
Calcul d'allocations de l'enquête pratiques culturales en viticulture 2019

Maxime SIMONOVICI

Insee, Direction Régionale Centre-Val de Loire¹

`maxime.simonovici@insee.fr`

Mots-clés. (6 maximum) : sondage, échantillonnage, Bootstrap, charge statistique.

Domaines. Théorie des sondages amont, échantillonnage particulier.

Résumé

Au Service de la Statistique et de la Prospective (SSP) du Ministère de l'Agriculture et de l'Alimentation, les enquêtes « Pratiques culturales » et « Pratiques phytosanitaires » ont été mises en place dans le cadre du suivi du plan Ecophyto. Elles constituent la première et seule source d'informations sur l'utilisation des produits phytopharmaceutiques. L'indicateur de fréquence de traitement (IFT) est utilisé pour la mesurer. Dans le cadre de certaines politiques publiques, il constitue un outil d'évaluation via le calcul d'IFT de référence sur certaines zones géographiques. L'obtention d'une bonne précision est nécessaire. Celle visée sur la moyenne de l'IFT est de 2,5 % dans chaque bassin viticole.

La précision des estimations effectuées à partir d'une enquête est liée à de multiples dimensions. Les erreurs non dues à l'échantillonnage ne seront pas étayées dans cette contribution. Les bilans des quatre précédentes enquêtes n'ont pas révélé d'erreur ni de couverture ni de mesure (mauvaise conception du questionnaire, biais de l'intervieweur, etc.). L'imprécision de l'estimation est donc principalement due à l'ampleur de l'erreur d'échantillonnage. Elle est généralement mesurée par la variance d'échantillonnage dépendante de différents facteurs (méthodes d'échantillonnage et d'estimation, taille de l'échantillon, variabilité de la caractéristique d'intérêt). Par ailleurs, la non-réponse totale est historiquement très faible (5 % maximum) dans les enquêtes pratiques phytosanitaires ou pratiques culturales en viticulture.

Dans le cadre de la préparation de la cinquième enquête de ce type en viticulture, Pratiques culturales en viticulture en 2019 (PKViti 2019), une nouvelle demande a été émise lors du Comité des Utilisateurs de l'enquête. Il s'agit d'étudier la possibilité d'estimer l'IFT moyen des parcelles menées selon des pratiques biologiques.

1. Au moment de la rédaction de ce papier, M. Simonovici travaillait au Ministère de l'Agriculture et de l'Alimentation, Service de la Statistique et de la Prospective (SSP).

L'analyse des plans de sondage des quatre précédentes enquêtes permet de tirer différentes conclusions afin de concevoir celui de PKViti 2019. Ces enquêtes sont complexes à plusieurs égards. L'unité statistique, la parcelle culturale, n'est pas identifiable dans la base de sondage. Un échantillon de sous-parcelles cadastrales est tiré selon un sondage stratifié de taille fixe à probabilités inégales. Les probabilités d'inclusion reposent sur une approximation. Leur mode de calcul est celui d'un tirage de Poisson à taille aléatoire. Les allocations ont été prévues en recherchant un taux de sondage uniforme dans chaque strate selon leur importance en termes de superficie. En 2010, un panel est établi jusqu'en 2016. Néanmoins, de nouveaux bassins de tailles plus petites ont intégré l'enquête, engendrant des différences dans les niveaux des poids des unités répondantes retenues.

Parmi les conclusions les plus importantes, la stratification sur le type de vin ne semble pas efficace en raison de la mauvaise qualité de l'information dans la base de sondage. L'amélioration potentielle la plus apparente concerne le niveau des allocations via la méthode de prévision de la précision des IFT moyens. Les allocations ne semblent pas toujours efficaces. La méthode employée pour les déterminer est perfectible car elle n'est pas individualisée par bassin viticole. Comme le montre le niveau très bas (inférieur à 1 %) de la précision dans certains bassins viticoles, le budget de l'enquête ne semble pas alloué de manière optimale.

Comme piste de plan de sondage de l'enquête PKViti 2019, celle d'un tirage à deux degrés (exploitation viti-vinicole puis sous-parcelle cadastrale) est abandonnée pour un nouveau tirage de taille fixe à probabilités inégales. Dans neuf bassins viticoles candidats sur les vingt-deux enquêtés, deux strates sont envisagées : une pour les parcelles culturales menées selon des pratiques conventionnelles et une autre selon des pratiques biologiques. Un des objectifs est de calculer des allocations garantissant une précision de 2,5 % de l'IFT moyen par bassin viticole. Les données de l'enquête 2016 issues de parcelles culturales menées selon des pratiques biologiques n'étant pas assez nombreuses, les premiers calculs de précision aboutissent à déterminer des allocations élevées.

Des simulations de Monte-Carlo sont ainsi mises en place pour prendre en compte ces considérations. Certaines permettent d'estimer l'impact d'une baisse de l'allocation sur la précision de l'IFT moyen dans chaque bassin viticole. D'autres ont pour but de prévoir la précision de l'IFT moyen des parcelles culturales menées selon des pratiques biologiques, en fonction du nombre de répondants. Le Bootstrap est la méthode d'estimation de précision par réplication choisie. En se basant sur le principe de cette méthode, des Bootstrap « à façon » sont programmés sous R pour simuler au maximum l'ensemble des travaux méthodologiques (tirage, redressement, calage) traditionnellement effectués dans ces enquêtes. Une minimisation des allocations est approchée sous la contrainte d'obtenir des précisions sur l'IFT moyen de 3,5 % et de 2,5 % respectivement sur la strate des parcelles culturales menées selon des pratiques biologiques et sur l'ensemble du bassin viticole.

En plus de réduire la charge statistique des répondants, la limitation du nombre de questionnaires enquêtés a été particulièrement opportune lors de la collecte, dont la seconde moitié a été effectuée pendant le premier confinement lié à la pandémie de Covid-19.

Abstract

The primary objective of the French surveys on viticulture is to evaluate the use of pesticides. The treatment frequency indicator (TFI) measures the quantity of doses of plant protection products applied to a crop during a wine-growing. The target precision of the average TFI is 2,5 % on each wine-growing area. A new aim of the next survey (PKViti2019) is to determine the TFI of the vines which were grown according to the organic farming practices.

This article focuses on the empirical approach used to determine the size of the sample to achieve these purposes. Firstly, the sample design of the previous surveys are analysed. The sampling design of the last survey is complex : a stratified sampling with unequal probability. Then, several bootstrap methods are studied and tested. Like the precision is not homogeneous according to the wine-growing areas, a custom method is set up to get the size of the sample on each wine-growing area. Even if the survey was disrupted by the lockdown in the second quarter of 2020, the objectives have mostly been achieved.

Introduction

Les plans Ecophyto I, II et II+² ont pour objectifs de « rendre l'agriculture moins dépendante aux pesticides » et répondent à une obligation européenne fixée par la directive 2009/128/CE, qui nécessite notamment de mesurer l'utilisation des produits phytopharmaceutiques. L'indicateur de fréquence de traitement (IFT) est l'indicateur de suivi utilisé pour effectuer cette mesure. Il a été établi au Danemark puis développé en France par l'INRA et le Ministère de l'Agriculture [1]. Il vaut le nombre de doses de référence appliquées à une culture pendant une campagne. Certaines politiques publiques s'en servent et il constitue l'un de leurs principaux outils d'évaluation.

Le déterminer avec précision constitue donc un enjeu important, au regard des nombreuses conséquences engendrées. Les enquêtes pratiques culturales et phytosanitaires permettent de collecter des données sur les travaux des agriculteurs et de calculer l'IFT sur certaines zones géographiques (région ou bassin viticole). La précision visée sur la moyenne de l'IFT est de 2,5 % dans chaque bassin viticole.

Dans le cadre de la préparation de la cinquième enquête de ce type en viticulture, Pratiques culturales en viticulture en 2019 (PKViti 2019)³, une nouvelle demande a été émise lors du Comité des Utilisateurs de l'enquête. Il s'agit d'étudier la possibilité d'estimer l'IFT moyen des parcelles menées selon des pratiques biologiques. Lors de la préparation du plan de sondage, une analyse de ceux des quatre précédentes enquêtes est effectuée ainsi que des simulations pour intégrer cette demande tout en maintenant la précision de l'IFT moyen à 2,5 % dans chaque bassin viticole.

2. La page de présentation du plan : <https://agriculture.gouv.fr/ecophyto>

3. La page de présentation de la source et du questionnaire du site agreste du SSP : https://agreste.agriculture.gouv.fr/agreste-web/methodon/S-PK_Viticulture_2019/methodon/

1 L'estimation de l'IFT dans les enquêtes précédentes

Dans un premier temps, un historique des travaux méthodologiques des quatre enquêtes précédentes est fait. Il permet de détailler la construction de l'estimateur. Dans un second temps, la précision de l'IFT moyen est analysée de manière complémentaire. De premiers enseignements sur l'amélioration de l'estimation de l'IFT moyen en découlent.

1.1 Les enquêtes pratiques culturales et phytosanitaires

1.1.1 Objectifs des enquêtes

Les enquêtes sur les pratiques culturales sont effectuées dans quatre filières : les grandes cultures, les fruits, les légumes et la vigne. Jusqu'en 2018, les enquêtes complètes sur les pratiques culturales (dites « PK ») avaient lieu tous les 5-6 ans par filière. Des enquêtes allégées sur les pratiques phytosanitaires (dites « Phyto ») intermédiaires, au questionnaire moins développé et centré sur l'utilisation des produits phytopharmaceutiques (notés PPP) étaient conduites entre deux enquêtes PK. À compter de 2019, seules des enquêtes PK quinquennales sont conduites.

Ces enquêtes permettent de collecter des données sur l'itinéraire technique des agriculteurs. Plusieurs thèmes sont abordés : caractéristiques de la parcelle, opérations mécaniques, fertilisation, traitements phytosanitaires, etc⁴. Ces derniers constituent un enjeu majeur de développement durable en raison de leur impact en termes de risques environnementaux et phytosanitaires. Les enquêtes PK et Phytos ont été mises en place dans le cadre du suivi des plans Ecophyto qui ont pour objectif de « rendre l'agriculture moins dépendante aux pesticides ». Ces enquêtes permettent également d'apporter des éléments pour évaluer les plans d'actions relatifs à la directive européenne « Nitrates ». Les résultats des enquêtes sont enfin valorisés dans des études et permettent de répondre à des demandes extérieures (chercheurs, instituts techniques, etc).

En viticulture, deux enquêtes PK (en 2006 et 2013) et deux enquêtes Phytos (en 2010 et en 2016) ont été menées. En 2019, la cinquième était en cours de préparation, avec un début de la collecte prévu en janvier 2020.

1.1.2 L'indicateur de fréquence de traitement

L'indicateur de fréquence de traitement (IFT) est la variable phare de ces enquêtes. Il vaut le nombre de doses de référence de PPP appliquées à une culture pendant une campagne. Pour un PPP donné, une dose de référence (DR) correspond à une dose recommandée pour une culture et une cible de traitement (ex : le mildiou). Le calcul de l'IFT vaut pour chaque traitement, le ratio entre la dose appliquée (DA) et la dose de référence du produit utilisé, multiplié par la part de surface traitée en % (PST). L'IFT d'un traitement t est égal à : $IFT_t = \frac{DA_t}{DR_t} \cdot PST_t$.

L'IFT d'un traitement en plein (sur l'intégralité de la surface de la parcelle) ne peut pas être inférieur à 0,1 ou supérieur à 2. Il s'agit d'une convention qui a pour but d'éviter les erreurs de saisies et l'utilisation de doses incohérentes avec les réalités agronomiques. En dehors de cet intervalle, l'IFT est redressé par l'IFT moyen du segment de la cible (ex : fongicide pour le mildiou) du bassin viticole concerné.

4. Les questionnaires et résultats sont disponibles pour chaque enquête sur le site agreste : <https://agreste.agriculture.gouv.fr/agreste-web/methodon/P/?searchurl/listeTypeMethodon/>.

L'IFT d'une unité statistique, ici d'une parcelle viticole c , vaut :

$$IFT(c) = \sum_{t \in c} IFT_t = \sum_{t \in c} \frac{DA_t}{DR_t} \cdot PST_t$$

L'IFT peut être décliné en segments définis selon quatre grandes catégories d'usage : fongicide (IFTF), herbicide (IFTH), insecticide-acaricide (IFTI) et autre (IFTA).

Ces quatre IFT par segment sont sommables par parcelle ou par bassin viticole, selon la formule : $IFT = IFTF + IFTH + IFTI + IFTA$. Un exemple de calculs des différents IFT est présent en Annexe 1.

L'IFT d'un bassin viticole (ou du vignoble français dans son ensemble) correspond à la moyenne pondérée des IFT de l'ensemble des parcelles répondantes du bassin viticole et de la période concernée. La méthodologie de calcul des IFT a été revue en 2015, afin de tenir compte des cibles dans le calcul des doses de référence, à la suite de la réforme du catalogue des usages de PPP⁵. Les IFT des enquêtes 2010, 2013 et 2016 ont été calculés en tenant compte de cette nouvelle méthodologie.

