HAL open science

Analyse de la dynamique à long terme des rendements et de la mortalité sur un réseau de parcelles viticoles dans la région du Cognac

Céline Bial

To cite this version:

Céline Bial. Analyse de la dynamique à long terme des rendements et de la mortalité sur un réseau de parcelles viticoles dans la région du Cognac. Agronomie. 2020. hal-04051735

HAL Id: hal-04051735

https://hal.inrae.fr/hal-04051735v1

Submitted on 30 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACADEMIE DE MONTPELLIER

UNIVERSITE DE MONTPELLIER

MASTER

STATISTIQUE POUR LES SCIENCES DE LA VIE

MEMOIRE sur le stage

Analyse de la dynamique à long terme des rendements et de la mortalité sur un réseau de parcelles viticoles dans la région du Cognac

• effectué du 25/02/2020 au 31/08/2020

• à l'Institut national de recherche pour l'agronomie, l'alimentation et l'environnement, INRAE

• sous la direction de Nathalie SMITS

• par Céline BIAL

• soutenu le 10/09/2020

• devant la commission d'examen

J.N. BACRO, C. REYNES, R. SABATIER.

Remerciements

A l'issue de mon stage, je tiens à remercier toutes les personnes qui ont contribué à son bon déroulement.

En tout premier lieu, je souhaite remercier mon encadrante de stage, Nathalie Smits pour l'accompagnement prévenant, la connaissance et la sympathie qu'elle m'a dispensés tout au long du stage.

Je remercie également grandement Christian Gary pour son implication et les conseils qu'il m'a prodigués lors des nombreuses discussions et réunions que nous avons eues.

Je souhaite remercier Nadine Hilgert pour le temps qu'elle m'a accordé, les échanges enrichissants et les conseils statistiques apportés.

Un grand merci à Joseph Stoll, Vincent Dumot et à Jean-Baptiste Morineau, du BNIC, pour leur aide dans l'appropriation de la base de données ainsi que les conseils apportés.

Je remercie aussi les membres de l'équipe de l'UMR System pour leur gentillesse et leur écoute face à mes diverses interrogations.

J'adresse mes remerciements à l'équipe enseignante du master Statistiques des Sciences de la Vie de Montpellier pour la qualité de la formation reçue lors de ces deux années de master.

Table des matières

Introduction 2
I/ Contexte et objectif 3
1. La viticulture française 3
2. La viticulture à Cognac 3
3. Le dépérissement viticole 4
4. Le cycle reproductif de la vigne 5
II/ Matériel et méthodes 5
1. Organisation initiale des données 5
2. Variables climatiques 7
3. Création d'un lexique de variables 8
4. Regroupement de variables et tri initial de la base de données 8
5. Méthodes descriptives et sélection des variables 8
6. Méthodes statistiques 9
A. Modèle de régression linéaire 9
B. Arbre de décision 10
7. Sélection d'années particulières 12
III/ Résultats 12
1. Création et ajout de variables 12
2. La base de données mise en forme 14
3. L'âge des parcelles du réseau 16
4. Dynamique du dépérissement viticole au cours du temps 18
A. Les indicateurs du rendement 18
B. Les indicateurs de la mortalité 24
C. Sélection des années particulières 26
5. Etudes annuelles des indicateurs de rendement et de mortalité 27
A. L'année 1999 27
B. L'année 2004 28
C. L'année 2007 29
D. Comparaison des caractéristiques 30
IV/ Discussion 32
Conclusion 35
Bibliographie 36

Introduction

Pour conduire la vigne selon les objectifs de production souhaités, les viticulteurs adaptent leurs modes de production et le matériel employé dans leurs exploitations. Toutefois, la baisse de rendement et de longévité des ceps de vigne se fait de plus en plus ressentir par les professionnels de la viticulture.

Pour comprendre les déterminants de ce dépérissement viticole et en limiter l'impact, FranceAgriMer et le Comité national des interprofessionnels du vin (CNIV) ont lancé en 2015, le Plan National Dépérissement du Vignoble (PNDV). A travers ce projet, la filière veut agir contre la diminution des rendements et la hausse de la mortalité des ceps de vigne, à tous les niveaux de la viticulture, de manière concertée, en proposant notamment des financements pour les projets de recherche et de développement qui abordent le sujet (Aubert et al., 2017; des ROBERT et al., 2016). Le Bureau d'Informations et de Prévisions Economiques (BIPE), FranceAgriMer et la CNIV ont montré que dans ce combat, il faut prendre en compte l'ensemble des agressions biologiques possibles. Mais il faut également regarder l'ensemble des stress environnementaux, des pratiques culturales, des stratégies d'exploitation et des contraintes technico-socio-économiques. Ce regard apporte une autre dimension au projet PNDV. Il ne s'agit pas de trouver une solution unique pour certaines pathologies mais d'apporter une réponse adaptée pour chaque type de déterminant du dépérissement et chaque région viticole (Baldassini and Despey, 2015).

Le mémoire qui suit s'inscrit dans un projet de recherche appliquée sur les déterminants du dépérissement viticole, le projet « Trajectoires de Dépérissement des Vignobles » (TraDeVi). Ce projet est porté par l'unité de recherche UMR System de Montpellier. Il fait intervenir une association de plusieurs organismes de recherche, d'enseignement et de développement. Le projet TraDeVi a pour but d'identifier les facteurs et les interactions de facteurs biotiques, abiotiques et techniques qui contribuent à l'affaiblissement et au dépérissement des ceps de vigne(Aubert et al., 2017; Gary, 2017). Cette étude et caractérisation du dépérissement viticole peut être réalisée par le biais d'analyses de données relevées sur de longues périodes par les organismes de développement viticoles. Pour réaliser cette étude, je dispose d'un jeu de données constituée par le Bureau National Interprofessionnel du Cognac (BNIC) sur des parcelles de la région de Cognac.

L'objectif de ce mémoire est de caractériser le jeu de données mise à disposition, de porter un diagnostic sur les évolutions du rendement et de la mortalité, d'identifier les facteurs pouvant être liés à d'éventuelles diminutions du rendement viticole et hausses de la mortalité des ceps de vigne au cours du temps. Les problématiques de ce mémoire sont donc :

Est-ce que la base de données fournie par le BNIC permet d'identifier un dépérissement viticole ? Quels sont les facteurs explicatifs et sur quelles périodes agissent-ils ?

Pour ce faire, dans un premier temps nous verrons le contexte de l'étude. Ensuite nous examinerons et adapterons le jeu de données fourni pour le rendre utilisable pour l'étude. On emploiera les méthodes d'analyses statistiques nécessaires. Et enfin, nous examinerons les résultats obtenus pour répondre à la problématique.

