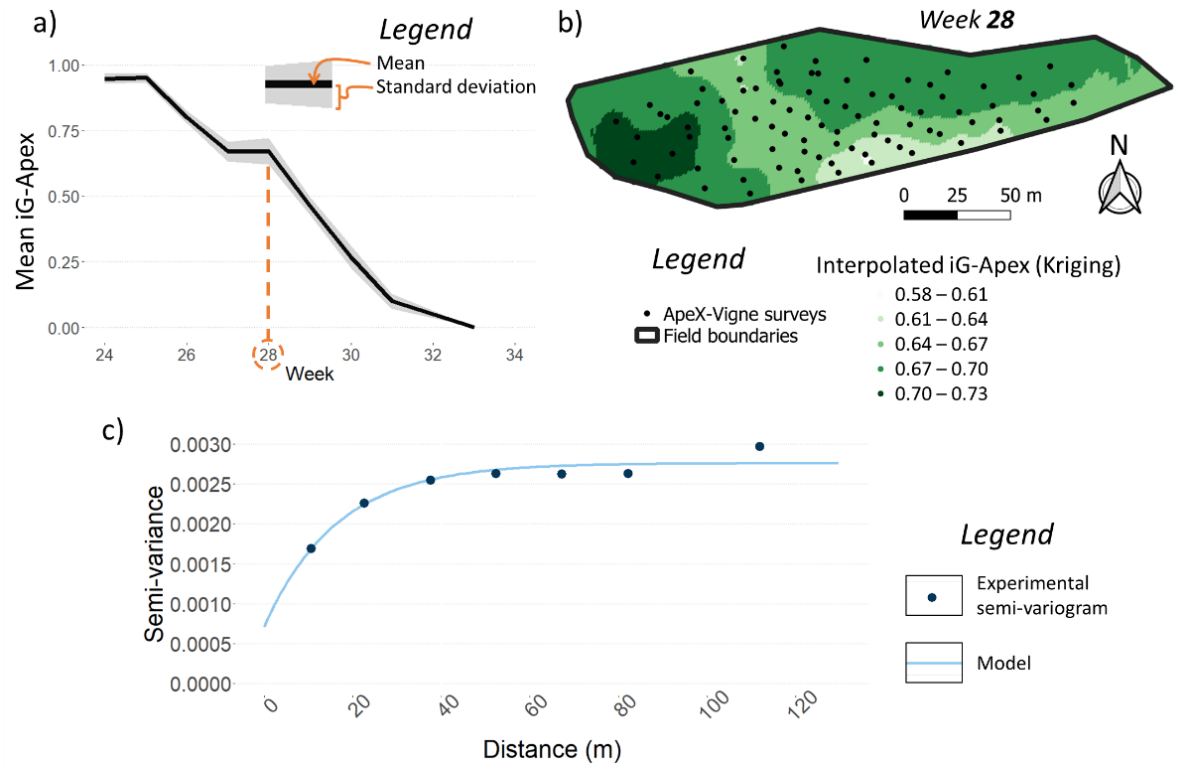


Figure 39: The ApeX-Vigne application was used by farmers and advisers for monitoring vines at the within-field scale: (a) Mean and standard deviation evolutions for the index of Growing Apex (iG-Apex) collected at the within-field scale in 2020 by a user of the ApeX-Vigne application, (b) example of an iG-Apex kriged map obtained from data collected in week 28, 2020 and, (c) the corresponding iG-Apex semi-variogram at this stage.

The second example illustrated vineyard monitoring at an inter-field scale. At this scale the temporal data collected permitted the 2020 season's shoot growth dynamics to be observed and compared for three selected fields (Figure 40). Field 1 had the latest and slowest decrease in iG-Apex, indicating that water restriction may have been less severe in this vineyard or at least that the water restriction occurred more slowly over the season. For Field 2, the iG-Apex started decreasing very early in the season, highlighting a very early water restriction. Field 3 had an intermediate behaviour, iG-Apex started decreasing later than Field 2; however, it showed the sharpest slope of decrease in iG-Apex indicating that once water restriction started, it proceeded at a very fast rate.

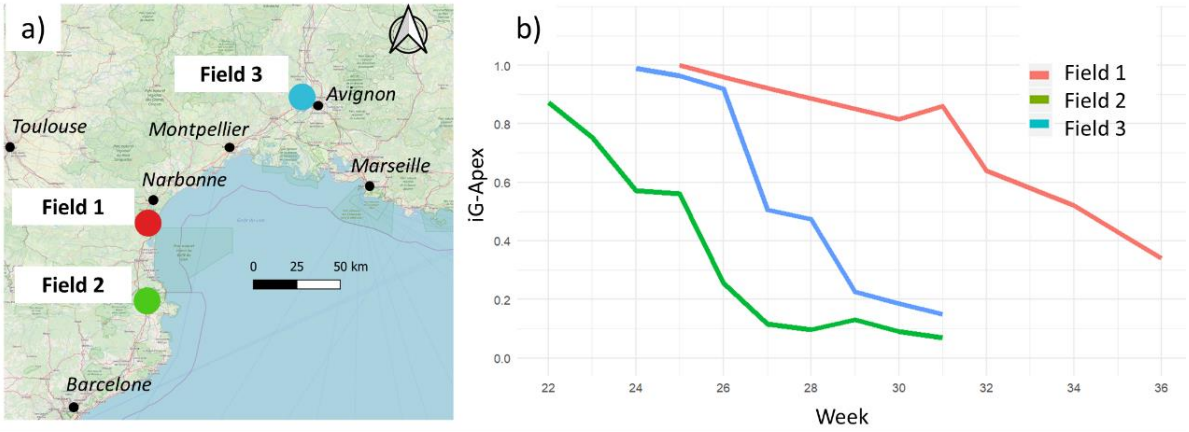


Figure 40: The ApeX-Vigne application as used by farmers and advisers for monitoring the evolution of vine water status at the field scale. (a) Examples from three individual fields in different wine-growing regions of the study zone (b) showing the evolution of the index of Growing Apex (iG-Apex) score from week 22 to week 36 in 2020.

In these Mediterranean climate conditions, and according to Martinez-De-Toda et al. (2010), the differences in Figure 40 indicated a greater or lesser sensitivity of the three fields to water restriction. It illustrated a potential to establish a benchmarking option for the app so that this information could be used by a farmer or an advisor to understand different fields dynamics at the vineyard, the denomination or the region scale. This information would be useful for operational decision support, for example by identifying groups of homogeneously functioning fields. During the season, analyses such as these will make it possible for growers to target vineyards or within-vineyard zones where observations need to be made with more spatial and temporal precision. This may permit them to better deploy more precise but expensive and laborious measurements, e.g. on sap flow or leaf water potentials, to monitor more precisely the vine water status and to target management interventions.

At the study zone scale, the collection of a large number of surveys and their synchronization on a centralized server made it possible to consider the understanding of the phenomenon over a large scale. As an illustration, the results presented in Figure 41 consider the changes of mean and standard deviation for the whole study zone dataset over the 2019 and 2020 seasons.

At the beginning of the period of interest (week 24), the iG-Apex was high over the entire study zone in both seasons (Figure 41). The iG-Apex then decreased towards values close to 0 around week 35. From week 24 to week 27 and from week 29 to week 35 the average iG-Apex was significantly lower in 2020 than in 2019. It is worth noting that week 28 was an exception in this global dynamic with the same average iG-Apex for both seasons (Figure 41). An event interfering with the shoot growth stop dynamics had apparently occurred around this period during one or both of the seasons.

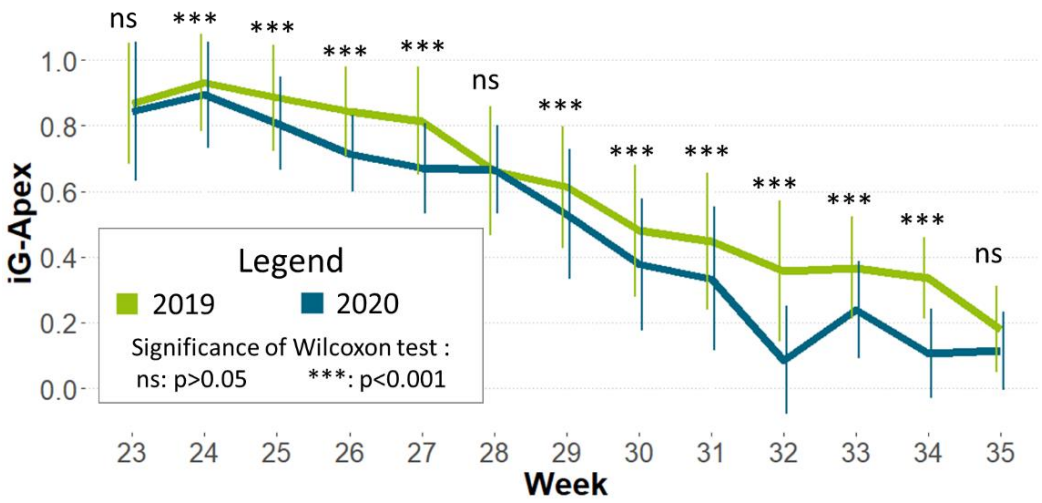


Figure 41: iG-Apex mean and standard deviation per week over the study zone during the 2019 and 2020 seasons.

These results indicated that between weeks 24 and 35, shoot growth gradually stopped over the study area. Over the same period, very little rainfall and relatively high temperatures were observed. This shoot growth stop dynamic can be explained by the gradual appearance of water restriction. These results are consistent with the Mediterranean climate of the study zone that Schultz (2017) characterized as a water-limited region. According to Schultz (2017), during the summer period, water constraints appear progressively because the evaporative demand largely exceeds precipitation. These results were also consistent with other research that have studied the shoot growth stop when the vine is exposed to water restriction, in both experimental (Pellegrino et al. 2005) and real conditions (Martinez-De-Toda et al. 2010)

The results here also indicated that the shoot growth stop was earlier in the 2020 season. Numerous papers have shown that high temperatures tend to accelerate vine phenology (Parker et al. 2020; Yzarra et al. 2015). Even though the design of this work did not allow for a formal interpretation of these results, it is possible that the higher temperatures in the spring of the 2020 season accelerated phenology and favoured earlier shoot growth stop that year. The particularly hot temperatures observed around weeks 27-28 of the 2019 season may have played a role in the same values of iG-Apex observed in week 28 for both seasons. Although the consequences on vine physiology of such temperatures observed so early in the season are still poorly documented, Ben Salem-Fnayou et al. (2011) showed that they have consequences on cell structure and photosynthesis that could be equivalent to those of a drought. Therefore, it is possible that these phenomena caused a sudden shoot growth stop in part of the study zone and that this resulted in a general decrease of iG-Apex.

At the study zone scale, the data collected with the ApeX-Vigne application helped to understand the evolution of the variance in iG-Apex values (Figure 42). In both seasons, the total variance was at first relatively small (around 0.02 for weeks 24-26), then it increased (around 0.04 to 0.06 for weeks 28-30) and remained at relatively high levels (around 0.04 for weeks 30-32). The spatial structure of this variance also evolved throughout the period of interest for both seasons but with a time lag of around two weeks between 2019 and 2020. Figure 42 shows the four semi-variograms that best illustrate this evolution of the spatial structure for each season. Considering the time

lag, weeks 24 to 30 and 26 to 32 were chosen for the 2019 and 2020 seasons respectively. The spatial structure was initially relatively low, then increased to reach its highest level at weeks 28 and 30 in 2019 and 2020 respectively, and then decreased again at the end of the period of interest. The semi-variograms computed for week 28 of 2019 season and for week 30 of 2020 season showed the strongest spatial auto-correlation in the iG-Apex data (Figure 42). Both phenomena were second order stationary and had exponential semi-variogram models with respective ranges of 21.8 km and 14.9 km and a nugget-to-sill ratio of ~0.13.

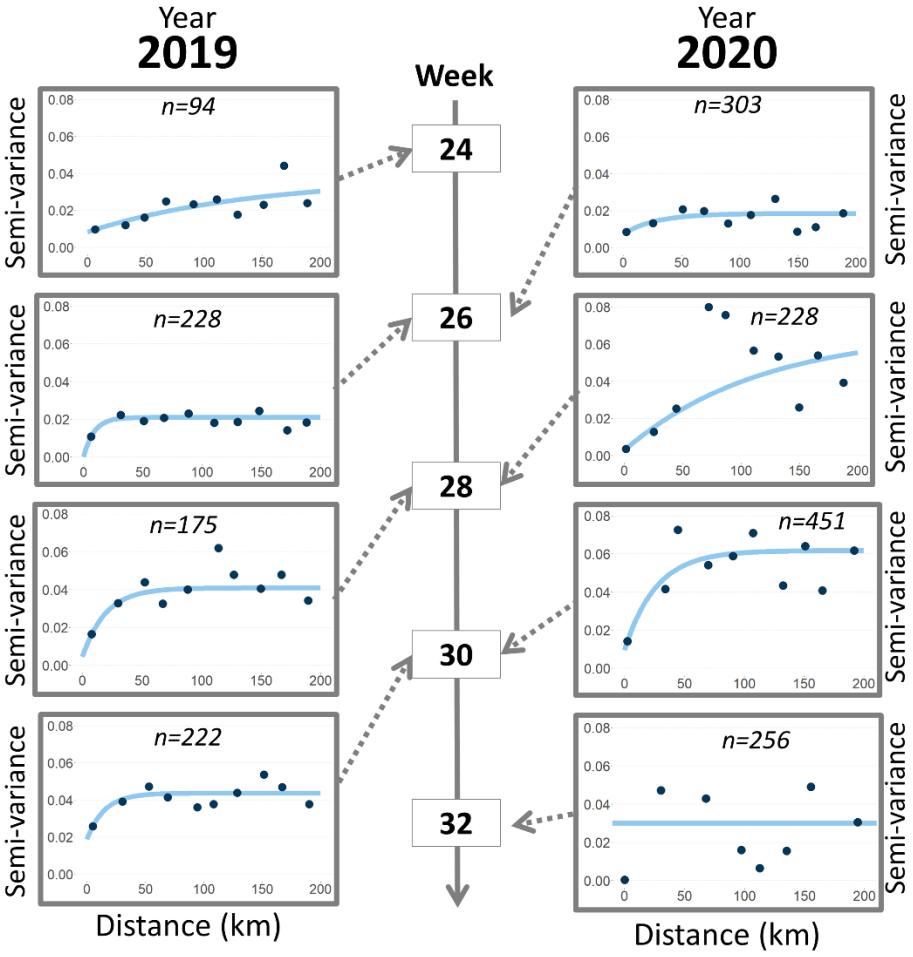


Figure 42: Fitted variograms of weekly iG-Apex data illustrating the evolution over time of the spatial structure in these data for the 2019 and 2020 seasons. Arrows indicate the shift associated with an apparent two-week time-lag in response between the two seasons.

Available data did not allow a detailed interpretation of these dynamics, but some hypotheses for interpreting them may be formulated based on the different known sources of variation in the vine water status:

- At the beginning of the period of interest (weeks 24-26), iG-Apex data ranged from 0.8 and 1 (Figure 42) indicating that all vineyards were still actively growing. The low spatial structure at these dates may be interpreted by the fact that no spatially organised factor was limiting shoot growth.

- The stronger spatial structure observed in weeks 28 (2019 season) and 30 (2020 season) corresponded to a period where mean iG-Apex was between 0.4 and 0.6 for both years. These iG-Apex mean values indicated that there was a factor(s) limiting shoot growth and the observed spatial structure indicated that this limiting factor was spatially organised with a range of 15 to 20 km. Factors that may limit shoot growth and that are spatially organised include soil (Van Leeuwen et al. 2018), topography (Santesteban et al. 2012) or local climate conditions (Deloire et al. 2004). It can be hypothesised that the spatial structure observed in the iG-Apex data at these dates was related to the spatial structure of one or a combination of several of these factors. Note that nugget effects at regional scale were respectively around 0.006 and 0.007. Most of the surveys collected using the ApeX-Vigne application corresponded to the sampling of 50 observations spread over a field. The observed nugget effect may be partly explained by the within-field variability of shoot growth. According to the example shown in Figure 39c, this within-field variability is ~0.003. The rest of the nugget effect, ~0.003 to 0.004, can be explained by factors that may influence shoot growth at the inter-field scale over very short distances, such as cultivar (Bota et al. 2001), row orientation (Hunter et al. 2020) or short range variability of soil characteristics (Reynolds and Willwerth 2020).
- At the end of the period of interest (weeks 30-32), the spatial structure exhibited a lower nugget-to-sill ratio (0.52) in 2019 and had almost disappeared in 2020 (Figure 42). During this period, mean iG-Apex ranged from 0.1 to 0.4 (Figure 42). This result may be interpreted by the saturation of the iG-Apex. When this index reached a value of 0, it did not evolve anymore, even if the water restriction kept on increasing because of summer water deficit. This limitation in the iG-Apex itself may have masked partially or totally the observed spatial structure of the phenomenon.

These results showed that the aggregation of observations collected by ApeX-Vigne may provide access to a new source of spatially structured information at regional scales. It is likely that in the coming years, with the high level of smartphone equipment among farmers (Michels et al. 2020) and their interest in collective monitoring approaches (Dehnen-Schmutz et al. 2016), the ApeX-Vigne application will be adopted more widely. A wider adoption of the ApeX-Vigne application should make it possible to study the spatial variability of vine water status at even more scales, including intermediate scales of a few kilometres to a few tens of kilometres. This spatial scale is particularly interesting because it corresponds to a scale of collective vineyard management (cooperatives, organisations of controlled denomination of origin, etc.) (Leibar et al. 2018). The proposed approach may be relevant as it would make it possible to identify areas with homogeneous behaviour and areas that are particularly prone to high water restriction. Studying the temporal evolution of a simple indicator based on the nugget-to-sill ratio seems relevant to define the best date to meet this objective. This information may be important to support the logistics of agriculture water supply across a whole region. In the longer term, this information could also support vineyard adaptation strategies to climate change at regional scale, such as the choice of the most suitable sites or cultivars (Naulleau et al. 2021).

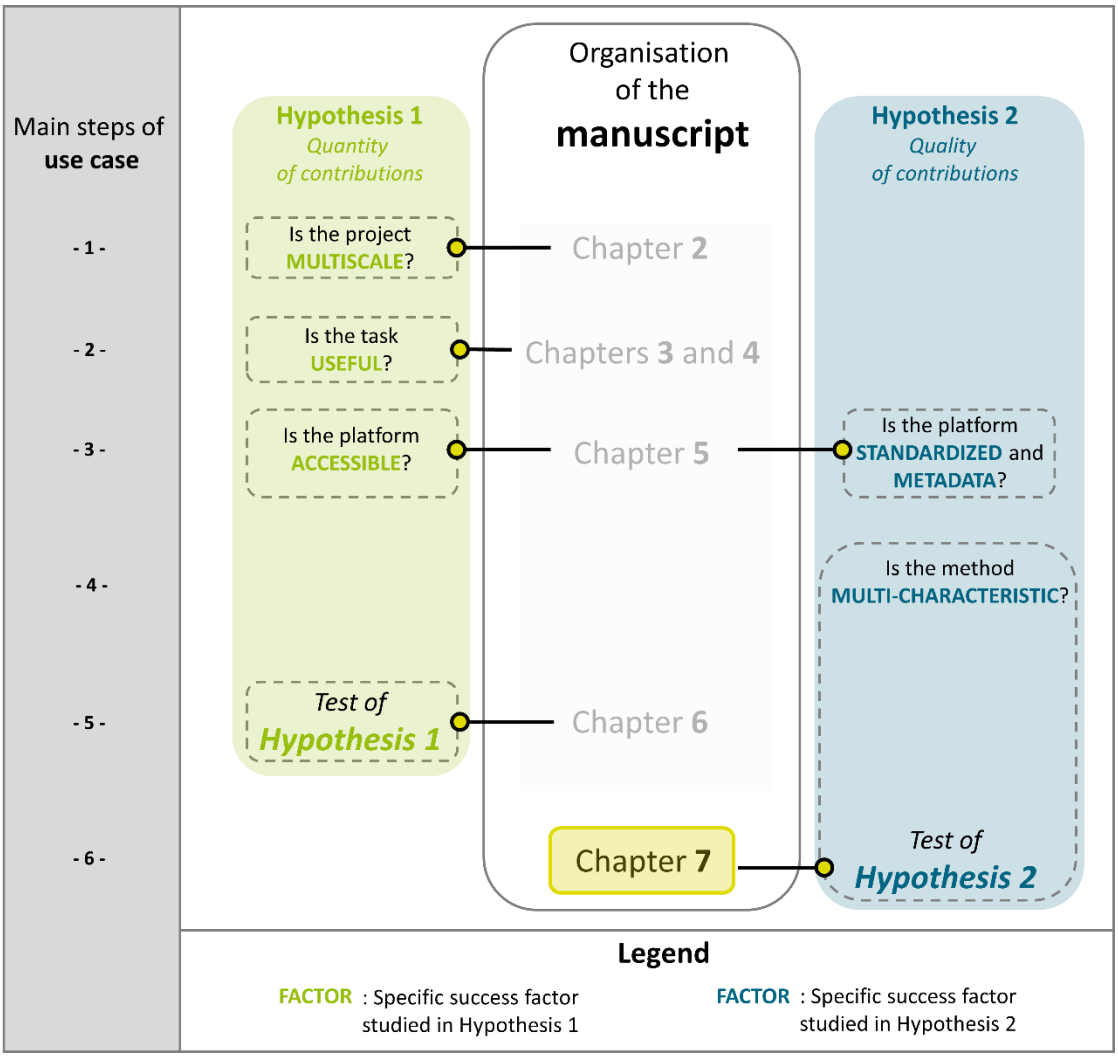
Nevertheless, the use of ApeX-Vigne data for decision making may raise issues of data quality as reported in literature for projects involving a wide variety of operators in gathering observations (Senaratne et al. 2016). The first major issue is data quality evaluation (Touya et al. 2017). This

challenge will be all the more important since, at these scales, finding reference data may be an issue (Severinsen et al. 2019). To assess its quality, it will be necessary to propose approaches that rely solely on the characteristics of the Apex-Vigne data, such as the spatio-temporal consistency of the collected data (Senaratne et al. 2016) or the trustworthiness of the operator (Fogliaroni et al. 2018). The second issue is data quality improvement (Goodchild and Li 2012). Ancillary data, such as sources of reference information (e.g. sap flow or pressure chamber) or high spatial resolution information (e.g. Sentinel 2 satellite imagery), may offer an opportunity to improve the quality of the collected data. However, this will require new approaches to be defined for the conjoint use of these sources of information with various characteristics (quality, spatial and temporal resolution). Both issues should benefit from further research.

6.5 Conclusion

The ApeX-Vigne mobile application was developed to monitor shoot growth and estimate vine water restriction using a simple and robust method that can be used by stakeholders with very different skills. This study presented the results obtained after two seasons of use of this application in the south of France. It showed that the ApeX-Vigne application was downloaded and used on a seasonal basis. The number of downloads indicated that it was a rather popular application in southern France with a growing number of users and an increasing number of surveys from 2019 and 2020. The study also highlighted the limitations of metrics as proposed in the literature to assess mobile application adoption in a seasonal context, as is often encountered in agriculture. The development of metrics that account for the seasonality of use and the specific behaviour of professional users is still an issue and may be the subject for future research. Through real examples, the study has shown the different scales that the ApeX-Vigne application can be applied to: i) at the within-field scale to identify potential management zones with differences in vine water restriction and ii) at the inter-field scale to compare and benchmark shoot growth dynamics and vine water restriction within the same farm or region. Although the use of the ApeX-Vine application at both scales is possible, inter-field use is the most frequently implemented by farmers and their advisers. This study has also shown that the regional aggregation of observations collected by ApeX-Vigne provided access to spatially structured information. It is likely that this may provide information on pedo-climatic phenomena influencing water restriction at this regional scale.

Chapter 7: Evaluating the quality of contributions in a farmsourcing project of spatial observations



The first objective of chapter 7 is to explore an ex post quality management strategy based on attribute, spatial and temporal characteristics of observations (MULTI-CHARACTERISTIC). The second objective is to test hypothesis 2 in order to generalise conclusions about the specific success factors concerning the quality of contributions. The questions explored in this chapter are: How to propose a method based on the attribute, spatial and temporal characteristics of observations to automatically identify outliers and surprising observations in a FSO dataset? How well does this proposed approach perform?

7.1 Details about the corresponding paper

7.1.1 Title of the paper

This chapter will be submitted as a full paper in the journal *Computers and Electronics in Agriculture* under the title "Automatic method for detecting outliers and surprising observations in farmsourcing datasets".

7.1.2 Authors:

Léo Pichon[1], Stefano Leggerio[12], Bruno Tisseyre[1]

[1]*ITAP, Univ. of Montpellier, Institut Agro Montpellier, INRAE, Montpellier, France*

[2]*Università Degli Studi Di Torino, Torino, Italy*

7.1.3 Abstract

The identification of outliers and surprising observations is a major issue for ensuring the quality of data collected in crowdsourcing projects of spatial observations. In agriculture, these projects have particular characteristics that have led some authors to coin the term farmsourcing. In farmsourcing of spatial observations (FSO) projects, the phenomena studied are dynamic and seasonal and the participants are mainly professionals. These specificities impact the implementation of existing methods to identify outliers and surprising observations. The objective of this paper is i) to propose an approach that can be automated and adapted to the specificities of FSO projects and ii) to implement this approach on the real use case in order to evaluate its performances. The studied use case is the ApeX-Vigne project, which aims at monitoring vine water status at regional scale based on shoot growth observations. The proposed approach considers attribute, spatial and temporal characteristics of the observations. It identifies the cluster corresponding to good quality observations by studying the density with a density-based clustering approach (DBSCAN). Observations which do not belong to this cluster are labelled as outliers or surprising observations. The comparison of results with labelling done by experts showed that this approach is robust with a good specificity and a good sensitivity to outliers. The approach was able to automatically identify surprising observations. These surprising observations were corresponding to outliers not previously identified or to fields that were particularly early or late compared to their spatio-temporal neighbourhood. These results show that the approach can be generalised to the identification of outliers and surprising observations in all FSO projects.