1.1.3 Millésimes retenus

Quatre enquêtes sur les pratiques culturales ou phytosanitaires ont été réalisées à ce jour par le SSP. Le plan de sondage (PDS) de l'enquête de 2006 est étudié car ses travaux de préparation constituent le fondement des enquêtes suivantes. Pour analyser l'estimation de l'IFT, les trois dernières enquêtes (PhytoViti2010, PKViti2013 et PhytoViti2016) sont retenues. L'IFT y est calculé de la même façon, *modulo* une différence entre le millésime de l'enquête et celui des doses de référence⁶ des PPP. Ces trois enquêtes ont des unités statistiques répondantes communes (panel). Le champ géographique s'est progressivement étendu depuis 2010.

1.2 Historique des travaux méthodologiques réalisés

1.2.1 Revue des plans de sondage passés

En 2006, le méthodologue étudie de nombreux aspects dans la préparation du PDS de l'enquête PKViti2006. Les principaux enseignements sont :

- l'unité statistique enquêtée, la parcelle culturale, n'est pas identifiable dans la base de sondage, le Casier Viticole Informatisé (CVI), formée de sous-parcelles cadastrales ;
- il en résulte la nécessité de construire un lien entre la sous-parcelle cadastrale tirée et la parcelle culturale interrogée ;
- les calculs de probabilités d'inclusion et de double inclusion doivent être mis à jour en conséquence après la collecte ;
- le type de vin (ou appellation : AOP, IGP, sans IG, etc) n'est pas connu pour toutes les sous-parcelles cadastrales, d'où la présence de strates composées uniquement de sous-parcelles au type de vin manquant ;
- les tirages se font par croisement département \otimes type de vin dans certains bassins viticoles

5. Pour en savoir plus : le guide méthodologique est disponible sur le site du ministère : <https://www.data.gouv.fr/fr/datasets/doses-de-reference-indicateur-de-frequence-de-traitements-phytosanitaires/>.

6. Pour les calculs des IFT des enquêtes 2010 et 2013, les doses de référence (DR) de 2015 sont utilisées. Cela engendre un biais dans l'estimation de l'IFT qui n'existe pas en 2016 (référentiel de la même année que celle de la campagne collectée : 2016).

- parmi les dix enquêtés⁷ ;
- les allocations n_h ont été prévues en recherchant un taux de sondage uniforme dans chaque strate S_h selon leur superficie (en ha) ;
 - les probabilités de tirage des exploitations viti-vinicoles (EVV) et les coefficients d'extrapolation sont calculés mais plusieurs problèmes sont soulevés.

En résumé, l'échantillon est tiré selon un sondage stratifié à taille fixe avec des probabilités de tirage proportionnelles à $surf(p)$ qui est la surface de la sous-parcelle cadastrale p associée à la parcelle culturale enquêtée c .

La probabilité d'inclusion π_c d'une parcelle culturale c appartenant à la strate S_h vaut alors :

$$\pi_c \approx \sum_{p \in c} \pi_p = \sum_{p \in c} n_h \cdot \frac{surf(p)}{\sum_{p \in S_h} surf(p)} = n_h \cdot \frac{\sum_{p \in c} surf(p)}{\sum_{p \in S_h} surf(p)} = n_h \cdot \frac{surf(c)}{\sum_{p \in S_h} surf(p)}$$

Cela signifie que la probabilité de tirage d'une parcelle culturale est approximativement égale à la somme des probabilités de tirage des sous-parcelles cadastrales qui la composent. Il a été considéré que le calcul effectué dans le cas d'un tirage de Poisson [2] à taille aléatoire pour obtenir la probabilité d'inclusion reste valable dans le cas d'un sondage à taille fixe. Cette approximation a été préférée à la méthode généralisée du partage des poids qui ne semblait pas appropriée au problème d'identification des parcelles de cette enquête. Le coefficient d'extrapolation associé vaut donc : $COEF1(c) = \frac{1}{\pi_c}$.

Le méthodologue évoque une distinction à effectuer entre les variables intensives et extensives. Une variable extensive possède la propriété suivante : la valeur prise sur la réunion de deux parcelles est la somme des valeurs prises sur chacune des parcelles. Par exemple, la surface de deux parcelles est la somme des surfaces de chaque parcelle. En revanche, l'IFT (exprimé par ha) est une variable intensive. Pour estimer la moyenne d'une variable intensive X ou la moyenne pondérée par la surface sur un sous-champ donné (ex : un bassin), il préconise de multiplier la valeur X par le coefficient d'extrapolation et la surface. Il n'y a pas eu de traitement de la non-réponse totale car elle était rare et des parcelles remplaçantes avaient été utilisées. Un calage a été effectué sur les surfaces totales par strate de la Statistique agricole annuelle (SAA).

En 2010, un panel est établi sur un périmètre de treize bassins viticoles. La même méthode de tirage est appliquée, *modulo* quelques différences dans la stratification relatives aux types de vin. La méthode employée pour déterminer les allocations sera décrite dans le paragraphe suivant. L'allocation fixée par bassin viticole est répartie proportionnellement à la superficie de la strate. Un traitement de la non-réponse totale est effectué par strate. Il est léger car le taux de réponse est excellent (supérieur à 95 %). Pour la diffusion des résultats, il faut raisonner en superficie et non en nombre de parcelles. Tous les calculs de moyenne sont rapportés à l'hectare. Un second coefficient d'extrapolation $COEF2(c)$ est créé pour chaque parcelle culturale c , tel que $COEF2(c) = COEF1(c) \cdot surf(c)$. Pour les calculs de quantité, le premier coefficient doit être utilisé. Un calage est effectué sur les surfaces viticoles par bassin via la SAA. Il aboutit à des coefficients $COEF2$ très homogènes par strate et par bassin (d'un facteur 3 maximum).

En 2013, la même logique d'enquête est poursuivie avec l'intégration de quatre nouveaux bassins et l'éclatement du bassin Midi-Pyrénées en trois bassins distincts (Cahors, Gaillac et Gers). Dans les bassins du panel, un faible nombre de nouvelles parcelles sont introduites pour lutter contre l'érosion du champ des répondants. La somme des coefficients $COEF2(c)$ vaut la

7. Le nombre de bassins viticoles se porte à vingt-et-un en 2016. Leurs caractéristiques sont présentées page 9.

superficie du bassin viticole. Après traitement de la non-réponse totale, on a :

$$COEF2(c) = COEF1(c) \cdot surf(c) = \frac{surf(c)}{\pi_c \cdot \hat{p}_r}$$

avec $\hat{p}_r = \frac{n_{r,h}}{n_h}$ avec $n_{r,h}$ répondants dans la strate S_h et n_h l'allocation de la même strate.

Le calage ne s'est pas fait sur la SAA mais sur une autre version du CVI en raison de problèmes de définition de la base de sondage. Les coefficients $COEF2$ sont moins homogènes par strate et par bassin, avec l'apparition de fortes différences jusqu'à un facteur supérieur à 30 (notamment dans le bassin des Pyrénées-Orientales).

En 2016, le panel est maintenu et deux nouveaux bassins (Bugey-Savoie et Corse) intègrent l'enquête. Les acquis de 2013 sont conservés et deux nouveautés méthodologiques sont intégrées pour réduire la variance des estimateurs. Une post-stratification est effectuée à partir du type de vin récupéré par l'enquête. Elle permet un calage plus fin sur la SAA par département et type de vin déclaré. Soixante zones de calage z sont ainsi constituées. Le calage s'est avéré difficile à effectuer, avec des premiers résultats insatisfaisants (forte variation des $COEF2$ dans certaines zones géographiques ; perte de l'égalité entre la somme des $COEF2$ et la surface du bassin viticole). Pour résoudre ces problèmes, l'équation de calage a été modifiée. On recherche de nouveaux poids w_c , proches des poids de départ $COEF2$ qui vérifient les équations de calage suivantes (où z est la zone de calage) :

$$\sum_{c \in s} w_c \cdot \mathbb{1}_{c \in z \cap s} = surf(z)$$

Cela a permis un gain notable de précision dans le bassin des Pyrénées-Orientales. Une plus faible dispersion des rapports de poids (poids postcalage/poids précalage) atteste de la meilleure qualité du calage. De plus, les coefficients $COEF2$ sont moins hétérogènes par bassin par rapport à 2013 (facteur 10 maximum). Ils sont très homogènes par zone de calage. Ils sont égaux dans les croisements $z \otimes S_h$.

Enfin, s'agissant de PKViti2019, dont la collecte est prévue début 2020, un changement de PDS est envisagé. Un tirage à deux degrés est prévu : l'EVV en unité primaire puis la sous-parcelle cadastrale en secondaire afin de répondre à une demande du Cnis concernant la limitation de la charge des enquêtés. Un nouveau tirage permettra d'interroger de nouvelles EVV et de fixer un nombre de questionnaires par EVV à ne pas dépasser. Vu le taux annuel d'arrachage de vignes (2,5 %), un panel ne semble plus adapté à la fréquence quinquennale des nouvelles enquêtes PK. L'attrition serait trop importante entre la première et la dernière enquête d'un nouveau cycle.

1.3 Étude de la précision de l'IFT

1.3.1 Allocations et objectifs de précision des quatre précédentes enquêtes

Les allocations de PKViti2006 avaient été prévues en recherchant un taux de sondage uniforme dans chaque strate selon leur importance en termes de superficie.

En 2010, les résultats de l'enquête PKViti2006 ont été utilisés pour construire un abaque indiquant approximativement la taille de l'échantillon nécessaire par bassin pour obtenir une précision donnée de l'IFT herbicide (IFTH) moyen. Il s'agit du coefficient de variation exprimé en pourcentage. Elle est calculée à partir d'une macro SAS Calva (Calcul de variance) utilisée au SSP. Une régression linéaire simple est effectuée pour calculer la précision en fonction de la taille d'échantillon ainsi qu'une borne à 95 % de cette précision sans détails sur la méthode employée

pour obtenir cette valeur (cf. tableau 1). Ces calculs aboutissent à fixer un minimum de 350 questionnaires dans chacun des trois nouveaux bassins pour atteindre un objectif de précision de 5 % sur les IFTH et IFTA (IFT autre) moyens.

TABLE 1 – Précision de l’IFT herbicide moyen en fonction de la taille de l’échantillon pour la préparation de l’enquête PhytoViti2010

Taille de l’échantillon	Précision attendue de l’IFTH moyen (en %)	Borne supérieure à 95 % de la précision de l’IFTH moyen (en %)
100	7,5	9,3
150	6,1	7,5
200	5,3	6,5
250	4,7	5,9
300	4,3	5,4
350	4,0	5,0
400	3,7	4,7
450	3,5	4,4
500	3,3	4,2

Note de lecture : la précision de l’IFT herbicide moyen vaut $100 \cdot \frac{\sqrt{V(\widehat{IFTH})}}{\widehat{IFTH}}$. Elle est estimée par

$100 \cdot \frac{\sqrt{V(\widehat{IFTH})}}{\widehat{IFTH}}$ où \widehat{IFTH} est l’estimateur au sens d’Horvitz-Thompson de l’IFTH moyen.

Source : Agreste - Enquête sur les pratiques culturales en viticulture 2006.

La même démarche est reconduite **en 2013** en analysant la précision de l’IFT moyen des données de 2010 (cf. tableau 2). L’abaque indique que 200 questionnaires suffisent pour assurer une précision de 2,5 % sur l’IFT moyen par bassin.

TABLE 2 – Précision de l’IFT moyen en fonction de la taille de l’échantillon pour la préparation de l’enquête PKViti2013

Taille de l’échantillon	Précision attendue de l’IFT moyen (en %)	Borne supérieure à 95 % de la précision de l’IFT moyen (en %)
100	2,78	3,37
150	2,27	2,84
200	1,96	2,52
250	1,76	2,3
300	1,60	2,14
350	1,48	2,02
400	1,39	1,92
450	1,31	1,83
500	1,24	1,76

Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2010.

En 2016, une régression est faite à partir des données des bassins communs aux deux précédentes enquêtes. L'abaque est mis à jour (cf. tableau 3) et le nombre minimal de questionnaires est porté à 250.

TABLE 3 – Précision de l'IFT moyen selon la taille de l'échantillon (Plan de sondage de l'enquête PhytoViti2016)

Taille de l'échantillon	Précision attendue de l'IFT moyen (en %)
100	3,09
150	2,90
200	2,72
250	2,53
300	2,34
350	2,16
400	1,97
450	1,79
500	1,60

Note de lecture : si 350 questionnaires sont collectés, la précision de l'IFT moyen sera en moyenne de 2,16 %. Sources : Agreste - Enquêtes sur les pratiques phytosanitaires (2010) et culturales (2013) en viticulture.

Cette méthode s'avère perfectible. Il n'y a en effet pas de relation linéaire entre n (le nombre de répondants) et la précision de l'estimateur. Pour obtenir un degré de précision demandé, la taille de l'échantillon augmente à un taux qui diminue à mesure qu'augmente la taille de la population [3]. À ce titre, le nombre de sous-parcelles par strate est très variable dans la base de sondage : de moins de 5 000 en Corse à plus de 328 000 dans le Bordelais.