I/ Contexte et objectif

1. La viticulture française

La France compte 750.000 hectares de vignobles au total, soit 10% de la surface mondiale de vignes de cuve. Ces vignobles se trouvent dans 80 départements français (Vin-Vigne.com, 2018). La France est le deuxième producteur mondial de vin en volume et le premier exportateur en valeur brute. Elle est également le deuxième pays consommateur de vin au monde devant l'Italie et derrière les Etats-Unis. Toutefois, une étude de FranceAgriMer montre une diminution de la consommation de vin par les français (FranceAgriMer, 2017). Ainsi, la consommation est passée de 100 litres par an par personne en 1975 à 40 litres par an par personne en 2020 (AFP, 2020).

Pour pallier cette diminution drastique de consommation et donc de vente, les vignobles se sont tournés vers une production favorisant la qualité à la quantité. Ainsi, les cépages de mauvaise qualité ont été au fil du temps remplacés par d'autres cépages plus qualitatifs. Cette modification des vignobles a été boostée par la mise en place de primes de l'Union Européenne pour l'arrachage de vigne (Barnier and Woerth, 2009). Toutefois, dans les dernières années la profession viticole a constaté un phénomène qualifié de dépérissement dans les vignobles, dont les causes multiples sont mal élucidées, et qui se manifeste comme une diminution de rendement subie des ceps et une augmentation de leur mortalité. Cette diminution de rendement à été estimé à 4.6hL/ha/an (CNIV and FranceAgriMer, 2018). Ainsi la viticulture doit dorénavant lutter contre le dépérissement précipité des vignes (Aubert et al., 2017; Baldassini and Despey, 2015).

2. La viticulture à Cognac

La viticulture en Poitou Charente est la 4ème plus grande superficie française destinée à la viticulture, avec un vignoble de plus de 80 385 ha dont 92,7% correspond à l'AOC Cognac (Préfecture de Charente, 2019). Le Cognac produit dans les vignobles de Poitou Charente, est un produit majoritairement destiné à l'exportation. Initialement, il était acheté par les navires hollandais qui venaient chercher le sel de la côte et aujourd'hui encore il est expédié pour plus de 90% vers des pays étrangers (Coussié, 1996).

La production de Cognac a subi de nombreuses crises aux cours des derniers siècles, remettant en cause la pérennité des vignobles de Charente. La crise du phylloxéra en est un bon exemple. Le phylloxéra est un insecte qui attaque la vigne en absorbant la sève des racines. Ils sont passés de 280 000 ha en 1877 à 42 581 ha en 1895. La reconstitution du vignoble s'effectue encore aujourd'hui, par le biais de porte-greffe invulnérables au

phylloxéra. Les cépages traditionnels ont peu à peu laissé place au cépage Ugni Blanc, plus résistant. Aujourd'hui, l'Ugni Blanc est utilisé à plus de 98% pour la production du Cognac. Il s'agit d'un cépage tardif d'origine italienne, très productif et résistant également à la pourriture grise. Pour une bonne pérennité des vignobles, une législation a été mise en place. La zone géographique de production du Cognac a été délimitée en 1909. Le Cognac a été reconnu comme une Appellation d'Origine Contrôlée (AOC) en 1936. Pour rentrer dans l'AOC il faut impérativement suivre le cahier des charges de l'AOC Cognac homologué par décret (Comité national des appellations d'origine relatives aux and vins et aux boissons alcoolisées, et des boissons spiritueuses, 2018).

3. Le dépérissement viticole

Le phénomène complexe du dépérissement viticole se retrouve dans toutes les régions viticoles de France. Il affecte la performance de production des exploitations viticoles en entrainant une diminution de la productivité des ceps et leur mort prématurée de manière progressive ou brutale (CNIV, 2017).

La hausse du dépérissement sur l'ensemble du vignoble français entraîne un manque à gagner conséquent qui ne cesse pas d'augmenter. A l'heure actuelle, on estime la perte de rendement liée au dépérissement à 4,6 hL/ha/an en France (CNIV and FranceAgriMer, 2018). La perte financière a été estimée autour du milliard d'euros pour l'année 2014 (des ROBERT et al., 2016). Cette perte de revenu est une des raisons qui fait passer la lutte contre le dépérissement viticole comme prioritaire. Quelques déterminants possibles du dépérissement ont été identifiés lors d'une étude réalisée par le BIPE (BIPE, 2015) (tableau 1). Nous favoriserons l'entrée de ces variables dans l'analyse des données.

Tableau 1 : Déterminants possibles du dépérissement viticole

Catégorie	Familles de déterminants possibles	Exemple de variables possibles
Facteurs techniques	Variables de production	*Rendement, Mortalité, Titre alcoométrique volumique probable*
	Variables de végétation	*Type de taille, Enherbement, Surface externe du couvert végétal*
	Variables d'amendement	*Fumures potassique, azotée, phosphaté*
	Variables de plantation	*Type de porte-greffe, Densité de plantation, Exposition de la parcelle*
Facteurs biotiques	Variables phytosanitaires	*Nombre de traitement contre des ravageurs et des pathogènes*
Facteurs abiotiques	Variables climatiques	*Précipitation, Nombre de jours de gel, Température*
	Variables d'analyse de sol	*Concentration du sol en divers nutriments*

Les variables phytosanitaires prennent en compte les contraintes liées aux maladies de la vigne qui affectent le dépérissement. Il en existe de deux types, les maladies du bois, qui attaquent les organes pérennes des ceps, et les maladies annuelles comme le mildiou, l'oïdium et le botrytis, qui ne persistent pas en hiver.

4. Le cycle reproductif de la vigne

Afin de caractériser le rendement et la mortalité des parcelles viticoles du réseau, il est nécessaire de comprendre le cycle de reproduction de la vigne. La vigne est une plante pérenne, son rendement est élaboré sur plus d'un an. La majorité des études affirme que le rendement est constitué au cours de deux années à partir de l'induction florale (Clingeleffer et al., 2001; Guilpart et al., 2014; Howell, 2001; Vasconcelos et al., 2009). De ce fait, l'analyse de deux cycles reproducteurs est nécessaire pour mettre en relief les facteurs modifiant le rendement. Il faut également prendre en compte les équilibres entre compartiments de la vigne. En effet, il existe une régulation interne de la physiologie du cep qui limite le développement de la partie reproductrice de la vigne en cas de développement particulièrement intense de la partie végétative (Carbonneau et al., 1977; Keller et al., 2010). Cette croissance végétative de la vigne dépend entre autres de la température, qui favorise ou limite celle-ci mais également l'accélère ou la ralentit (Grimaldi, 2018). L'impact du rayonnement est difficilement dissociable de la température mais les deux facteurs influencent l'élaboration du rendement. Le rayonnement a une incidence sur toutes les composantes du rendement, aussi bien grâce à la photosynthèse que par les effets direct sur le bourgeon (M. Carmo Vasconcelos et al., 2009).

II/ Matériel et méthodes

1. Organisation initiale des données

Les données utilisées sont été mises à notre disposition par le Bureau National Interprofessionnel du Cognac, le BNIC. Les données retracent le suivi de 55 parcelles par an de la zone géographique de Cognac, pour une période allant de 1977 à 2019.