7.1.4 Keywords

Agriculture, DBSCAN, Crowdsourcing, Density-based approach, Spatio-temporal

7.2 Introduction

Crowdsourcing is an approach to answering a question defined by an organisation (research laboratory, company, etc.) that relies on the collective intelligence of a community of contributors (Brabham 2008). Over the last decade, crowdsourcing projects have been developed in fields as diverse as astronomy (Raddick et al. 2010), medicine (Paulsgrove et al. 2021) or indoor localisation (Zhou et al. 2018). Depending on the scientific communities and the characteristics of the projects (passive or active contribution; spatial or aspatial data), they have been described by different keywords such as 'Volunteered Geographic Information', 'Citizen science' or 'Participatory sensing' (See et al. 2016). Among crowdsourcing projects, Minet et al. (2017) defined three types according to the nature of the contributions: knowledge gathering (Bruce, 2016), task allocation (Fritz et al. 2009) and collection of spatial observations by sensors (Kuhlgert et al. 2016) or operators (Van Etten, Beza, et al. 2019). This last type of crowdsourcing, which can be grouped under the term crowdsourcing of spatial observations, has been particularly developed in the field of environmental monitoring, where numerous research projects have been carried out, often with the keyword 'Citizen science' (Conrad and Hilchey 2011). This is the case, for example, for dynamic monitoring of the spatial distribution of plants or animal species (Sullivan et al. 2009).

Crowdsourcing projects of spatial observations are effective when they mobilise a large number of contributors to make it possible to collect a large volume of data. This provides an opportunity for scientists and land managers to monitor dynamic and spatially structured phenomena that would be difficult or even impossible to monitor by other means. In agriculture, the crowdsourcing of spatial observations is currently less developed than in the field of environmental monitoring (Minet et al. 2017), but it presents interesting opportunities. Many agricultural issues need to be monitored and controlled over spatial areas larger than the field or farm, ranging from a few kilometres to sometimes several hundred kilometres. This is the case, for example, for monitoring diseases (Trivellone et al. 2016), pests (Malek et al. 2018) or monitoring abiotic stresses, such as the effect of heavy rainfall or extreme temperatures (Reinermann et al. 2019). These events are generally dynamic and the period during which they must be observed in order to adapt to them is often short. Moreover, they are conditioned by environmental factors that are not randomly distributed (topography, geology, pedology, etc.). Therefore, the phenomena of interest are often spatially structured on scales that encompass several hundreds, tens or even thousands of fields, which often makes the monitoring of these phenomena a collective issue. As a result, crowdsourcing of spatial observations is a particularly suitable approach for the dynamic and spatial monitoring of many agricultural phenomena.

Nevertheless, the contribution of many participants with different levels of involvement and expertise is a double-edged sword because it also reduces the quality of the collected data (Chittilappilly et al. 2016). Managing data quality is a major issue for crowdsourcing projects of spatial observations and for the entire processing chain that uses these data. This issue has been widely addressed by the scientific community (Wiggins et al. 2011), particularly for the classification of surprising or unexpected observations. These observations have received particular attention as they can be in some cases outliers that need to be eliminated or, alternatively, informative observations characterising a new phenomenon, such as the colonisation of a new habitat by a species (Balčiauskas, 2020). In the literature, existing methods generally follow a

common approach of measuring a distance between an observation and a reference and then considering the observation as surprising if this distance is above a certain threshold (Li et al. 2020). However, the way in which the distance, reference and threshold are defined depends strongly on the application under consideration and on the characteristics of the resulting dataset.

The specificities of crowdsourcing projects of spatial observations in agriculture impact the implementation of these methods. These specificities led Minet et al. (2017) to coin the term 'farmsourcing' to describe these projects. First of all, in farmsourcing projects of spatial observations (FSO), the observed phenomena are dynamic and often seasonal (Maistrello et al. 2018). The range of dates during which an event can be observed (development stage of a plant or its pests, appearance of symptoms, etc.) is often very limited. This characteristic is likely to lead to the appearance of new types of outliers corresponding to values that are unlikely to be observed at given time periods. Secondly, the participants in FSO projects are almost exclusively professionals from the agricultural sector (Minet et al., 2017). Their level of expertise and the factors motivating their participation are different from those of amateurs and the general public, who are frequently involved in environmental monitoring projects (Van de Gevel et al. 2020). These differences are likely to influence the distribution of observations in time and space as well as the metrological quality of the measurement itself. Depending on the case, this influence may or may not be favourable to the quality of the final dataset, but very little work has been done to date on this issue.

FSO projects are beginning to emerge with some research work mentioning the development of tools for collecting this type of data. This is the case, for example, for the regional monitoring of septoria in wheat (Michel et al. 2016) or vine water status (See Chapter 5:). These tools offer new opportunities for monitoring dynamic and spatially structured agricultural phenomena. Nevertheless, as for environmental monitoring projects, their use for decision support requires the implementation of adapted pre-processing methods in order to assess the quality of the data and to identify possible surprising observations. Moreover, the phenomena under study are generally of short duration. Therefore, many observations are collected in a short period of time and their analysis must be carried out quickly in order to be used for decision support. As a result of this limitation, automation will be a key aspect of any new proposed methods for analysing FSO data. However, to date there has been no published work that has focused on the implementation and automation of such methods that are capable of adapting to the characteristics of FSO datasets.

To address this knowledge gap, the objective of this paper is i) to clearly pose a problem statement for developing quality control metrics for FSO data, ii) to briefly review the existing approaches to outlier and surprising observation detection in these data, iii) to propose an automatable method to classify surprising observations in FSO datasets, iv) to implement this method to identify outliers and surprising observations on the real use case of the ApeX-Vigne mobile application by simultaneously taking into account the attribute, spatial and temporal components of the observations, and v) to evaluate the performances of this method.

7.3 Material and methods

7.3.1 Problem statement

7.3.1.1 *Farmsourcing of spatial observations: Characteristics of datasets*

FSO datasets have some characteristics that are common to every crowdsourcing projects and some specific characteristics. It is important to formalise the specific characteristics of FSO projects as they will guide the data processing methods to be implemented.

- *Phenomena are structured in time and space*: Phenomena studied, such as the presence of diseases, pests or abiotic stresses are generally structured in time and space (Malek et al., 2018). This means that two observations collected on close dates and at close sites have a higher probability of having close attribute values than two observations that are distant in time and space. As a result, the datasets collected for the monitoring of these phenomena show strong temporal and spatial covariances. However, the characteristics of these covariances and their variations in space and time are not known a priori.

- *Phenomena are dynamic and seasonal*: Studied phenomena generally evolve over time. Most of them depend on climatic conditions. Therefore, they tend to evolve seasonally and to be observable only at certain times of the year (e.g. Maistrello et al. 2018). Participants trying to observe these phenomena are gathering observations during these periods. As a result, the number of observations varies over time.

- *Observations are only made in restricted areas*: Observations are only made on agricultural fields. Phenomena under study can only be observed in restricted areas that correspond to the spatial distribution of the crops under consideration.

- *The number of observations is important*: FSO projects generate a large number of observations. In order to be used for decision support, it is necessary that the approaches developed to analyse these data be automatable.

- *Data are asynchronous and heterotopic*: Observations are collected at sites and dates that are only determined by participants according to their individual constraints. These sites and dates are not known *a priori* and their distribution is not necessarily optimal for characterising the phenomenon of interest. The resulting datasets are asynchronous, i.e. observations made on two different sites are generally made on two different dates, and conversely, they are heterotopic, i.e. observations made on two different dates are generally located on two different sites.

- *Data are mostly of good quality*: Making an observation requires a significant effort for the operators. They generally make every effort to carry out observations that are useful to them and allow them to effectively monitor the phenomenon of interest. As a result, it can be hypothesised that the vast majority of observations collected are of good quality (Mehdipoor et al. 2015).

- *Reference data are difficult to obtain*: The phenomena under study are often dynamic. Reference data that are available in some crowdsourcing projects, such as state data

(Arsanjani et al. 2015), would be extremely costly to acquire for FSO projects over large areas. Therefore, FSO projects rarely have reference data (or if they are available, they are only available very locally). As a result, the observations made by the participants are in the vast majority of cases the only data available to describe the phenomena under study (Sullivan et al., 2009).

7.3.1.2 *Outliers and surprising observations: Existing approaches*

Approaches for classifying outliers and surprising observations in crowdsourcing datasets of spatial observations have been widely explored in the literature (Senaratne et al., 2016). Figure 43 provides a general scheme that summarises the main steps. Each observation is assessed for an indication (distance) of being an outlier. If this indicator does not meet the defined condition (Condition 1), the observation is considered as an outlier. Then, the level of surprise of the remaining observations (non-outliers) is evaluated and the second condition (Condition 2) is tested. If this condition is not met, the observation is considered as a surprising observation. These surprising observations are usually evaluated by experts, who then classify them as either outliers or as surprising observations (Sullivan et al. 2009).

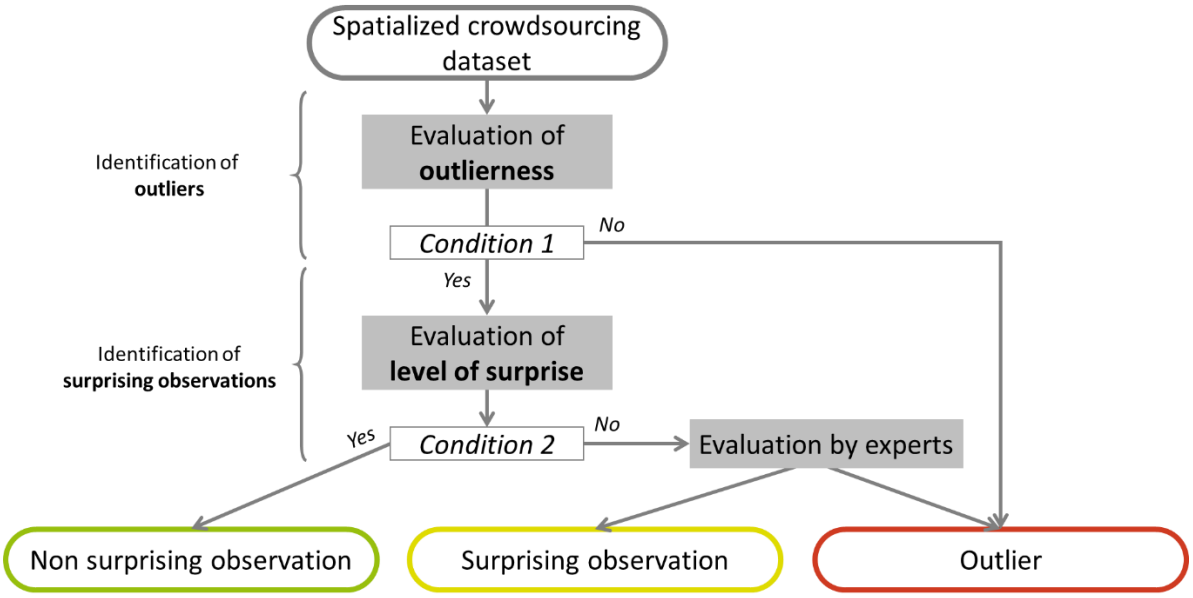


Figure 43: Main steps of existing approaches for classifying outliers and surprising observations in spatialized crowdsourcing datasets.

The assessment of outliers, and the selection of a threshold value, is often done in a simple way. This step is seen by different authors as a pre-processing step (Li et al., 2020). The research work focuses mainly on the evaluation of the level of surprise. This is generally evaluated by comparing attribute values of the observations with expected attribute values. The comparison is sometimes carried out according to one attribute (Li et al., 2020) or according to several (Schlieder and Yanenko 2010). Expected values are often estimated based on the correlation between the variable of interest and auxiliary variables (Mehdipoor et al. 2015) or on the spatial and/or temporal autocorrelation of this variable (Simoes and Peterson 2018). Condition 2 is often defined by a threshold level of surprise beyond which it is considered that the observation does

not behave like the others. This threshold can be set by experts (Bonter and Cooper 2012) or calculated by automatic methods (Mehdipoor et al. 2015). This global approach is common to the majority of existing methods, but the implementation varies according to the application considered and the characteristics of the dataset. The approach developed in this study is based on this framework, widely validated in the literature, and proposes to adapt it to the specificities of FSO projects.

7.3.2 Formalization of the approach

7.3.2.1 *Definitions and notations*

The regional monitoring of a dynamic and spatially structured agricultural phenomena by FSO can be seen as the sampling of a random variable Z auto-correlated in time and space. The attribute value observed at site s_i and at date t_j can be noted $z(s_i, t_j)$. The set of these observations, i.e. the set of pairs (s_i, t_j) for which $z(s_i, t_j)$ exists can be noted $\mathcal{ST}$. Among the observations belonging to $\mathcal{ST}$, those collected at sites close to s_i belong to its spatial neighbourhood, noted $\mathcal{SN}_i$. Those collected at dates near t_j belong to its temporal neighbourhood, noted $\mathcal{TN}_j$. The seasonality of the observations implies the definition of a suitable period for observations, noted $\mathcal{T}$. Similarly, relevant observations can only be made in a restricted area corresponding to the locations of the agricultural fields. This area will be called area of interest and is noted $\mathcal{AI}$. Table 8 summarizes these main notations.

Table 8: Main notations and definitions

Notation	Definition
$z(s_i, t_j)$	Observation collected at site s_i and date t_j
$\mathcal{ST}$	Set of collected observations
$\mathcal{SN}_i$	Set of observations in the spatial neighbourhood of s_i
$\mathcal{TN}_j$	Set of observations in the temporal neighbourhood of t_j
$\mathcal{T}$	Relevant period of observation
$\mathcal{AI}$	Area of interest

7.3.2.2 *Classification of outliers*

Variable Z can potentially be sampled at an infinite number of sites and at an infinite number of dates. Nevertheless, most observations $z(s_i, t_j)$ are made in such a way that $t_j \in \mathcal{T}$ because participants are trying to observe the phenomenon as it occurs. Observations for which $t_j \notin \mathcal{T}$ are considered as outliers as they do not provide any information about this phenomenon. They were considered to be the result of another mechanism (e.g. an operator testing the application outside the relevant time period) (Hawkins, 1980). Similarly, observations are only informative if they are collected within $\mathcal{AI}$. All observations for which $s_i \notin \mathcal{AI}$ were also considered as outliers. Condition 1 can be written as:

$$Condition\ 1\text{: } s_i \in \mathcal{AI} \text{ et } t_j \in \mathcal{T} \qquad \text{Equation 4}$$

Observations that did not meet this Condition 1 were considered as outliers.

7.3.2.3 *Classification of surprising observations*

Z is auto-correlated in time and space. Therefore, it is expected that $z(s_i, t_j)$ is close to the values observed in its spatio-temporal neighbourhood. In order to take into account the asynchronous and heterotopic character of the data, the spatio-temporal neighbourhood of $z(s_i, t_j)$ was defined as the set of observations collected on sites spatially close to s_i and in a time interval Δ. A function F was defined to study the attribute difference between an observation $z(s_i, t_j)$ and the observations in its spatio-temporal neighbourhood. This function F depends on the interval Δ. It is therefore noted F^{Δ}. In the proposed approach, it was considered that Condition 2 was fulfilled if the attribute difference between an observation and its spatiotemporal neighbourhood was sufficiently small. Condition 2 can therefore be written as:

$$Condition\ 2: F^{\Delta}\left(z(s_i, t_j)\right) \leq \alpha \qquad \text{Equation 5}$$

Where α represents the maximum acceptable attribute distance between $z(s_i, t_j)$ and its spatio-temporal neighbourhood

7.3.3 Implementation of the approach

7.3.3.1 *Considered characteristics for testing Conditions 1 and 2*

The first part of Condition 1 ($s_i \in \mathcal{AI}$) was simply tested by considering the spatial characteristic of each observation in order to evaluate its membership to $\mathcal{AI}$. The second part of Condition 1 ($t_j \in \mathcal{T}$) was tested by considering two characteristics for each observation: its attribute value and its date. The observations were represented in a two-dimensional space $z(s_i, t_j)\ vs\ Time$. The date of each observation was expressed in number of days since the beginning of the year and the coordinates on the two axes were centred and reduced.

Condition 2 was tested by studying the similarity of each remaining observation with its spatial neighbours over three successive time intervals, in order to compensate for the asynchronous and heterotopic nature of datasets. These three time intervals were defined in order to include observations collected respectively just before, at the same time and just after the observation considered. For an observation collected at date t_j, the three intervals considered were noted $\Delta_j - 1$, Δ_j and $\Delta_j + 1$, respectively. The function F^{Δ} was used to study the attribute difference between an observation $z(s_i, t_j)$ and the observations belonging to each of these three intervals. In a first approach, it was defined as the average of the attribute differences between an observation and its spatial neighbours. This average was weighted according to a membership index (w_{ik}) of each site s_k to the spatial neighbourhood of site s_i. This membership index was defined based on inverse distance weighting (Shepard 1968). The function F^{Δ} can be written as :

$\forall\ s_k \in \mathcal{SN}_i,$
$\forall\ t_l \in \Delta :$

$$F^{\Delta}(z(s_i, t_j)) = \frac{1}{\sum_{k=1}^{m} w_{ik}} \cdot \sum_{k=1}^{m} w_{ik} \cdot \left| z(s_i, t_j) - z(s_k, t_l) \right| \qquad \textit{Equation 6}$$

With

$$w_{ik} = \frac{1}{d(s_i, s_k)}$$

Condition 2 was tested by investigating for each observation three characteristics: $F^{\Delta_j-1}, F^{\Delta_j}, F^{\Delta_j+1}$. Observations were represented in a three-dimensional space defined by these three characteristics.

7.3.3.2 *Method implemented for testing Conditions 1 and 2*

The first part of Condition 1 ($s_i \in \mathcal{AI}$) was tested using the CORINE Land Cover 2018 database proposed by the Copernicus programme. The area of interest $\mathcal{AI}$ was defined as the area encompasing all agricultural land (codes 21 to 33 of CORINE Land Cover 2018). The second part of Condition 1 ($t_j \in \mathcal{T}$) and Condition 2 were tested based on the specificity of the FSO datasets, which contained mostly quality data. It was assumed that the vast majority of observations met Conditions 1 and 2 and that they had very similar characteristics. Therefore, it was considered that these reliable observations constituted a cluster in the studied dimensions and that all observations that did not belong to this cluster were outliers (case of Condition 1) or surprising data (case of Condition 2).

In order to identify these outliers and surprising observations, methods based on the density of observations, which have been widely addressed in the literature (Wang et al. 2019), were chosen. As a first approach, the DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise) was chosen and implemented as it is relatively simple and widely documented. DBSCAN needs to define two parameters ε and $MinPts$. For each observation x_i, DBSCAN considers that all other observations located at a distance less than ε from x_i are part of its neighbourhood. If the number of these observations is greater than $MinPts$ then x_i is considered as belonging to the core of a cluster. If this condition is not fulfilled but one of x_i's neighbours is an observation that itself belongs to the core of a cluster, then x_i is considered as a cluster border observation. In this study, border observations were considered as belonging to the cluster. If none of these conditions are met, then x_i is considered as noise. In the proposed approach, observations classified by DBSCAN as noise were considered as outliers and surprising observations when testing conditions 1 and 2, respectively. Implementing the DBSCAN algorithm to test Conditions 1 and 2 required defining specific optimal values: these parameter pairs were called $(\varepsilon_1, MinPts_1)$ and $(\varepsilon_2, MinPts_2)$ for Conditions 1 and 2 respectively.

7.3.3.3 *Automation of the implemented method*

Optimal values for $(\varepsilon_1, MinPts_1)$ and $(\varepsilon_2, MinPts_2)$ were defined automatically based on the characteristics of the dataset following the approach proposed by Leroux et al. (2018a). According to these authors, it is expected that good quality observations are grouped within the same cluster and that the most frequently observed distance between two observations is characteristic of these quality observations. In this case, Leroux et al. (2018a) proposed to use this distance as the optimal value of ε. These same authors then proposed to study the distribution of the number of observations present in a neighbourhood defined by this optimal distance ε. A break in the distribution indicates the separation between two clusters. The first local minimum was used to define the optimal value of $MinPts$ for identifying outliers. This method was implemented to estimate the optimal values of respective parameters for both Conditions 1 and 2. Axes of the

two-dimensional space $z(s_i, t_j)\ vs\ Time$ in which condition 1 was tested were dimensionless numerical quantities because these variables were centred and reduced. Therefore, distances calculated for condition 1 were also dimensionless numerical quantities. For condition 2, the distances calculated in the three-dimensional space $F^{\Delta_j-1}, F^{\Delta_j}, F^{\Delta_j+1}$ were in the same dimension as Z (Equation 6).

7.3.3.4 *Flow chart of the implemented approach*

Figure 44 summarises the different steps of the proposed approach. It presents the conditions tested and their chronology.

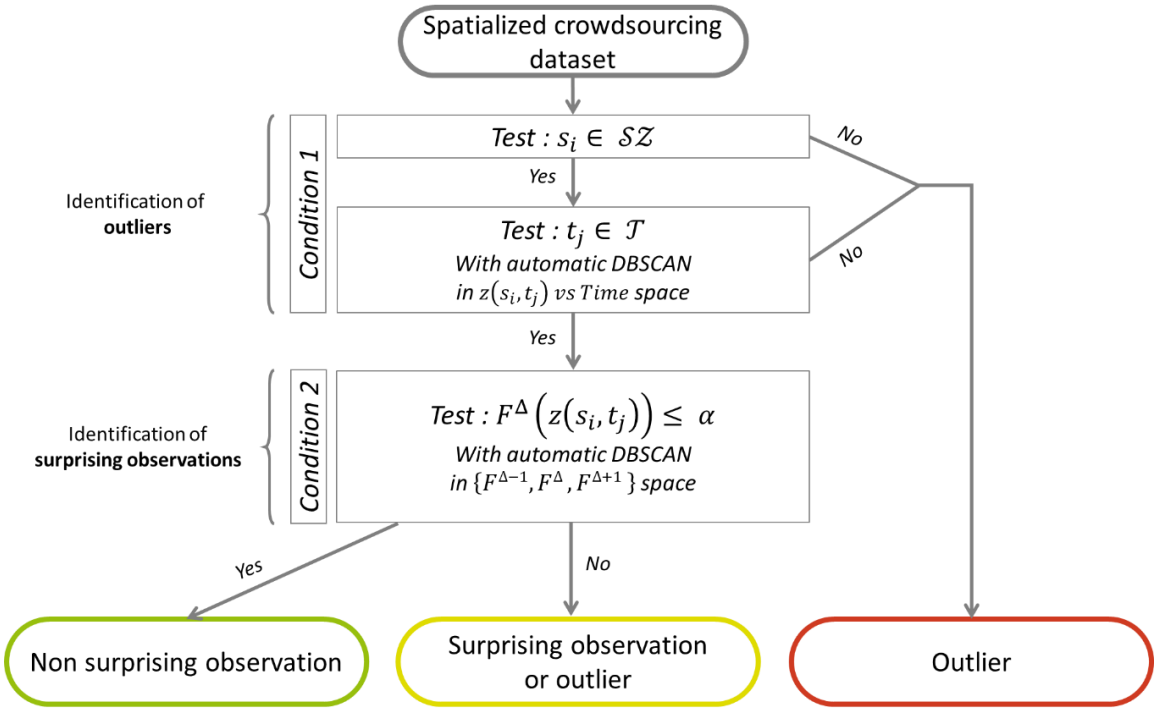


Figure 44: Main steps of the implemented approach for identifying outliers and surprising observations in spatialized crowdsourcing datasets.

7.3.4 Use case

7.3.4.1 *ApeX-Vigne dataset*

The approach was tested on a dataset collected with the ApeX-Vigne application (see Chapter 5:). This mobile application was used by professionals to estimate the occurrence of vine water restriction at the field level. It is based on the gathering of simple observations on vine shoot growth by professionals (farmers, advisors). ApeX-Vigne summarises the observations with an index, iG-Apex, which varies between 1 (full shoot growth) and 0 (total cessation of shoot growth) (See Chapter 3:). This index is a dimensionless numerical quantity. The application dates and georeferences iG-Apex observations and sends them to a central server (Chapter 5:). Apex-Vigne is available for free download from the Google playstore on Android (https://play.google.com/store/apps/details?id=ag.GB.apex&hl=fr&gl=US). Participants use the application as a decision support tool at the field level. The contribution to a collective cartography at the regional scale is therefore a secondary use for them.

Chapter 6 showed that the phenomenon described by iG-Apex was dynamic, seasonal and structured in time and space (Pichon et al. 2021). It was considered that the gathering and sharing of observations with the ApeX-Vigne application was representative of a FSO process. Therefore, the iG-Apex variable was chosen as a use case of the generic variable Z described above. The dataset corresponded to the 4404 observations collected using the ApeX-Vine application during the 2020 season within the study zone (see next section). The interval Δ was chosen to be equal to one week as this is to the most frequently observed interval between two observations on the same field in this dataset.