1.3.2 Impacts de l'échantillonnage

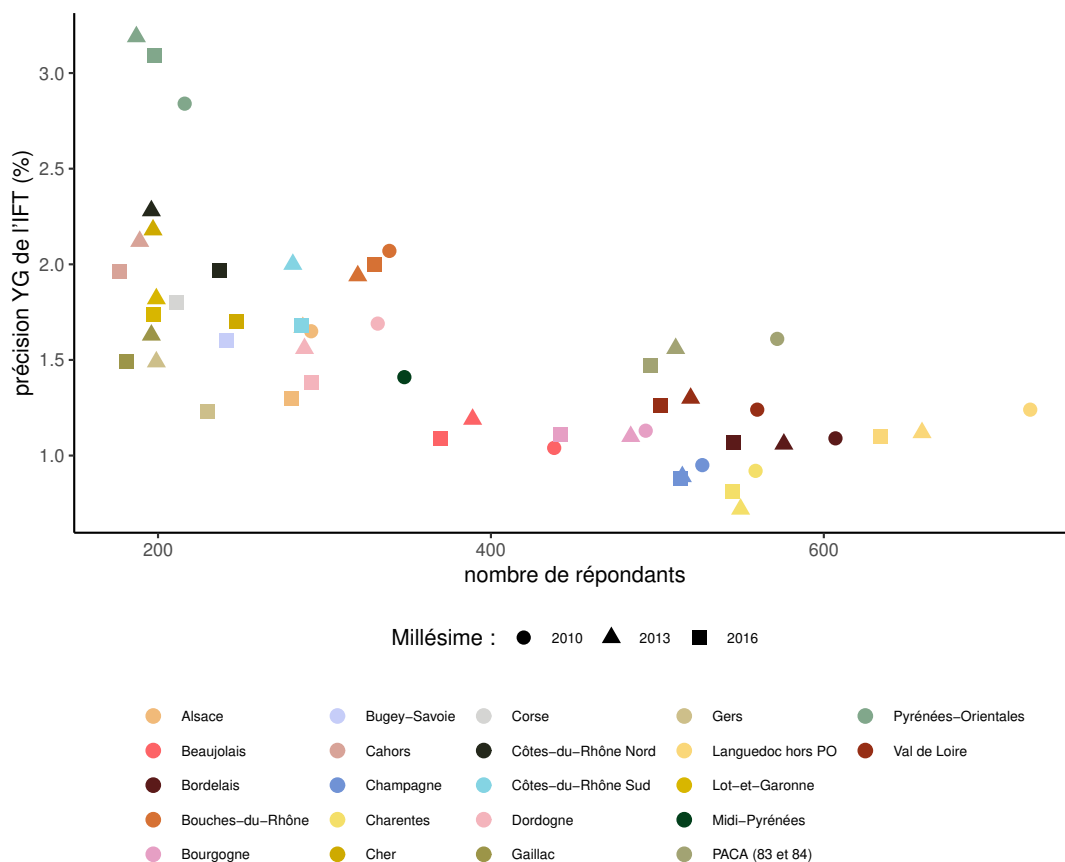
En examinant la méthode utilisée jusqu'en 2016 sous d'autres angles, ses limites sont plus claires. Les résultats sont présentés par millésime et par bassin viticole (cf. illustration 1).

La précision étudiée est celle calculée en utilisant la variance de Yates-Grundy car elle est légèrement plus faible que celle obtenue à partir de la variance d'Horvitz-Thompson. S'il n'y a pas réellement d'effet millésime sur la précision de l'IFT moyen, un effet bassin apparaît clairement avec une hétérogénéité des situations (ex : Pyrénées-Orientales). Cela incite à étudier la précision des IFT au niveau de chaque bassin viticole enquêté. Les tableaux en Annexe 2 résument les principales caractéristiques des vingt-un bassins enquêtés en 2016. L'introduction de petits bassins en terme de surface a créé des différences dans les taux de sondage, en nombre de sous-parcelles et de surface. Il n'est plus uniforme par strate. La précision de l'IFT moyen est plus élevée dans les bassins comprenant davantage de zones de calage ou un ratio plus élevé entre le plus grand et le plus petit coefficient d'extrapolation. À l'opposé, le nombre de strates n'est pas significatif.

Ces premières conclusions sont à prendre avec précaution car issues de régressions effectuées à partir de seulement vingt-et-un bassins (et cinquante-deux IFT moyens), ce qui est insuffisant pour construire un abaque. En revanche, ils permettent de tirer plusieurs enseignements utiles dans l'élaboration de futurs PDS des enquêtes PK en viticulture : la stratification par type de vin ne semble pas efficace ; une trop forte dispersion entre les poids dans un même bassin est à éviter ; les écarts-types de l'IFT par bassin sont du même ordre de grandeur mais une approche

par bassin permettra de prendre en compte certaines particularités locales, en lien avec des réalités agronomiques.

FIGURE 1 – Précision de l'IFT moyen par millésime et par bassin viticole



Note de lecture : avec 200 répondants, la précision de l'IFT moyen est de 1,5 % dans le Gers en 2013.
Sources : Agreste - Enquêtes sur les pratiques phytosanitaires (2010, 2016) et culturales (2013) en viticulture.

1.3.3 Recherche de nouveaux cofacteurs dans la base de sondage

Outre leur localisation et dénombrement, la base de sondage fournit pour chaque sous-parcelle cadastrale les variables suivantes :

- le cépage. Dans la littérature agronomique, la sensibilité au cépage est reconnue. Plus un bassin viticole possède de cépages différents, plus la dispersion de l'IFT devrait être grande, ainsi que la précision de l'IFT moyen ;
- l'année de plantation (date de la première feuille) ;
- la distance entre les rangs et entre les ceps ;
- la taille de l'EVV (la superficie en vigne à raisin de cuve de l'EVV possédant la parcelle).

Plusieurs analyses effectuées via des statistiques descriptives et des régressions linéaires livrent les conclusions suivantes :

- le nombre de cépages différents d'un bassin est la seule variable significative en excluant les bassins Languedoc hors Pyrénées-Orientales et Pyrénées-Orientales ;
- il n'y a pas de relation linéaire entre la taille de l'EVV et l'IFT, même par bassin ;
- parmi les EVV multi-interrogées, la dispersion de l'IFT de parcelles de cépages différents est plus grande que de celle des IFT des parcelles de même cépage.

1.3.4 Autres informations mobilisables

D'autres analyses ont été conduites, notamment sur :

- le **climat** : si elles sont favorables au développement des parasites, certaines conditions climatiques de la campagne enquêtée peuvent entraîner une augmentation de l'utilisation de pesticides. Les résultats de la dernière enquête les détaillent dans un dossier, avec une fiche par bassin viticole comprenant des indicateurs de dispersion des IFT [4]. Le niveau de l'IFT moyen est effectivement très variable d'un bassin à l'autre et s'étend en 2016 de 9,3 dans les Bouches-du-Rhône à 23,5 en Champagne. Un des défis des enquêtes PK est donc de mesurer l'IFT corrigé des différences dues à la pression parasitaire. Un premier essai a été effectué au SSP sur les données des enquêtes PhytoViti2010 et PKViti2013 et ses résultats [5] soulignent les difficultés pour estimer les facteurs climatiques. Un autre essai a été effectué en construisant un indicateur d'hétérogénéité des pressions parasitaires ressenties issues des données de PhytoViti2016. Les traitements mis en place aboutissent à une absence de lien significatif entre ces indicateurs et la précision de l'IFT.
- la **diversité des produits appliqués** : les doses de référence et les conditions d'utilisation des PPP diffèrent d'un produit à l'autre. Une plus grande diversité de produits illustre une différence entre les pratiques. La corrélation avec l'IFT est assez significative : le coefficient de corrélation ajusté montant jusqu'à 0,34 en 2013. Néanmoins, ce niveau de lien n'est pas stable dans le temps (0,14 en 2016). En outre, cette information ne peut pas être prise en compte dans l'élaboration de PDS futurs car elle est inconnue pour chaque sous-parcelle et par bassin avant la collecte.

1.4 Enseignements pour les futurs essais d'élaboration de plan de sondage

La précision de l'IFT moyen est expliquée principalement par l'échantillonnage : nombre de répondants et dispersion des poids. Le plan de sondage actuel est complexe mais apporte des résultats satisfaisants au niveau de la précision. La stratification par type de vin n'est pas forcément à reproduire car ce qui y est déclaré est un potentiel de type de vin revendicable, qui peut s'avérer différent une fois la parcelle récoltée. De plus, l'information sur le type de vin *a priori* est manquante dans la base de sondage, à un pourcentage parfois élevé dans certains bassins. Cela ajoute une étape de post-stratification, parfois en proportion élevée (ex : le bassin Corse en 2016). Dans certaines zones de calage (département \otimes type de vin) avec un nombre faible de parcelles, la post-stratification augmente la dispersion des coefficients d'extrapolation.

La limite la plus apparente des plans de sondages passés concerne le niveau des allocations et la méthode de prévision de précision de l'IFT moyen. Il serait préférable de l'étudier au niveau de chaque bassin viticole en raison de leur hétérogénéité. La méthode de régression tous bassins confondus est trop approximative et les allocations obtenues ne semblent pas efficaces. Comme le montre leur niveau très bas (inférieur à 1 %) dans certains bassins, le budget de l'enquête ne semble pas utilisé de manière optimale. D'autres objectifs pourraient être atteints au même coût, comme la prise en compte d'une contrainte supplémentaire.

Dans le contexte d'un tirage à deux degrés, il pourrait être opportun d'intégrer le cépage dans les unités primaires. Au lieu de créer des strates par taille d'exploitation comme cela avait été envisagé, les unités primaires pourraient être des croisements EVV \otimes cépage. Il n'y a effectivement pas de relation linéaire entre la taille d'une exploitation et l'IFT.

2 Préparation du plan de sondage de PKViti2019

2.1 Premiers essais

Lors des travaux de préparation du plan de sondage (PDS), un tirage à deux degrés [6] a d'abord été envisagé. Les EVV devaient constituer les unités primaires selon des strates par taille d'exploitation, par analogie avec les PDS d'autres enquêtes récentes du SSP (ex : enquête pratiques culturales en grandes cultures et prairies en 2017). Les unités secondaires prévues étaient les sous-parcelles cadastrales. Mais il a été abandonné après plusieurs essais du pôle méthodes statistiques (BMIS) car il en résultait un nombre trop élevé de questionnaires pour assurer une précision de 2,5 % sur la moyenne de l'IFT par bassin. Cela aurait eu pour conséquence d'alourdir la charge de certains répondants. L'absence de lien entre la taille de l'exploitation et le niveau de la variable d'intérêt (IFT) a également incité à abandonner ce type de PDS.

Le tirage choisi pour PKViti2019 est finalement le même que celui des enquêtes précédentes : un tirage à probabilités inégales, proportionnelles à la surface des sous-parcelles cadastrales. La stratification par type de vin, opérée depuis 2006 dans certains bassins, est abandonnée suite aux enseignements de la partie 1 (voir page 11). Outre le report de la charge sur de nouveaux enquêtés, le retraitage complet des unités et l'abandon du panel offrent l'opportunité d'améliorer la détermination des allocations.

Dans le cadre du plan de sondage de PKViti2019, l'objectif est de maintenir une précision de l'IFT moyen de 2,5 % dans chaque bassin. Les allocations doivent se répartir d'une manière différente dans un contexte double : l'intégration d'une nouvelle demande et la maîtrise des coûts.

2.2 Prise en compte de la nouvelle demande

Afin de répondre à une demande émise lors du Comité des Utilisateurs de l'enquête, la possibilité d'estimer l'IFT moyen des parcelles menées en Agriculture Biologique (AB) est étudiée. Les données de PhytoViti2016 n'étaient pas suffisantes pour diffuser un IFT moyen en AB, même au niveau national. En employant une méthode d'estimation avec un score de propension, la différence avec les parcelles menées selon des pratiques conventionnelles a été établie [7]. La précision de l'IFT moyen des parcelles menées en AB (cf. tableau 4), assimilées à un domaine (noté d) dans certains bassins, est beaucoup moins bonne (de 4,4 % à 12,1 %). C'est principalement dû à un faible nombre d'enquêtés dans ces domaines (de 10 à 85). Le PDS n'avait pas été prévu pour le faire et les IFT moyens en AB ont été calculés à titre expérimental, sans réutilisation possible.

En parallèle, des travaux à partir d'autres critères sont entrepris conjointement avec le pôle méthodes statistiques (BMIS) car la base de sondage ne contient pas d'information sur le mode de conduite des sous-parcelles. Un fichier transmis par l'Agence Bio permet d'obtenir la superficie totale de vigne engagée en AB par EVV. L'appariement de ce fichier avec le CVI adjoint, pour chaque EVV, l'information « AB » (si une surface viticole de l'EVV est retrouvée dans le fichier de l'Agence Bio) ou « non AB » (dans le cas contraire). Un seuil est ensuite fixé. Ainsi, il est proposé qu'une EVV mène toutes ses parcelles selon des pratiques dites « biologiques » si la superficie viticole figurant dans le fichier de l'Agence Bio représente au moins 80 % de la superficie de l'ensemble des sous-parcelles cadastrales viticoles figurant dans le CVI. L'ensemble des sous-parcelles cadastrales de cette EVV sont alors considérées comme menées selon des pratiques biologiques et classées dans la strate « biologiques ». Sinon, l'ensemble des sous-parcelles cadastrales de l'EVV seront qualifiées de conventionnelles et versées dans la strate « conventionnelles ».

À partir de cet appariement, plusieurs indicateurs ont été calculés par bassin viticole : nombre de sous-parcelles cadastrales des EVV en AB, proportion d'EVV en AB retrouvées, surface menée en AB identifiée dans le CVI et part de l'AB dans l'ensemble du bassin. Ces résultats permettent de dresser une première liste de neuf bassins candidats à une interrogation spécifique en AB. Ces bassins ont une superficie en AB non négligeable (souvent supérieure à 10 % de la superficie en vignes du bassin), un nombre important d'EVV en AB et par voie de conséquence un nombre important de sous-parcelles cadastrales pouvant être échantillonnées, donc un risque faible d'avoir une même EVV interrogée plusieurs fois (sauf en Corse).