Le jeu de données est sous la forme de 16 fichiers Excel qui possèdent des périodes de mesure plus ou moins longues et des échelles de mesure variables selon le fichier (tableau 2). Dans ces fichiers nous retrouvons les mesures faites sur 55 parcelles par an. Lorsqu'une parcelle quitte le réseau, la parcelle remplaçante récupère le numéro de la parcelle sortante et le numéro de génération de la nouvelle parcelle est celui de la parcelle sortante incrémenté de un. Chaque parcelle est donc identifiable individuellement grâce à l'association du numéro de parcelle et du numéro de génération.

Dans ce jeu de données, 348 variables sont suivies de manière continue ou intermittente. Ces variables donnent des informations sur les caractéristiques de la parcelle, sur sa production en grappes, sur son état phytosanitaire, sur l'activité agronomique de l'exploitant et sur la composition du sol, des mouts et des feuilles. Les variables ont été regroupées en quatre grands types : caractéristiques de la parcelle fixes dans le temps, variables agronomiques, pratiques du viticulteur sur la parcelle et caractéristiques commerciales, qui ont étés numérotées respectivement de 1 à 4. Toutefois, aucune variable climatique n'est renseignée dans la base de données.

Tableau 2: Liste des fichiers de la base de données fournies par le BNIC.
Légende : Nom : nom du fichier Excel ; Echelle de mesure : unité de mesure minimale pour chaque ligne du fichier ; PLA : numéro de placette ; PAR : numéro de parcelle; GEN : numéro de génération; AN : année; ORGA : organe examiné; EPOQ : période de mesure; Cep : numéro de cep ; Début : première année de mesure ; Fin : dernière année de mesure

Nom	Echelle de mesure	Début	Fin	Informations présentes
GEN	PAR/AN	*1986*	*2018*	Identification des parcelles
QMPAR	PAR/GEN	/	/	Caractéristiques des parcelles
DENS	PAR/GEN	/	/	Taille des parcelles
AS	PAR/GEN/AN	*1990*	*2016*	Analyse de sol
NBG	PLA/PAR/GEN/AN	*1979*	*2019*	Comptage des grappes
QMAN	PAR/GEN/AN	*1988*	*2019*	Informations de récolte
REC1	PAR/GEN/AN	*1977*	*1993*	Analyses du mout
REC2	PAR/GEN/AN	*1994*	*2019*	Analyses du mout
SFE	PAR/GEN/AN	*1995*	*2014*	Surface foliaire exposée
AF1	PAR/GEN/ORGA/EPOQ/AN	*1990*	*1998*	Analyses de feuilles
AF2	PAR/GEN/ORGA/EPOQ/AN	*1999*	*2011*	Analyses de feuilles
ESCAC	Cep/PAR/GEN/AN	*1995*	*2019*	Taux d'infection Esca et de Black Dead Arm
EUTYC	Cep/PAR/GEN/AN	*1993*	*2019*	Taux d'infection Eutypiose
Phyto18	PAR	*2018*	*2018*	Informations phytosanitaires annuelles
Phyto19	PAR	*2019*	*2019*	Informations phytosanitaires annuelles
Tabrec18	AN	*1979*	*2018*	Tableau récapitulatif annuel
CODQM	/	/	/	Lexiquepartielde QMAN et QMPAR

Les variables mesurées sont quantitatives ou qualitatives. On retrouve des variables fixes ou évolutives dans le temps. Chaque variable peut être mesurée de une à plusieurs fois par an. Les variables des fichiers ESCAC et EUTYC sont les seules variables mesurées à l'échelle du cep de vigne. Le fichier NBG mesure les variables à l'échelle de la placette. Dans les autres fichiers les variables sont mesurées à l'échelle de la parcelle. Une placette correspond à l'emplacement de cinq ceps voisins sur le même rang et la parcelle représente le champ.

2. Variables climatiques

Des variables décrivant le climat de la zone de Cognac ont été ajoutées (tableau3). Les données climatiques sont extraites de la base climatique SAFRAN fournie par Météo-France. Des indices bioclimatiques propres à la vigne sont également calculés (Desoutter et al., 2020; Grimaldi, 2018). Ces variables permettront ainsi de relier les données agronomiques et les données météorologiques.

Les données climatiques SAFRAN sont basées sur une technique d'interpolation optimale grâce à un découpage en zones climatologiques homogènes de la carte de France. L'interpolation optimale consiste à déterminer la meilleure estimation linéaire au sens des moindres carrées pour les différents emplacements sans mesures. Les données sont analysées par massif et par zone d'altitude, puis une interpolation spatiale est faite pour projeter les données sur une grille régulière de 8km (CNRM, 2014; Gandin, 1966; Quintana-Seguí et al., 2008). Chaque parcelle du réseau est associée à la zone de la grille la plus proche.

Tableau 3: Variables climatiques ajoutées à la base de données, calculées pour chaque parcelle.

Variables ajoutées	Définition de la variable ajoutée
Jchauds	nombre de jours dans l'année où la température a dépassé 30°C
JTchauds	nombre de jours dans l'année où la température a dépassé 35°C
DJsup25	cumul du surplus de températures moyennes journalières supérieures à 25°C
Jgel	nombre de jours où la température est inférieure à 0°C pour la période allant de septembre de l'année n-1 à août de l'année n
Pluie_annee	cumul de précipitations sur l'année calendaire
Pluie_saison	cumul de précipitations sur la saison (du 10/04 au 30/09)
Froid_cum	cumul de la différence de température entre la température moyenne et 5°C pour les jours ayant une température inférieure à 5°C
Huglin	indice héliothermique de Huglin

L'indice héliothermique de Huglin est utilisé en viticulture pour classer les climats en catégories allant du plus froid au plus chaud (annexe 1). Il s'agit de la moyenne des températures maximales et moyennes journalières de l'air supérieures à 10°C, cumulées sur la période entre le 1er avril et le 30 septembre, et affectées d'un coefficient k variant selon la latitude. On prend un seuil de 10°C car il s'agit de la température minimale à partir de laquelle une plante se développe. La période de l'année choisie correspond aux dates du cycle reproducteur de la vigne (Desoutter et al., 2020; Grimaldi, 2018).

$$\text{Indice de Huglin} = IH = \sum ([(Tm-10)+(Tx-10)]*k)/2 \qquad \text{(équation 1)}$$

avec Tm = Température moyenne, Tx =Température maximale et k le coefficient de longueur du jour variant de 1,02 à 1,06 entre40 et 50 degrés de latitude.

3. Création d'un lexique de variables

Pour une bonne appropriation du jeu de données et pour faciliter la communication sur mes travaux, j'ai commencé par identifier toutes les variables ainsi que les modalités présentes pour toutes les variables. Ainsi, j'ai créé un lexique d'identification des variables et des modalités. Dans ce fichier, j'ai également spécifié le type de donnée (quantitatif ou qualitatif), le type d'information, la dynamique de la variable (fixe ou évolutive dans le temps) et le nom du fichier source (annexe 2). Cette classification a permis notamment d'accéder plus facilement à l'ensemble des variables et de leurs caractéristiques, par exemple pour échanger avec les encadrants de ce stage.