7.3.4.2 *Study zone*

The study zone was situated in the south of France. It encompasses the vineyards of the Languedoc, Provence and Côtes du Rhone viticulture regions (Figure 45a).

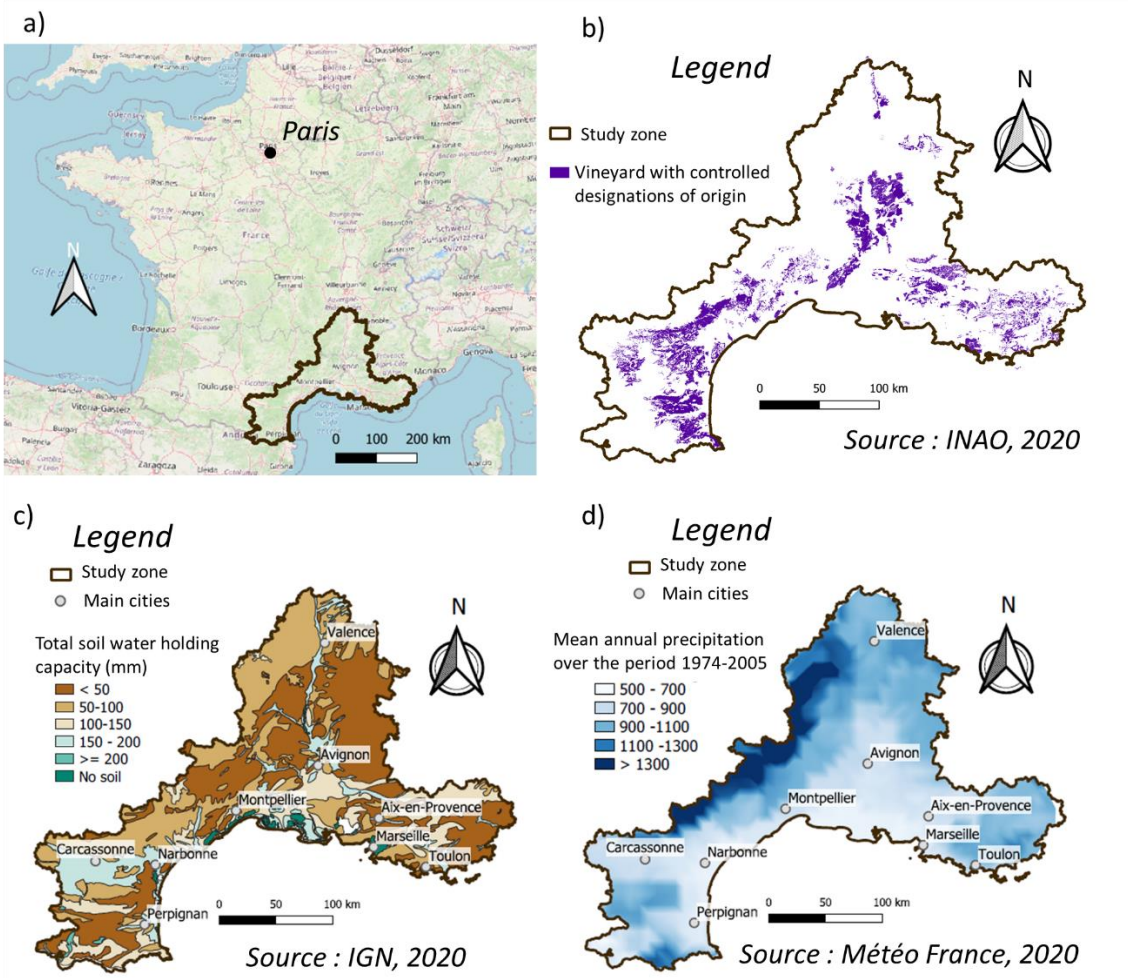


Figure 45: The study zone is a wine-growing region with a large variability in soil and climate conditions and in which the ApeX-Vigne application is relevant and widely used. (a) Location of the study zone within France, (b) location of vineyards within the study zone that have a controlled denomination of origin (INAO, 2020). Two principal pedo-climatic characteristics of this zone: (c) total soil water holding capacity (calculated from soil depth and texture) (IGN, 2020) and (d) mean annual precipitation over the period 1974-2005 (Météo France, 2020).

The study zone covers an area of ~49 500 km^2 along the Mediterranean Sea (Figure 45a) and encompasses >300 000 ha of vines (Figure 45b) and >26 000 growers (Agreste, 2010). The majority of the vineyards are non-irrigated and the climate is Mediterranean with hot, dry summers. The soils mostly have low soil water holding capacity (<50 mm) (Figure 45c) and the mean annual precipitation over the last 30 years varies from relatively low values (500 – 700 mm yr^{-1}) in coastal areas to relatively high values (>1 300 mm yr^{-1}) in the hinterland (Figure 45d).

This study zone is favourable to the use of the ApeX-Vigne application because i) it is one of the largest contiguous wine-growing regions in France, ii) the pedoclimatic conditions make the monitoring of vine water restriction a major issue for vineyard management, and iii) the iG-Apex method is known and has been used for several years by wine-growers within the study zone and their advisors.

7.3.5 Validation of the approach

7.3.5.1 *Selection of experts*

For FSO projects, as for crowdsourcing projects in general, validation of outlier detection approaches is often a complex issue as reference data are rarely available (Fonte et al. 2017). Foody et al. (2013) demonstrated that collecting the opinions of three experts could be an alternative to reference data. The performance of the proposed approach was evaluated using this method. It was considered that the experts' opinion could be influenced by i) their knowledge of the study zone, ii) their knowledge of the ApeX-Vigne application and its uses, and iii) their familiarity with FSO datasets. These three criteria were chosen in order to select experts that were as different as possible (Table 9:). Expert 1 was an agricultural advisor with strong field experience, expert 2 was an agricultural engineer specialised in developing services and conducting experiments in the region and expert 3 was a researcher in the field of digital viticulture.

Table 9: Description of the experts and their characteristics (1 = low expertise; 2 = medium expertise and 3 = high expertise)

	Expert 1	Expert 2	Expert 3
Knowledge of the study zone	3	1	2
Knowledge of the ApeX-Vigne app	3	2	1
Familiarity with farmsourcing datasets	1	3	2

7.3.5.2 *Validation of the approach for the classification of outliers*

The objective was to evaluate the ability of the proposed approach to test Condition 1 and to identify outliers in FSO datasets. The first part of Condition 1 ($s_i \in \mathcal{AI}$) was considered valid as it was tested using reference data (CORINE Land Cover 2018). For the second part of Condition 1 ($t_j \in \mathcal{T}$), the classification made by the proposed approach was compared to the classification made by the experts. The whole dataset was presented to the experts to identify all observations they considered to be outliers. The observations were presented on a 2-dimensional graph $iG\text{-}Apex\ vs\ Time$. Time was represented by the number of days since the beginning of the year. If an observation was classified as an outlier by at least two of the three experts, it was considered as an outlier. The comparison between the expert classification and the classification made by the

proposed approach was carried out using a confusion matrix. This confusion matrix was used to evaluate the sensitivity (share of correctly classified outliers) and specificity (share of correctly classified non-outliers) of the approach.

7.3.5.3 *Evaluation of the robustness of the approach for the classification of outliers*

The proposed approach was based on the estimation of DBSCAN parameters from the characteristics of the dataset under study. As a result, the approach was necessarily sensitive to this dataset. In order to assess the robustness of the approach, a bootstrap was used to resample the data. This method makes it possible to estimate the confidence intervals of estimators without making any assumptions about their distribution (Efron, 1982). In this study, it was used to estimate the confidence intervals of the confusion matrix elements. There were 1000 bootstrap iterations performed using the original dataset. Each iteration included the pre-processing step of testing for Conditions 1 and 2 before the DBSCAN analysis. With each iteration, all the resampled observations were classified as an outlier or non-outlier and compared to the expert classification. The bootstrap results were used to estimate the median of the 1000 iterations and the confidence intervals associated with each element of the confusion matrix.

7.3.5.4 *Validation of the approach for the classification of surprising observations*

The objective was to evaluate the ability of the approach to identify surprising observations (Condition 2). According to Figure 43, these observations can be outliers or surprising observations. In the case of outliers, the expert classification was used as a reference. However, for surprising observations no reference was available and the results obtained could only be simply discussed. As a first step, the classification made by the proposed approach was compared to the outlier/non-outlier classification made by the experts (see previous sections). This step was used to check whether the real outliers that had not been identified by the test of Condition 1 were identified by the test of Condition 2. In a second step, the objective was to compare the behaviour of the observations identified as surprising to the rest of the non-surprising observations in order to discuss the results. A temporal comparison was made by comparing the surprising observations to the rest of the observations and in particular to two reference fields representing fields with early and late cessation of shoot growth, respectively (see next section). A spatial comparison was made by comparing the surprising observations to a map of the non-surprising observations (see section 7.3.5.6).

7.3.5.5 *Temporal comparison of surprising observations: reference fields*

Two fields were chosen to illustrate a particularly early and a particularly late cessation of shoot growth (Figure 46). These two fields were used to define the minimum and maximum expected iG-Apex values in the region and used as a reference for interpreting observations classified as surprising.

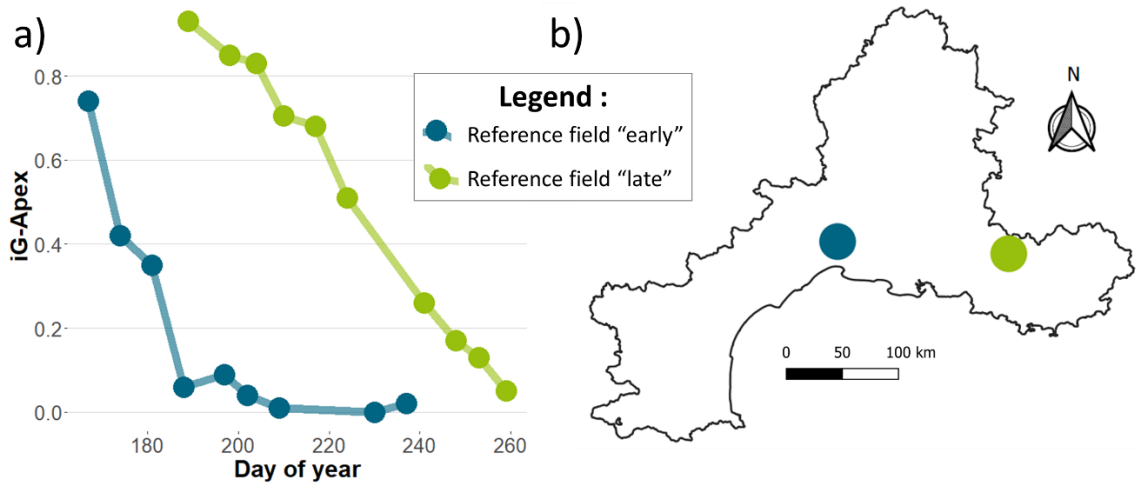


Figure 46: (a) Evolution of the shoot growth index (iG-Apex) during the season in an early and late developing vineyard and (b) location within the study zone of the "early" (9 observations) and "late" (10 observations) reference fields.

The observations from these fields were considered of good quality as they were collected by experienced and trusted operators. Both fields were located within the study zone in regions where many other observations were collected (Figure 46b). They were labelled reference field "early" and reference field "late" respectively.

7.3.5.6 *Spatial comparison of surprising observations: map of non-surprising observations*

A map of the attribute values of non-surprising observations was made over the whole study zone in order to represent the global dynamics of iG-Apex values. The surprising observations were compared to this global dynamic in order to illustrate the possible attribute differences with their spatial neighbours. This map was made from the observations collected during week 30 because this was the date having the most important number of observations. It was made by interpolating iG-Apex values by ordinary kriging on a regular 10 km grid.

7.3.6 Graphs and maps:

All calculations and graphics were performed using R 3.6.1 (R Core Team, 2020). Maps were made with Qgis 3.4.8-Madeira (QGIS Development Team, 2020) and kriging with GeoFIS (Leroux et al. 2018b).

7.4 Results

7.4.1 Classification of outliers :

4381 observations out of the 4404 of the FSO dataset (~99.5%) met the first part of Condition 1 ($s_i \in \mathcal{AI}$). These observations are distributed over all the vineyard regions of the study zone (Figure 47).

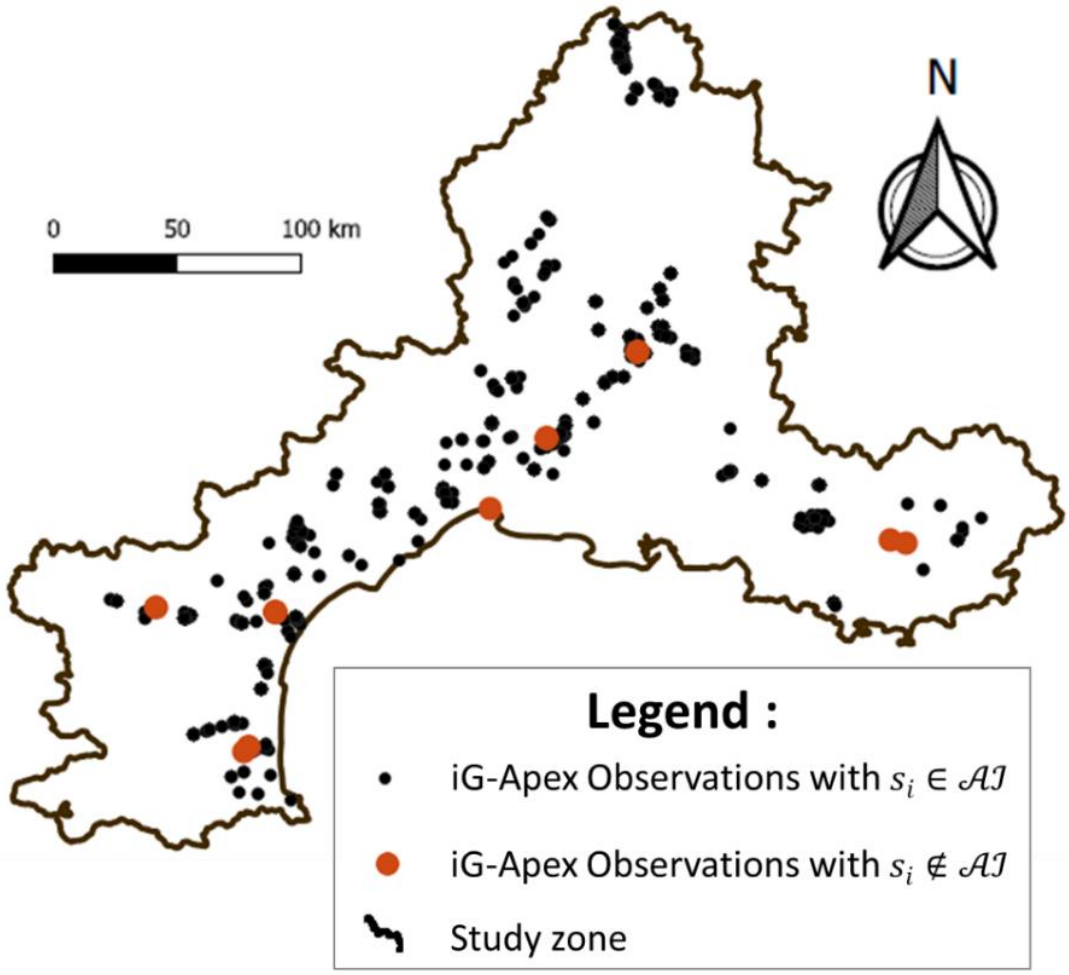


Figure 47: Location of the 4381 iG-Apex observations having $s_i \in \mathcal{AI}$ and of the 23 observations having $s_i \notin \mathcal{AI}$. Points are overlapping. Therefore, at this wide spatial scale, very close observations may appear as a single point.

The representation of these observations in the $iG\text{-}Apex\ vs\ Time$ space showed several groups of observations with varying density (Figure 48a).The period between 150 and 250 days had a particularly high density of observations compared to the periods before and after it. The density curve of the distances between two observations indicated that the most frequently observed distance was equal to 0.84 (dimensionless numerical quantity) (Figure 48b). This value was automatically chosen for the DBSCAN parameter ϵ_1. For a neighbourhood of $\epsilon_1 = 0.84$, the number of neighbours for each observation varied between 0 and nearly 2000 (Figure 48c). The first density minimum was observed for 113 neighbours. This value of 113 neighbours corresponded to the threshold between observations with relatively few neighbours and other observations. This value was chosen for the parameter $MinPts_1$.

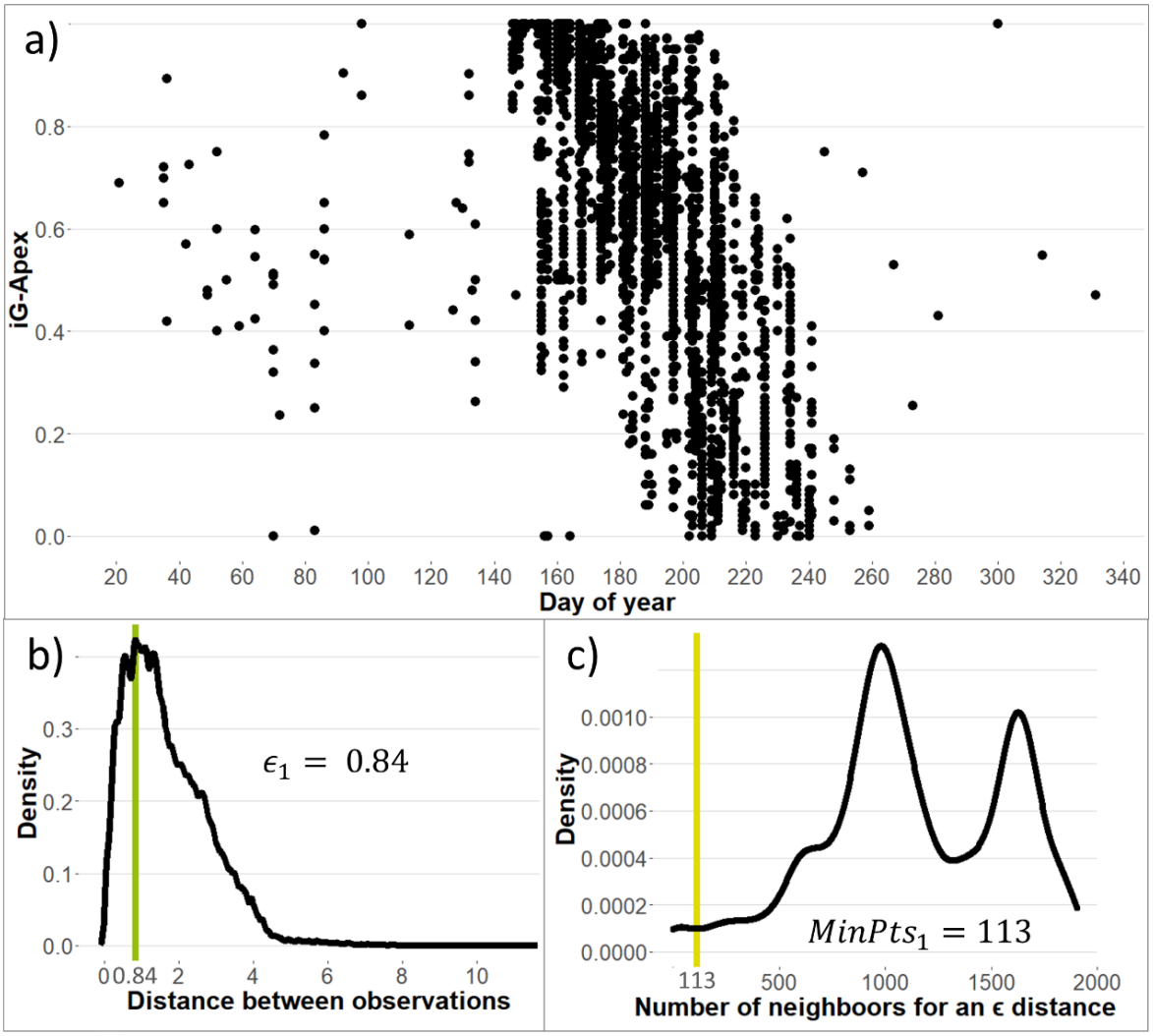


Figure 48: (a) Representation of the ApeX-Vigne dataset in 2-dimensional space $iG\text{-}Apex\ vs\ Time$ *and estimation of parameters* ϵ_1 *and* $MinPts_1$ *for classification of outliers from: (b) density of observed distance between two observations and (c) density of number of neighbours observed for this distance. Green and yellow lines represent the values chosen for parameters* ϵ_1 *and* $MinPts_1$ *respectively.*

These results indicated that the proposed approach was able to automatically estimate DBSCAN parameters in the $iG\text{-}Apex\ vs\ Time$ space. These parameters were used to classify the observations as outliers or non-outliers and to compare this classification with the experts' opinion. Firstly, it is interesting to note that the experts estimated that only 1.6% (1.3% + 0.3%) of the observations were outliers (Table 10). This indicated that the FSO dataset was globally of good quality and that outliers were rare. There were 99.7% (1.3% + 98.4%) of observations correctly classified (Table 10). The sensitivity of this approach was 81.2% and its specificity 100%.

Table 10: Outliers' classification confusion matrix for the proposed approach compared to expert reference classification on the ApeX-Vigne original dataset.

		Experts	
		Outlier	Non-outlier
DBSCAN	Outlier	56 1.3%	0 0%
	Non-outlier	13 0.3%	4312 98.4%

These results indicated that the outlier classification performance of the approach was good. Nevertheless, the method tended to be rather cautious and to classify as outliers only observations that were actually outliers. The main consequence of this was that a percentage of the outliers (0.3%) identified by the experts were not detected by the proposed method.

Observations classified as outliers corresponded mainly to those collected before day 150 and after day 260 (Figure 49). Observations collected between these two dates were almost all classified as non-outlier by the experts and by the proposed approach.

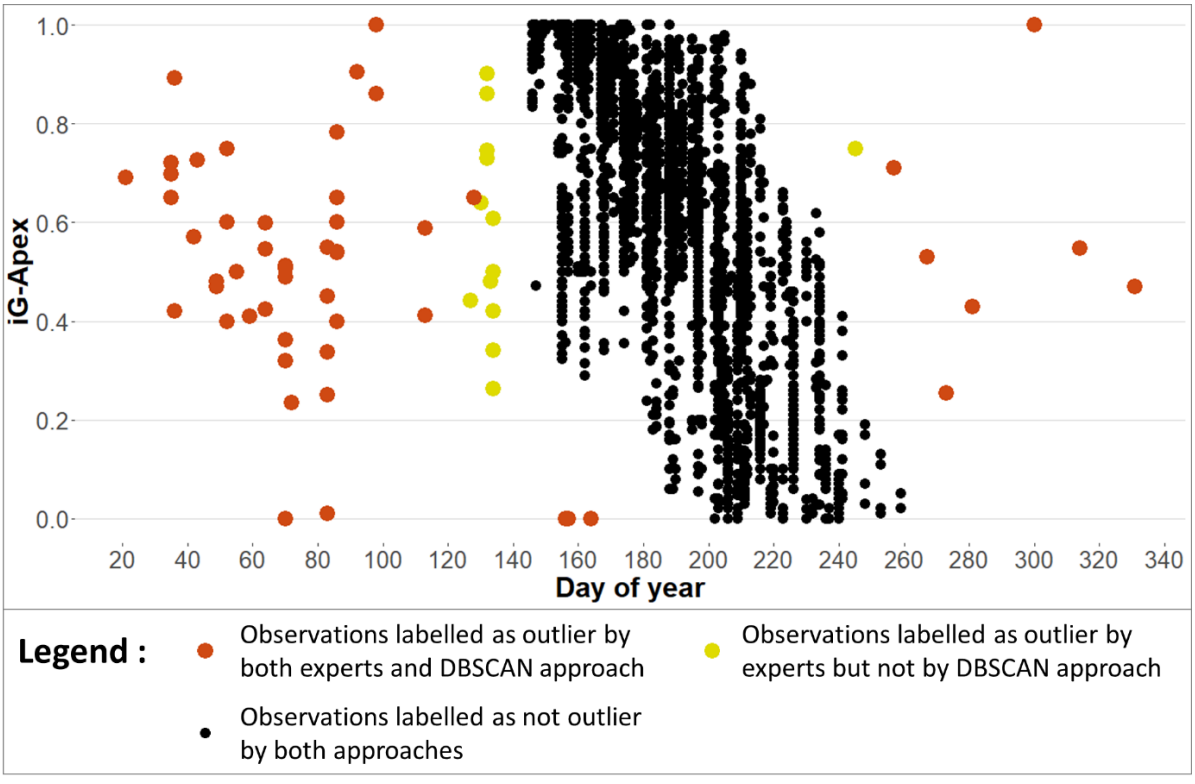


Figure 49: Location of observations correctly and wrongly classified as outliers by the proposed approach according to expert reference classification plotted in the 2-dimensional space $iG\text{-}Apex\ vs\ Time$.

These results were consistent with vine phenology as the period between days 150 and 260 corresponded to the summer period during which shoot growth was likely to be affected by water restriction. Therefore, this was the period during which iG-Apex observations were preferentially gathered by winegrowers and their advisors. Some observations collected during this period were nevertheless classified as outliers by both the experts and the proposed approach. This was the case, for example, for the observations collected around day 160, with iG-Apex values very close to 0. According to the experts, strong water restrictions at these dates are unrealistic and the observations were considered as outliers. The proposed approach correctly classified these observations. The outliers that were not identified corresponded mainly to observations made around day 140 (Figure 49). It is likely that the date of these observations, close to the high density zone, did not allow these observations to be identified as outliers.