TABLE 4 – Informations selon le type de pratiques sur une sélection de bassins

Bassin viticole	Type de pratiques	Nombre de répondants PhytoViti2016	Dispersion de l'IFT	Précision de l'IFT moyen (en %)
Alsace	Conventionnelles	236	8,95	1,47
	Biologiques	44	20,00	4,43
Bordelais	Conventionnelles	502	17,58	1,14
	Biologiques	44	29,74	5,13
Languedoc hors PO	Conventionnelles	580	13,83	1,70
	Biologiques	54	25,70	5,14
Pyrénées-Orientales	Conventionnelles	162	16,83	3,92
	Biologiques	36	18,19	9,00
PACA (Var-Vaucluse)	Conventionnelles	422	11,87	1,97
	Biologiques	74	16,35	4,76
Val de Loire hors Sancerre	Conventionnelles	453	19,26	1,93
	Biologiques	49	28,34	6,81
Bouches-du-Rhône	Conventionnelles	245	11,67	2,20
	Biologiques	85	12,88	5,03
Côtes-du-Rhône Sud	Conventionnelles	238	12,53	1,73
	Biologiques	48	13,48	4,95
Corse	Conventionnelles	183	21,48	2,16
	Biologiques	28	20,57	9,42

Note de lecture : la dispersion de l'IFT vaut dans chaque croisement M (bassin \otimes type de pratiques) :

$$\frac{1}{m-1} \sum_{c \in M} \left(IFT_c - \overline{IFT}^M \right)^2$$

avec m répondants dans le croisement M en 2016.

Ce tableau est un extrait des calculs de dispersion et de précision selon le type de pratiques par bassin. Ne sont représentés que les neuf bassins candidats à une stratification selon le type de pratiques.

Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

Pour PKViti2019, la solution envisagée est de stratifier selon le mode de conduite de la parcelle : selon des pratiques conventionnelles ou biologiques.

2.3 Calculs théoriques d'allocations

Une revue de la littérature a été réalisée sur la détermination des allocations dans le cadre d'un tirage à probabilités inégales, proportionnelles à la taille d'une variable auxiliaire. Mais ces recherches n'ont pas abouti à la possibilité de calculer n dans chaque strate. Afin de conduire un tirage à taille fixe et à probabilités de tirage inégales, la même approximation des probabilités d'inclusion que celle historiquement faite est reprise. Le PDS est à grande entropie car le tirage est similaire à un tirage de Poisson de taille fixe. Un article sur l'estimation de variance quand les probabilités d'inclusion sont inégales incite à utiliser l'estimateur de Hájek [8]. Une approximation de la variance de l'estimateur de l'IFT moyen (avec Y l'IFT pour faciliter les écritures) peut s'écrire dans chaque domaine d :

$$v(\widehat{Y}) = \frac{1}{n_d^2} \sum_{i \in s \cap d} c_i e_i^2 \text{ où } e_i = \frac{y_i}{\pi_i} - \hat{B} \text{ avec } \hat{B} = \frac{\sum_{i \in s \cap d} a_i \frac{y_i}{\pi_i}}{\sum_{i \in s \cap d} a_i}, c_i = a_i = \frac{n_d}{n_d - 1} (1 - \pi_i)$$

Néanmoins cette approche a une limite : on ne peut pas faire augmenter le nombre de répondants n_d du domaine. Le calcul dans le cadre d'un tirage aléatoire simple donne les allocations les plus coûteuses (cf. tableau 5) pour atteindre l'objectif de précision dans chaque strate en termes de superficie et de nombre de sous-parcelles. Elles sont effectivement trop élevées au vu des précisions passées obtenues dans les bassins entiers (dont certains comprennent un nombre non négligeable mais insuffisant de parcelles menées en AB).

TABLE 5 – Allocations calculées dans le cadre d'un sondage aléatoire simple par strate pour obtenir une précision relative de 2,5 % sur l'IFT moyen

Bassin viticole	Type de pratiques	Allocations en fonction de	
		la surface	N
Alsace	Conventionnelles	239	243
	Biologiques	480	577
Beaujolais	Toutes confondues	273	277
Bordelais	Conventionnelles	360	361
	Biologiques	678	712
Bouches-du-Rhône	Conventionnelles	700	732
	Biologiques	899	961
Bourgogne	Toutes confondues	341	345
Champagne	Toutes confondues	264	266
Charentes	Toutes confondues	223	223
Dordogne	Toutes confondues	375	383
Languedoc hors PO	Conventionnelles	426	427
	Biologiques	1 082	1 101
Pyrénées Orientales	Conventionnelles	900	931
	Biologiques	1 050	1 187
PACA (Var-Vaucluse)	Conventionnelles	670	675
	Biologiques	1005	1044
Val de Loire hors Sancerre	Conventionnelles	486	490
	Biologiques	1 212	1 431
Gers	Toutes confondues	211	213
Gaillac	Toutes confondues	257	264
Cahors	Toutes confondues	412	439
Cher	Toutes confondues	436	479
Côtes-du-Rhône Nord	Toutes confondues	559	594
Côtes-du-Rhône Sud	Conventionnelles	433	440
	Biologiques	567	664
Lot-et-Garonne	Conventionnelles	383	400
Corse	Conventionnelles	547	535
	Biologiques	680	795
Bugey-Savoie	Toutes confondues	351	396
Jura	Toutes confondues	423	504

Note de lecture : ces allocations représentent le nombre de questionnaires nécessaires pour obtenir une précision de 2,5 % sur l'IFT moyen en utilisant la moyenne et la dispersion de l'IFT des données de PhytoViti2016. Le calcul selon la surface considère le dénombrement total comme la surface du bassin ; celui selon N considère le dénombrement total comme le nombre de sous-parcelles cadastrales du bassin. Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

3 Calculs empiriques d'allocations

Afin de déterminer des allocations en adéquation avec la réalité de la collecte, des simulations sont effectuées à partir d'une méthode de réplcation : celle du Bootstrap introduite par Efron en 1979 [9].

De nombreux essais ont été réalisés progressivement. Le principe général est présenté dans un cadre théorique. D'une revue bibliographique approfondie sur le sujet, deux références ont servi de guides pour réaliser ces calculs empiriques. Il s'agit de la thèse de G. Chauvet [10] et d'un article de D. Haziza, C. Leger et Z. Mashreghi [11]. Ils constituent un panorama complet des différentes méthodes de Bootstrap existantes.

3.1 Méthode de Bootstrap : rappels théoriques

Soit X_1, \dots, X_n un échantillon i.i.d. (indépendantes et identiquement distribuées) d'une loi F inconnue. Pour estimer θ , un paramètre⁸, le principe du plug-in consiste à utiliser la fonction de répartition empirique \hat{F} de l'échantillon pour remplacer F . L'estimateur Bootstrap de la variance de θ associé est une variance conditionnelle aux valeurs de X_1^*, \dots, X_n^* prises dans un échantillon tiré dans X_1, \dots, X_n selon \hat{F} :

$$V^* = V^*(\hat{\theta}^* | X_1^*, \dots, X_n^*)$$

Pour approcher cette variance, on va procéder à une simulation de Monte-Carlo [12]. L'algorithme général est :

1. Pour b allant de 1 à B (le nombre d'itérations choisi) :
 - (a) tirer un échantillon i.i.d. de taille n selon la loi \hat{F} conditionnellement à l'échantillon de départ X_1, \dots, X_n ;
 - (b) calculer $\hat{\theta}_b$;
2. Calculer l'estimateur Bootstrap de θ : $\widehat{\theta}_{Boot} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b$;
3. Calculer l'estimateur Bootstrap de la variance de θ : $\widehat{V}_{Boot} = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b - \widehat{\theta}_{Boot})^2$

Par rapport à la construction de l'intervalle de confiance (IC), la méthode dite percentiles est choisie. Pour un niveau $1 - 2\alpha$, l'IC vaudra $[G_{Boot}^{-1}(\alpha); G_{Boot}^{-1}(1 - \alpha)]$ où G_{Boot} est la fonction de répartition de $\hat{\theta}$. Cet IC est préconisé par rapport à d'autres, comme ceux appelés t-Bootstrap. Il est le mieux adapté dans le cadre d'estimation de variance dans des sondages complexes comme celui de cette enquête. Il l'est également pour l'estimation de quantile [10].

Afin de mesurer la qualité des méthodes de Bootstrap programmées et de s'assurer que le nombre d'itérations choisi est suffisant, plusieurs indicateurs sont calculés. Il s'agit :

- du biais relatif (RB) : $BR(\widehat{\theta}_{Boot}) = 100 \cdot \frac{\widehat{\theta}_{Boot} - \hat{\theta}}{\hat{\theta}}$;
- de l'erreur quadratique moyenne (MSE) : $MSE(\widehat{\theta}_{Boot}) = E(\hat{\theta}_b - \hat{\theta})^2$;
- du coefficient de variation des estimations Bootstrap de variance (CV(\hat{V})), un indicateur

8. L'objectif du PDS étant une précision sur l'IFT moyen, le paramètre majoritairement testé a été la moyenne. D'autres tests ont été effectués sur un percentile d'un IFT spécifique : le 70^e percentile de IFT_HBC , l'IFT excluant les produits de biocontrôle.

de stabilité :

$$CV(\widehat{V}_{Boot}) = \frac{\sqrt{V(\widehat{V}_{Boot})}}{E(\widehat{V}_{Boot})}.$$

Bootstrap de type pseudo-population :

Pour un plan de sondage à probabilités inégales, D. Haziza, C. Léger et Z. Mashreghi proposent un algorithme très approprié à notre problème [11]. Cet algorithme décrit un Bootstrap de la classe des méthodes de pseudo-population (parmi trois classes formalisées par les auteurs). Une pseudo-population, notée U^* , est une reconstruction d'une population de taille N à partir des n unités ré pondantes. Pour résumer cette opération, on va répliquer chaque unité k de l'échantillon s autant de fois que son poids $\frac{1}{\pi_k}$. Comme la littérature le souligne [10], un des problèmes est l'obtention d'inverses de probabilités d'inclusion entières. Pour y arriver, la pseudo-population a été créée en deux sous-parties. La première l'est en dupliquant $\lfloor \frac{1}{\pi_k} \rfloor$ fois chaque parcelle k (où $\lfloor \cdot \rfloor$ est la partie entière); la seconde via un tirage de Poisson, avec pour chaque k la probabilité d'inclusion $\frac{1}{\pi_k} - \lfloor \frac{1}{\pi_k} \rfloor$.

Après une simplification des écritures, le résumé de la méthode est :

1. Pour c allant de 1 à C (le nombre de créations de pseudo-population choisi) :
 - (a) constituer en deux temps la pseudo-population U^* (voir ci-dessus);
 - (b) pour b allant de 1 à B , le nombre d'itérations choisi;
 - i. tirer un échantillon selon le PDS de l'enquête;
 - ii. calculer $\hat{\theta}_b$;
 - (c) calculer : $v_{Boot}^c = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b - \widehat{\theta}_{Boot})^2$ avec $\widehat{\theta}_{Boot} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b$;
2. Calculer l'estimateur Bootstrap de la variance de θ : $\widehat{V}_{Boot} = \frac{1}{C} \sum_{c=1}^C v_{Boot}^c$

En prenant B et C grands, on en déduit l'IC à un niveau de confiance de 95 % :

$$IC_{0,95}(\theta) = \left[\hat{\theta} - 1,96\sqrt{\widehat{V}_{Boot}}; \hat{\theta} + 1,96\sqrt{\widehat{V}_{Boot}} \right].$$

3.2 Estimer l'impact de la baisse des allocations sur la précision

3.2.1 Objectif

Comme évoqué en conclusion de la première partie, la précision est trop inégale d'un bassin viticole à un autre. Une meilleure répartition des allocations est possible, notamment en réduisant celle des bassins où la précision est très basse (ex : 0,8 % en Charentes). L'objectif recherché des simulations est, dans chaque bassin viticole, de :

prévoir l'impact d'une baisse de l'allocation sur la précision de l'IFT moyen.

3.2.2 Résumé des simulations successives

Dans un premier temps, les contraintes du futur PDS (stratifier selon le type de vin) ne sont pas intégrées. Certains bassins, comme le Bordelais, présentent des caractéristiques idéales. En comportant une seule strate en 2016, aucune post-stratification n'est effectuée et le COEF2 est constant dans tout le bassin, sans différence de rapports de poids avant et après calage entre les

observations. **Il s'agit d'estimer dans quelle mesure la précision de l'IFT moyen aurait été supérieure avec un nombre de questionnaires exploités de PhytoViti2016 plus bas.** Avec 546 questionnaires exploités, la précision de l'IFT moyen diffusé était de 1,07 % dans ce bassin.

La première classe de Bootstrap essayée est celle des **Bootstrap directs** [11]. Ils consistent à sélectionner les échantillons Bootstrap directement dans l'échantillon d'origine. Les premiers essais ont été des Bootstrap en sous-échantillonnant sans remise.