4. Regroupement de variables et tri initial de la base de données

Avant de commencer à analyser les données présentes dans la base, un travail de "nettoyage" et d'homogénéisation était nécessaire. J'ai réalisé l'importation des tableaux Excel dans le logiciel R, version 3.5.3 (R Core Team, 2019). J'ai relevé les différents problèmes présents dans les tableaux et je les ai corrigés. J'ai ainsi retiré les lignes et colonnes entièrement vides. J'ai homogénéisé la notation des valeurs manquantes, des unités de mesures dans une même colonne et des modalités des différentes variables. J'ai modifié le nom des variables lorsque deux variables mesuraient des choses similaires ou différentes. J'ai fusionné les variables mesurant les mêmes informations sur des périodes différentes avec le même protocole de mesure. Ainsi mise en forme, la base de données contient un total de 256 variables différentes.

5. Méthodes descriptives et sélection des variables

Les données fournies doivent être organisées, regardées et représentées graphiquement pour faire une sélection permettant de rassembler l'information utile sous une forme plus aisée à manipuler et à communiquer. Pour les différentes variables quantitatives du jeu de données, j'ai réalisé des statistiques descriptives générales telles que la valeur minimale, le premier quartile, la médiane, la moyenne, le troisième quartile, la valeur maximale, le taux de valeurs manquantes et la variance. Pour représenter la répartition en effectif des parcelles dans les différentes modalités des variables qualitatives, j'ai opté pour des représentations en diagrammes circulaires. Suite à cette étape, des variables ont été retirées pour cause de très faible variabilité inter-parcelle ou d'un taux de valeurs manquantes trop élevé.

Pour une sélection plus importante des variables explicatives, des discussions ont été réalisées avec mes encadrants et des personnes travaillant dans le domaine de la viticulture. Ainsi de nouvelles variables explicatives ont été créées et dans chaque catégorie une sélection plus sévère a été réalisée (annexe 3).

6. Méthodes statistiques

Dans un premier temps, nous avons cherché à comprendre les dynamiques des principales variables caractérisant le rendement et la mortalité des ceps de vigne. Par la suite, nous nous concentrerons sur certaines années pour faire ressortir les liens entre variables et vérifier s'ils diffèrent selon les années. Afin de pouvoir sélectionner des années particulières pour la suite de l'analyse, nous allons nous concentrer sur les évolutions au cours du temps des variables caractéristiques du dépérissement (rendement et mortalité) grâce à la régression linéaire. Puis on regardera plus en détail les années sélectionnées avec des arbres de décision. Toutes les analyses sont réalisées à l'aide du logiciel R.

A. Modèle de régression linéaire

La régression linéaire simple est une méthode statistique employée pour évaluer la significativité du lien entre deux variables quantitatives. On note Y la variable à expliquer et X la variable explicative. La variable Y s'écrit comme une combinaison linéaire de la variable X. Le principe de la régression linéaire simple est de trouver une équation de droite qui passe au plus près des points créés par l'association entre Y et X, en pivotant la droite selon un point fixe qui est le point associant la moyenne de X et la moyenne de Y. Pour calculer les paramètres de cette droite, on utilise la méthode des moindres carrés, c'est-à-dire qu'on calcule la distance entre chaque point et chaque droite éventuelle, qu'on élève au carré. Puis on sélectionne la droite qui passe au plus près de l'ensemble des points et qui minimise la somme de ces distances au carré. Les distances entre les points et la droite ainsi estimée constituent les résidus du modèle de régression.

Le modèle de régression s'écrit sous la forme

$$Y = a + bX \qquad \text{(équation2)}$$

avec a et b les paramètres définissant la droite de régression. On peut également obtenir a et b grâce aux formules dérivées de la méthode du maximum de vraisemblance :

$$a = \overline{y} - b\overline{x} \qquad \text{(équation3)}$$

où $\overline{y}$ est la moyenne des y et $\overline{x}$ est la moyenne des x et

$$b = \frac{\sum_{i=1}^{n}(y(i)-\overline{y})(x(i)-\overline{x})}{\sum_{i=1}^{n}(x(i)-\overline{x})^2} \qquad \text{(équation4)}$$

où n est le nombre de points mesurés.

La pente de la droite de régression caractérise la variation de Y selon X. Le signe de la pente nous indique si la relation est croissante ou décroissante. La valeur de la pente renseigne sur la vitesse d'évolution de Y selon X. La significativité de la relation linéaire dépend de l'incertitude de la pente qui est mesurée par son erreur standard. Cette erreur mesure la dispersion moyenne des observations autour de la droite. Ainsi, plus les points sont éloignés

de la droite de régression plus l'incertitude de la pente est grande. La statistique T permet d'évaluer la significativité de la relation linéaire en rapportant la valeur absolue de la pente à son erreur standard. La statistique T suit une loi de Student à n-2 degrés de liberté (Montgomery and Peck, 1982), où n est le nombre de points observés. Le package employé pour l'utilisation de la méthode est « car »(Fox and Weisberg, 2019).

Evaluation des hypothèses de validité des résultats

Dans un premier temps, il faut vérifier que la relation entre les variables de l'analyse est globalement linéaire. Cette vérification peut s'effectuer graphiquement avec un diagramme de dispersion. Ensuite, pour valider les conclusions du modèle de régression linéaire, il est impératif de vérifier que les résidus :

- sont indépendants : mise en évidence par un « lag plot » et par un test de Durbin-Watson. Comme notre variable X est une variable temporelle, il est nécessaire d'ajouter une structure de corrélation des résidus afin de faire apparaitre cette dépendance pour pouvoir réaliser le modèle linéaire. On ajoute donc une matrice de variance covariance de type autorégressive 1 (AR-1).Le package employé est « nlme » (Pinheiro et al., 2018).

- sont distribués selon une loi normale de moyenne nulle : évaluation graphique avec un QQplot et réalisation d'un test de Shapiro-Wilk ;

- ont une variance constante : on utilise le test de Breush-Pagan et de manière visuelle on peut utiliser un graphe montrant la racine carrée des résidus standardisés, le « residuals vs fitted plot » avec comme fitted les réponses prédites par le modèle pour les valeurs observées de la variable prédictive.