In general, these results showed that the proposed approach was relevant for identifying outliers. The confusion matrix performed on the 1000 bootstrap iterations (Table 11) confirmed that the proposed method was also robust. The median value observed on these 1000 iterations was very close to the value observed from the initial dataset. The confidence intervals ($\alpha = 0.1$) were very small for the observations classified as outliers by the experts. However, they were quite large for the observations classified as non-outliers by the experts. According to this result, the specificity of the method may possibly be sensitive to the collected dataset.

Table 11: Outliers' classification confusion matrix for the proposed approach compared to expert reference classification over 1000 bootstrap's iterations on the original ApeX-Vigne dataset. Bold values indicate the median value from the bootstrapping while values in parentheses are confidence intervals for $\alpha = 0.1$

		Experts	
		Outlier	Non-outlier
DBSCAN	Outlier	1.4% [1.14% - 1.66%]	0% [0% – 31.9%]
	Non-outlier	0.2% [0% - 0.33%]	98.4% [66.8% – 100%]

These wide confidence intervals ($\alpha = 0.1$) can be explained by the organisation of the point cloud classified as non-outliers in the iG-Apex vs. time space. This point cloud is very dense at its core and less dense at its periphery. Depending on the resampling carried out during the bootstraping, it was possible that, in certain cases, the density of the zones of the $iG\text{-}Apex\ vs\ Time$ graph that were located at the periphery of the period $\mathcal{T}$ may have been lower than in the initial dataset. When this phenomenon was too pronounced, it may be that observations located in the period $\mathcal{T}$ were classified as outliers. However, these cases are unlikely to happen in FSO datasets where the gathering of observations is costly for participants, which tends to favour a very high rate of observations within period T and thus limits the impact of this problem.

7.4.2 Classification of surprising observations:

Observations that were identified as non-outliers presented varying densities in the three-dimensional space ($F^{week_j-1}, F^{week_j}, F^{week_j+1}$) (Figure 50a). The zone near the origin was particularly dense indicating that observations with attribute values close to those of their spatial and temporal neighbours were frequent. The most frequently observed distance was equal to 7.10^{-4} (dimensionless numerical quantity) (Figure 50b). This value was chosen for the parameter ϵ_2. For a neighbourhood of $\epsilon_2 = 7.10^{-4}$, the number of neighbours for each observation varied between 0 and over 3000 (Figure 50c). There was a minimum density observed for 664 neighbours that distinguished observations with relatively few neighbours from others. This value was chosen for the parameter $MinPts_2$.

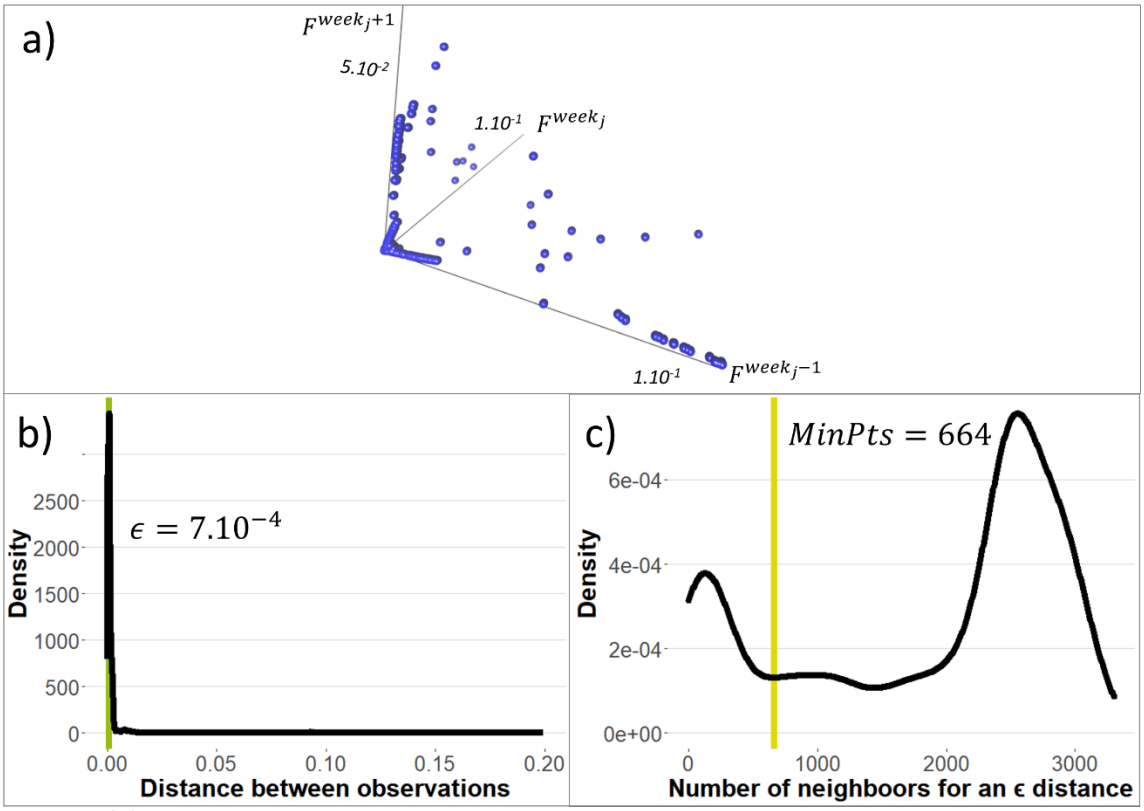


Figure 50: (a) Representation of the non-outliers observations of the ApeX-Vigne dataset in the 3 dimensional space ($F^{week_j-1}, F^{week_j}, F^{week_j+1}$), and estimation of parameters ϵ_2 and $MinPts_2$ for classification of surprising observations from: (b) density of observed distance between two observations and (c) density of number of neighbours observed for the most frequent distance. Green and yellow lines represent the values chosen for parameters ϵ_2 and $MinPts_2$ respectively.

These results showed that the proposed approach could estimate the parameters ϵ_2 and $MinPts_2$ automatically from the ApeX-Vigne dataset. These parameters were then used to test Condition 2 and to classify observations as surprising or not surprising. Figure 51 shows that all outliers that were not identified by the test of Condition 1 were classified as surprising observations. This result indicated that tests of Conditions 1 and 2 were complementary and identified all outliers.

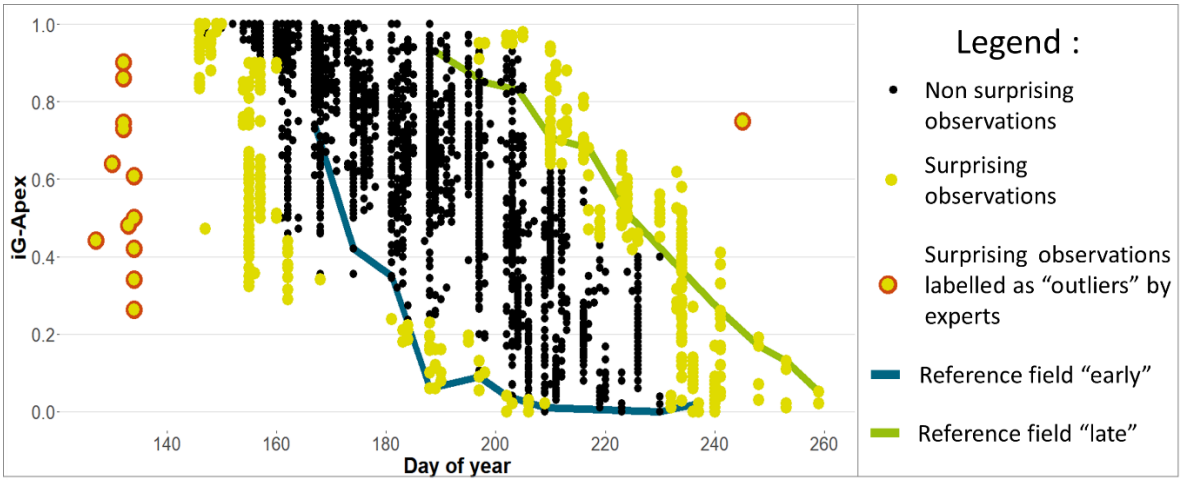


Figure 51: Classification of surprising observations plotted in the 2-dimensional space $iG\text{-}Apex\ vs\ Time$.

The representation over time of other surprising observations suggested that there are two populations of surprising observations (Figure 51); one corresponding to observations close to the trajectory of the reference field "early" and a second corresponding to observations close to the trajectory of the reference field "late". These observations can be interpreted as fields whose cessation of growth was earlier or later than the fields in their spatial and temporal neighbourhoods. This interpretation was confirmed by the spatial representation of the surprising observations at week 30 (Figure 52).

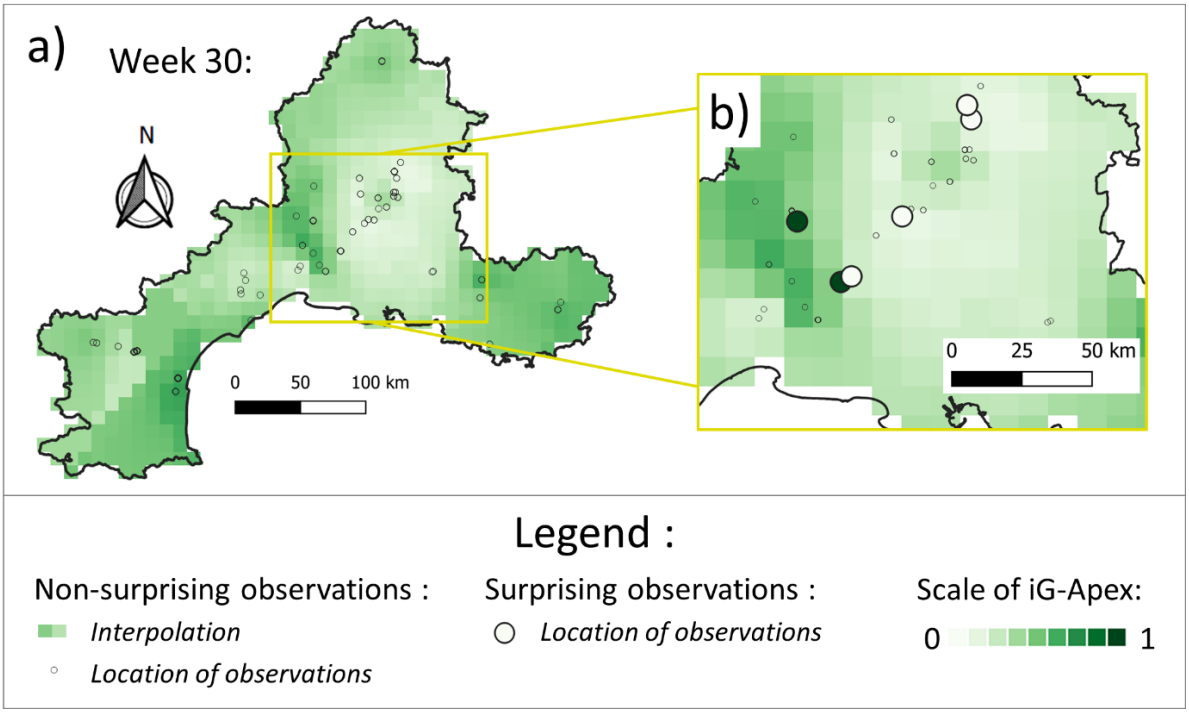


Figure 52: (a) Interpolated map of iG-Apex during week 30 at the study zone scale made from the non-surprising observations. (b) Detail of this map representing the surprising observations on the same iG-Apex scale.

The map of non-surprising observations showed that it is possible to produce a map of iG-Apex at the study zone scale (Figure 52a). This map exhibited a spatial organisation of iG-Apex at this scale confirming the results of Pichon et al. (2021). It also indicated that the surprising observations have iG-Apex values that were either much higher or much lower than their spatial neighbourhood (Figure 52b). This result confirmed that surprising observations corresponded to fields whose cessation of growth was earlier or later than the vast majority of the other observations in their neighbourhood.

7.5 Discussion

The results of this study showed that taking into account attribute, spatial and temporal characteristics of observations makes it possible to adapt the framework proposed in the literature to the specificities of FSO datasets and to identify surprising observations and outliers. The formulated hypothesis on the very high proportion of good quality data in FSO datasets proved to be valid, which is consistent with what has been observed in crowdsourcing projects from other sectors (Mehdipoor et al. 2015). The validation of this hypothesis allowed the implementation of an approach to identify outliers and surprising observations based on the identification of a single cluster corresponding to good quality observations. The use of the DBSCAN method, based on the study of the density of observations, proved to be relevant because it was robust and fully automatable. This approach has the advantage of requiring little knowledge of the phenomenon studied and of automatically adapting thresholds to the dataset under study (Wang et al., 2019).

The adaptation of the existing framework to the characteristics of FSO datasets allowed to obtain particularly good performances for the detection of outliers since they were all identified after the test of Conditions 1 and 2. Those that were not identified by the test of Condition 1 correspond to outliers that had the closest behaviour to the majority of good quality observations and are therefore necessarily the most difficult to discriminate. To address this limitation, other automation approaches for the choice of DBSCAN parameters could be explored. Indeed, these parameters have a strong influence on the classification result (Leroux et al., 2018a), especially in the cluster border zones where outliers not identified by Condition 1 are located. New methods for automating these parameters, which are the subject of research in computer science (Starczewski et al. 2020), could make it possible to find compromises that are even more favourable to the identification of outliers by Condition 1.

For the identification of surprising observations, the performances of the proposed approach were satisfactory but their evaluation was more difficult to achieve. Indeed, in the case of this study, as in most FSO projects (Van Etten, Beza, et al. 2019), no reference data were available. Therefore, if this approach were to be improved, it would be necessary to set up an experimental design to obtain a classification of surprising observations that could be used as a reference. The labelling of surprising observations by experts, as implemented in this study for outliers, seems to be a relevant approach. However, it would require an experimental design that would allow several experts to evaluate each observation individually. Given the large number of observations considered, micro-working platforms that allow unitary tasks to be proposed to a large number of people (e.g. defining whether an observation is surprising or not) seems to be a possible approach that has already been implemented in other domains (Aguinis et al. 2021). Furthermore, one of the limitations of the proposed approach was that it did not take into account the spatial structure

of the studied phenomenon in defining the spatial neighbourhood of an observation. It is possible that the use of data-driven stochastic approaches to define the spatial neighbourhood of an observation could improve the detection of surprising observations. Such stochastic approaches could be implemented, for example, by using a semi-variance model of the variable under consideration (Negreiros et al. 2010) or by studying co-variables (Gao et al. 2020). Another limitation of the proposed approach is the binary labelling of surprising observations. This approach has the advantage of being simple but it may generate threshold effects for observations whose behaviour is intermediate between surprising and expected. It is possible that the exploration of approaches defining a level of surprise for each observation, rather than a binary classification, will allow the complexity of the phenomenon studied to be taken into account and thus better support users in their decision-making. Approaches implementing clustering methods with fuzzy logic (Gosain and Dahiya 2016) seem to be interesting avenues to explore.

The proposed approach automatically adapted to the attribute, spatial and temporal characteristics of the considered dataset. It has been implemented and tested on the ApeX-Vigne use case but it can be generalised to all FSO projects. The challenge of this generalisation lies mainly in the choice of the time interval Δ. In the case where the phenomenon studied evolves rapidly and where the level of asynchronism of the observations is particularly strong, for example for pest monitoring (Jiang et al. 2008), it is possible that this definition of the temporal neighbourhood needs to be improved. Future work could favour the generalisation of the approach for this type of use cases by considering time in a continuous manner and by defining a degree of membership to the temporal neighbourhood using, for example, a reasoning similar to the one implemented for the spatial neighbourhood. Finally, in a more general way, the proposed approach can allow the identification of outliers and surprising observations in any dataset resulting from the dynamic monitoring of a phenomenon structured in time and in space and containing many observations, even if this dataset is neither asynchronous nor heterotopic. For example, the identification of surprising pixels in an image dataset resulting from the dynamic monitoring of agricultural fields from remote sensing images is an important issue (Mouret et al. 2021). The approach proposed for FSO datasets could be extended to this type of use cases.

7.6 Conclusion

This study posed the problem statement of developing quality control metrics for FSO data and reviewed existing approaches for detecting outliers and surprising observations. It proposed an approach for adapting the existing framework to FSO datasets taking into account attribute, spatial and temporal characteristics of observations. Results showed that the approach based on the identification of a single cluster representing quality observations was automatable. The implementation of this approach on a real dataset from observations collected with the mobile application ApeX-Vigne demonstrated an effective identification of outliers and surprising observations. For the identification of outliers, results showed that the approach i) was robust, i.e. it maintained good performance when the dataset varied, ii) had good sensitivity, i.e. it was able to correctly label outliers and ii) had good specificity, i.e. it was also able to correctly label observations that were non-outliers. The results of this study also showed that the proposed approach was able to identify surprising observations. In some rare cases, these observations were outliers that had not been previously identified. In the majority of cases, these surprising

observations corresponded to fields that were particularly early or late in relation to their spatial and temporal neighbourhood. Finally, the proposed approach can be generalised to the identification of outliers and surprising observations in other FSO datasets and more widely in any dataset resulting from the dynamic monitoring of a phenomenon structured in time and space.

Chapter 8: Conclusion and perspectives

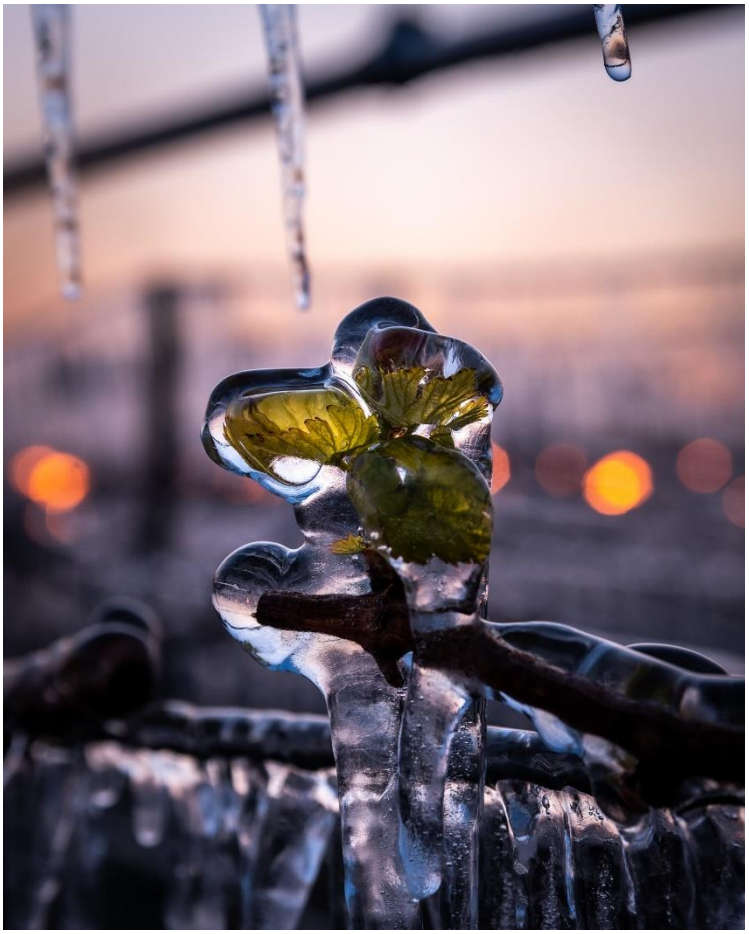


Illustration: Vine in Chablis – April 2021

The objective of chapter 8 is to conclude on validation and generalization of the two formulated hypotheses. This chapter also opens up some directions for future research work.

8.1 Looking back to the studied use case

In this thesis, monitoring of vine water status by characterising vine shoot growth has been chosen as a use case for FSO projects. The outputs of this thesis are firstly related to this use case. The first of them is the development of the ApeX-Vigne mobile application (Chapter 5). This application is an operational tool that can be used for decision support by wine industry's stakeholders. This thesis has also reinforced the knowledge concerning the method of shoot growth characterisation (Chapter 3). It has demonstrated the relationship between this method and a reference method based on several cultivars, several vintages and several fields. It also made it possible to define the range of validity of the method, its specificity and its sensitivity to vine water status. Finally, this thesis demonstrated that the uncertainty associated with this method was low enough to characterise the within-field variability of vine water status if observations were collected by the same operator (Chapter 4).
Agronomic knowledge related to this use case could complement the generic approach proposed for the identification of outliers and surprising observations (Chapter 7). For example, the shoot growth dynamic tend to decrease during the season while water stress sets in (Chapter 3). Therefore, the study of time series of observations made on the same field should make it possible to identify cases where the shoot growth dynamic accelerates, which could correspond to surprising observations. From a physiological point of view, this case can only occur if there is a supply of water to the field by irrigation or rainfall.
In an operational context, access to contextual information such as agronomic practices of vine growers (e.g. irrigation) or local soil and climate conditions could help to interpret this surprising behaviour. This perspective seems interesting when the information can be collected without any effort from the contributor, such as access to weather conditions via the location of the observation and the interrogation of dedicated databases. On the other hand, when access to this information requires an action from the contributor (e.g. is the field irrigated?), it becomes necessary to reassess his or her motivation to contribute and the levers that can be used to influence this motivation. In the future, evolutions concerning the participants' contribution in the ApeX-Vigne project will have to be designed by considering the methodological framework proposed in this thesis and by making sure the trade-off between the quantity of contextual information collected and the motivation of the contributor is well balanced.

8.2 Looking back to the proposed approach

FSO projects have specificities because of their characteristics concerning i) the phenomenon under study (STRUCTURED and SEASONAL), ii) the observation gathering process (PROFESSIONALS, ONCE), iii) the resulting dataset (ASYNCHRONOUS) and iv) the available reference data (REFERENCE). In this thesis, specific success factors of FSO projects have been studied. Two hypotheses on the factors promoting on the one hand quantity and on the other hand quality of collected observations were tested through a systemic approach implemented with the real FSO project, ApeX-Vigne. Hypothesis 1 make the assumption that FSO projects must i) address an agricultural issue at both farm and regional scales (MULTISCALE) (Chapter 2), ii) propose to participants a task that is simple and useful for them (USEFUL) (Chapters 3 and 4) and iii) rely on a platform accessible to the maximum of potential participants (ACCESIBLE) (Chapter 5) to generate a significant amount of observations. Hypothesis 2 make the assumption that these projects must i) rely on a platform allowing contributions to be collected according to a standardised protocol (STANDARDIZED) by recording metadata on acquisition conditions (METADATA) (Chapter 5) and ii) implement ex-post quality strategies based on attribute, spatial

and temporal characteristics of observations (MULTI-CHARACTERISTIC) (Chapter 7) to generate contributions of good quality. The ApeX-Vigne project was a relevant case study to test these two hypothesis because it had all these characteristics.

The analysis of 2019 and 2020 data of the ApeX-Vigne project showed that the observations collected were numerous (Chapter 6) and of good quality (Chapter 7). Therefore, the two hypotheses were validated for this use case. Moreover, the ApeX-Vigne project can be considered as representative of FSO projects because it was studied over several seasons and over a large spatial area of several hundred kilometres, encompassing many vineyards grown under various conditions. Therefore, results obtained on this use case can be generalised and hypotheses 1 and 2 can be considered as valid also for other FSO projects.

Nevertheless, the proposed approach has some limitations. Firstly, each of the specific success factors have not been tested independently. For instance, the validation of hypothesis 1 indicates that the presence of the three specific success factors is resulting in a significant number of contributions. However, it is not possible to distinguish the relative importance of these three factors or even to guarantee that the three of them are essential. The same consideration can also be extended to the three specific success factors tested by hypothesis 2. In the future, individual study of each of these specific success factors could be facilitated by the emergence of a larger number of FSO projects having more diverse characteristics. Secondly, the tests of hypotheses 1 and 2 rely on the evaluation of the quantity of observations collected and on the quality of these observations. However, the methods and metrics existing in the literature to evaluate this quantity and quality are poorly adapted to the specificities of FSO projects such as the seasonality of the studied phenomena (SEASONAL) or the asynchronous and heterotopic characteristics of datasets (ASYNCHRONOUS). In the implemented approach, the quantity and quality levels obtained have been simply compared to existing values in the literature for other crowdsourcing projects. In the future, the design of approaches that take into account these specificities in the evaluation of these quantity and quality levels would allow more precise hypotheses to be formulated and tested.

The knowledge generated in this thesis helped to better understand the specific success factors of FSO projects. It also opens up new research perspectives regarding i) the improvement of some steps of the method that have been simplified as first approach and ii) possible usages of FSO data that are made possible by this new knowledge. The following sections detail these perspectives.

8.3 Looking forward to future research work

First of all, in FSO projects, it is common that several observations are made at the same field and by the same operator throughout the season. However, in the ex-post strategy explored in this thesis for identifying surprising observations, attribute values of observations were simply compared to those of their spatio-temporal neighbours without considering whether these observations belong to a time series. It is possible that, for a given observation, the study of the time series to which it belongs may provide information on its most likely attribute value. This expected value could be compared to the observed value in order to estimate the level of surprise of an observation. Therefore, in the future, exploring the relevant characteristics of time series for assessing the level of surprise of an observation may help to improve ex-post quality methods.