Lors de ces estimations de l'IFT moyen (la variable d'intérêt), la précision de l'IFT moyen peut être calculée à chacune des B itérations grâce à l'utilisation de la fonction CalvaR en R. La précision étudiée est celle calculée en utilisant la variance de Yates-Grundy car elle est légèrement plus faible que celle obtenue à partir de la variance d'Horvitz-Thompson mais celle-ci est également calculée car elle est disponible en sortie de la fonction CalvaR⁹. Cette dernière permet de prendre en compte les propriétés du PDS dans les calculs de variance et de précision.

Ces simulations par méthode de réplcation sur les données de PhytoViti2016 sont possibles tout en mesurant leur qualité (biais relatif et erreur quadratique moyenne de l'estimateur de l'IFT moyen).

Ces estimations reposent sur deux hypothèses :

- la non-réponse est négligeable. Cette méthode ne la prend pas en compte. Cela se justifie principalement par les taux de réponse historiquement très élevés (95 % et plus).
- la composition des B sous-échantillons dont les calculs sont issus a un effet de structure par rapport au type de vin. Dans un quart des bassins, cela ne posera pas de problème car un seul type de vin existe. Dans les autres, on émet l'hypothèse que les proportions des différents types de vins déclarés en 2016 (AOP, IGP, etc) seront conservées en 2019.

Afin d'adapter cette démarche à des bassins plus complexes (comprenant plusieurs strates, une post-stratification) et pour se rapprocher de plus en plus du PDS de 2016 puis de celui de 2019, se sont enchaînés des Bootstrap, via des méthodes dites directes, effectués :

- via des sous-échantillonnages (sans et avec remise) ;
- via des sur-échantillonnages ;
- avec intégration de propriétés du PDS de l'enquête (post-stratification, calage, propriété des poids dans certains croisements).

Dans le cadre de **Bootstrap directs**, il est préconisé de choisir un nombre d'itérations proche de l'inverse du taux de sondage par bassin $k = \frac{N}{n}$ pour les Bootstrap via des tirages aléatoires simples sans remise. Néanmoins, ces k sont trop petits (≤ 600) et aboutissent à des erreurs quadratiques moyennes et à des coefficients de variation des estimations Bootstrap de variance trop élevés. Des tests ont été effectués pour 500, 1 000, 2 000 et 5 000 itérations. Le compromis entre précision des résultats et temps de calcul va en faveur des 2 000 itérations. Par souci de parcimonie, les résultats de ces tests intermédiaires ne sont pas présentés.

Dans le cadre de **Bootstrap de type pseudo-population**, les résultats sont élaborés à partir de C=100 (C le nombre de pseudo-population créées) et B=1 000 (B le nombre d'échantillons tirés dans chaque pseudo-population C). Dans sa thèse [10], G. Chauvet propose un algorithme simplifié où tous les rééchantillons sont tirés dans la même pseudo-population selon le PDS de l'enquête. De plus, cette méthode est rapide à exécuter : une heure pour les calculs des vingt-et-un

9. En 2016, cette fonction en R remplace la macro SAS Calva (Calcul de variance). Elle devient usuelle dans la plupart des calculs méthodologiques des enquêtes du SSP.

bassins. Néanmoins, cette méthode aboutit à deux limites dans des bassins complexes de cette enquête, repérées lors de l'enregistrement de tables d'information sur les B échantillons :

- le nombre de répondants par strate est instable (une limite à relativiser car elle permet de se rapprocher de la diversité des cas possibles à l'issue de la collecte) ;
- l'absence de recalcul des poids pour intégrer la post-stratification et le calage dans les simulations.

De ces deux points en découle également la perte de l'égalité de la somme des poids par bassin et dans des sous-ensembles des bassins (obtenue grâce au calage).

Bootstrap à façon pour prendre en compte la non-réponse et du calage :

Afin de prendre en compte le plus possible le PDS de PhytoViti2016, un Bootstrap à façon est créé. Il permet notamment d'intégrer la post-stratification et le calage via une hypothèse simplificatrice : conserver la valeur de la surface agrégée dans chaque croisement $g = \text{Strate} \otimes \text{zone de calage}$. L'algorithme est le suivant :

1. Pour b allant de 1 à B (le nombre d'itérations choisi) :
 - (a) par rapport à la taille t fixe de chaque échantillon s_b tiré, calculer proportionnellement dans chaque croisement $g = \text{Strate} \otimes \text{zone de calage}$, t_g le nombre de parcelles à tirer sans remise (ou avec remise dans le bassin des Pyrénées-Orientales) ;
 - (b) dans chaque croisement g :
 - i. tirer un échantillon i.i.d. de taille t_g (tirage aléatoire simple sans ou avec remise) ;
 - ii. recalculer le poids $COEF2(c)$ en $w_2(c)$ de tous les individus de l'échantillon afin de garantir l'égalité :

$$\sum_{c \in s_b} w_2(c) \cdot \mathbb{1}_{c \in g \cap s_b} = surf(g)$$

- (c) calculer via la fonction CalvaR : $\hat{\theta}_b$ l'estimateur Bootstrap propre à b ainsi que les variances et précisions au sens d'Horvitz-Thompson et de Yates-Grundy ;
2. calculer $\widehat{\theta}_{Boot}$ et son intervalle de confiance à 95 % ;
3. calculer les indicateurs RB, MSE et $CV(\hat{V})$;
4. produire le graphique et le tableau des données associées des précisions : favorable (précision de l'estimateur correspondant au quantile 2,5 %), moyenne et défavorable (précision de l'estimateur correspondant au quantile 97,5 %).

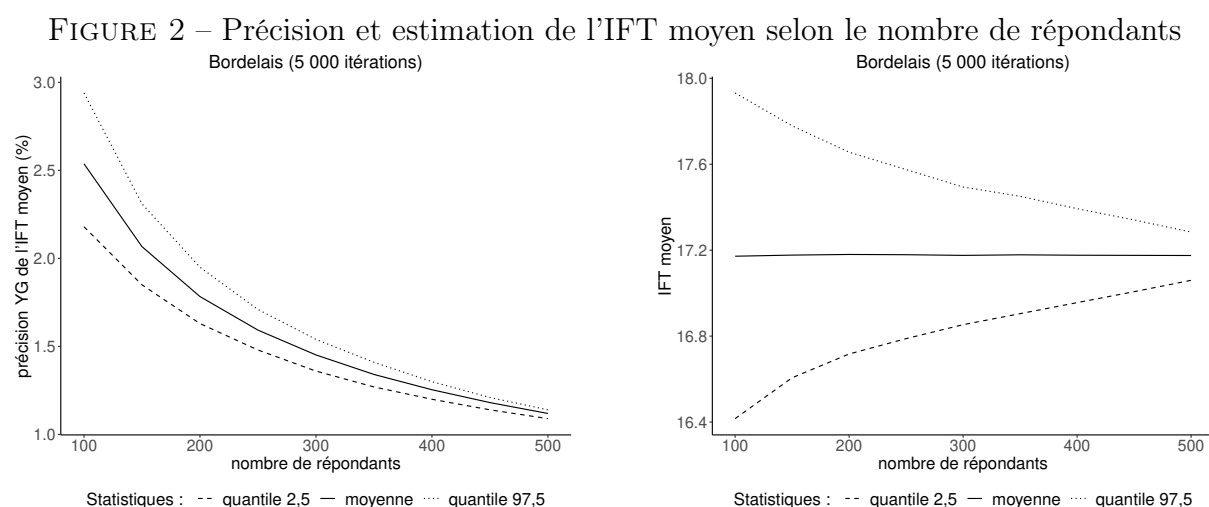
Plusieurs nombres d'itérations ont été testés (de 500 à 10 000) et les indicateurs de stabilité se sont avérés indispensables pour s'adapter aux différents bassins viticoles. Dans certains, il y a peu de gain et de différences entre 1 000 et 2 000 itérations. Néanmoins, le temps de calcul par bassin est long (plusieurs heures) même s'il offre plus de résultats par rapport au Bootstrap de type pseudo-population présenté dans le paragraphe précédent.

Avec 1 000 itérations, les biais relatifs et les erreurs quadratiques moyennes sont bas, ce qui atteste d'un nombre d'itérations suffisant pour avoir des résultats de qualité. Cette méthode est à employer avec précaution, notamment dans les croisements g avec un effectif très petit dans l'échantillon de départ s . Par rapport à la méthode de pseudo-population, elle permet d'intégrer un aléa supplémentaire : la post-stratification. En revanche, cela se fait toujours à partir des mêmes unités et dans le même ordre de grandeur.

3.2.3 Programmation et qualité

Même si deux packages du logiciel R (boot et bootstrap) permettent de faire du Bootstrap via différentes options, le Bootstrap a été codé manuellement. La raison est de se rapprocher le plus possible du PDS complexe et très particulier de ces enquêtes.

Les programmes sont conçus pour produire des résultats par bassin sous différentes formes : plusieurs graphiques, un tableau résumé et un enregistrement de la structure des B sous-échantillons Bootstrap. Un exemple pour le bassin Bordelais est présenté ci-dessous (cf. illustration 2). Le graphique de gauche permet de visualiser, en fonction du nombre de répondants, le niveau que devrait avoir la précision au sens de YG en moyenne et entre quelles valeurs elle devrait être comprise dans 95 % des cas, notamment celui d'une situation défavorable (quantile 97,5). Celui de droite représente l'intervalle de confiance (à 95 %) de l'IFT moyen associé.



Note de lecture : l'estimation de l'IFT moyen et de sa précision de Yates-Grundy sont issues d'un Bootstrap à façon (tirage proportionnel par g) de 5 000 itérations.

Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

Un autre avantage est de pouvoir calculer les indicateurs que l'on souhaite et de mettre en place des procédures spécifiques pour :

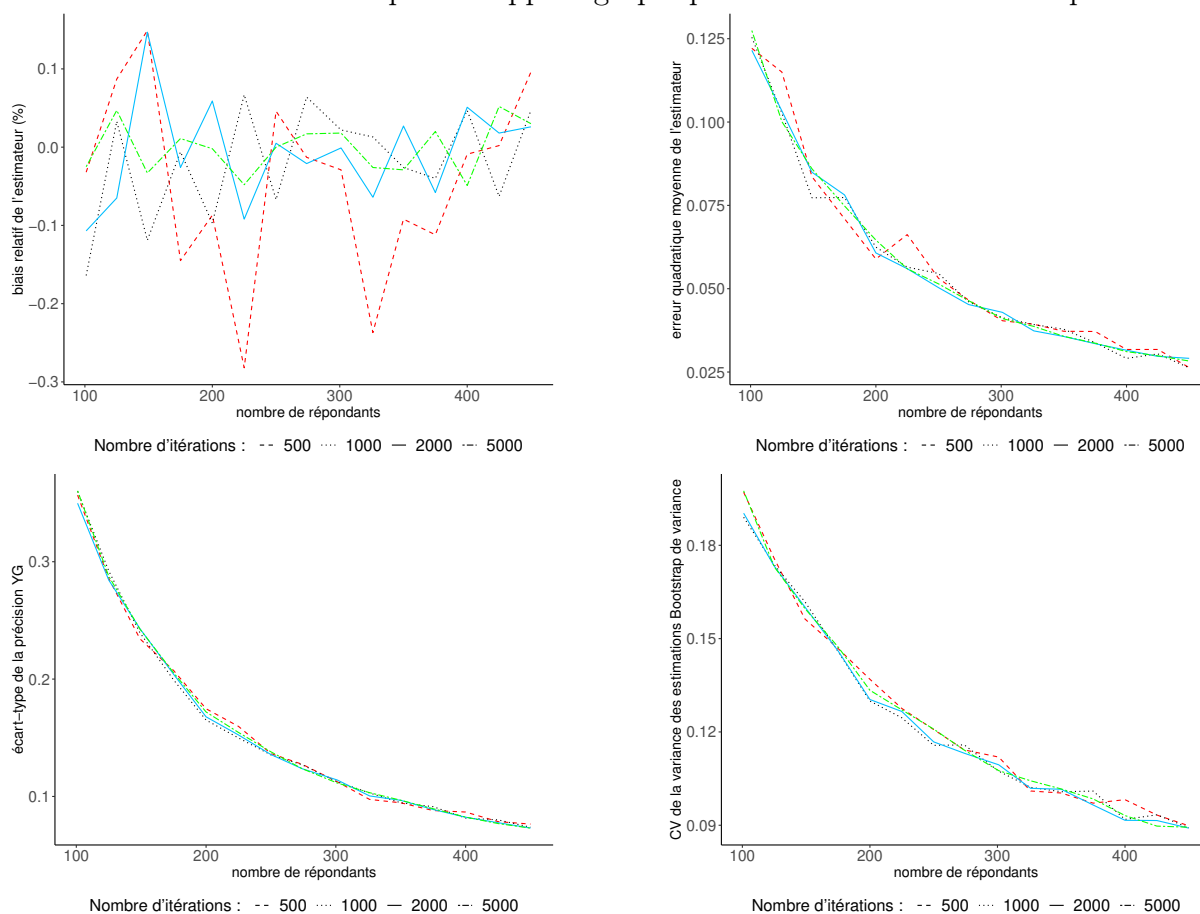
- effectuer des sauvegardes régulières quand le fardeau de calcul devient difficile à gérer ;
- créer un rapport graphique et vérifier efficacement la stabilité des simulations.