B. Arbre de décision

Les arbres de décision sont une des méthodes les plus intuitives du data mining. On construit un arbre grâce à des données dont on connait la variable à expliquer. Les arbres sont utilisés pour déceler des critères permettant de séparer les individus d'une population en classes. L'arbre part d'une racine principale contenant tous les individus de la population, et à chaque nœud, il va séparer les individus de la population en deux sous-ensembles. La séparation en deux sous-ensembles à chaque nœud se fait à l'aide d'une variable. Chaque variable est testée pour utiliser la variable donnant des branches avec le plus d'individus du même groupe possible. L'objectif est d'avoir à la fin des branches des individus d'un même groupe. Ainsi on peut déterminer quel chemin d'action engendre une valeur donnée de rendement ou de mortalité. Comme nos variables à expliquer sont quantitatives, nous réaliserons des arbres de décision de type régression (Tufféry, 2012).

a. CART

La méthode CART permet de créer des arbres de décision. Cette méthode utilise le critère de Gini (Weymark, 1981),qui mesure l'inégalité de distribution, pour séparer à chaque nœud les individus de la population en deux sous-ensembles. Cette méthode permet de faire de la sélection de variables en plus de la régression. L'algorithme de construction d'un arbre

de régression CART se fait en trois étapes. Il s'agit de la construction de l'arbre de régression maximal, de l'élagage de l'arbre et de la sélection finale. Le package employé pour l'utilisation de la méthode est « rpart » (Therneau and Atkinson, 2018).

Lorsque la méthode CART a été utilisée, le pourcentage d'explication des variables à expliquer était faible, avant même l'élagage de l'arbre. Cela peut s'expliquer par le nombre important de variables explicatives. De ce fait, cette technique n'a pas été gardée.

b. Random forest

La méthode de Random forest est une généralisation des arbres présentés avec la méthode CART. Cette méthode peut être employée puisque nous avons un grand nombre de variables explicatives. La particularité de cette méthode est que chaque modèle est créé à partir d'un sous-ensemble de la population et de variables explicatives sélectionnées aléatoirement. Cette double randomisation a pour effet de diminuer la corrélation des modèles et donc la variance du modèle associé. On ne pourra toutefois pas implanter de règles de décision car il y a un effet boite noire. Le package R employé pour l'utilisation de la méthode est « randomForest » (Liaw and Wiener, 2002).

Pour chacune des années sélectionnées, cinq random forest ont été réalisés, trois pour les trois méthodes d'évaluation du rendement et deux pour les deux variables qui expriment la mortalité (tableau 6 page 15). Comme nous l'avons vu précédemment, certaines variables mesurées ont une influence sur les ceps de vigne de l'année suivante, de ce fait les variables analysées dans les random forest comprennent les mêmes variables mesurées l'année n et l'année n-1. Pour les arbres de régression du rendement, les deux autres variables indicatrices du rendement n'ont pas été incluses dans l'analyse. De même, pour les arbres de régression de la mortalité les indicateurs de la mortalité ont été retirés.

Pour cette méthode, il faut définir le nombre de variables sélectionnées par nœuds (mtry) et le nombre d'arbres (ntree). Pour avoir un modèle aussi fiable que possible, j'ai fait varier les deux paramètres optimisants de la random forest afin d'avoir un « out-of-bag » et un pourcentage de variance expliqué déterminants. L'out-of-bag est un indicateur qui donne une information de performance calculée à partir des échantillons non utilisés lors de la construction des arbres. Plus ce critère est petit plus l'arbre est performant. Le pourcentage de la variance expliqué est une mesure de la façon dont les variables expliquent la variabilité de la variable à expliquer. Ainsi cette métrique sert également à évaluer la performance d'un modèle (Jeong et al., 2020).

Pour une lecture plus aisée de l'implication des variables explicatives, j'ai réalisé un arbre par analyse de random forest grâce au package « reprtree » (Dasgupta, 2014). Cette représentation correspond à l'arbre représentant l'arbre le plus proche de la moyenne des arbres de la random forest (Banerjee et al., 2012). Chaque arbre est comme un point dans un espace de grande dimension et les arbres sont regroupés selon une certaine mesure de proximité. La métrique d2 employée capture la similitude de prédiction. Deux arbres sont similaires si les prédictions des deux arbres sont les mêmes pour tous les sujets.

7. Sélection d'années particulières

Pour déterminer les variables explicatives dominantes et voir si les facteurs des variables à expliquer évoluent à travers le temps, nous avons choisi de réaliser une étude plus poussée sur des années présentant des rendements et/ou mortalités particulières. On comparera ensuite les variables explicatives dominantes des différentes années sélectionnées. Pour une analyse du dépérissement dans ce réseau de parcelles viticoles, nous avons décidé de sélectionner des années présentant :

- soit une forte hausse ou diminution du rendement sachant que cette évolution ne doit pas être une simple compensation d'une forte évolution de rendement l'année antérieure ;

- soit une forte hausse ou diminution de la mortalité ;

- soit une association entre les deux points ci-dessus.

III/ Résultats

1. Création et ajout de variables

La base de données contient de nombreuses variables explicatives et à expliquer. Toutefois, certaines variables indispensables à l'étude ne sont pas présentes, ou sont présentes sous une forme qui les rend difficiles à exploiter. Il a donc fallu créer à partir des informations à notre disposition des variables plus adaptées à notre étude.

Les variables ajoutées permettant d'apporter des informations pour identifier les parcelles sont CODE et AGE. La variable CODE correspond à un code d'identification individuel des parcelles. Il s'agit de la concaténation du numéro de parcelle et du numéro de génération de la parcelle avec l'ajout d'un « P » devant le numéro de parcelle et d'un « G » devant le numéro de génération. Par exemple, la parcelle 8 génération 2 a pour CODE « P8G2 ». Cette variable constitue donc un identifiant unique des parcelles. La variable AGE correspond à l'âge de la parcelle au moment de la mesure. Il s'agit de la soustraction entre l'année de mesure et l'année de plantation des ceps dans la parcelle.

Pour rendre l'utilisation des informations des fichiers ESCAC et EUTYC possible, un changement de forme a été nécessaire. Initialement, ces fichiers avaient en colonne l'année de mesure et en ligne l'emplacement du cep mesuré (tableau 4).

Tableau 4 : Forme initiale du fichier ESCAC (extrait)

PAR	GEN	FIN	COTE	HB	CEP	ESC05	ESC06	ESC07	ESC08	ESC09	ESC10	ESC11
53	1		1	2	4	ES0	ES0	ES0	0	0	0	0
53	1		1	2	5	ES0	ES0	ES0	0	0	0	0
53	1		2	1	1	ES0	EL2	ES0	BEF2	J	J	J
53	1		2	1	2	EL1	EF1	ES0	0	BEF2	M	M
53	1		2	1	3	ES0	ES0	BEF1	BEF1	BEL2	BEF2	0

Le contenu de chaque cellule était un code qui indiquait l'état de l'emplacement en question. Cet état prenait les modalités : vivant, mort, sain ou malade. Ce code donnait également des informations sur le desséchement du pied de vigne, sur la forme et la gravité de la maladie du cep de vigne.