Secondly, in the approach proposed in this work, the level of surprise of an observation did not take into account the person who collected it. However, this information about the participant is quite classically used in other crowdsourcing projects. In this case, it is often the reputation of the participant that is taken into account, i.e. the trust that the community places in him/her (Fogliaroni et al. 2018). These approaches have mainly been developed on collaborative mapping applications such as OpenStreetMap (Zhao et al. 2016). The reputation of a contributor is usually estimated based on the consistency between his observations and the observations that other contributors have made on the same objects. However, in the case of FSO projects, the same object can only be observed once (Specificity ONCE). Therefore, existing methods in the literature cannot be applied as they stand and other approaches must be explored to assess the reputation of participants. Specificities of FSO projects such as the seasonality of the phenomenon under study (Specificity SEASONAL) or the contribution of participants in the context of their professional activity (Specificity PROFESSIONALS) are structuring the behaviour of all participants by inducing cyclical contribution and participation at regular intervals respectively. Participants trying to observe the phenomenon under the right conditions all have the same type of behaviour. Therefore, it seems possible to evaluate the reputation of a participant according to the gap between this expected behaviour and his or her actual behaviour and according to his or her changes in behaviour. Understanding these behaviours and exploring the analyses that can be made of them to estimate the reputation of participants may lead to a better understanding of both *ex post* and *ex ante* strategies for managing the quality of collected observations.

Finally, the results of this thesis work provide tools and methods to collect many good quality observations in FSO projects. These observations can therefore be considered as a potential source of information complementary to other existing approaches for the study of agricultural phenomena at regional scale. For example, the inversion of mechanistic models such as water balance model (Roux et al. 2019) is an interesting approach because it makes it possible to estimate agronomic variables of interest such as FTSW (Fraction of Transpirable Soil Water) over large spatial areas. However, this approach relies on field measurement of parameters such as predawn leaf water potential (ψ_{PD}), which are often expensive to gather. Observations from FSO projects could potentially be used instead of these reference measurements. The use of these imprecise, asynchronous and heterotopic datasets would constitute an originality compared to existing approaches in the literature. The scientific question that arises is whether the large number of observations of FSO projects can balance these imperfections in order to implement a mechanistic model inversion approach. This question will soon be explored through a thesis which has received funding from the #Digitag convergence lab and the Occitanie region. It will start in autumn 2021.

8.4 Opening up new perspectives towards on-farm experimentation

With FSO projects, it is possible to collect and share spatialized, dated and objectified observations. These projects and the corresponding methods could be used by groups of farmers and/or researchers to collect and share their experimentations. This could help to design and set up collective experimentations under real conditions and in various pedoclimatic situations. FSO projects and associated knowledge could help to i) involve farmers, ii) collect and pre-process data, iii) analyse and discuss the results, and iv) repeat these processes in order to produce knowledge. FSO projects can therefore be seen as a means of accompanying agriculture in its

transition by helping agricultural stakeholders to organise themselves collectively to answer to the complex and sometimes contradictory issues they are facing. This approach, which aims to promote the construction of knowledge by relying on experiments conducted by farmers in real conditions, is developing more and more in the scientific community. It is called On-Farm Experimentation (OFE) (Cook et al. 2013). OFE raises new questions about approaches for involving farmers (MacMillan and Benton 2014), methods for analysing such data (Laurent et al. 2019) and agronomic knowledge that can be extracted from it (Bullock et al. 2019). Therefore, this emerging scientific community (https://ispag.org/Communities/On-Farm) is at the crossroads of social sciences, agricultural sciences and data sciences. The community is in the process of federating and will organise a first scientific conference dedicated to OFE in autumn of 2021 (https://ofe2021.com/). In the future, the tools and methods developed during this thesis and more broadly for FSO projects might contribute to this new scientific community.

References

Acevedo-Opazo, C., Ortega-Farias, S., & Fuentes, S. (2010). Effects of grapevine (Vitis vinifera L.) water status on water consumption, vegetative growth and grape quality: An irrigation scheduling application to achieve regulated deficit irrigation. *Agricultural Water Management*, *97*(7), 956–964. https://doi.org/10.1016/j.agwat.2010.01.025

Acevedo-Opazo, C., Tisseyre, B., Ojeda, H., Ortega-Farias, S., & Guillaume, S. (2008). Is it possible to assess the spatial variability of vine water status? *Journal International des Sciences de la Vigne et du Vin*, *42*(4), 203–220. https://doi.org/10.20870/oeno-one.2008.42.4.811

Agreste, Recensement agricole (Census of Agriculture) (2010). [web] (consulted on 30/09/2020) https://www.agreste.agriculture.gouv.fr/

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk Research: Review and Recommendations. *Journal of Management*, *47*(4), 823–837. https://doi.org/10.1177/0149206320969787

Altwaijry, H., Trulls, E., Hays, J., Fua, P., & Belongie, S. (2016). Learning to Match Aerial Images with Deep Attentive Architectures. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3539–3547). Las Vegas, NV, USA. https://doi.org/10.1109/CVPR.2016.385

Améglio, T., Archer, P., Cohen, M., Valancogne, C., Daudet, F. A., Dayau, S., & Cruiziat, P. (1999). Significance and limits in the use of predawn leaf water potential for tree irrigation. *Plant and Soil*, *207*(2), 155–167. https://doi.org/10.1023/A:1026415302759

Arsanjani, J. J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In P. M. and M. H. J. Arsanjani, A. Zipf (Ed.), *OpenStreetMap in GIScience: experiences, research, applications* (pp. 37–58). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-14280-7_3

Asingizwe, D., Murindahabi, M. M., Koenraadt, C. J. M., Poortvliet, P. M., van Vliet, A. J. H., Ingabire, C. M., et al. (2019). Co-Designing a Citizen Science Program for Malaria Control in Rwanda. *Sustainability (Switzerland)*, *11*(24), 1–17. https://doi.org/10.3390/su11247012

Awuor, F., Otanga, S., Kimeli, V., Rambim, D., & Abuya, T. (2019). E-Pest Surveillance: Large Scale Crop Pest Surveillance and Control. In *2019 IST-Africa Week Conference, IST-Africa 2019* (pp. 1–8). Nairobi, Kenya: IST-Africa Institute and Authors. https://doi.org/10.23919/ISTAFRICA.2019.8764824

Balčiauskas, L., Balčiauskienė, L., Litvaitis, J.A., Tijušas, E. (2020) Adaptive monitoring: using citizen scientists to track wolf populations when winter-track counts become unreliable. *Wildlife Research, 48*, 76-85. https://doi.org/10.1071/WR19180

Baralon, K., Payan, J. C., Salançon, E., & Tisseyre, B. (2012). Spider: Spatial extrapolation of the vine water status at the whole denomination scale from a reference site. *Journal International des Sciences de la Vigne et du Vin*, *46*(3), 167–175. https://doi.org/https://doi.org/10.20870/oeno-one.2012.46.3.1517

Batson, C. D., Ahmad, N., & Tsang, J. A. (2002). Four motives for community involvement. *Journal of Social Issues*, *58*(3), 429–445. https://doi.org/10.1111/1540-4560.00269

Bécart, V., Vernière, A., Leonet, A., & Lacroix, R. (2020). *Le journal du stress hydrique des côtes du Rhône*. https://www.syndicat-cotesdurhone.com/liste-articles/referentiel-stress-hydrique

Bellon-Maurel, V., & Huyghe, C. (2017). Putting agricultural equipment and digital technologies at the cutting edge of agroecology. *OCL - Oilseeds and fats, Crops and Lipids*, *24*(3). https://doi.org/10.1051/ocl/2017028

Ben Salem-Fnayou, A., Bouamama, B., Ghorbel, A., & Mliki, A. (2011). Investigations on the leaf anatomy and ultrastructure of grapevine (Vitis vinifera) under heat stress. *Microscopy Research and Technique*, *74*(8), 756–762. https://doi.org/10.1002/jemt.20955

Beza, E., Steinke, J., Van Etten, J., Reidsma, P., Fadda, C., Mittra, S., et al. (2017). What are the prospects for citizen science in agriculture? Evidence from three continents on motivation and mobile telephone use of resource-poor farmers. *PLoS ONE*, *12*(5), 1–26. https://doi.org/10.1371/journal.pone.0175700

Billaud, O., Vermeersch, R. L., & Porcher, E. (2021). Citizen science involving farmers as a means to document temporal trends in farmland biodiversity and relate them to agricultural practices. *Journal of Applied Ecology*, *58*(2), 261–273. https://doi.org/10.1111/1365-2664.13746

Bonter, D. N., & Cooper, C. B. (2012). Data validation in citizen science: A case study from Project FeederWatch. *Frontiers in Ecology and the Environment*, *10*(6), 305–307. https://doi.org/10.1890/110273

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M., & Rampini, A. (2014). On predicting and improving the quality of Volunteer Geographic Information projects. *International Journal ofDigital Earth*, *9*(2), 134–155. https://doi.org/10.1080/17538947.2014.976774

Bota, J., Flexas, J., & Medrano, H. (2001). Genetic variability of photosynthesis and water use in Balearic grapevine cultivars. *Annals of Applied Biology*, (138), 353–361. https://doi.org/https://doi.org/10.1111/j.1744-7348.2001.tb00120.x

Bota, J., Tomás, M., Flexas, J., Medrano, H., & Escalona, J. M. (2016). Differences among grapevine cultivars in their stomatal behavior and water use efficiency under progressive water stress. *Agricultural Water Management*, *164*(August), 91–99. https://doi.org/10.1016/j.agwat.2015.07.016

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, *14*(1), 75–90. https://doi.org/10.1177/1354856507084420

Brovelli, M. A., Minghini, M., Molinari, M., & Mooney, P. (2017). Towards an Automated Comparison of OpenStreetMap with Authoritative Road Datasets. *Transactions in GIS*, *21*(2), 191–206. https://doi.org/10.1111/tgis.12182

Brunel, G., Pichon, L., Taylor, J., & Tisseyre, B. (2019). Easy water stress detection system for vineyard irrigation management. In *Precision Agriculture 2019 - Papers Presented at the 12th European Conference on Precision Agriculture, ECPA 2019* (pp. 935–942). https://doi.org/10.3920/978-90-8686-888-9_115

Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., et al. (2019). The data-intensive farm management project: Changing agronomic research through on-

farm precision experimentation. *Agronomy Journal*, *111*(6), 2736–2746. https://doi.org/10.2134/agronj2019.03.0165

Caravia, L., Collins, C., Petrie, P. R., & Tyerman, S. D. (2016). Application of shade treatments during Shiraz berry ripening to reduce the impact of high temperature. *Australian Journal of Grape and Wine Research*, *22*(3), 422–437. https://doi.org/10.1111/ajgw.12248

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior and Organization*, *90*, 123–133. https://doi.org/10.1016/j.jebo.2013.03.003

Chittilappilly, A. I., Chen, L., & Amer-Yahia, S. (2016). A Survey of General-Purpose Crowdsourcing Techniques. *IEEE Transactions on Knowledge and Data Engineering*, *28*(9), 2246–2266. https://doi.org/10.1109/TKDE.2016.2555805

Choné, X., Van Leeuwen, C., Dubourdieu, D., & Gaudillère, J. P. (2001). Stem water potential is a sensitive indicator of grapevine water status. *Annals of Botany*, *87*(4), 477–483. https://doi.org/10.1006/anbo.2000.1361

Cifre, J., Bota, J., Escalona, J. M., Medrano, H., & Flexas, J. (2005). Physiological tools for irrigation scheduling in grapevine (Vitis vinifera L.): An open gate to improve water-use efficiency? *Agriculture, Ecosystems and Environment*, *106*(2-3 SPEC. ISS.), 159–170. https://doi.org/10.1016/j.agee.2004.10.005

Colaço, A. F., Molin, J. P., Rosell-Polo, J. R., & Escolà, A. (2019). Spatial variability in commercial orange groves. Part 2: relating canopy geometry to soil attributes and historical yield. *Precision Agriculture*, *20*(4), 805–822. https://doi.org/10.1007/s11119-018-9615-0

Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment*, *176*(1–4), 273–291. https://doi.org/10.1007/s10661-010-1582-5

Cook, S., Cock, J., Oberthür, T., & Fisher, M. (2013). On-farm experimentation. *Better Crops with Plant Food*, *97*(4), 17–20.

Crall, A. W., Newman, G. J., Jarnevich, C. S., Stohlgren, T. J., Waller, D. M., & Graham, J. (2010). Improving and integrating data on invasive species collected by citizen scientists. *Biological Invasions*, *12*(10), 3419–3428. https://doi.org/10.1007/s10530-010-9740-9

Criscuolo, L., Carrara, P., Bordogna, G., Pepe, M., Zucca, F., Seppi, R., et al. (2016). Handling quality in crowdsourced geographic information. In *European Handbook of Crowdsourced Geographic Information* (Ubiquity P., pp. 57–74). London. https://doi.org/10.5334/bax.e

De Souza Filho, H. M., Vinholis, M. M. B., Carrer, M. J., & Bernardo, R. (2021). Determinants of adoption of integrated systems by cattle farmers in the State of Sao Paulo, Brazil. *Agroforestry Systems*, *95*(1), 103–117. https://doi.org/10.1007/s10457-020-00565-8

Dehnen-schmutz, K., Foster, G. L., Owen, L., & Persello, S. (2016). Exploring the role of smartphone technology for citizen science in agriculture. *Agronomy for sustainable development*, *36*(25), 1–9. https://doi.org/10.1007/s13593-016-0359-9

Dehnen-Schmutz, K., Foster, G. L., Owen, L., & Persello, S. (2016). Exploring the role of smartphone technology for citizen science in agriculture. *Agronomy for Sustainable Development*, *36*(2), 25. https://doi.org/10.1007/s13593-016-0359-9

Dell'Acqua, F., Iannelli, G. C., Torres, M. A., & Martina, M. L. V. (2018). A novel strategy for very-large-scale cash-crop mapping in the context of weather-related risk assessment, combining global satellite multispectral datasets, environmental constraints, and in situ acquisition of geospatial data. *Sensors (Switzerland)*, *18*(2). https://doi.org/10.3390/s18020591

Deloire, A., Carbonneau, A., Wang, Z., & Ojeda, H. (2004). Vine and water a short review. *Journal International des Sciences de la Vigne et du Vin*, *38*(1), 1–13. https://doi.org/10.20870/oeno-one.2004.38.1.932

Ebitu, L., Avery, H., Mourad, K. A., & Enyetu, J. (2021). Citizen science for sustainable agriculture – A systematic literature review. *Land Use Policy*, *103*, 105326. https://do*i.org/10.1016/j.landusepol.2021.105326*

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plan*s. CBMS-NSF Regional Conference Series in Applied Mathematics,* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics. 85 p.

Fogliaroni, P., D'Antonio, F., & Clementini, E. (2018). Data trustworthiness and user reputation as indicators of VGI quality. *Geo-Spatial Information Science*, *21*(3), 213–233. https://doi.org/10.1080/10095020.2018.1496556

Fonte, C. C., Antoniou, V., Bastin, L., Estima, J., Arsanjani, J. J., Laso Bayas, J.-C., et al. (2017). Assessing VGI Data Quality. In V. (eds. . Foody, G, See, L, Fritz, S, Mooney, P, Olteanu-Raimond, A-M, Fonte, C C and Antoniou (Ed.), *Mapping and the Citizen Sensor* (pp. 137–163). London: Ubiquity Press. https://doi.org/https://doi. org/10.5334/bbf.g

Foody, G. M., See, L., Fritz, S., Van der Velde, M., Perger, C., Schill, C., & Boyd, D. S. (2013). Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project. *Transactions in GIS*, *17*(6), 847–860. https://doi.org/10.1111/tgis.12033

Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., et al. (2009). Geo-wiki.org: The use of crowdsourcing to improve global land cover. *Remote Sensing*, *1*(3), 345–354. https://doi.org/10.3390/rs1030345

Gao, H., Zhang, X., Wang, L., He, X., Shen, F., & Yang, L. (2020). Selection of training samples for updating conventional soil map based on spatial neighborhood analysis of environmental covariates. *Geoderma*, *366*(February), 114244. https://doi.org/10.1016/j.geoderma.2020.114244

Garcia, L., Celette, F., Gary, C., Ripoche, A., Valdés-Gómez, H., & Metay, A. (2018). Management of service crops for the provision of ecosystem services in vineyards: A review. *Agriculture, Ecosystems and Environment*, *251*(October 2017), 158–170. https://doi.org/10.1016/j.agee.2017.09.030

Gardner, C. M. K., Robinson, D., Blyth, K., & Cooper, J. D. (2001). Soil Water Content. In K. A. Smith & C. E. Mullins (Eds.), *Soil and Environmental Analysis : Physical methods* (pp. 1–64). New York, New York, USA: Dekker, Marcel. https://doi.org/10.1081/e-ess3-120042638

Gaudillère, J. P., Van Leeuwen, C., & Ollat, N. (2002). Carbon isotope composition of sugars in grapevine, an integrated indicator of vineyard water status. *Journal of Experimental Botany*, *53*(369), 757–763. https://doi.org/10.1093/jexbot/53.369.757

Gaudin, R., Kansou, K., Payan, J. C., Pellegrino, A., & Gary, C. (2014). A water stress index based on water balance modelling for discrimination of grapevine quality and yield. *Journal International des Sciences de la Vigne et du Vin*, *48*(1), 1–9. https://doi.org/10.20870/oeno-one.2014.48.1.1655

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. https://doi.org/10.1016/j.spasta.2012.03.002

Gosain, A., & Dahiya, S. (2016). Performance Analysis of Various Fuzzy Clustering Algorithms: A Review. *Procedia Computer Science*, *79*, 100–111. https://doi.org/10.1016/j.procs.2016.03.014

Griffith C. 2017. Mobile App Development with Ionic Revised Edition Cross Platform Apps with Ionic Angular & Cordova. Sebastopol, California (USA): O'Reilly Media, Inc.

Guilpart, N., Metay, A., & Gary, C. (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*, 9–20. https://doi.org/10.1016/j.eja.2013.11.002

Haklay, M. (2010). How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design*, *37*(4), 682–703. https://doi.org/10.1068/b35097

Hardie, W. J., & Martin, S. R. (2000). Shoot growth on de-fruited grapevines: A physiological indicator for irrigation scheduling. *Australian Journal of Grape and Wine Research*, *6*(1), 52–58. https://doi.org/10.1111/j.1755-0238.2000.tb00162.x

Hawkins, D. 1980. Identification of Outliers, London, UK: Chapman & Hall. https://doi.org/10.1007/978-94-015-3994-4

Hekkert P., van Dijk M. B. 2011. Vision in design: A guidebook for innovators. Amsterdam, The Netherdlands: BIS Publishers B.V.

Hsiao, T. C. (1973). Plant responses to water stress. *Annual Review of Plant Physiology*, *24*, 519–570. https://doi.org/10.1093/aob/mcf159

Hu, Z. (2020). What socio-economic and political factors lead to global pesticide dependence? A critical review from a social science perspective. *International Journal of Environmental Research and Public Health*, *17*(21), 1–22. https://doi.org/10.3390/ijerph17218119

Hunter, J. K., Tarricone, L., Volschenk, C., Giacalone, C., Melo, M. S., & Zorer, R. (2020). Grapevine physiological response to row orientation-induced spatial radiation and microclimate changes. *Oeno One*, (April), 411–433. https://doi.org/10.20870/oeno-one.2020.54.2.3100

IGN, Institut national de l'information géographique et forestière (National Institute of Geographic and Forest Information) (2020) [web] (consulted on 16/07/2020) https://www.ign.fr/institut/identity-card

INAO, Institut National de l'origine et de la qualité (National Institute of Origin and Quality) (2020) [web] (consulted on 16/07/2020) https://www.inao.gouv.fr/eng/

Jabeur, N., Karam, R., & Melchiori, M. (2018). A comprehensive reputation assessment framework for volunteered geographic information in crowdsensing applications. *Personal*

and Ubiquitous Computing, *23*, 669–685. https://doi.org/https://doi.org/10.1007/s00779-018-1122-9

Jasse, A., Berry, A., Aleixandre-Tudo, J. L., & Poblete-Echeverría, C. (2021). Intra-block spatial and temporal variability of plant water status and its effect on grape and wine parameters. *Agricultural Water Management*, *246*(December), 106696. https://doi.org/10.1016/j.agwat.2020.106696

Jiang, J. A., Tseng, C. L., Lu, F. M., Yang, E. C., Wu, Z. S., Chen, C. P., et al. (2008). A GSM-based remote wireless automatic monitoring system for field information: A case study for ecological monitoring of the oriental fruit fly, Bactrocera dorsalis (Hendel). *Computers and Electronics in Agriculture*, *62*(2), 243–259. https://doi.org/10.1016/j.compag.2008.01.005

Joly, A., Bonnet, P., Goëau, H., Barbe, J., Selmi, S., Dufour-kowalski, S., et al. (2016). A look inside the Pl @ ntNet experience. *Multimedia Systems*, *22*(6), 751–766. https://doi.org/https://doi.org/10.1007/s00530-015-0462-9

Jones, H. G., Stoll, M., Tiago, S., De Sousa, C., Chaves, M. M., & Grant, O. M. (2002). Use of infrared thermography for monitoring stomatal closure in the field: application to grapevine. *Journal of Experimental Botany*, *53*(378), 2249–2260. https://doi.org/10.1093/jxb/erf083

Jonietz, D., Antonio, V., See, L., & Zipf, A. (2017). Highlighting current trends in volunteered geographic information. *ISPRS International Journal of Geo-Information*, *6*(7), 1–8. https://doi.org/10.3390/ijgi6070202

Kaufmann, N., Schulze, T., & Veit, D. (2011). More than fun and money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk. In *Proceedings of the Seventeenth Americas Conference on Information Systems* (pp. 1–11). https://doi.org/10.1145/1979742.1979593

Kazmierski, M., Glemas, P., Rousseau, J., & Tisseyre, B. (2011). Temporal stability of within-field patterns of ndvi in non irrigated mediterranean vineyards. *Journal International des Sciences de la Vigne et du Vin*, *45*(2), 61–73. https://doi.org/10.20870/oeno-one.2011.45.2.1488

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., et al. (2019). Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity. *BioScience*, *69*(3), 170–179. https://doi.org/10.1093/biosci/biz010

Kerry, R., and Oliver, M.A. (2007). Sampling requirements for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*, *140*, 383–396. https://doi.org/10.1016/j.geoderma.2007.04.019

Kosonen, M., Gan, C., Olander, H., & Blomqvsit, K. (2013). My idea is our idea! Supporting user-driven innovation activities in crowdsourcing communities. *International Journal of Innovation Management*, *17*(03), 1340010. https://doi.org/10.1142/S1363919613400100

Kuhlgert, S., Austic, G., Zegarac, R., Osei-Bonsu, I., Hoh, D., Chilvers, M. I., et al. (2016). MultispeQ Beta: A tool for large-scale plant phenotyping connected to the open photosynQ network. *Royal Society Open Science*, *3*(10). https://doi.org/10.1098/rsos.160592

Land-Zandstra, A. M., Devilee, J. L. A., Snik, F., Buurmeijer, F., & Van Den Broek, J. M. (2016). Citizen science on a smartphone: Participants' motivations and learning. *Public Understanding of Science*, *25*(1), 45–60. https://doi.org/10.1177/0963662515602406

Laurent, A., Kyveryga, P., Makowski, D., & Miguez, F. (2019). A framework for visualization and analysis of agronomic field trials from on-farm research networks. *Agronomy Journal*, *111*(6), 2712–2723. https://doi.org/10.2134/agronj2019.02.0135

Lebon, E., Dumas, V., Pieri, P., & Schultz, H. R. (2003). Modelling the seasonal dynamics of the soil water balance of vineyards. *Functional Plant Biology*, *30*(6), 699–710. https://doi.org/10.1071/FP02222

Lebon, E., Pellegrino, A., Louarn, G., & Lecoeur, J. (2006). Branch development controls leaf area dynamics in grapevine (Vitis vinifera) growing in drying soil. *Annals of Botany*, *98*(1), 175–185. https://doi.org/10.1093/aob/mcl085

Leibar, U., Unamunzaga, O., & Fernández-gómez, M. J. (2018). Benefit of ancillary data acquired at the cooperative level to study soil type and climatic zone influence on berry composition : a case study in Rioja appellation. *Oeno One*, *52*(2), 119–133. https://doi.org/10.20870/oeno-one.2018.52.2.1851

Leroux, C., Jones, H., Pichon, L., Guillaume, S., Lamour, J., Taylor, J., et al. (2018b). GeoFIS: An open source, decision-support tool for precision agriculture data. *Agriculture (Switzerland)*, *8*(6). https://doi.org/10.3390/agriculture8060073

Leroux, Corentin, Jones, H., Clenet, A., Dreux, B., Becu, M., & Tisseyre, B. (2018a). A general method to filter out defective spatial observations from yield mapping datasets. *Precision Agriculture*, *19*(5), 789–808. https://doi.org/10.1007/s11119-017-9555-0

Li, J. S., Hamann, A., & Beaubien, E. (2020). Outlier detection methods to improve the quality of citizen science data. *International Journal of Biometeorology*, *64*, 1825–1833. https://doi.org/https://doi.org/10.1007/s00484-020-01968-z

Li, S., Kang, S., Li, F., & Zhang, L. (2008). Evapotranspiration and crop coefficient of spring maize with plastic mulch using eddy covariance in northwest China. *Agricultural Water Management*, *95*(11), 1214–1222. https://doi.org/10.1016/j.agwat.2008.04.014

Liu, X., Li, H., Lu, X., Xie, T., Mei, Q., Feng, F., & Mei, H. (2018). Understanding Diverse Usage Patterns from Large-Scale Appstore-Service Profiles. *IEEE Transactions on Software Engineering*, *44*(4), 384–411. https://doi.org/10.1109/TSE.2017.2685387

MacMillan, T., & Benton, T. G. (2014). Engage farmers in research. *Nature*, *509*, 25–27. https://doi.org/https://doi.org/10.1038/509025a

Mahmud, D. M., & Abdullah, N. A. S. (2015). Reviews on agile methods in mobile application development process. In *9th Malaysian Software Engineering Conference (MySEC)* (pp. 161–165). https://doi.org/10.1109/MySEC.2015.7475214

Maistrello, L., Dioli, P., Dutto, M., Volani, S., Pasquali, S., & Gilioli, G. (2018). Tracking the spread of sneaking aliens by integrating crowdsourcing and spatial modeling: The Italian invasion of halyomorpha halys. *BioScience*, *68*(12), 979–989. https://doi.org/10.1093/biosci/biy112

Malek, R., Tattoni, C., Ciolli, M., Corradini, S., Andreis, D., Ibrahim, A., et al. (2018). Coupling traditional monitoring and citizen science to disentangle the invasion of Halyomorpha halys. *ISPRS International Journal of Geo-Information*, *7*(5), 171. https://doi.org/10.3390/ijgi7050171

Martinez-De-Toda, F., Balda, P., & Oliveira, M. (2010). Estimation of vineyard water status

(Vitis Vinifera L. Cv. Tempranillo) from the developmental stage of the shoot tips. *Journal International des Sciences de la Vigne et du Vin*, *44*(4), 201–206. https://doi.org/https://doi.org/10.20870/oeno-one.2010.44.4.1476

Martinez-Vergara, A., Payan, J. C., Salançon, E., & Tisseyre, B. (2014). Spiderδ: An empirical method to extrapolate grapevine (vitis vinifera l.) water status at the whole denomination scale using δ13c as ancillary data. *Journal International des Sciences de la Vigne et du Vin*, *48*(2), 129–140.