L'exemple de rapport graphique ci-après (cf. figure 3) concerne une simulation de type Bootstrap à façon pour estimer la moyenne de l'IFT dans le bassin des Bouches-du-Rhône, à partir d'échantillons de taille inférieure à celui issu de la collecte de PhytoViti2016. Les indicateurs sont présentés pour 500 (tirets rouges), 1 000 (pointillés noirs), 2 000 (traits continus bleus) et 5 000 itérations (alternance de tirets et points verts). Une analyse de ces graphiques montre que :

- le Bootstrap de 5 000 itérations a le biais relatif le plus faible ;
- les erreurs quadratiques moyennes sont très proches pour les Bootstrap de taille 1 000, 2 000 et 5 000 ;
- seulement ceux de 2 000 et 5 000 itérations sont stables (présence de légères hausses du $CV(\widehat{V}_{boot})$, quand le nombre de répondants augmentent, pour 1 000 itérations).

Dans cet exemple, un Bootstrap de 2 000 itérations est de qualité suffisante.

FIGURE 3 – Exemple de rapport graphique d’une méthode Bootstrap



Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

3.3 Calculs affinés d’allocations individualisés par bassin viticole pour PKViti2019

La mise en place de cette méthode visait d’abord à reproduire les différentes étapes du PDS de PhytoViti2016 (notamment la stratification par type de vin) afin de s’assurer que l’IFT moyen aurait bien été calculé sans biais et sans une détérioration trop importante de sa précision à partir d’un nombre de répondants plus petit que celui réellement obtenu à l’issue de la collecte.

Si les niveaux de précision semblent encore satisfaisants avec un nombre de répondants très bas, on ne pourra pas diminuer le nombre de questionnaires autant que les résultats ne le suggèrent. Comme une analyse le montre, la précision se détériore sur certaines déclinaisons de l’IFT (par segment : fongicide, etc) et sur les quantiles. Il faut donc garder une certaine marge également pour assurer une précision satisfaisante sur ces IFT.

De nouvelles simulations sont effectuées via la conception d’une nouvelle méthode de Bootstrap à façon, conçue pour se conformer au PDS de PKViti2019. Deux types de simulations sont nécessaires :

- dans les bassins non stratifiés ;
- dans les bassins candidats à la stratification selon le mode de conduite des parcelles (conventionnelle ou biologique).

Afin de simuler le PDS 2019, les coefficients doivent être recalculés dans ces deux types de simulation. Le calage qui sera effectué post-collecte n'est pas programmé mais approché par une simplification. Celle-ci consiste, par strate, à établir le *COEF2* comme la surface de la strate divisée par l'allocation de la strate. Sous des hypothèses vraisemblables de taux de réponse élevé et de faible post-stratification, le calage respectant une seule équation devrait aboutir à une transformation similaire des poids. Les résultats fournissent des allocations plus réalistes d'un point de vue opérationnel. Les simulations et calculs de précisions sont effectuées au sein de chaque strate le cas échéant et au niveau du bassin entier.

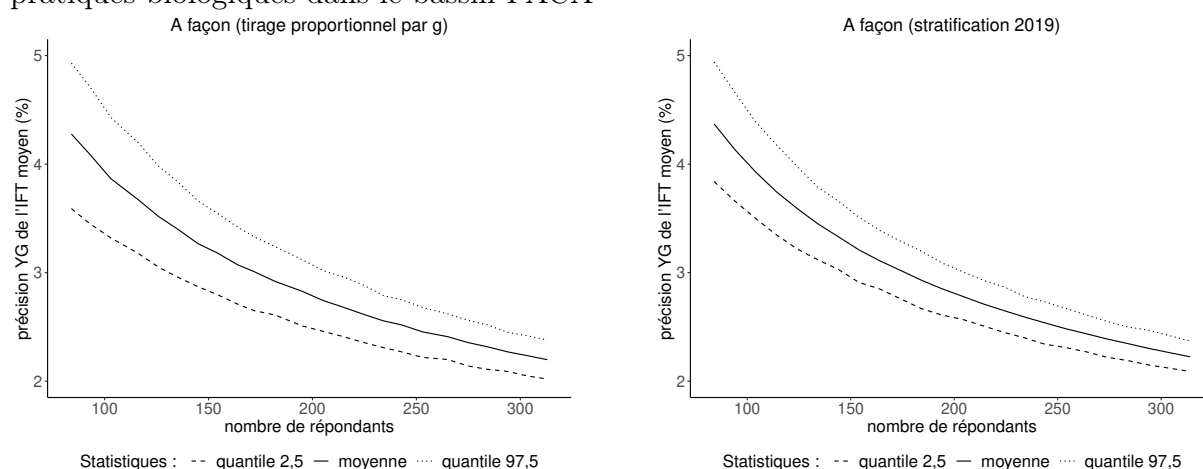
En revanche, les indicateurs de stabilité (biais relatif, erreur quadratique moyenne, coefficient de variation des estimations Bootstrap de variance) ne peuvent plus être calculés même en simulant les résultats de PhytoViti2016 via le PDS de PKViti2019. L'estimateur de l'IFT moyen sera en effet différent de celui calculé à partir de l'échantillon entier selon un PDS différent. Néanmoins, **le niveau que devrait avoir la précision au sens de YG en moyenne et dans certains cas (notamment celui d'une situation défavorable relative au quantile 97,5) peut être approché en fonction du nombre de répondants** selon une approche empirique similaire.

3.4 Un travail spécifique dans les bassins stratifiés

Comme évoqué en 2.1, la précision de l'IFT moyen des parcelles menées en Agriculture Biologique est plus élevée en raison d'un nombre trop faible d'unités répondantes (cf. tableau 4). Dans les neuf bassins candidats à la stratification selon le mode de conduite des parcelles, ce faible nombre d'unités répondantes s'accompagne également d'une plus grande variabilité de la caractéristique d'intérêt. La dispersion de l'IFT peut être plus de deux fois plus importante que celle des IFT des parcelles menées selon des pratiques conventionnelles.

Dans le fichier de données de PhytoViti2016, le bassin des Bouches-du-Rhône présente la double particularité d'avoir un nombre de parcelles menées selon des pratiques biologiques important (85, soit un quart du bassin) dont la dispersion de l'IFT est proche de celles menées selon des pratiques conventionnelles (12,9 contre 11,7). Une simulation via un nouveau type de Bootstrap à façon, y est testée dans la strate des parcelles menées selon des pratiques biologiques. Le double graphique ci-dessous (cf. illustration 4) montre le niveau de la précision de l'IFT moyen par rapport au nombre de répondants selon les deux méthodes, avec des résultats très proches.

FIGURE 4 – Évolution de la précision de l'IFT moyen des parcelles conduites selon des pratiques biologiques dans le bassin PACA

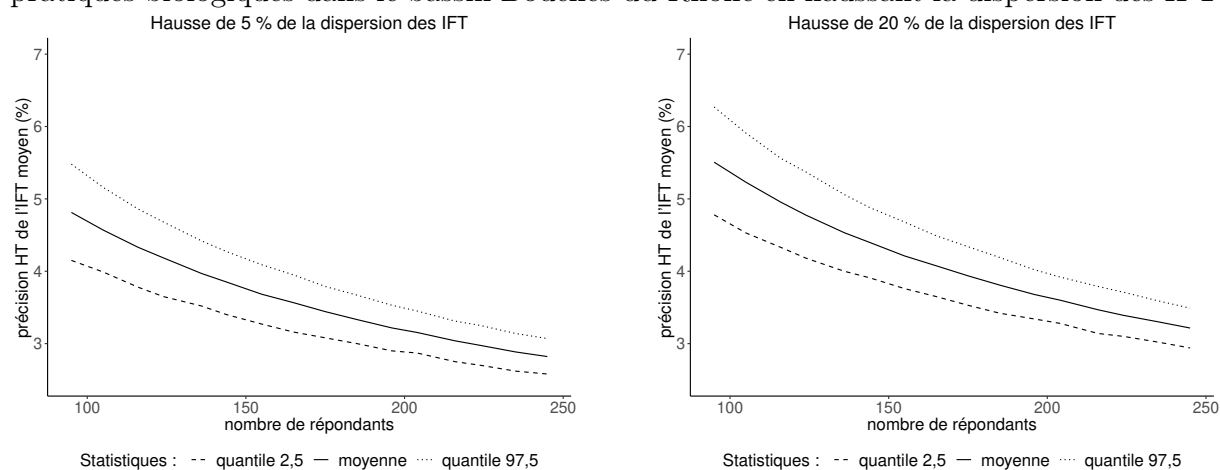


Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

Néanmoins, ces deux essais ont une limite importante. Les tirages avec remise à partir d'un nombre trop bas de valeurs peuvent engendrer des écarts dans les calculs d'estimation de variance. L'expérimentation s'est donc poursuivie avec l'intégration d'une modification de la dispersion des IFT, sans modification de la moyenne de l'IFT, par une transformation linéaire des IFT des parcelles menées selon des pratiques biologiques. Ainsi, l'hypothèse de similitude de la dispersion des futurs IFT de 2019 avec ceux de 2016 peut être relâchée. L'impact sur la précision de l'IFT moyen est mesuré.

Dans la plupart des neuf bassins candidats, la dispersion sera diminuée. À l'inverse, une hausse de la dispersion a été testée dans le bassin des Bouches-du-Rhône (cf. illustration 5) par précaution, en faisant une hypothèse de variabilité plus élevée des IFT des parcelles menées selon des pratiques biologiques que celle observée dans les unités répondantes de PhytoViti2016.

FIGURE 5 – Evolution de la précision de l'IFT moyen des parcelles conduites selon des pratiques biologiques dans le bassin Bouches-du-Rhône en haussant la dispersion des IFT



Note de lecture : en moyenne pour 200 répondants, la précision de l'IFT moyen sera respectivement de 3,15 % et de 3,6 % si la dispersion des IFT augmente de 5 % et de 20 %.

Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

À partir des résultats des différentes simulations, des premières allocations avancées ont été établies dans les vingt-et-un bassins déjà enquêtés en 2016. Avec peu d'information préalable et une taille de bassin très petite en superficie et en nombre de parcelles, le nouveau bassin du Jura obtient une allocation de 260 (cf. tableau 6). Les allocations issues des simulations ont été augmentées de 5 % en prévision d'une non-réponse historiquement proche de 5 % dans les précédentes enquêtes.

Ces premières allocations calculées aboutissent à des baisses notables de la charge des répondants dans certains bassins par rapport à la précédente enquête PhytoViti2016. C'est particulièrement le cas dans les grands bassins viticoles non stratifiés (Bordelais, Champagne, Charentes et Bourgogne), jusqu'à - 22 % dans le bassin Charentes.

Les allocations dans certains des neuf bassins candidats à la stratification selon le mode de conduite des parcelles semblent trop élevées au niveau du bassin entier. Eu égard à leur taille (superficie du vignoble), ce nombre est conséquent dans les bassins Corse et Bouches-du-Rhône, avec une probabilité forte d'avoir des répondants multi-interrogés dans de fortes proportions.

TABLE 6 – Allocations calculées pour le dossier de présentation au Comité du Label

Bassin viticole	Strate	Allocations	Précision attendue (en %)
Alsace	Conventionnelles	240	< 2,5
	Biologiques	160	2,5
	Ensemble	400	2,5
Beaujolais	Ensemble	320	< 2,5
Bordelais	Conventionnelles	420	< 2,5
	Biologiques	240	2,5
	Ensemble	660	< 2,5
Bouches-du-Rhône	Conventionnelles	260	2,5
	Biologiques	240	3
	Ensemble	500	2,5
Bourgogne	Ensemble	370	< 2,5
Champagne	Ensemble	420	< 2,5
Charentes	Ensemble	420	< 2,5
Dordogne	Ensemble	240	2,5
Languedoc hors PO	Conventionnelles	530	2,5
	Biologiques	210	2,5
	Ensemble	740	2,5
Pyrénées-Orientales	Conventionnelles	370	< 2,5
	Biologiques	260	3
	Ensemble	630	2,5
PACA (Var-Vaucluse)	Conventionnelles	420	< 2,5
	Biologiques	290	3
	Ensemble	710	2,5
Val de Loire	Conventionnelles	370	< 2,5
	Biologiques	180	2,75
	Ensemble	550	2,5
Gers	Ensemble	220	< 2,5
Gaillac	Ensemble	210	< 2,5
Cahors	Ensemble	220	< 2,5
Cher	Ensemble	240	< 2,5
Côtes-du-Rhône Nord	Ensemble	260	< 2,5
Côtes-du-Rhône Sud	Conventionnelles	180	2,5
	Biologiques	200	2,5
	Ensemble	380	2,5
Lot-et-Garonne	Ensemble	210	< 2,5
Corse	Conventionnelles	200	< 2,5
	Biologiques	240	3
	Ensemble	440	2,5
Bugey-Savoie	Ensemble	240	< 2,5
Jura	Ensemble	260	2,5
ENSEMBLE		8 640	< 2,5

Note de lecture : les allocations ci-dessus tiennent compte d'un sur-échantillonnage de 5 % en prévision d'une non-réponse historiquement de 5 % dans les précédentes enquêtes.

Les travaux se sont poursuivis dans le but de les affiner dans les neuf bassins candidats à la stratification. La démarche a consisté à estimer la précision de l'IFT moyen en faisant varier les allocations des deux strates de parcelles menées selon les différentes pratiques. Le tableau 7 présente un exemple des niveaux de précisions obtenus à partir de quelques couples d'allocations sous la forme d'un « intervalle de confiance » des précisions. Il ne s'agit pas d'intervalle de confiance au sens strict du terme mais l'abréviation « IC » est ici utilisée par abus de langage¹⁰.