Variables de mortalité (données présentes pour les années postérieures à 1994)

Lors de la remise en forme, j'ai dans un premier temps associé une ligne à un emplacement de cep dans une parcelle pour une année donnée. Puis j'ai créé les variables de nombre de pieds malades de l'esca et de l'eutypiose selon le fichier. Pour ce faire, j'ai utilisé une variable booléenne intermédiaire qui indiquait la présence ou l'absence de maladie sur le cep, puis j'ai fait une somme des ceps malades selon la parcelle et l'année de mesure. De la même manière, j'ai obtenu le nombre de ceps morts annuel par parcelle, à partir du fichier ESCAC, créant le fichier ESCAC2 (tableau 5).

Tableau 5: Forme du fichier ESCAC2 (extrait)

CODE	AN	Nb_MALADE_ESCA	Nb_MORT
P11G0	2018	1	0
P11G0	2019	1	0
P12G0	1995	4	0
P12G0	1996	2	0
P12G0	1997	2	0

D'autres variables indicatrices de la mortalité ont été créées. Il s'agit des variables taux de mortalité, AGE_1er_MORT, NEW_MORT et le taux_mort_annuel. La variable taux de mortalité est le pourcentage de ceps morts dans une parcelle lors de l'année n, sans préjuger du moment où ils sont morts. Cette variable est calculée en divisant le nombre de ceps notés morts l'année n par le nombre maximum de pieds de vigne sur la parcelle. La variable AGE_1er_MORT correspond à l'âge des parcelles lorsque le premier cep mort est observé. Pour ce faire, je regarde, pour chaque parcelle entrée après 1995, la valeur minimale de l'année pour un nombre de morts différent de zéro, puis je calcule l'âge de la parcelle. Pour les parcelles entrées avant 1995, je regarde les parcelles n'ayant pas eu de mort de 1995 à 2015, puis je calcule l'âge de la parcelle lorsqu'elle présente un mort. La variable NEW_MORT indique le nombre de ceps morts de l'année par parcelle, c'est-à-dire la mortalité survenue entre les observations de l'année n-1 et de l'année n. Pour sa création j'ai créé une variable booléenne indiquant si chaque cep était mort l'année n mais pas l'année n-1, puis j'ai sommé selon les parcelles pour chaque année. Enfin la variable taux_mortalité_annuel est le pourcentage de ceps morts au cours de l'année dans une parcelle. Cette variable est calculée en divisant le nombre de ceps morts de l'année par le nombre de cep maximum sur la parcelle. Une variante de cette dernière variable a été créée pour différencier le taux de mortalité d'une année et le taux de mortalité moyen annuel sur une plage de temps. Cette nouvelle variable est « taux_mort_annuel_global » qui correspond au taux de ceps morts par an entre le début des suivis de mortalité et l'année considérée.

Variables de rendement

Pour analyser le rendement des parcelles du réseau, il a fallu créer les variables Nb_etablie, Tx_pas_etablie, NBG_1P, NBG_1C, Rdt_placette et Rdt_cep. Les variables Nb_etablie et Tx_pas_etablie ont été créées à partir du fichier ESCAC. Nb_etablie correspond au nombre de pieds de vigne productifs (produisant des grappes de raisin) pour une année donnée. Pour ce faire, j'ai compté par parcelle et année les ceps qui n'étaient pas absents, morts, jeunes, recépés, en provignage ni entièrement secs. La variable Tx_pas_etablie indique le taux de cep non productifs de la parcelle pour l'année mesurée. Les ceps non productifs correspondent à des ceps absent, morts, jeunes, secs, recépé ou en provignage. Cette variable est calculée annuellement pour chaque parcelle en divisant le nombre de pieds non productifs par le nombre maximum de pieds sur la parcelle.

Les variables NBG_1P et NBG_1C ont été créées à partir des informations du fichier NBG, on y retrouve le comptage du nombre de grappes sur une à quatre placettes par parcelle et année. Une placette correspond à l'emplacement de cinq ceps voisins sur le même rang. NBG_1P est le nombre de grappes moyen sur une placette pour une parcelle et une année données. Pour la création de NBG_1P, j'ai réalisé une moyenne du nombre de grappes, selon le nombre de placettes, l'année de mesure et la parcelle. La variable NBG_1C correspond au nombre de grappes moyen présent par cep productif dans la parcelle pour une année donnée. Pour la création de cette variable, j'ai dû regarder le nombre de ceps vivants et productifs par parcelle, le diviser par le nombre de placettes dans la parcelle pour avoir le nombre moyen de ceps de vigne vivants par placette. Puis, j'ai divisé NBG_1P par le nombre moyen de ceps vivants et productifs par placette. La variable Rdt_placette correspond au rendement moyen produit dans une placette d'une parcelle lors d'une année donnée. Il s'agit de la multiplication du poids de grappe moyen par parcelle par le nombre de grappes présentes dans la placette. La variable Rdt_cep représente le rendement moyen par cep productif dans une parcelle une année donnée. Cette fois, il s'agit de la multiplication du poids de grappe moyen par parcelle parle nombre de grappes présentes sur un cep productif.

2. La base de données mise en forme

Après la mise en forme et la sélection des variables, la base de données finale contient 95 colonnes pour 1869 lignes. Ainsi on retrouve 95 variables différentes mesurées sur 55 parcelles (54 parcelles en 1986) par an de 1986 à 2019. Comme les parcelles peuvent sortir du réseau et être remplacées pour de nombreuses raisons, on retrouve au total 107 parcelles différentes.

Nous disposons ainsi de variables explicatives, quantitatives et qualitatives, dont on souhaite déterminer si elles permettent de proposer des hypothèses sur le dépérissement viticole au sein du réseau. Les variables à expliquer représentant le rendement et la mortalité des ceps de vigne sont au nombre de cinq (tableau 6).

Tableau 6: Variables à expliquer pour les différentes thématiques du dépérissement

Rendement	**Mortalité**
- VP : rendement de la parcelle déclaré par l'exploitant (en hL/ha) - Rdt_placette : rendement moyen par placette calculé (en g/placette) - Rdt_cep : rendement moyen par cep productif calculé (en g/cep)	- taux_mortalité_annuel : taux de ceps morts dans l'année - taux_mort_annuel_global : taux de ceps morts par an entre le début des suivis de mortalité et l'année considérée

Période de mesure des variables à expliquer

La variable de rendement fournie par les exploitants, VP, a été mesurée de 1991 à 2019. Les variables de rendements calculées nécessitent des informations issues de variables mesurées pour la période de 1988 à 2019 pour le rendement placette, et de 1995 à 2019 pour le rendement moyen des ceps productifs. Pour la mortalité des ceps de vigne, nous possédons des informations pour la période allant de 1995 à 2019. Toutes les années antérieures ne pourront donc pas être analysées en ce qui concerne ces thématiques. Lors des analyses suivantes nous nous limiterons donc à une période allant de 1995 à 2019.

Fiabilité des variables de rendement calculées

Deux des trois variables à expliquer pour le rendement sont calculées. Pour vérifier la cohérence des différents modes de prise en compte du rendement à notre disposition, nous les avons comparés entres elles selon l'année (figure 1). Comme elles ne possèdent pas d'unités de mesure similaires, nous avons homogénéisé les valeurs pour les ramener à une même dimension. Ainsi, la moyenne des trois différents rendements a été uniformisée.