Massad, I., & Dalyot, S. (2018). Towards the crowdsourcing of massive smartphone assisted-GPS sensor ground observations for the production of digital terrain models. *Sensors (Switzerland)*, *18*(3), 898. https://doi.org/10.3390/s18030898

Medrano, H., Escalona, J. M., Cifre, J., Bota, J., & Flexas, J. (2003). A ten-year study on the physiology of two Spanish grapevine cultivars under field conditions: Effects of water availability from leaf photosynthesis to grape yield and quality. *Functional Plant Biology*, *30*(6), 607–619. https://doi.org/10.1071/FP02110

Mehdipoor, H., Zurita-milla, R., Rosemartin, A., Gerst, K. L., & Weltzin, J. F. (2015). Developing a Workflow to Identify Inconsistencies in Volunteered Geographic Information : A Phenological Case Study. *PLoS ONE*, *10*(10), 1–14. https://doi.org/10.1371/journal.pone.0140811

Meissner, G., Athmann, M., Fritz, J., Kauer, R., Stoll, M., & Schultz, H. R. (2019). Conversion to organic and biodynamic viticultural practices: impact on soil, grapevine development and grape quality. *Oeno One*, *53*(4), 639–659. https://doi.org/10.20870/oeno-one.2019.53.4.2403

Météo France (French Institute of Meteorology and Climatology) (2020) [web] (consulted on 16/07/2020): https://meteofrance.com/

Michel, L., Brun, F., Piraux, F., & Makowski, D. (2016). Estimating the incidence of Septoria leaf blotch in wheat crops from in-season field measurements. *European Journal of Plant Pathology*, *146*(1), 17–35. https://doi.org/10.1007/s10658-016-0887-9

Michelot, P. (2020). *Etat hydrique des sols - N°7 - 28/07/2020*. https://herault.chambre-agriculture.fr/productions-techniques/viticulture/nos-missions-specialisees/irrigation/

Michels, M., Fecke, W., Feil, J. H., Musshoff, O., Pigisch, J., & Krone, S. (2020). Smartphone adoption and use in agriculture: empirical evidence from Germany. *Precision Agriculture*, *21*(2), 403–425. https://doi.org/10.1007/s11119-019-09675-5

Minet, J., Curnel, Y., Gobin, A., Goffart, J.-P., Mélard, F., Tychon, B., et al. (2017). Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach. *Computers and Electronics in Agriculture*, *142*, 126–138. https://doi.org/10.1016/j.compag.2017.08.026

Mirás-avalos, J. M., & Intrigliolo, D. S. (2017). Grape Composition under Abiotic Constrains : Water Stress and Salinity. *Frontiers in Plant Science*, *8*, 851. https://doi.org/10.3389/fpls.2017.00851

Mirás-Avalos, J. M., & Silva Araujo, E. (2021). Optimization of Vineyard Water Management : Challenges , strategies and perspectives. *Water*, *13*(6), 746. https://doi.org/10.3390/w13060746

Mishra, U., Lal, R., Liu, D., & Van Meirvenne, M. (2010). Predicting the Spatial Variation of the Soil Organic Carbon Pool at a Regional Scale. *Soil Science Society of America Journal*, *74*(3), 906–914. https://doi.org/10.2136/sssaj2009.0158

Moneyron, A., Lallemand, J. F., Schmitt, C., Perrin, M., Soustre-Gacougnolle, I., & Masson, J. E. (2017). Linking the knowledge and reasoning of dissenting actors fosters a bottom-up design of agroecological viticulture. *Agronomy for Sustainable Development*, *37*(5). https://doi.org/10.1007/s13593-017-0449-3

Montes de Oca Munguia, O., Pannell, D. J., & Llewellyn, R. (2021). Understanding the Adoption of Innovations in Agriculture: A Review of Selected Conceptual Models. *Agronomy*, *11*(1), 139. https://doi.org/10.3390/agronomy11010139

Mouret, F., Albughdadi, M., Duthoit, S., Kouamé, D., Rieu, G., & Tourneret, J. Y. (2021). Outlier detection at the parcel-level in wheat and rapeseed crops using multispectral and sar time series. *Remote Sensing*, *13*(5), 1–25. https://doi.org/10.3390/rs13050956

Naulleau, A., Gary, C., Prévot, L., & Hossard, L. (2021). Evaluating Strategies for Adaptation to Climate Change in Grapevine Production – A Systematic Review. *Frontiers in Plant Science*, *11*, 1–20. https://doi.org/10.3389/fpls.2020.607859

Neethling, E., Barbeau, G., Coulon-Leroy, C., & Quénol, H. (2019). Spatial complexity and temporal dynamics in viticulture: A review of climate-driven scales. *Agricultural and Forest Meteorology*, *276–277*, 107618. https://doi.org/10.1016/j.agrformet.2019.107618

Negreiros, J. G., Painho, M. T., Aguilar, F. J., & Aguilar, M. A. (2010). A comprehensive framework for exploratory spatial data analysis: Moran location and variance scatterplots. *International Journal of Digital Earth*, *3*(2), 157–186. https://doi.org/10.1080/17538940903253898

Nguyen, Q. N., Frisiello, A., & Rossi, C. (2017). Co-design of a crowdsourcing solution for disaster risk reduction. In *I-TENDER 2017 - Proceedings of the 2017 1st CoNEXT Workshop on ICT Tools for Emergency Networks and DisastEr Relief* (pp. 7–12). https://doi.org/10.1145/3152896.3152898

Ojeda, H., Deloire, A., & Carbonneau, A. (2001). Influence of water deficits on grape berry growth. *Vitis*, *40*(3), 141–145. https://doi.org/https://doi.org/10.5073/vitis.2001.40.141-145

Ojeda, H., & Saurin, N. (2014). L'irrigation de précision de la vigne : méthodes, outils et stratégies pour maximiser la qualité et les rendements de la vendange en économisant de l'eau. *Innovations Agronomiques*, *38*, 97–108.

Oliver, M.A. and Webster, R. (2015). Basic Steps in Geostatistics: The Variogram and Kriging., Cham, Switzerland: Springer International Publishing. 1–99.

Parker, A. K., Cortázar-atauri, I. G. De, Trought, M. C. T., Destrac, A., & Agnew, R. (2020). Adaptation to climate change by determining grapevine cultivar differences using temperature-based phenology models. *Oeno One*, *54*(4), 955–974. https://doi.org/10.20870/oeno-one.2020.54.4.3861

Paulsgrove K, Miller E, Seidel K, Kinter S, Tse R. (2021) Crowdsourcing to Assess Speech Quality Associated With Velopharyngeal Dysfunction. *The Cleft Palate-Craniofacial Journal*, 58(1), 25-34. https://doi.org/10.1177/1055665620948770

Payan, J.-C. (2020). Peut-on réduire la sensibilité au changement climatique sans irriguer ? (Can sensitivity to climate change be reduced without irrigation?). In *Entretiens de la vigne et du vin – Languedoc Roussillon*.

Pebesma, E.J. (2004). Multivariable geostatistics in S: the gstat package. Computers \& *Geosciences*, *30*, 683-691. https://doi.org/10.1016/j.cageo.2004.03.012

Pellegrino, A., Lebon, E., Simonneau, T., & Wery, J. (2005). Towards a simple indicator of water stress in grapevine (Vitis vinifera L.) based on the differential sensitivities of vegetative growth components. *Australian Journal of Grape and Wine Research*, *11*(3), 306–315. https://doi.org/10.1111/j.1755-0238.2005.tb00030.x

Picaut, J., Fortin, N., Bocher, E., Petit, G., Aumond, P., & Guillaume, G. (2019). An open-science crowdsourcing approach for producing community noise maps using smartphones. *Building and Environment*, *148*, 20–33. https://doi.org/10.1016/j.buildenv.2018.10.049

Pichon, L., Brunel, G., Payan, J. C., Taylor, J., Bellon-Maurel, V., & Tisseyre, B. (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), 608–626. https://doi.org/10.1007/s11119-021-09797-9

Pichon, L., Leroux, C., Geraudie, V., Taylor, J., & Tisseyre, B. (2019). Investigating the harmonization of highly noisy heterogeneous datasets hand-collected over the same study domain. In *Precision Agriculture 2019 - Papers Presented at the 12th European Conference on Precision Agriculture, ECPA 2019*. https://doi.org/10.3920/978-90-8686-888-9_91

Pons, P. J., Truyols, M., Flexas, J., Cifre, J., & Medrano, H. (2008). Sap flow technique as a tool for irrigation schedule in grapevines : Control of the plant physiological status. *Drought management: scientific and technological innovations : CIHEAM*, *378*(80), 375–378.

Prieto, J. A., Lebon, É., & Ojeda, H. (2010). Stomatal behavior of different grapevine cultivars in response to soil water status and air water vapor pressure deficit. *Journal International des Sciences de la Vigne et du Vin*, *44*(1), 9–20. https://doi.org/10.20870/oeno-one.2010.44.1.1459

Prudic, K. L., McFarland, K. P., Oliver, J. C., Hutchinson, R. A., Long, E. C., Kerr, J. T., & Larrivée, M. (2017). eButterfly: Leveraging massive online citizen science for butterfly consevation. *Insects*, *8*(2), 53. https://doi.org/10.3390/insects8020053

QGIS Development Team, (2020). QGIS Geographic Information System. Open Source Geospatial Foundation. URL: http://qgis.osgeo.org

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org

Raddick, M.J., Bracey, G., Gay, P.L., Lintott, C.J., Murray, P., Schawinski, K., Szalay, A.S., Vandenberg, J., (2010). Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review, 9* (1), 010103. http://dx.doi.org/10.3847/AER2009036

Rahmati, A., & Zhong, L. (2013). Studying smartphone usage: Lessons from a four-month field study. *IEEE Transactions on Mobile Computing*, *12*(7), 1417–1427. https://doi.org/10.1109/TMC.2012.127

Raykar, V. C., & Yu, S. (2012). Eliminating Spammers and Ranking Annotators for Crowdsourced Labeling Tasks. *Journal of Machine Learning Research*, *13*, 491–518. https://doi.org/10.1.1.413.3825

Rechenberger, T., Jung, V. M. E., Schmidt, N., & Rosenkranz, C. (2015). Utilizing the Crowd – A Literature Review on Factors influencing Crowdsourcing Initiative Success. In *PACIS 2015 Proceedings. 250* (p. 250). http://aisel.aisnet.org/pacis2015%0Ahttp://aisel.aisnet.org/pacis2015/250

Reinermann, S., Gessner, U., Asam, S., Kuenzer, C., & Dech, S. (2019). The effect of droughts on vegetation condition in Germany: An analysis based on two decades of satellite earth observation time series and crop yield statistics. *Remote Sensing*, *11*(15), 1783. https://doi.org/10.3390/rs11151783

Rey D. and Neuhäuser M. (2011). Wilcoxon-Signed-Rank Test. In: M. Lovric (ed) *International Encyclopedia of Statistical Science*. Berlin, Heidelberg, Germany: Springer. https://doi.org/10.1007/978-3-642-04898-2_616

Reynolds, A. G., & Willwerth, J. J. (2020). Spatial variability in Ontario Riesling Vineyards : I . Soil , vine water status and vine performance. *Oeno One*, *54*(2), 327–349. https://doi.org/https://doi.org/10.20870/oeno-one.2020.54.2.2401

Rienth, M., & Scholasch, T. (2019). State-of-the-art of tools and methods to assess vine water status. *Oeno One*, *53*(4), 619–637. https://doi.org/10.20870/oeno-one.2019.53.4.2403

Rodrigues, F., Pereira, F., & Ribeiro, B. (2013). Learning from multiple annotators: Distinguishing good from random labelers. *Pattern Recognition Letters*, *34*(12), 1428–1436. https://doi.org/10.1016/j.patrec.2013.05.012

Rodriguez Lovelle, B., Trambouze, W., & Jacquet, O. (2009). Évaluation de l´état de croissance végétative de la vigne par la méthode des apex (Evaluation of vine vegetative growth by the apex method). *Progrès Agricole et Viticole*, *126*, 77–88.

Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., & Laredo, J. A. (2013). CrisisTracker: Crowdsourced social media curation for disaster awareness. *IBM Journal of Research and Development*, *57*(5), 4:1-4:13. https://doi.org/10.1147/JRD.2013.2260692

Román Dobarco, M., Arrouays, D., Lagacherie, P., Ciampalini, R., & Saby, N. P. A. (2017). Prediction of topsoil texture for Region Centre (France) applying model ensemble methods. *Geoderma*, *298*, 67–77. https://doi.org/10.1016/j.geoderma.2017.03.015

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., et al. (2012). Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 217–226). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2145204.2145238

Roux, S., Gaudin, R., & Tisseyre, B. (2019). Why does spatial extrapolation of the vine water status make sense? Insights from a modelling approach. *Agricultural Water Management*, *217*(June 2018), 255–264. https://doi.org/10.1016/j.agwat.2019.03.013

Ryan, S. F., Adamson, N. L., Aktipis, A., Andersen, L. K., Austin, R., Barnes, L., et al. (2018). The role of citizen science in addressing grand challenges in food and agriculture research. In *Proceedings of the Royal Society B* (Vol. 285). Royal Society. https://doi.org/http://dx.doi.org/10.1098/rspb.2018.1977

Salembier, C., Segrestin, B., Sinoir, N., Templier, J., Weil, B., & Meynard, J. M. (2020). Design of equipment for agroecology: Coupled innovation processes led by farmer-designers. *Agricultural Systems*, *183*, 102856. https://doi.org/10.1016/j.agsy.2020.102856

Santesteban, L. G., Guillaume, S., Royo, J. B., & Tisseyre, B. (2012). Are precision agriculture tools and methods relevant at the whole-vineyard scale? *Precision Agriculture*, *14*(1), 2–17. https://doi.org/10.1007/s11119-012-9268-3

Saoud, Z., Fontaine, C., Loïs, G., Julliard, R., & Rakotoniaina, I. (2020). Miss-identification detection in citizen science platform for biodiversity monitoring using machine learning. *Ecological Informatics*, *60*, 101176. https://doi.org/10.1016/j.ecoinf.2020.101176

Schlieder, C., & Yanenko, O. (2010). Spatio-Temporal Proximity and Social Distance: A Confirmation Framework for Social Reporting. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks* (pp. 60–67). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1867699.1867711

Scholander, P. F., Hammel, H. T., Brandstreet, E. T., & Hemmingsen, E. (1965). Sap pressure in vascular plants. *Science*, *148*, 339–346. https://doi.org/https://doi.org/10.1126/science.148.3668.339

Scholasch, T., & Rienth, M. (2019). Review of water deficit mediated changes in vine and berry physiology; Consequences for the optimization of irrigation strategies. *Oeno One*, *53*(3), 409–422. https://doi.org/https://doi.org/10.20870/oeno-one.2019.53.3.2407

Schultz, H. R. (2017). Issues to be considered for strategic adaptation to climate evolution Is atmospheric evaporative demand changing ? *Oeno One*, *51*(2), 107–114. https://doi.org/10.20870/oeno-one.2016.0.0.1619

Schultz, H. R., & Matthews, M. A. (1988). Vegetative growth distribution during water deficits in Vitis vinifera L. *Australian Journal of Plant Physiology*, *15*, 641–656. https://doi.org/https://doi.org/10.1071/PP9880641

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., et al. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*, *5*(5), 55. https://doi.org/10.3390/ijgi5050055

Senaratne, H., Mobasheri, A., Ali, A. L., & Capineri, C. (2016). A Review of Volunteered Geographic Information Quality Assessment Methods. *International Journal of Geographical Information Science*, *31*(1), 139–167. https://doi.org/10.1080/13658816.2016.1189556

Severinsen, J., de Roiste, M., Reitsma, F., & Hartato, E. (2019). VGTrust: measuring trust for volunteered geographic information. *International Journal of Geographical Information Science*, *33*(8), 1683–1701. https://doi.org/10.1080/13658816.2019.1572893

Shepard, D. (1968). A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. In *Proceedings of the 1968 23rd ACM National Conference* (pp. 517–524). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/800186.810616

Simoes, M., & Peterson, A. T. (2018). Utility and limitations of climate-matching approaches in detecting different types of spatial errors in biodiversity data. *Insect Conservation and Diversity*, *11*, 407–414. https://doi.org/https://doi.org/10.1111/icad.12288

Slimi, C., Prost, M., Cerf, M., & Prost, L. (2021). Exchanges among farmers' collectives in

support of sustainable agriculture: From review to reconceptualization. *Journal of Rural Studies*, *83*, 268–278. https://doi.org/https://doi.org/10.1016/j.jrurstud.2021.01.019

Starczewski, A., Goetzen, P., & Er, M. J. (2020). A New Method for Automatic Determining of the DBSCAN Parameters. *Journal of Artificial Intelligence and Soft Computing Research*, *10*(3), 209–221. https://doi.org/10.2478/jaiscr-2020-0014

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird : A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

Surowiecki, J. (2005). *The Wisdom of Crowds* (Anchor.). Anchor.

Taylor, J. A., Acevedo-Opazo, C., Ojeda, H., & Tisseyre, B. (2009). Identification and significance of sources of spatial variation in grapevine water status. *Australian Journal of Grape and Wine Research*, *16*(1), 218–226. https://doi.org/10.1111/j.1755-0238.2009.00066.x

Taylor, J.A., Tisseyre, B., Acevedo-Opazo, C., & Lagacherie, P. (2009). Field-scale model of the spatio-temporal vine water status in a viticulture system. In *Precision Agriculture* (pp. 537–544). Wageningen Academic Publishers.

Taylor, James A., Acevedo-Opazo, C., Pellegrino, A., Ojeda, H., & Tisseyre, B. (2012). Can within-season grapevine predawn leaf water potentials be predicted from meteorological data in non-irrigated mediterranean vineyards? *Journal International des Sciences de la Vigne et du Vin*, *46*(3), 221–232. https://doi.org/https://doi.org/10.20870/oeno-one.2012.46.3.1521

Tisseyre, B., Ojeda, H., Carillo, N., Deis, L., & Heywang, M. (2005). Information and Technology for Sustainable Fruit and Vegetable Production Precision Viticulture and Water Status : Mapping the Predawn Water Potential to Define within Vineyard Zones. In H. R. Shultz (Ed.), *Proceedings of 14th GIESCO congress* (pp. 12–16). Montpellier, France.

Tisseyre, B., Ojeda, H., & Taylor, J. A. (2007). New technologies and methodologies for site-specific viticulture. *Journal International des Sciences de la Vigne et du Vin*, *41*(2), 63–76. https://doi.org/10.20870/oeno-one.2007.41.2.852

Touya, G., Antoniou, V., Olteanu-Raimond, A. M., & Van Damme, M. D. (2017). Assessing crowdsourced POI quality: Combining methods based on reference data, history, and spatial relations. *ISPRS International Journal of Geo-Information*, *6*(3), 80. https://doi.org/10.3390/ijgi6030080

Trambouze W., Rodriguez-Lovelle B. and Jacquet O., (2009). The shoot tip method: a practical tool to evaluate the vine growth status. In: *Viticulture & Enology, XVIth international Giesco Symposium*. Davis, California.

Treude, C., Barzilay, O., & Storey, M. A. (2011). How do programmers ask and answer questions on the web? (NIER track). In *Proceedings - International Conference on Software Engineering* (pp. 804–807). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/1985793.1985907

Trivellone, V., Filippin, L., Narduzzi-Wicht, B., & Angelini, E. (2016). A regional-scale survey to define the known and potential vectors of grapevine yellow phytoplasmas in vineyards South of Swiss Alps. *European Journal of Plant Pathology*, *145*(4), 915–927.

https://doi.org/10.1007/s10658-016-0880-3

Van de Gevel, J., Van Etten, J., & Deterding, S. (2020). Citizen science breathes new life into participatory agricultural research. A review. *Agronomy for Sustainable Development*, *40*(5). https://doi.org/10.1007/s13593-020-00636-1

Van Etten, J., Beza, E., Calderer, L., Van Duijvendijk, K., Fadda, C., Fantahun, B., et al. (2019). First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (Tricot). *Experimental Agriculture*, *55*(S1), 275–296. https://doi.org/10.1017/S0014479716000739

Van Etten, J., de Sousa, K., Aguilar, A., Barrios, M., Coto, A., Dell'Acqua, M., et al. (2019). Crop variety management for climate adaptation supported by citizen science. In B. L. Turner (Ed.), *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 116, pp. 4194–4199). Tempe, USA. https://doi.org/10.1073/pnas.1813720116

Van Leeuwen, C., Pichon, L., & Prévot, L. (2019). *Vers un observatoire national de la contrainte hydrique : Quels outils ?* Gruissan: LACCAVE - INRAE.

Van Leeuwen, C., Roby, J. P., & De Rességuier, L. (2018). Soil-related terroir factors : a review. *Oeno*, *52*(2), 173–188. https://doi.org/10.20870/oeno-one.2018.52.2.2208

Van Leeuwen, C., & Seguin, G. (1994). Incidence de l'alimentation en eau de la vigne, appreciee par l'etat hydrique du feuillage, sur le developpement de l'appareil vegetatif et la maturation du raisin (Vitis vinifera variété Cabernet franc, Saint-Emilion 1990). *Journal international des sciences de la vigne et du vin*, *28*(2), 81. https://doi.org/10.20870/oeno-one.1994.28.2.1152

Van Leeuwen, C., Tregoat, O., Choné, X., Bois, B., Pernet, D., & Gaudillére, J. P. (2009). Vine water status is a key factor in grape ripening and vintage quality for red bordeaux wine. How can it be assessed for vineyard management purposes? *Journal International des Sciences de la Vigne et du Vin*, *43*(3), 121–134. https://doi.org/https://doi.org/10.20870/oeno-one.2009.43.3.798

Vieira, S. R., Tillotson, P. M., Biggar, J. W., & R., N. D. (1997). Scaling of semivariograms and the kriging estimation of field-measured properties. *Revista Brasileira de Ciência do Solo*, *21*(June), 525–533. https://doi.org/https://doi.org/10.1590/S0100-06831997000400001

Vincent, C., Isaacs, R., Bostanian, N. J., & Lasnier, J. (2012). Principles of Arthropod Pest Management in Vineyards. In I. R. Bostanian N., Vincent C. (Ed.), *Arthropod Management in Vineyards*. 309–338. Dordrecht, The Netherland: Springer. https://doi.org/10.1007/978-94-007-4032-7_1

Walter, T. P., & Back, A. (2011). Towards Measuring Crowdsourcing Success: An Empirical Study on Effects of External Factors in Online Idea Contests. In *6th Mediterranean Conference on Information Systems (MCIS 2011)* (pp. 1–12). Atlanta, USA: AIS Electronic Library (AISeL).