TABLE 7 – Simulations de précisions de l'IFT moyen en fonction des allocations dans le bassin des Pyrénées-Orientales

Strate des pratiques conventionnelles		Strate des pratiques biologiques		Bassin entier	
Allocation	« IC » de la précision	Allocation	« IC » de la précision	Allocation	« IC » de la précision
150	[2,67 ; 3,35]	150	[3,36 ; 4,25]	300	[2,34 ; 2,90]
200	[2,36 ; 2,86]	150	[3,39 ; 4,25]	350	[2,08 ; 2,49]
200	[2,34 ; 2,85]	200	[2,93 ; 3,57]	400	[2,05 ; 2,46]
250	[2,12 ; 2,52]	200	[2,93 ; 3,60]	450	[1,86 ; 2,20]
250	[2,11 ; 2,53]	250	[2,63 ; 3,15]	500	[1,85 ; 2,18]
300	[1,94 ; 2,28]	250	[2,62 ; 3,15]	550	[1,70 ; 1,98]
300	[1,94 ; 2,87]	300	[2,40 ; 3,83]	600	[1,69 ; 1,97]
370	[1,77 ; 2,04]	260	[2,58 ; 3,07]	630	[1,56 ; 1,78]

Note de lecture : 250 réponses pour les parcelles menées selon des pratiques conventionnelles et 200 pour celles menées selon des pratiques biologiques aboutiraient pour 95 % des échantillons à une précision de Horvitz-Thompson comprise entre 1,86 % et 2,2 % dans le bassin.
Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

La dernière étape de la démarche a pour objectif de déterminer le couple d'allocations le moins coûteux garantissant une précision de 2,5 % dans le bassin et de 3,5 % maximum dans la strate des parcelles menées selon des pratiques biologiques. En appliquant cette méthode dans les neuf bassins qui comprendront deux strates, les allocations peuvent être revues à la baisse (cf. tableau 8).

TABLE 8 – Récapitulatif de la révision des allocations dans les bassins pluri-strates

Bassin viticole	Strate des pratiques conventionnelles		Strate des pratiques biologiques		Bassin entier	
	Allocation	« IC » de la précision	Allocation	« IC » de la précision	Allocation	« IC » de la précision
Alsace	120	[1,60 ; 2,17]	120	[2,19 ; 3,41]	240	[1,42 ; 1,88]
Bordelais	290	[1,36 ; 1,64]	140	[2,74 ; 3,39]	430	[1,28 ; 1,54]
Bouches-du-Rhône	200	[1,78 ; 2,32]	200	[2,87 ; 3,49]	400	[1,78 ; 2,32]
Languedoc hors PO	310	[1,39 ; 1,69]	210	[2,71 ; 3,47]	520	[1,29 ; 1,55]
Pyrénées-Orientales	220	[2,27 ; 2,78]	220	[2,85 ; 3,48]	440	[1,98 ; 2,40]
Corse	220	[1,82 ; 2,37]	220	[2,72 ; 3,44]	440	[1,68 ; 2,15]
PACA (83 et 84)	190	[2,23 ; 2,87]	170	[2,93 ; 3,49]	360	[1,93 ; 2,43]
Val de Loire hors Sancerre	340	[1,47 ; 1,68]	340	[2,44 ; 3,44]	680	[1,37 ; 1,57]
Côtes-du-Rhône Sud	160	[1,93 ; 2,43]	130	[2,62 ; 3,45]	290	[1,77 ; 2,22]

Note de lecture : les précisions sont calculées à partir des données de PhytoViti2016 en simulant le plan de sondage de PKViti2019. Les allocations tiennent compte d'une augmentation de 5 % (en prévision de la non-réponse). Dans le bassin Languedoc hors Pyrénées-Orientales, 310 questionnaires de parcelles conventionnelles et 210 menées selon des pratiques biologiques aboutiraient pour 95 % des échantillons à une précision de HT comprise entre 1,29 % et 1,55 % au niveau du bassin entier.
Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.

10. Ce sont les précisions qui seraient obtenues dans les cas les plus favorables et défavorables de 95 % des échantillons à l'issue de ces simulations.

Avec ces allocations, le coût total de l'enquête serait inchangé pour vingt-et-un bassins : 7 170 questionnaires contre 7 156 en 2016 (cf. tableau 9).

TABLE 9 – Allocations de PKViti2019 recalculées pour minimiser le coût sous contraintes

Bassin viticole	Strate	Allocations	Précision attendue (en %)
Alsace	Conventionnelles	120	< 2,5
	Biologiques	120	3,5
	Ensemble	240	2
Beaujolais	Ensemble	320	< 2,5
Bordelais	Conventionnelles	290	< 2
	Biologiques	140	< 3,5
	Ensemble	430	< 2
Bouches-du-Rhône	Conventionnelles	200	< 2,5
	Biologiques	200	< 3,5
	Ensemble	400	< 2,5
Bourgogne	Ensemble	370	< 2,5
Champagne	Ensemble	420	< 2,5
Charentes	Ensemble	420	< 2,5
Dordogne	Ensemble	240	2,5
Languedoc hors PO	Conventionnelles	310	< 2
	Biologiques	210	< 3,5
	Ensemble	520	< 2
Pyrénées-Orientales	Conventionnelles	220	< 3
	Biologiques	220	< 3,5
	Ensemble	440	< 2,5
PACA (Var-Vaucluse)	Conventionnelles	190	< 3
	Biologiques	170	< 3,5
	Ensemble	360	< 2,5
Val de Loire	Conventionnelles	340	< 2
	Biologiques	340	< 3,5
	Ensemble	680	< 2
Gers	Ensemble	220	< 2,5
Gaillac	Ensemble	210	< 2,5
Cahors	Ensemble	220	< 2,5
Cher	Ensemble	240	< 2,5
Côtes-du-Rhône Nord	Ensemble	260	< 2,5
Côtes-du-Rhône Sud	Conventionnelles	160	< 2,5
	Biologiques	130	< 3,5
	Ensemble	290	< 2,5
Lot-et-Garonne	Ensemble	210	< 2,5
Corse	Conventionnelles	220	< 2,5
	Biologiques	220	< 3,5
	Ensemble	440	< 2,5
Bugey-Savoie	Ensemble	240	< 2,5
Jura	Ensemble	260	2,5
ENSEMBLE		7 430	< 2,5

Note de lecture : les allocations ci-dessus tiennent compte d'un sur-échantillonnage de 5 % en prévision d'une non-réponse historiquement de 5 % dans les précédentes enquêtes.

3.5 Derniers ajustements et allocations officielles

Suite à ces travaux préparatoires et en amont de la collecte, certaines modifications ont été apportées au plan de sondage de PKViti2019 :

- la stratification dans le bassin Corse est abandonnée, principalement en raison d'un risque trop élevé qu'il y ait une part importante d'exploitants interrogés pour plusieurs parcelles en raison du faible nombre d'EVV connu *ex ante* menant leurs parcelles selon des pratiques biologiques (malgré une proportion importante de la surface du vignoble en AB) ;
- la stratification dans le bassin des Pyrénées-Orientales (PO) est également annulée, majoritairement à cause de la charge supplémentaire qui aurait été très importante localement ;
- les allocations d'autres bassins (Alsace, Bordelais, Languedoc hors PO, PACA (Var-Vaucluse) et Côtes-du-Rhône Sud) ont été établies à un niveau inférieur à celui présenté dans le dossier présenté au Comité du Label mais supérieur à celui minimisé, par précaution.

Sept bassins sur vingt-deux sont donc stratifiés sur le critère du mode de conduite : ceux où la superficie de vignes menées selon des pratiques biologiques est la plus importante.

4 Collecte et précisions des données diffusées

4.1 Une collecte marquée par une part de non-réponse plus forte

La collecte de l'enquête PKViti2019 s'est déroulée au premier semestre 2021 à partir du 15 janvier. La seconde partie a été marquée par la crise sanitaire du Covid-19 avec pour conséquence l'impossibilité du maintien du protocole de collecte à partir de mi-mars. Le taux de non-réponse, historiquement faible dans les enquêtes PK, a donc été plus élevé que d'habitude (6,7 % avant apurement). Dans certains bassins, des difficultés supplémentaires ont été rencontrées.

La publication des résultats d'enquête [13] établit 4 % de parcelles hors champ et 7,8 % de questionnaires restés sans réponse (viticulteur injoignable, refus, réponses trop partielles). Au final, 6 849 questionnaires sont exploitables (soit 88,2 % de l'échantillon initial) dont 1 874 parcelles conduites selon le mode biologique.

Les traitements méthodologiques post-collecte attendus ont été effectués :

- correction de la non-réponse totale par l'augmentation du poids des répondants ;
- calage pour faire coïncider les superficies viticoles départementales estimées par l'enquête avec celles de la Statistique agricole annuelle 2019.

Un calage supplémentaire a été effectué pour faire correspondre la part des surfaces menées selon des pratiques biologiques par département avec ceux des fichiers de l'Agence Bio.

4.2 Les objectifs de précision atteints dans la majorité des strates et bassins

Les précisions de l'IFT moyen calculées à l'issue de la collecte (cf. tableau 10) sont majoritairement inférieures ou égales à 2,5 % : dans dix-huit bassins viticoles sur vingt-deux. La baisse des allocations a été maîtrisée, avec une légère hausse de la précision tout en restant inférieure à 2,5 %. Il en résulte une diminution considérable de la charge des répondants dans les grands bassins viticoles non stratifiés (Bordelais, Champagne, Charentes, Bourgogne et Beaujolais) où certaines EVV avaient été multi-interrogées lors des enquêtes précédentes. Dans ces bassins, entre 92 et 146 questionnaires exploitables ont été économisés par rapport à 2016 tout en conservant

une précision basse de l'IFT moyen (inférieure à 2 %). Des gains ont également été réalisés dans les bassins de plus petite surface (ex : - 73 questionnaires en Dordogne pour une précision similaire).

TABLE 10 – Précision de l'IFT moyen à l'issue de la collecte de PKViti2019

Bassin viticole	Strate	Allocations	Nombre de questionnaires exploitables	Taux de questionnaires exploitables (en %)	Précision de l'IFT moyen (en %)
Alsace	Conventionnelles	170	152	89,4	2,2
	Biologiques	170	164	96,5	3,5
	Ensemble	340	316	92,9	1,9
Beaujolais	Ensemble	320	277	86,6	1,7
Bordelais	Conventionnelles	360	318	88,3	1,4
	Biologiques	240	221	92,1	3,1
	Ensemble	600	539	89,8	1,3
Bouches-du-Rhône	Conventionnelles	240	179	74,6	3,0
	Biologiques	240	158	65,8	2,7
	Ensemble	480	337	70,2	2,3
Bourgogne	Ensemble	370	350	94,6	1,5
Champagne	Ensemble	420	392	93,3	1,4
Charentes	Ensemble	420	399	95,0	1,1
Dordogne	Ensemble	240	219	91,3	2,2
Languedoc hors PO	Conventionnelles	420	337	80,2	1,9
	Biologiques	210	184	87,6	3,8
	Ensemble	630	521	82,7	1,8
Pyrénées-Orientales	Ensemble	325	256	78,8	2,7
PACA (Var-Vaucluse)	Conventionnelles	310	246	79,4	2,1
	Biologiques	230	219	95,2	5,5
	Ensemble	540	465	86,1	2,0
Val de Loire	Conventionnelles	360	318	88,3	2,1
	Biologiques	260	244	93,8	3,0
	Ensemble	620	562	90,6	1,9
Gers	Ensemble	220	203	92,3	1,7
Gaillac	Ensemble	210	196	93,3	2,1
Cahors	Ensemble	220	190	86,4	2,8
Cher	Ensemble	240	222	92,5	2,2
Côtes-du-Rhône Nord	Ensemble	260	243	93,5	2,3
Côtes-du-Rhône Sud	Conventionnelles	180	154	85,6	4,3
	Biologiques	180	172	95,6	4,8
	Ensemble	360	326	90,6	3,8
Lot-et-Garonne	Ensemble	210	193	91,9	2,1
Corse	Ensemble	240	193	80,4	2,7
Bugey-Savoie	Ensemble	240	217	90,4	1,9
Jura	Ensemble	260	233	89,6	2,1
ENSEMBLE		7 765	6 849	88,2	0,6

Source : SSP - Agreste - Enquête sur les pratiques culturales en viticulture 2019.

Parmi les quatre bassins où la précision de l'IFT moyen a été supérieure à 2,5 %, celui des Pyrénées-Orientales est encore présent avec une précision de 2,7 % pour 256 questionnaires exploitables. Depuis l'enquête de 2010, la précision est moins bonne dans ce bassin que dans les autres (cf. partie 1). Le taux de questionnaires exploitables (78,8 %) est le second plus faible. L'objectif de précision de 2,5 % aurait vraisemblablement été atteint avec un taux de questionnaires exploitables compris entre 85 % et 90 %. Les seuils minimaux de nombre de questionnaires à interroger par bassin des enquêtes précédentes (respectivement 200 et 250 pour PKViti2013 et PhytoViti2016) n'auraient pas été suffisants, ce bassin nécessitant un suréchantillonnage.