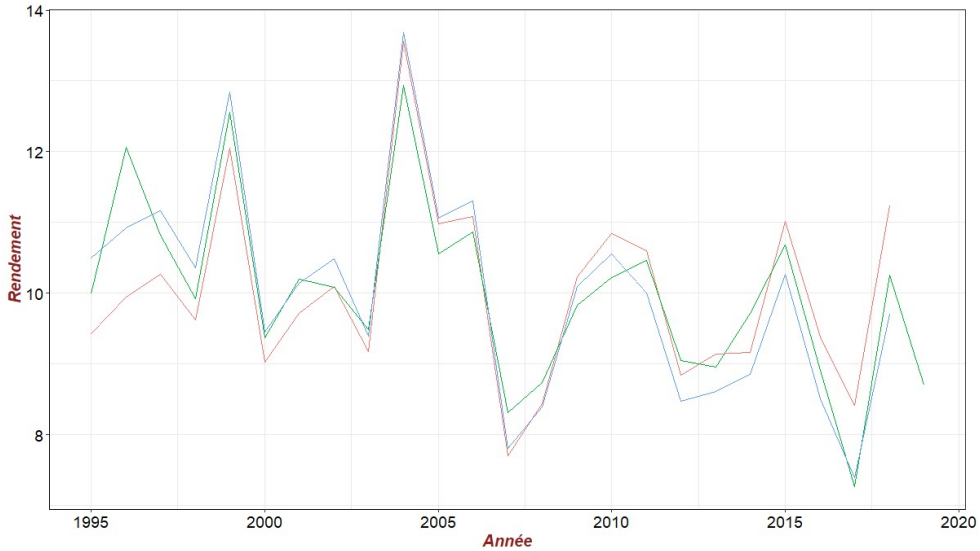


Figure 1 : Cohérence entre les rendements calculés à partir de mesures sur les placettes et déclarés à la parcelle par les viticulteurs.
Légende des courbes: rouge : rendement des ceps productifs ; vert : rendement déclaré par les exploitants ; bleu : rendement calculé à la placette

Le rendement déclaré et le rendement à la placette devraient donner une information similaire, puisqu'ils représentent le rendement global des ceps qu'ils soient productifs ou non, tandis que le rendement par cep se concentre uniquement sur les rendements des ceps productifs, et peut donc être légèrement supérieur aux deux précédents. La figure 1montre que les différents rendements ont une même dynamique selon les années. Le rendement des ceps productifs semble présenter des valeurs de rendement supérieures aux deux autres indicateurs de rendement pour une large période, comme par exemple toutes les années après 2015. Toutefois, le rendement des ceps productifs est inférieur au rendement déclaré et au rendement placette pour les années comprises entre 1995 et 2003. Cette discordance pourrait s'expliquer par une modification du protocole d'identification des ceps productifs ou encore par l'échantillonnage : les grappes sont récoltées sur toute la parcelle et des grappes plus petites et plus légères que la moyenne ont pu avoir été prélevées.

3. L'âge des parcelles du réseau

Au cours du temps, des parcelles sortent du réseau et sont remplacées. L'entrée d'une nouvelle parcelle peut s'effectuer selon des critères tels que l'emplacement de la parcelle, la disponibilité de l'exploitant ou encore l'âge de la parcelle. La figure 2 présente l'âge des parcelles au moment de leur entrée dans le réseau.

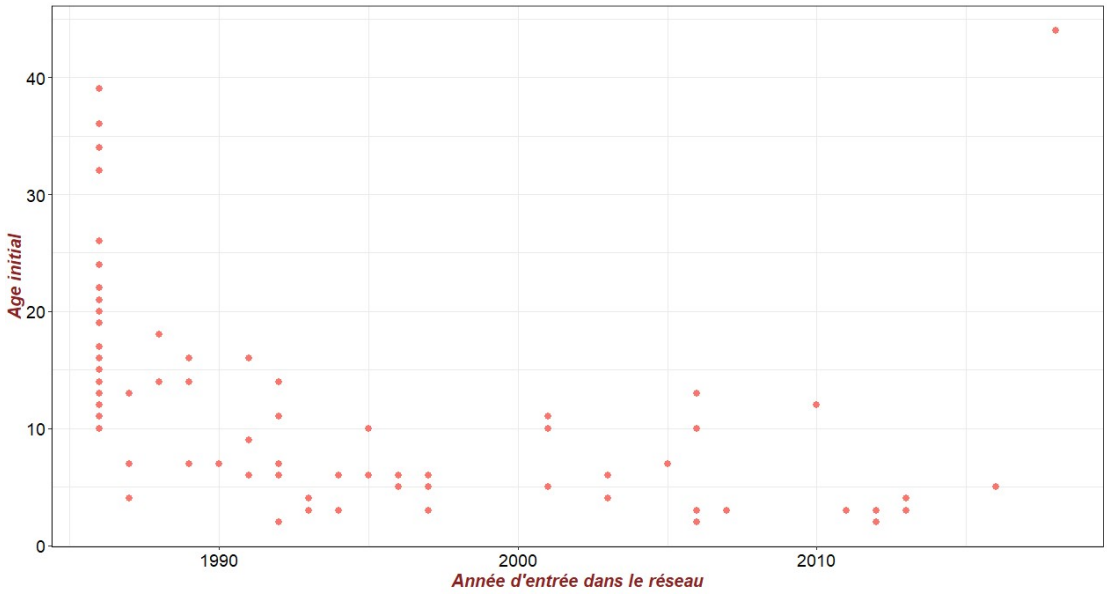


Figure 2 : Age des parcelles lors de leur entrée dans le réseau de Cognac
Légende : Chaque point correspond à une parcelle

Lors de la création du réseau, la plage d'âge des parcelles entrantes était beaucoup plus grande que pour les années suivantes. En effet, des parcelles d'âges variables ont été sélectionnées pour créer le réseau. Initialement, les parcelles avaient un âge d'entrée compris entre 10 et 40 ans ; ensuite, pour remplacer les parcelles quittant le réseau, les parcelles entrantes ont été choisies jeunes, plantées depuis moins de 20 ans. On retrouve tout de même une parcelle entrante à 44ans en 2018. De manière globale, la moyenne d'âge des parcelles entrantes est de 24 ans mais si on exclut les années 1986 et 2017, la moyenne d'âge d'entrée est de 7 ans.

Figure 3 : Âge moyen des parcelles au cours du temps

La figure 3 présente l'âge moyen des 55 parcelles au cours du temps. On remarque que l'âge moyen des parcelles du réseau n'a cessé d'augmenter, alors que les parcelles entrantes sont jeunes. L'âge moyen des parcelles, toutes années confondues, est de 24 ans.