Wang, H., Bah, M. J., & Hammad, M. (2019). Progress in Outlier Detection Techniques: A Survey. *IEEE Access*, *7*, 107964–108000. https://doi.org/10.1109/ACCESS.2019.2932769

Weddagala, W. M. T. B., Basnayake, B. M. R. L., Wijesekara, H. M. L., Dharmathilaka, N. R. D. S., Kiriveldeniya, K. K. A., Karunaratne, A. S., & De Silva, D. A. M. (2020). Crowd

sourcing for value chain management: a case of market decision support system in Sri Lankan agriculture market. In *Acta Horticulturae* (pp. 173–178). International Society for Horticultural Science (ISHS), Leuven, Belgium. https://doi.org/10.17660/ActaHortic.2020.1278.26

Westerman, W. (2009). Epistemology, the sociology of knowledge, and the wikipedia userbox controversy. In *Folklore and the Internet* (pp. 123–158). https://doi.org/https://doi.org/10.2307/j.ctt4cgrx5.9

Wiggins, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011). Mechanisms for Data Quality and Validation in Citizen Science. In *2011 Seventh IEEE International Conference on e-Science Workshops* (pp. 14–19). Stockholm, Sweden. https://doi.org/10.1109/eScienceW.2011.27

Yan, Y., Feng, C.-C., & Chang, K. (2017). Towards Enhancing Integrated Pest Management Based on Volunteered Geographic Information. *ISPRS International Journal of Geo-Information*, *6*(7), 224. https://doi.org/10.3390/ijgi6070224

Yzarra, W., Sanabria, J., Cáceres, H., Solis, O., & Lhomme, J.-P. (2015). Impact of climate change on some grapevine varieties grown in Peru for Pisco production. *Oeno One*, *49*(2), 103–112. https://doi.org/https://doi.org/10.20870/oeno-one.2015.49.2.90

Zhao, Yijiang, Zhou, X., Li, G., & Xing, H. (2016). A Spatio-Temporal VGI Model Considering Trust-Related Information. *ISPRS International Journal of Geo-Information*, *5*(2), 10. https://doi.org/10.3390/ijgi5020010

Zhao, Yuxiang, & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*, *16*(3), 417–434. https://doi.org/10.1007/s10796-012-9350-4

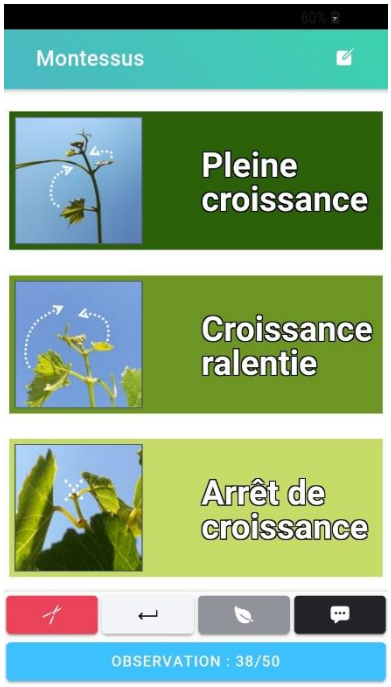
Zhou, X., Chen, T., Guo, D., Teng, X., Yuan, B. (2018). From one to crowd: a survey on crowdsourcing-based wireless indoor localization. *Frontiers in Computer Sciences, 12*, 423–450. https://doi.org/10.1007/s11704-017-6520-z

Zufferey, V., Spring, J. L., Verdenal, T., Dienes, A., Belcher, S., Lorenzini, F., et al. (2017). Influence of water stress on plant hydraulics, gas exchange, berry composition and quality of pinot noir wines in Switzerland. *Oeno One*, *51*(1), 37–57. https://doi.org/10.20870/oeno-one.2017.51.1.1314

Chapter 9: Appendices

9.1 Screenshots of the latest version of ApeX-Vigne mobile application (release of 2021 May 7th)

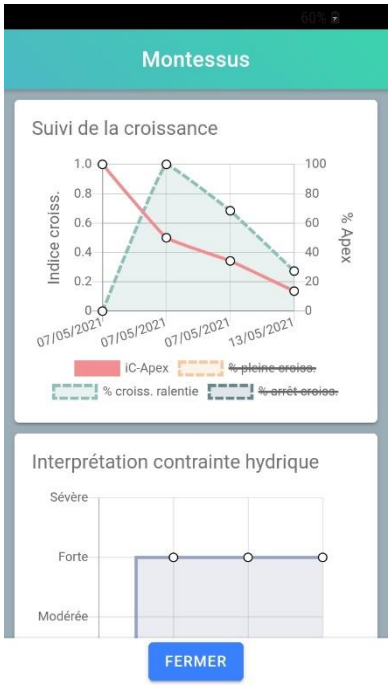
Input screen:



Output screen:



Screen for consulting field details after observation:



Screen for selecting development stage at the gathering of observation:



9.2 Press release sent to the wine press for the official launch of the ApeX-Vigne application

Communiqué de presse

ApeX Vigne : Une application mobile gratuite pour faciliter le suivi de la croissance des vignes et estimer la contrainte hydrique

L'Institut Français de la Vigne et du Vin (IFV) et Montpellier SupAgro, soutenus par la région Occitanie, lancent le 3 juin 2019 une application mobile gratuite basée sur la méthode des apex pour suivre l'évolution de la croissance des vignes et estimer le niveau de contrainte hydrique.

La méthode des apex est une méthode simple pour caractériser la croissance végétative de la vigne. Elle est basée sur l'observation de l'extrémité des rameaux, que l'on appelle les apex. Cette méthode existe depuis plusieurs années. Elle est couramment utilisée au vignoble par les techniciens, particulièrement en région méditerranéenne où l'on essaye de relier l'arrêt précoce de la croissance à des niveaux de contrainte hydrique importants. En revanche, sa diffusion a été jusqu'à présent freinée par la difficulté de calculer facilement un indice de croissance végétative directement à la parcelle pour une interprétation instantanée. C'est pour répondre à cet enjeu que l'IFV et Montpellier SupAgro ont développé l'application mobile ApeX Vigne.

ApeX Vigne peut se télécharger gratuitement sur le magasin d'applications. Elle permet grâce à un système de comptage simple de dénombrer les apex puis de calculer un indice de croissance végétative. L'application propose ensuite une synthèse et un suivi de l'évolution de la croissance des apex pour chaque parcelle. Enfin, un module de l'application, élaboré par l'IFV, propose une interprétation du niveau de contrainte hydrique en fonction des observations réalisées. Les informations sont stockées de manière sécurisée. L'utilisateur peut ensuite les récupérer pour des analyses complémentaires ou les partager avec d'autres personnes.

Cette application peut ainsi être utilisée par un viticulteur sur ses propres parcelles, par un conseiller sur les parcelles qu'il suit ou par un réseau de conseillers à l'échelle par exemple d'une coopérative ou d'une appellation.

L'application ApeX Vigne répond à une forte demande de la profession de pouvoir bénéficier d'outils simples de caractérisation de l'état des vignes. Dans un contexte de changement climatique, elle peut également offrir des clés de compréhension du phénomène de contrainte hydrique à l'échelle parcellaire comme territoriale.

Téléchargement sur le google play store : https://play.google.com/store/apps/details?id=ag.GB.apex

Les porteurs du projet

Le projet d'application ApeX Vigne est porté par l'Institut Français de la Vigne et du Vin et Montpellier SupAgro. Il est soutenu par la région Occitanie.

Contacts :

Jean-Christophe Payan - jean-christophe.payan@vignevin.com – 04 66 20 67 05

Léo Pichon – leo.pichon@supagro.fr – 06 83 69 15 15

9.3 Other peer-reviewed articles published during the thesis

Received: 3 May 2019 | Accepted: 26 February 2020 | Published: 10 March 2020
DOI:10.20870/oeno-one.2020.54.1.2481

OENO One
VINE AND WINE OPEN ACCESS JOURNAL

Using smartphone leaf area index data acquired in a collaborative context within vineyards in southern France

Leo Pichon*, James Taylor and Bruno Tisseyre

UME ITAP, Univ Montpellier, Montpellier SupAgro, Inrae, France

*corresponding author: leo.pichon@supagro.fr

ABSTRACT

Aim: The objective of this short communication is to study the opportunity of using a smartphone application for leaf area index (LAI) observations within vineyards in southern France in a farmsourcing context, i.e. when several operators make parallel acquisitions over the same area. In this context, several sources of variability are likely to affect measurement quality, such as the smartphone model or the operator. Understanding these sources of variability will enhance the ability to properly interpret LAI observations to produce relevant information for decision-making.
Methods and material: A study was specifically designed to evaluate the ability of a smartphone application to differentiate sites with different LAI and to determine the origin and the relative importance of different sources of variation in a context of farmsourcing data collection. This focused on the VitiCanopy application, which has been developed specifically for viticulture LAI measurements. Measurements were performed by 18 operators with 11 different smartphone models, on three different vines presenting controlled canopy size to evaluate the ability of the smartphone application to differentiate sites under varying acquisition conditions. Controlled repetitions over seven vines by seven operators with seven smartphone models were performed to further determine the sources of variation and their relative importance.
Results: LAI estimations made with VitiCanopy were consistent with the different levels of controlled vine size in the experiment. The operator and the smartphone model had a significant effect on the variance of the estimated LAI. The variance caused by the observation protocol was relatively low compared to the variability between plants within the observation site (seven vines).
Conclusions: This study showed that the VitiCanopy application was relevant for ordering or classifying vines according to LAI. In an operational context, the results of this study support the use of this smartphone application for relative measurements. However, the best results were achieved when smartphone model differences were minimised or avoided and with homogeneous acquisition conditions between operators. This last condition will require the organisation of group training sessions to minimise an observed operator effect on measurement variability.
Significance and impact of the study: This short communication demonstrated the potential of LAI observations collected with smartphones by several operators for decision-making in a context of farmsourcing. The results showed that this new source of observations, which is inexpensive to collect, made it possible to characterise vine size (LAI) differences in vineyards of southern France. This shows the potential of this app for large production areas such as cooperatives. Further investigations are needed to understand how different training systems may affect the measurement. This source of observations could be complementary to other information sources that are more precise or more accurate, but also more expensive (i.e. destructive methods).

KEYWORDS

precision viticulture, crowdsourcing, data quality, precision, accuracy

https://doi.org/10.20870/oeno-one.2020.54.1.2481

Precision Agriculture (2019) 20:1211–1230
https://doi.org/10.1007/s11119-019-09650-0

Automatic harmonization of heterogeneous agronomic and environmental spatial data

Corentin Leroux[1,2] · Hazaël Jones[2,3] · Léo Pichon[2] · James Taylor[2] · Bruno Tisseyre[2]

Published online: 20 March 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
The analysis and mapping of agronomic and environmental spatial data require observations to be comparable. Heterogeneous spatial datasets are those for which the observations of different datasets cannot be directly compared because they have not been collected under the same set of acquisition conditions, for instance within the same time period (if the variable of interest varies across time), with consistent sensors or under similar management practices (if the management practices impact the measured value) among others. When heterogeneous acquisition conditions take place, there is a need for harmonization procedures to make possible the comparison of such observations. This analysis details and compares four automated methodologies that could be used to harmonize heterogeneous spatial agricultural datasets so that the data can be analysed and mapped conjointly. The theory and derivation of each approach, including a novel, local spatial approach is given. These methods aim to minimize the occurrence of discrepancies (discontinuities) in the data. The four approaches were evaluated and compared with a sensitivity analysis on simulated datasets with known characteristics. Results showed that none of the four methods consistently delivered a better harmonization accuracy. The accuracy and preferred choice for the harmonization procedures was shown to be influenced by (i) within-field spatial structures of the datasets, (ii) differences in acquisition conditions between the heterogeneous spatial datasets, and (iii) the spatial resolution of the simulated data. The four approaches were used to harmonize real within-field grain yield datasets and a discussion to help users select an appropriate harmonization methodology proposed. Despite significant improvements in dataset harmonization, discontinuities were not entirely removed and some uncertainty remained.

Keywords Data harmonization · Discontinuity · Spatial autocorrelation · Yield

https://doi.org/10.1007/s11119-019-09650-0

Precision agriculture '19



Editor: John V. Stafford

Published: 2019 **Pages:** 1030

eISBN: 978-90-8686-888-9 | **ISBN:** 978-90-8686-337-2

https://doi.org/10.3920/978-90-8686-888-9

Book Type: Conference Proceedings

Investigating the harmonization of highly noisy heterogeneous datasets hand-collected over the same study domain

L. Pichon ⓘ, C. Leroux ⓘ, V. Geraudie ⓘ, J. Taylor ⓘ, B. Tisseyre ⓘ

Pages: 735 - 741

https://doi.org/10.3920/978-90-8686-888-9_91

Published Online: July 08, 2019

Abstract | References | Full-text

Abstract:

The objective of this paper is to propose an approach to harmonise noisy spatial data acquired by different operators using (low-cost) hand-held sensors over the same spatial domain. In such cases, datasets need to be harmonised first to be comparable before decision making. This work proposes a methodology to address this issue in the case of nested and noisy spatial data. First, it proposes the implementation of a non-parametric test of Kolmogorov-Smirnov to determine if harmonisation is needed. Then, it proposes an aspatial harmonization method based on standardization. The method was applied on grape sugar content datasets collected by 2 hand-held spectrometers. Results showed that harmonizing a less confident dataset is interesting solely if the size of the trusted one is too small.

Keywords: sugar content, spectrometer, precision viticulture, crowdsourcing

https://doi.org/10.3920/978-90-8686-888-9_91

Precision agriculture '19



Editor: John V. Stafford

Published: 2019 **Pages:** 1030

eISBN: 978-90-8686-888-9 | **ISBN:** 978-90-8686-337-2

https://doi.org/10.3920/978-90-8686-888-9

Book Type: Conference Proceedings

A collective framework to assess the adoption of precision agriculture in France: description and preliminary results after two years

N. Lachia ⓘ, L. Pichon ⓘ, B. Tisseyre ⓘ

Pages: 851 - 857

https://doi.org/10.3920/978-90-8686-888-9_105

Published Online: July 08, 2019

Abstract | References | Full-text

Abstract:

The article presents the results of a project that led to a collective organisation designed to assess and follow the level of adoption of precision agriculture in France. This organisation, called the French Observation Centre of Digital Agriculture adoption (FrOCDA) allowed the design and the set-up of an original and exhaustive study method mainly based on interviews of the main players in precision farming in France. The article presents the results of three studies. The first study dealt with remote sensing. It showed that the adoption of this technology is relatively high in field crops (about 1,000,000 ha) and more marginal in viticulture. The second dealt with soil mapping. The mapping remains marginal since it concerns 130,000 ha mainly in viticulture. Finally, the use of agricultural robotics in France will be presented.

Keywords: adoption, consortium, national survey, technology

https://doi.org/10.3920/978-90-8686-888-9_105

Precision Agriculture (2019) 20:278–294
https://doi.org/10.1007/s11119-019-09634-0

What relevant information can be identified by experts on unmanned aerial vehicles' visible images for precision viticulture?

Leo Pichon[1] · Corentin Leroux[1,2] · Catherine Macombe[1] · James Taylor[1] · Bruno Tisseyre[1]

Published online: 29 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
Unmanned aerial vehicles (UAV) offer interesting alternatives to satellites or airplanes regarding flight agility and image resolution. These sensor platforms may well be used to monitor vines field all throughout the vine's growing season at a very high spatial resolution. They could provide useful information, different to that normally considered in the literature. To identify the possible uses of UAV images in viticulture, a specific exploratory survey was put into place. This study aimed at identifying (i) relevant information that growers and advisers (G&A) can extract from UAV images and (ii) the added value that this information can have for both G&A's vineyard management decisions. This approach was conducted on an 11.3 ha commercial vineyard with soil, climate and a training system representative of vineyards in the south of France. UAV-based visible images (25 mm resolution) were acquired every two weeks from budburst to harvest by several UAV companies. Images were shown to a panel of G&As over six sessions during the growing season. Each of these sessions consisted of (i) an individual period during which images were first presented one at a time to each expert and then all together in chronological order from budburst to harvest, and (ii) a collective period during which G&As were asked to share and discuss their point of view. In this exploratory survey, the application of the proposed methodology demonstrated that most of the information on vine status, soil and vineyard environment could be extracted from UAV-based visible images by the experts, thus showing high interest in developing specific image processing techniques to extract this information from images. Results showed that this information was of great interest throughout the growing cycle of the vine, particularly for advisers, as a support to drive management strategies. The outputs of this exploratory study should be confirmed in other contexts than the Languedoc, France region to extrapolate the observed conclusions.

Keywords Industry view · Visible images · Decision making · Expertise gathering

https://doi.org/10.1007/s11119-019-09634-0

Precision Agric (2018) 19:957–971
https://doi.org/10.1007/s11119-018-9566-5

How to define the optimal grid size to map high resolution spatial data?

B. Tisseyre[1] · C. Leroux[1,3] · L. Pichon[1] · V. Geraudie[2] · T. Sari[1]

Published online: 10 March 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract The development and the release of sensors capable of providing data with high spatial resolution (>4 000 points ha^{-1}) in agriculture raises new questions as to how to represent this spatial information. The objective of this study was to propose a methodology to help define the optimal grid size to map high resolution data in agriculture. The geostatistical method finds the grid size which maximizes the sum of two components: (i) the proportion of nugget variance that is removed, and (ii) the proportion of sill variance that remains in the data. The optimum grid size was found to be dependent on the resolution of the available information and the spatial structure of the raw data. Experiments on simulated datasets with varying data resolution (from 500 to 2 000 pts.ha^{-1}) and spatial structure (range of variogram between 10 and 45 m) showed that the proposed methodology was able to define varying optimal grid sizes (from 5 to 12 m). The proposed geostatistical approach was then applied on a real dataset of total soluble solids/sugar content of table grape so that the optimal mapping grid size could be found. Once it was defined, two interpolation methods: simple averaging over blocks and block kriging, were applied to mapping the data. Results show that both methods help depict the within-field variability in the data. While the averaging procedure is easier to automate, the block kriging approach provides users with a level of uncertainty in the aggregated data. Both mapping approaches significantly impacted the within-field spatial structure: (i) the small-scale variations were ten times lower than in the raw data, and (ii) the signal-to-noise ratio of the aggregated data with the optimal grid was twice as high as that of the raw data. As the proposed geostatistical methodology is a first attempt to define the optimal grid size to map high resolution spatial data, areas for future development applications are also proposed.

Keywords Filtering · High spatial resolution · Mapping

https://doi.org/10.1007/s11119-018-9566-5

agriculture

MDPI

Article

GeoFIS: An Open Source, Decision-Support Tool for Precision Agriculture Data

Corentin Leroux [1,2,*], Hazaël Jones [1], Léo Pichon [1], Serge Guillaume [1], Julien Lamour [1,3], James Taylor [1], Olivier Naud [1], Thomas Crestey [1], Jean-Luc Lablee [1] and Bruno Tisseyre [1]

[1] ITAP (Information–Technologies-Environmental Analysis-Agricultural Processes, French Denomination), Montpellier SupAgro, Irstea, University of Montpellier, 34000 Montpellier, France; hazael.jones@supagro.fr (H.J.); leo.pichon@supagro.fr (L.P.); serge.guillaume@irstea.fr (S.G.); j.lamour@fruitiere.fr (J.L.); James.Taylor6@newcastle.ac.uk (J.T.); olivier.naud@irstea.fr (O.N.); thomas.crestey@supagro.fr (T.C.); jean-luc.lablee@irstea.fr (J.-L.L.); bruno.tisseyre@gmail.com (B.T.)

[2] SMAG, 34960 Montpellier, France

[3] Compagnie Fruitière, 13000 Marseille, France

* Correspondence: cleroux@smag-group.com; Tel.: +33-(0)3-2621-8420

Received: 25 April 2018; Accepted: 28 May 2018; Published: 30 May 2018

check for updates

Abstract: The world we live in is an increasingly spatial and temporal data-rich environment, and agriculture is no exception. However, data needs to be processed in order to first get information and then make informed management decisions. The concepts of 'Precision Agriculture' and 'Smart Agriculture' are and will be fully effective when methods and tools are available to practitioners to support this transformation. An open-source software called GeoFIS has been designed with this objective. It was designed to cover the whole process from spatial data to spatial information and decision support. The purpose of this paper is to evaluate the abilities of GeoFIS along with its embedded algorithms to address the main features required by farmers, advisors, or spatial analysts when dealing with precision agriculture data. Three case studies are investigated in the paper: (i) mapping of the spatial variability in the data; (ii) evaluation and cross-comparison of the opportunity for site-specific management in multiple fields; and (iii) delineation of within-field zones for variable-rate applications when these latter are considered opportune. These case studies were applied to three contrasting crop types, banana, wheat and vineyards. These were chosen to highlight the diversity of applications and data characteristics that might be handled with GeoFIS. For each case-study, up-to-date algorithms arising from research studies and implemented in GeoFIS were used to process these precision agriculture data. Areas for future development and possible relations with existing geographic information systems (GIS) software is also discussed.

Keywords: decision-making; GeoFIS; geostatistics; open source software; precision agriculture; spatial analysis

https://doi.org/10.3390/agriculture8060073

Résumé étendu

Chapitre 1 :

L'agriculture est en transition et fait actuellement face à des enjeux complexes et parfois contradictoires. Comment produire une nourriture saine et abondante à des prix accessibles tout en rémunérant dignement les agriculteurs ? Comment cultiver sous un climat dont les températures moyennes sont plus élevées de 1, voire 2 ou 3 degrés par rapport à la fin du siècle dernier et dans lequel les évènements climatiques extrêmes sont plus fréquents ? Comment limiter sa consommation d'énergie fossile et sa production de gaz à effet de serre ? Comment tirer parti des possibilités d'observation et de compréhension du vivant offertes par les outils numériques ? Les agriculteurs, qui sont confrontés à ces enjeux dans leur activité quotidienne, expérimentent seuls ou en groupe des solutions pour y répondre. Ils échangent régulièrement entre pairs et avec des acteurs de l'innovation et des chercheurs. Ils s'organisent collectivement pour partager leurs résultats et les confronter à ceux obtenus dans des contextes similaires. Les projets de crowdsourcing sont emblématiques de cette approche qui mêle agriculteurs et chercheurs pour générer de la connaissance et concevoir des réponses à des enjeux complexes.

Le crowdsourcing est une approche qui consiste à répondre à une question en s'appuyant sur l'intelligence collective d'une communauté de contributeurs. Il existe différents types de projets de crowdsourcing en fonction de la nature des contributions, de la thématique traitée et de la communauté scientifique qui l'étudie (Figure 53).

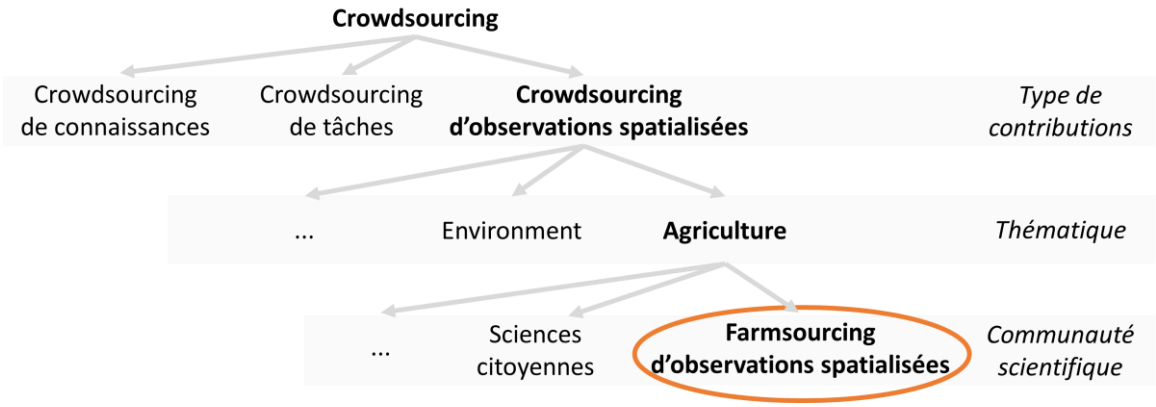


Figure 53: Du crowdsourcing au farmsourcing d'observations spatialisées : synthèse de l'imbrication des principaux concepts.

En agriculture, le crowdsourcing d'observations spatialisées joue un rôle particulièrement important. Ces projets comptent de nombreuses spécificités qui influent sur les facteurs favorisant leur succès. L'une de ces spécificités est le profil des participants, qui sont principalement des professionnels contrairement aux projets existants dans le domaine du suivi environnemental où les participants sont souvent des amateurs ou le grand public. Cette spécificité a fait émerger le terme de « farmsourcing » pour décrire les projets de crowdsourcing en agriculture. La thèse se concentre sur l'étude des projets de farmsourcing d'observations spatialisées (FOS) et les travaux menés se concentrent sur les facteurs de succès spécifiques de ces projets. Deux hypothèses principales ont été testées.

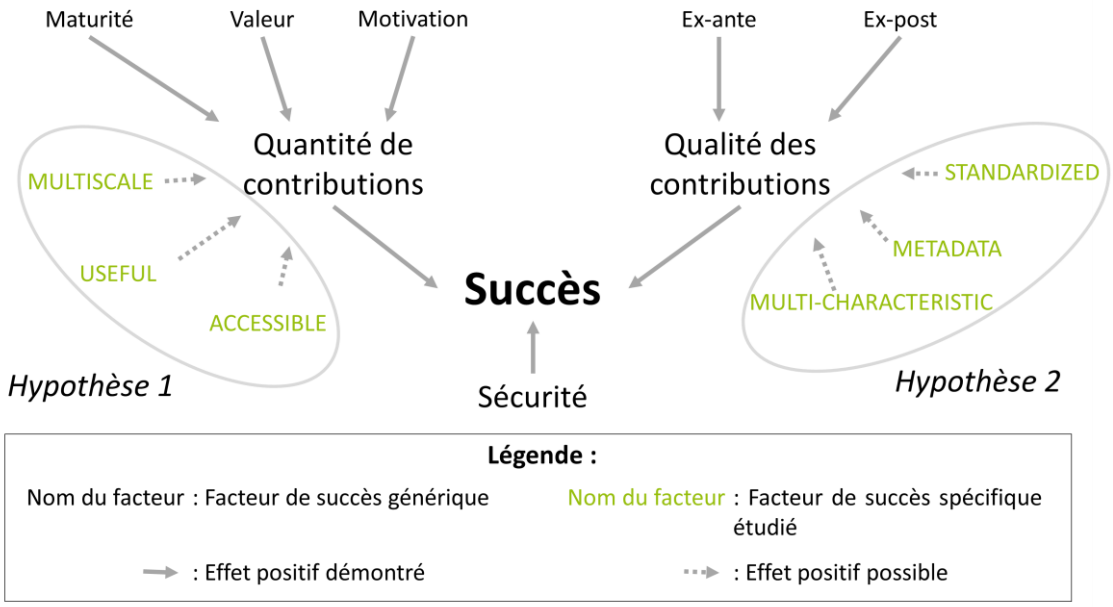


Figure 54: Facteurs de succès spécifiques des projets de FOS testés par les hypothèses 1 et 2 et relation avec les facteurs de succès génériques.