Dans les bassins Corse et Cahors, la précision de l'IFT moyen est très proche de 2,5 %, avec des taux de questionnaires exploitables figurant également parmi les plus bas (respectivement

80,4 % et 86,4 %). Le seul bassin où l'objectif de précision a été loin d'être atteint est celui des Côtes-du-Rhône Sud (3,8 %). Avec plus de questionnaires exploitables qu'en 2016 (326 contre 286), la précision de l'IFT moyen y est moins bonne (- 2,1 points). Le taux de questionnaires non exploitables du bassin (9,4 %) est légèrement plus bas que dans les autres bassins, mais est deux fois plus important que celui prévu (5 %) lors de la préparation du PDS. Ce niveau de précision reste bon quand on relativise par rapport à ceux généralement obtenus dans les enquêtes.

Concernant la précision de l'IFT moyen des parcelles menées selon des pratiques biologiques, l'objectif de 3,5 % maximum a été atteint dans quatre bassins, presque cinq (3,8 % en Languedoc hors PO). Comme pour la strate des parcelles menées selon des pratiques conventionnelles (4,3 %), la précision a été plus élevée que prévue dans le bassin des Côtes-du-Rhône Sud (4,8 %). Elle dépasse 5 % dans le bassin Provence (Var-Vaucluse) malgré 219 questionnaires exploitables.

Des IFT moyens ont également été diffusés pour les parcelles conduites en mode biologique (certifiées ou en conversion) dans certains bassins non stratifiés (ex : Cahors), avec des précisions comprises entre 4,9 % et 8,2 %. L'IFT moyen des parcelles menées selon des pratiques biologiques au niveau français intègre plus de bassins qu'initialement prévu et sa précision est de 1,6 %.

Conclusion

En définitive, ces calculs empiriques d'allocations s'avèrent encourageants. Les objectifs de précision ont été atteints dans la plupart des bassins viticoles, aux niveaux espérés. Le taux de questionnaires non-exploitables plus important que celui historiquement faible dans les enquêtes PK et la perturbation des conditions de collecte par le confinement n'ont pas été sans conséquence dans certains bassins où les niveaux de précision ont été dépassés. L'opportunité d'utiliser les données des enquêtes précédentes d'une manière plus développée via des méthodes de réplification peut être saisie en visant plusieurs objectifs. Les plus importants consistent à étudier la possibilité de construire une stratification plus efficace et d'affiner la détermination des allocations.

Des enseignements peuvent en être tirés. Le principal concerne l'individualisation des calculs d'allocations par bassin viticole. Les précisions calculées à l'issue de l'enquête valident de manière générale la pertinence de cette approche. Elle pourrait être conservée dans le cadre de la préparation des plans de sondage des futures enquêtes, quels que soient le tirage et la stratification choisis. En effet, un autre enseignement est que cette méthode permet de simuler l'intégration de nouvelles demandes ou objectifs d'enquête, comme l'interrogation en AB dans PKViti2019, selon un coût d'enquête similaire à celui des enquêtes précédentes. Cette méthode a l'avantage d'être opérationnelle assez rapidement avec des résultats explicites. Elle offre une alternative pour justifier le nombre de questionnaires à collecter par bassin, notamment auprès des services chargés de la collecte localement, tout en veillant à garantir la qualité des estimations.

Afin d'améliorer cette méthode, plusieurs pistes sont envisageables :

- utiliser les données de PKViti2019 pour améliorer la détermination des allocations des futures enquêtes, notamment dans les bassins stratifiés selon le mode de conduite des parcelles. Un nombre conséquent de données est désormais disponible en AB pour mieux étudier et prévoir la précision de l'IFT moyen selon ce mode de conduite ;
- calculer les allocations en prenant en compte la précision sur d'autres variables ou sur d'autres statistiques de l'IFT (ex : quantile) ;
- approfondir la connaissance des spécificités historiques (ex : bassin des Pyrénées-Orientales) de l'enquête et ajuster les allocations ;
- assurer une veille sur la qualité de la base de sondage et rechercher de nouvelles informations auxiliaires (ex : cépage) à mobiliser pour mettre en place une stratification plus efficace.

Bibliographie

- [1] Brunet, N., Guichard, L., Omon, B., Pingault, N., and Pley-Ber, E. et Seiler, A. (2008) L'indicateur de fréquence de traitements (IFT) : un indicateur pour une utilisation durable des pesticides, *Le Courrier de l'environnement de l'INRA*, déc 2008, volume 56, 131-141, 2008.
- [2] Ardilly, P., *Les techniques de sondage*, Technip, Paris, 2006.
- [3] Statistiques Canada, *Méthodes et pratiques d'enquête*, octobre 2003, 12-587-X chapitre 8, 2003.
- [4] Simonovici, M., *Enquête pratiques phytosanitaires en viticulture en 2016 : nombre de traitements et indicateurs de fréquence de traitement*, Agreste les Dossiers 2019-2, 2019.
- [5] Pujol, J., *Apports de produits phytosanitaires en viticulture et climat : une analyse à partir des enquêtes pratiques culturelles*, Agreste les Dossiers 39, 2017.
- [6] Tillé, Y., *Théorie des sondages : échantillonnage et estimation en populations finies*, Dunod, 2001.
- [7] Simonovici, M., *Hausse des traitements fongicides en viticulture entre 2010 et 2016 : une évaluation de l'impact des différentes pratiques culturelles*, Agreste les Dossiers 2021-1, 2021.
- [8] Haziza, D., Mecatti, F. et Rao, J.N.K., *Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design*, *Metron - International Journal of Statistics*. LXVI., 91-108, 2008.
- [9] Efron, B., *Bootstrap methods : another look at the jackknife*, *Annals of Statistics*, 7, 1-26, 1979.
- [10] Chauvet, G., *Méthodes de Bootstrap en population finie*, 2008.
- [11] Mashreghi, Z., Haziza, D. et Léger, C., *A survey of bootstrap methods in finite population sampling*, *Statistics Surveys*, volume 10, 1-52, 2016.
- [12] Fromont, M. et Vimond, M., *Bootstrap et rééchantillonnage. Atelier SFdS – Partie 1*, 2012.
- [13] Caray J. et Simonovici, M., *Enquête Pratiques culturelles en viticulture en 2019 - IFT et nombre de traitements*, Agreste Chiffres et Données 2021-19, 2021.

Annexe 1 - exemple de calculs d'IFT

Ci-dessous, un exemple fictif de calculs de l'IFT total et des IFT segments :

TABLE 11 – Première partie du tableau d'un exemple de calcul des IFT

n° Passage	Produit	Fonction	Cible	Biocontrôle	Part de surface traitée (en %)	Dose saisie	Unité
1	BARBARIAN XL	HERBICIDE	HER	0	30	2	L/HA
2	PREV-AM	FONGICIDE	MIL	1	100	1,6	L/HA
2	PROSPER	FONGICIDE	OID	0	100	0,6	L/HA
3	DITHANE NEOTEC	FONGICIDE	MIL	0	100	3	KG/HA
3	MILORD	FONGICIDE	OID	0	100	0,5	L/HA
4	NORDOX 75 WG	FONGICIDE	MIL	0	100	2	KG/HA
4	MAYANDRA	FONGICIDE	OID	0	100	0,5	L/HA
5	NORDOX 75 WG	FONGICIDE	MIL	0	100	2	KG/HA
5	PROSPER	FONGICIDE	OID	0	100	0,6	L/HA
5	FASTAC	INSECTICIDE_ACARICIDE	CIC	0	100	0,2	L/HA
5	PROCLAIM	INSECTICIDE_ACARICIDE	TORD	0	100	1,5	KG/HA
6	COLLIS	FONGICIDE	OID	0	100	0,4	L/HA
7	INOX	FONGICIDE	OID	0	100	0,6	L/HA
7	BOULLIE BORDELAISE RSR DISPERS	FONGICIDE	MIL	0	100	4	KG/HA
7	PROCLAIM	INSECTICIDE_ACARICIDE	TORD	0	100	1,5	KG/HA
8	BOULLIE BORDELAISE RSR NC	FONGICIDE	MIL	0	100	4	KG/HA

TABLE 12 – Seconde partie du tableau d'un exemple de calcul des IFT

n° Passage	Dose calculée (en plein)	Dose de référence	IFT avant redressement	Redressement	IFT moyen du segment	IFT traitement redressé
1	6,667	8	0,25	0	0,788	0,25
2	1,6	1,6	1	0	0,834	1
2	0,6	0,6	1	0	0,834	1
3	3	2	1,5	0	0,834	1,5
3	0,5	0,5	1	0	0,834	1
4	2	2	1	0	0,834	1
4	0,5	0,5	1	0	0,834	1
5	2	2	1	0	0,834	1
5	0,6	0,6	1	0	0,834	1
5	0,2	0,2	1	0	0,96	1
5	3,1	1,5	2,067	1	0,96	0,96
6	0,4	0,4	1	0	0,834	1
7	0,6	0,6	1	0	0,834	1
7	4	3,75	1,067	0	0,834	1,067
7	1,5	1,5	1	0	0,96	1
8	4	12	0,333	0	0,834	0,333

Les différents IFT par segment valent donc :

TABLE 13 – Valeurs des IFT liés à ce tableau de traitements

IFTH	IFTF	IFTF_HBC	IFTI	IFTI_HBC	IFTA	IFT_HBC	IFT
0,25	11,9	10,9	2,96	2,96	0	14,11	15,11

- IFTH : IFT herbicide concernant les produits permettant de détruire les mauvaises herbes ainsi que les produits utilisés pour l'épamprage chimique ;
- IFTF : IFT fongicide concernant les produits permettant de lutter contre les bactéries et les maladies cryptogamiques dues aux champignons microscopiques ;
- IFTF_HBC : IFT fongicide hors produits de biocontrôle ;
- IFTI : IFT insecticide - acaricide : concerne les produits permettant de lutter contre les insectes, les acariens et les nématodes ;
- IFTI_HBC : IFT insecticide - acaricide hors produits de biocontrôle ;
- IFTA : IFT autre concernant les produits telles que les substances de croissance et les stimulateurs de défense naturelle ;
- IFT_HBC : IFT total hors produits de biocontrôle ;
- IFT : IFT total, tous types de traitements confondus, c'est la somme des IFT par segment.

Annexe 2 - informations sur les bassins enquêtés en 2016

TABLE 14 – Informations sur les vingt-un bassins de l'enquête PhytoViti2016

Bassin viticole	Nombre de sous-parcelles	Nombre de répondants	Taux de sondage de sous-parcelles (en %)	Surface (en ha)	Taux de sondage en surface (en %)
Alsace	170 752	280	0,16	15 531	0,33
Beaujolais	73 590	370	0,5	18 685	1,19
Bordelais	328 800	546	0,17	114 076	0,51
Bourgogne	150 308	442	0,29	28 018	0,85
Champagne	325 853	514	0,16	30 514	0,53
Charentes	165 381	545	0,33	78 079	0,98
Languedoc hors PO	398 277	634	0,16	197 922	0,43
Pyrénées-Orientales	62 978	198	0,31	25 720	0,89
PACA (Var-Vaucluse)	187 873	496	0,26	78 249	0,65
Val de Loire hors Sancerre	166 868	502	0,3	46 346	0,91
Bouches-du-Rhône	18 201	330	1,81	10 313	4,61
Dordogne	30 622	292	0,95	11 515	2,12
Gers	37 386	230	0,62	18 172	1,42
Gaillac	18 593	181	0,97	6 548	2,12
Cahors	16 934	177	1,05	4 812	2,32
Lot-et-Garonne	16 422	197	1,2	5 914	2,66
Côtes-du-Rhône Nord	26 556	237	0,89	6 743	2,36
Côtes-du-Rhône Sud	53 761	286	0,53	19 553	1,13
Cher	26 720	247	0,92	4 169	2,8
Corse	4 746	211	4,45	5 848	13,95
Bugey-Savoie	20 206	241	1,19	2 694	2,93

Bassin viticole	Nombre de strates	Nombre de zones de calage	Ratio entre les COEF2	IFT moyen	Ecart-type IFT	Précision de l'IFT moyen (en %)
Alsace	1	2	1,06	14,93	3,27	1,3
Beaujolais	1	1	1	18,66	3,97	1,09
Bordelais	1	1	1	17,18	4,32	1,07
Bourgogne	1	1	1	19,34	4,58	1,11
Champagne	1	2	1,18	23,45	4,86	0,88
Charentes	1	2	1,02	18,02	3,44	0,81
Languedoc hors PO	4	9	6,48	13,97	3,9	1,1
Pyrénées-Orientales	4	3	10,18	10,41	4,17	3,09
PACA (Var-Vaucluse)	4	6	2,76	10,33	3,55	1,47
Val de Loire hors Sancerre	3	8	3,11	15,04	4,72	1,26
Bouches-du-Rhône	2	3	1,42	9,28	3,56	2
Dordogne	1	3	4,68	15,18	3,84	1,38
Gers	4	4	2,64	18,89	3,53	1,22
Gaillac	1	3	1,94	15,65	3,27	1,49
Cahors	1	3	1,53	15,78	4,27	1,96
Lot-et-Garonne	1	3	3,58	15,86	4,07	1,74
Côtes-du-Rhône Nord	2	1	1,16	13,6	4,29	1,97
Côtes-du-Rhône Sud	2	1	1,02	12,73	3,69	1,68
Cher	1	3	2,98	17,61	4,98	1,7
Corse	2	5	5,89	13,87	5,17	1,8
Bugey-Savoie	1	1	1	18,11	4,64	1,6

Source : Agreste - Enquête sur les pratiques phytosanitaires en viticulture 2016.