Selon la première hypothèse (Hypothèse 1) , pour générer un nombre important d'observations, les projets de FOS doivent adresser un enjeu agricole multi-échelles (MULTISCALE), proposer des tâches aux participants qui soient simples et utiles pour leur activité professionnelle (USEFUL) et s'appuyer sur une plateforme accessible au plus grand nombre (ACCESSIBLE) (Figure 54). Selon la seconde hypothèse (Hypothèse 2), pour que ces contributions soient de bonne qualité, les projets de FOS doivent s'appuyer sur un protocole de collecte standardisé (STANDARDIZED), sur l'enregistrement de métadonnées (METADATA) et sur la mise en place de stratégies de gestion de la qualité ex post basées sur les caractéristiques attributaire, spatial et temporelle des observations (MULTI-CARACTERISTIQUES) (Figure 54). Ces hypothèses ont été testées dans le cadre d'une approche systémique qui consiste à mettre en œuvre un projet de FOS réel, à tester les hypothèses au cours des différentes étapes du projet puis à généraliser les résultats à l'ensemble des projets de FOS. Pour pouvoir assurer cette généralisation, le projet doit avoir des caractéristiques représentatives de l'ensemble des projets de ce type.

Chapitre 2 :

Le chapitre 2 présente le cas d'utilisation choisi dans le cadre de cette démarche systémique : le suivi de l'état hydrique de la vigne à différentes échelles spatiales. L'état hydrique de la vigne influence à la fois la qualité de la récolte, le rendement et la quantité de végétation. Il varie fortement dans le temps et dans l'espace et présente de fortes structures spatiales et temporelles. Son suivi au cours de la saison et en de nombreux sites est donc un enjeu fort pour l'aide à la décision à l'échelle parcellaire. C'est également un enjeu à des échelles spatiales plus larges comme celles d'une aire d'apport d'une cave coopérative, d'un syndicat d'appellation, d'un département ou même d'une grande région. Il existe de nombreux outils pour mesurer ou estimer l'état hydrique de la vigne mais ils sont souvent complexes et/ou couteux à mettre en œuvre.

Chapitre 3 :

Le chapitre 3 explore les potentialités d'une méthode simple de suivi de l'état hydrique de la vigne : iG-Apex. Cette méthode s'appuie sur le fait que la croissance végétative est la première fonction physiologique affectée par la contrainte hydrique. Elle vise à caractériser cette croissance végétative en réalisant des observations visuelles simples sur 50 apex qui sont classés en 3 catégories (Figure 55). Ces observations sont ensuite synthétisées dans l'indice iG-Apex qui varie de 1 (pleine croissance) à 0 (arrêt total de croissance). Le chapitre 3 étudie la capacité de l'indice iG-Apex à servir d'outil d'aide à la décision opérationnel pour le suivi de l'état hydrique.



Figure 55: Exemples d'apex de vigne classés dans les catégories (a) « pleine croissance », (b) « croissance ralentie » et (c) « croissance arrêtée ».

L'étude a été réalisée en comparant des mesures d'iG-Apex réalisées dans des conditions de contrainte hydrique variées à des mesures de référence de potentiel hydrique foliaire de base (ψ_{PD}). Les mesures ont été réalisées sur 55 parcelles situés sur le vignoble de Tavel (Occitanie, France) durant 5 millésimes (de 2008 à 2012) et sur 4 cépages (Cinsault, Grenache, Mourvère, Syrah). Les résultats ont montré que le potentiel hydrique foliaire de base moyen (ψ_{PD}) pouvait être lié aux valeurs d'iG-Apex mesurées (Figure 56).

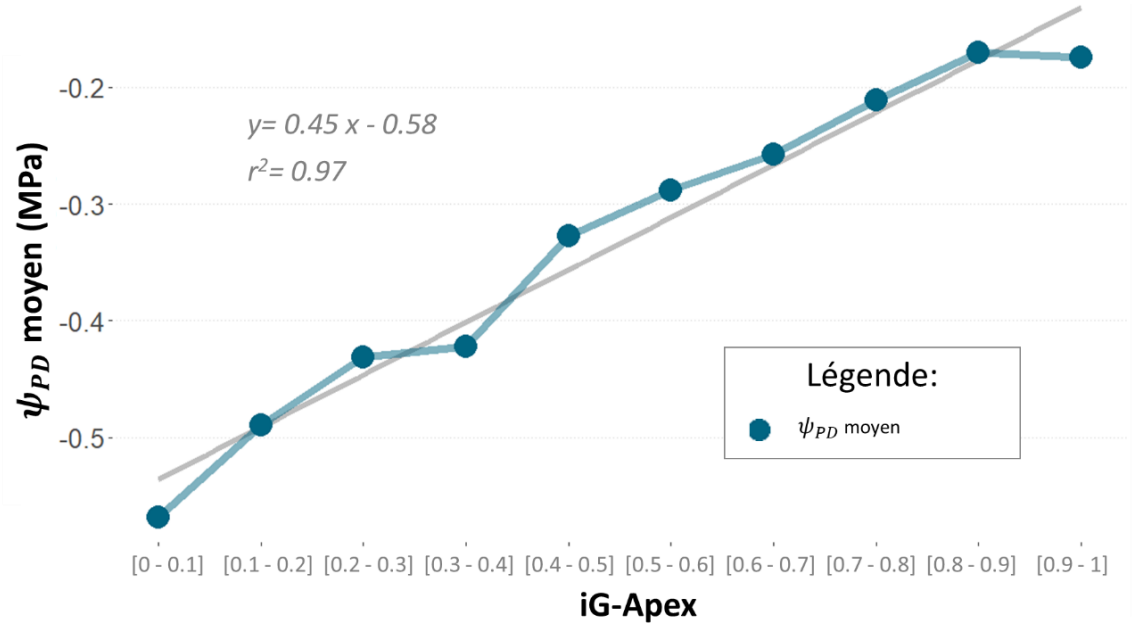


Figure 56 : Potentiel hydrique foliaire de base moyen (ψ_{PD}) en fonction des valeurs d'iG-Apex.

Les résultats ont montré que iG-Apex peut être utilisée comme un outil d'aide à la décision opérationnel à l'échelle parcellaire dans une gamme de contraintes hydriques allant de -0,2 MPa à -0,8 MPa. Néanmoins, des précautions doivent être prises lors de l'interprétation des résultats car la méthode n'est pas spécifique à l'état hydrique de la vigne et est également sensible à d'autres phénomènes. Enfin, les faibles contraintes opérationnelles de cette méthode permettent d'envisager de l'utiliser pour réaliser un suivi collectif au cours de la saison sur de grandes étendues spatiales.

Chapitre 4 :

Le chapitre 4 poursuit l'exploration des potentialités de la méthode iG-Apex en étudiant sa capacité à caractériser la structure spatiale de la croissance végétative à l'échelle intra-parcellaire. L'étude a été réalisée sur deux parcelles de vigne située dans les Corbières (Occitanie). Une centaine d'observations d'iG-Apex ont été réalisée chaque semaine sur chacune des parcelles tout au long de la saison 2020. L'étude basée sur des méthodes géostatistiques, a montré que la méthode iG-Apex peut être utilisée pour définir des zones intra-parcellaires homogènes (Figure 57) si les observations sont effectuées par un seul opérateur et pendant les périodes où la contrainte hydrique de la vigne est forte.

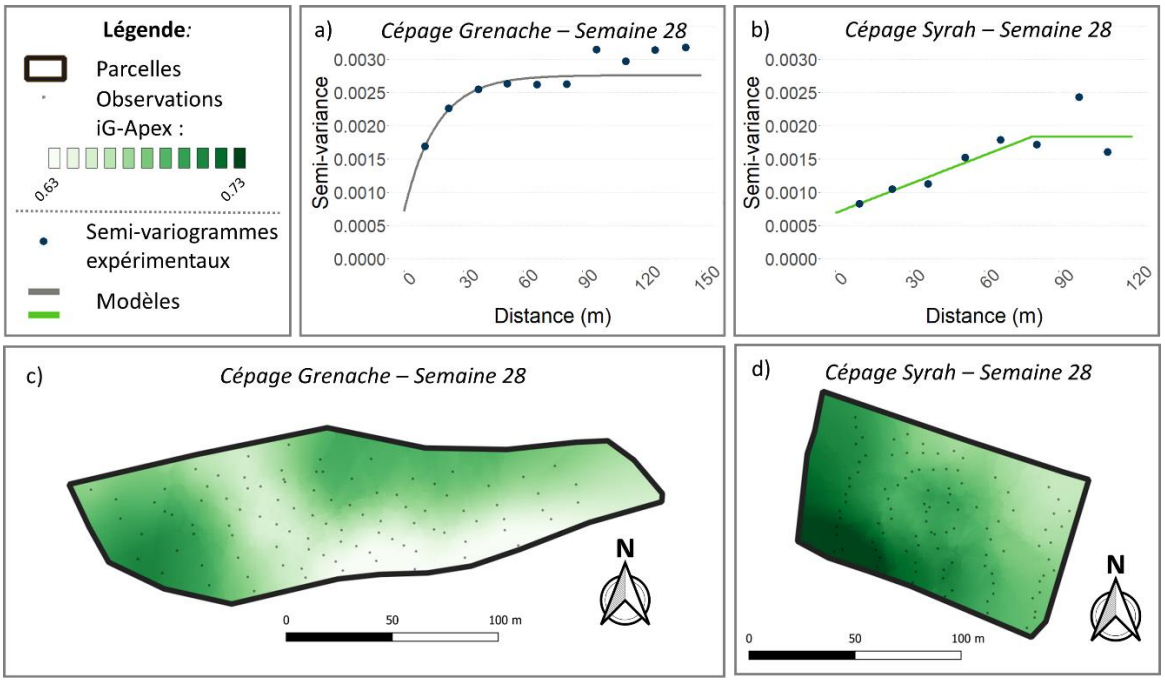


Figure 57: Les semi-variogrammes expérimentaux et modélisés des valeurs d'iG-Apex pour la semaine 28 sur les deux parcelles d'étude Grenache (a) et Syrah (b) mettent en évidence une organisation spatiale des valeurs d'iG-Apex. Les cartes correspondantes (c et d) confirment cette structure spatiale et illustrent les motifs spatiaux associés.

Chapitre 5 :

Le chapitre 5 présente les choix méthodologiques et technologiques mis en œuvre pour développer l'application mobile ApeX-Vigne. Cette application permet la mise en place du projet de FOS en proposant un outil de collecte collaborative d'observations d'iG-Apex. Ce chapitre montre que l'application ApeX-Vigne est facilement accessible et utilisable par les participants potentiels,

qu'elle permet de collecter des données selon un protocole standardisé et qu'elle enregistre des informations contextuelles (géolocalisation, date, opérateur, etc.) sur les observations recueillies (Figure 58). Les principaux utilisateurs de cette application sont les viticulteurs et les conseillers.

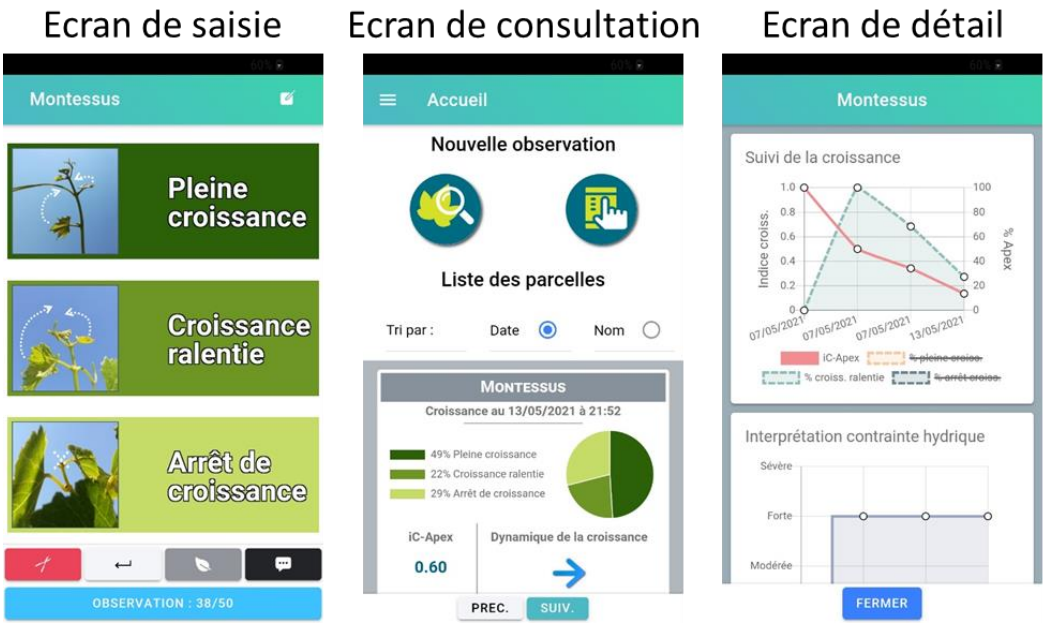


Figure 58: Principaux écrans de l'application mobile ApeX-Vigne.

Chapitre 6 :

L'objectif de ce chapitre est de tester l'hypothèse 1 en évaluant la quantité d'observation collectées avec l'application ApeX-Vigne. Ce chapitre présente les résultats obtenus avec les observations collectées lors des saisons 2019 et 2020 sur une zone d'étude englobant l'ensemble de l'arc méditerranéen français (Figure 59).

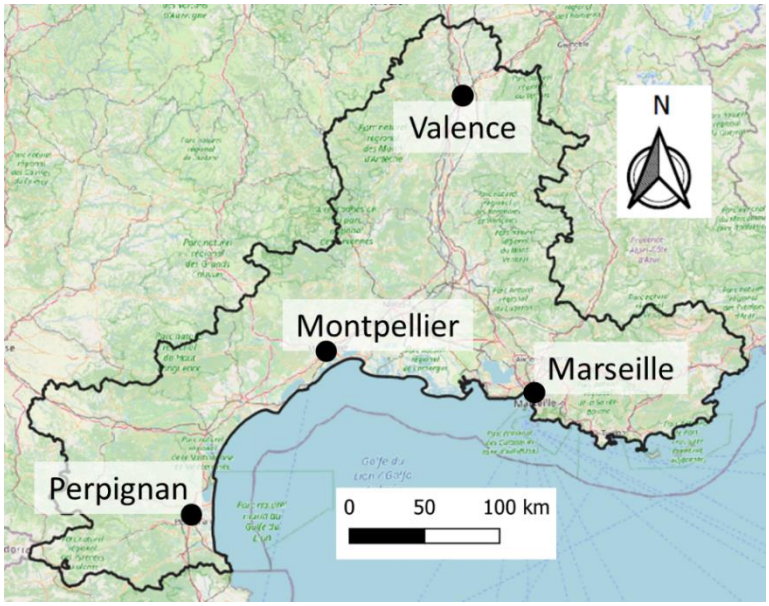


Figure 59 : Zone d'étude le long de l'arc méditerranéen français.

Les résultats montrent que l'application compte plus de 1200 téléchargements et 6000 observations effectuées au cours de la saison 2020. Le nombre de téléchargements et d'observations hebdomadaires est cyclique (Figure 60) ce qui rend les méthodes d'évaluation existantes difficiles à mettre en œuvre.

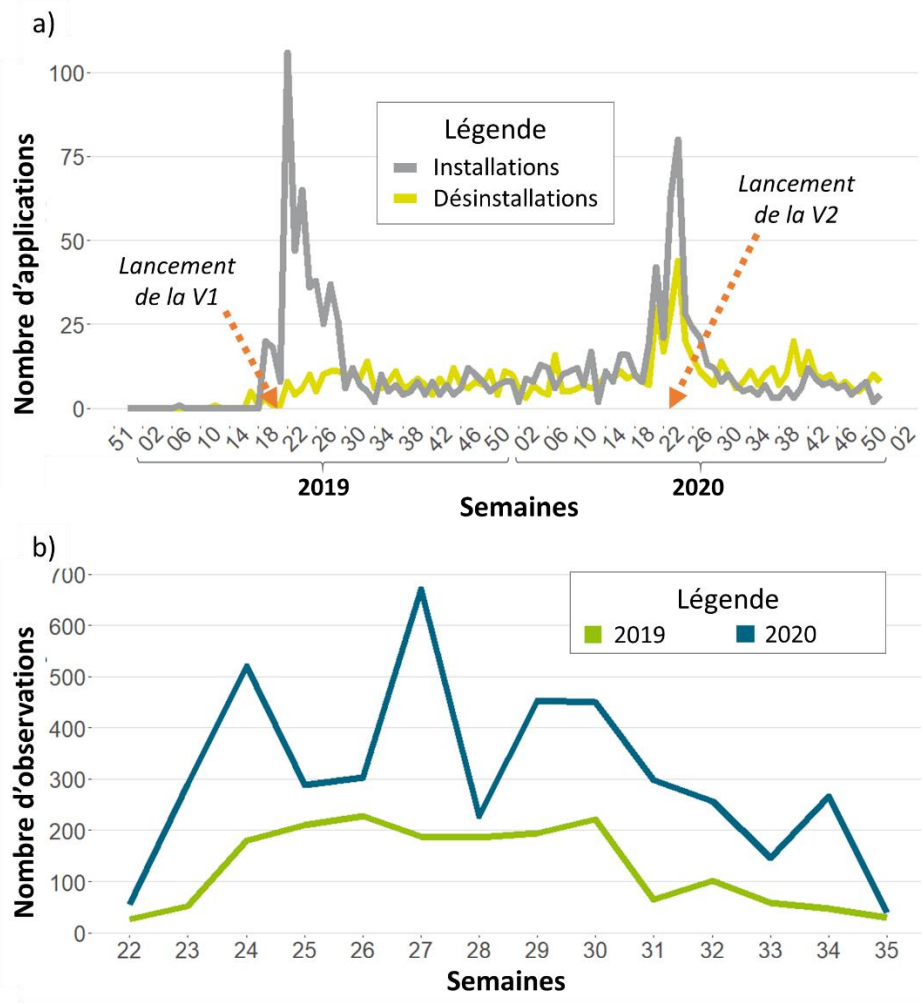


Figure 60 : Statistiques d'utilisation de l'application ApeX-Vigne. (a) Évolution du nombre d'installations et de désinstallations au cours des saisons 2019 et 2020 (lancement en semaine 21, 2019) et (b) nombre d'observations réalisées par semaine sur la période d'utilisation (semaines 22 - 35) de ces deux saisons.

Ces résultats montrent que l'application ApeX-Vigne est appréciée et adoptée par les acteurs de la filière viticole. Elle permet donc de collecter un nombre important d'observations. Ils indiquent également que les observations collectées peuvent être utilisées comme aide à la décision aux échelles intra-parcellaire, inter-parcellaire et régionale.

Chapitre 7 :

Le premier objectif de ce chapitre est de proposer une méthode de gestion de la qualité des observations collectées dans le cadre de projets de FOS. Le second objectif est de mettre en œuvre cette méthode sur les données collectées à l'aide de l'application ApeX-Vigne sur la zone d'étude en 2020 afin d'en évaluer leur qualité et de tester l'hypothèse 2. La méthode proposée

s'appuie sur les caractéristiques attributaires, spatiales et temporelles des observations pour identifier de façon automatique celles qui sont aberrantes ou surprenantes. Pour cela, elle identifie un cluster correspondant aux observations de bonne qualité en étudiant leur densité à l'aide de la méthode DBSCAN. Les observations qui n'appartiennent pas à ce cluster sont considérées comme des observations aberrantes ou surprenantes. La comparaison des résultats obtenus avec une labellisation réalisée par des experts a montré que cette approche était robuste, qu'elle avait une bonne spécificité et une bonne sensibilité aux observations aberrantes. L'approche a également été capable d'identifier automatiquement des observations considérées comme surprenantes. Ces observations surprenantes correspondaient à des valeurs aberrantes qui n'avaient pas été identifiées lors de la première étape ou à des parcelles particulièrement précoces ou tardives par rapport à leur voisinage spatio-temporel (Figure 61).

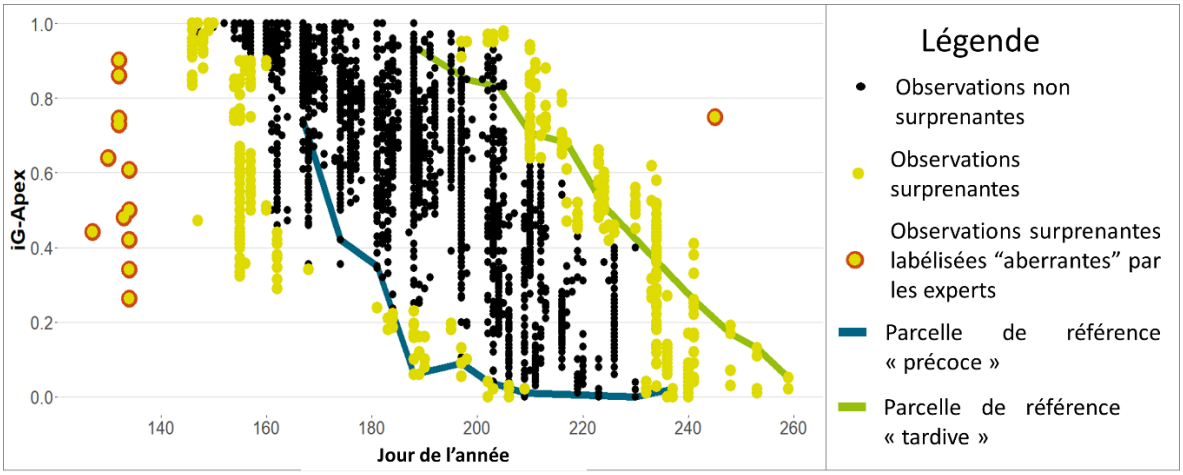


Figure 61: Représentation dans l'espace en 2 dimensions $iG\text{-}Apex\ vs\ Temps$ *des observations labélisées comme surprenantes par l'approche proposée.*

Chapitre 8 :

Ce chapitre conclut les travaux de la thèse et ouvre des perspectives pour de futurs travaux de recherche. Les deux hypothèses testées portant sur la quantité d'observations collectées (hypothèse 1) et la qualité de ces observations (hypothèse 2) ont été validées pour le projet ApeX-Vigne. Ce cas d'usage est représentatif des projets de FSO car il en partage les principales caractéristiques concernant i) le phénomène étudié, ii) le processus de collecte d'observations, iii) le jeu de données qui en résulte et iv) la disponibilité des données de référence. De plus, ce cas d'usage a été considéré sur une large emprise spatiale de plusieurs centaines de kilomètres, englobant de nombreux vignobles cultivés dans des conditions variées et au cours de plusieurs saisons. Les résultats obtenus sur le projet ApeX-Vigne peuvent donc être généralisés et les hypothèses 1 et 2 peuvent être considérées comme valides pour l'ensemble des projets de FSO.

Néanmoins, l'approche mise en œuvre possède certaines limites. Tout d'abord, chacun des facteurs spécifiques n'a pas pu être testé indépendamment. A l'avenir, l'étude individuelle de chacun des facteurs de succès spécifiques pourrait être favorisée par l'émergence d'un plus grand nombre de projets de FSO permettant d'étudier des projets aux caractéristiques plus divers. Ensuite, les tests des hypothèses 1 et 2 reposent sur des méthodes d'évaluation de la quantité d'observations collectées et de la qualité de ces observations mal adaptés aux spécificités des projets de FSO. Pour de futurs travaux, la conception d'approches qui prennent en compte ces

spécificités dans l'évaluation de la quantité et de la qualité des observations permettrait de formuler et de tester des hypothèses plus précises.

Les projets de FSO permettent de collecter des observations spatialisées, datées et objectivées puis de les partager. Ces projets et les méthodes associées pourraient être mis au service de collectifs regroupant des agriculteurs et/ou des chercheurs pour faciliter l'expérimentation en conditions réelles et dans des situations pédoclimatiques variées. Ils peuvent donc être vus comme un moyen d'accompagner l'agriculture dans sa transition en aidant les acteurs de l'agriculture à s'organiser collectivement pour répondre aux enjeux complexes et parfois contradictoires auxquels elle est confrontée. Cette approche visant à favoriser la construction de connaissances en s'appuyant sur des expérimentations menées par les agriculteurs en conditions réelles se développe de plus en plus dans la communauté scientifique où elle est appelée le On-Farm Experimentation. A l'avenir, les outils et les méthodes développés dans le cadre de cette thèse et plus largement dans le cadre des projets de FSO devraient pouvoir contribuer à cette communauté scientifique en émergence.