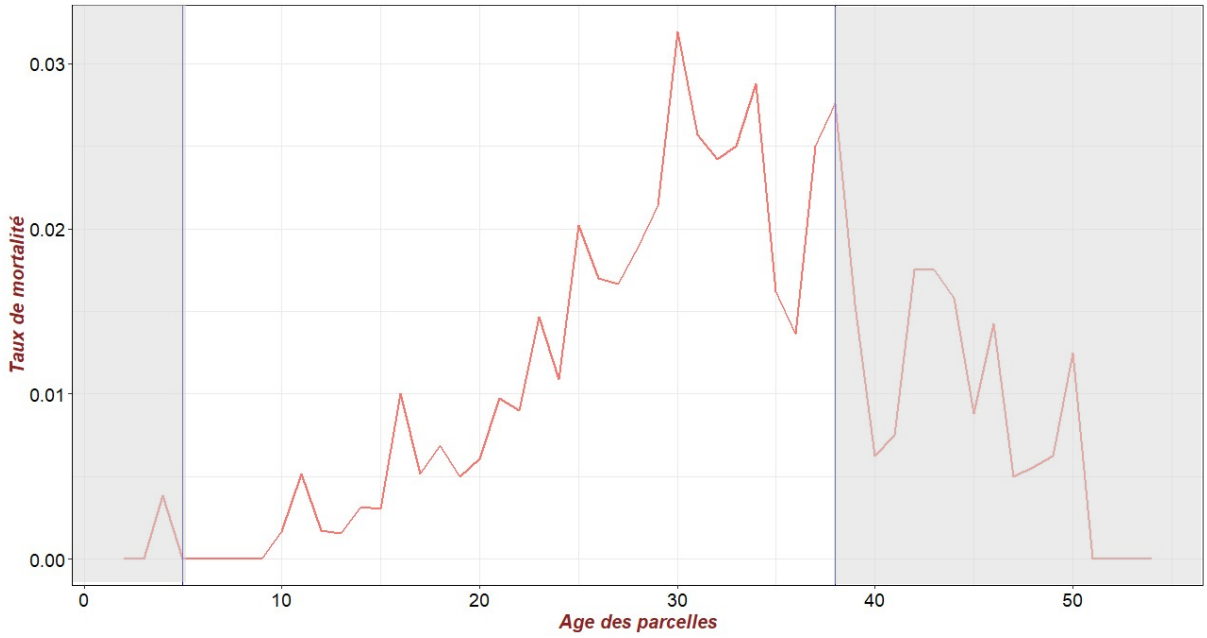


Figure 4 : Taux de mortalité selon l'âge des parcelles
Légende : bleu : séparations des zones d'âge analysées

Le vieillissement des parcelles est naturellement un facteur pouvant expliquer la mortalité des ceps qui y sont présents. On observe donc le taux de mortalité selon l'âge des parcelles (figure4). Ce taux de mortalité est la somme des ceps morts sur les parcelles d'un âge donné, divisé par le nombre total de ceps présents sur ces mêmes parcelles. Parmi la plage d'âge des parcelles présentées dans le graphique, on ne dispose pas de beaucoup de parcelles

de moins de 5 ans et ou plus de 38 ans (zones grisées). On dispose en moyenne de 9 parcelles par âge pour les parcelles de moins de 5 ans et de 12 parcelles par âge pour les parcelles de plus de 38 ans. De ce fait, on n'interprétera pas ces parties du graphique. Pour la plage centrale d'âge, il y a en moyenne 34 parcelles par âge.

De manière globale le taux de mortalité augmente selon l'âge des parcelles. A partir de 20 ans, on remarque une augmentation rapide du taux de mortalité, qui passe de 0,6 ceps morts sur 100 emplacements à son maximum de 3,19 ceps morts sur 100 emplacements à 30 ans. Ainsi le taux de mortalité a plus que quintuplé en l'espace de 10 ans. Entre 30 et 38 ans on observe une légère décroissance du taux de mortalité, puisque l'on passe à 2,33 ceps morts sur 100 emplacements pour les parcelles de 38 ans.

4. Dynamique du dépérissement viticole au cours du temps

A. Les indicateurs du rendement

Le rendement à la parcelle est le rendement présentant la plus grande représentativité réelle pour l'exploitant puisqu'il prend en compte tous les types de ceps présents sur la parcelle (vivants, morts, malades…). L'évolution du nombre de parcelles présentant une hausse ou une diminution du rendement parcellaire permet de se faire une idée sur l'évolution réelle du rendement au cours du temps. La figure 5présente pour chaque année le nombre de parcelles, parmi les 55 parcelles du réseau, dont le rendement est inférieur, supérieur ou égal à l'année précédente.

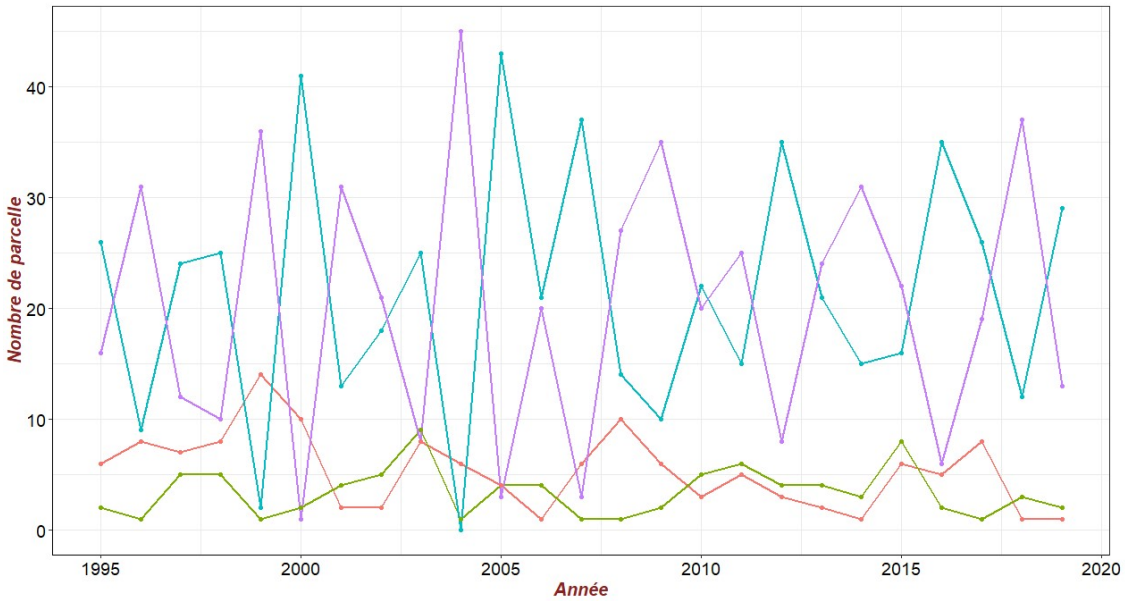


Figure 5 : Evolution du nombre de parcelle selon leurs classe de rendement
Légende des couleurs de courbe : Bleu : classe indiquant une diminution de rendement; Violet : hausse de rendement; Vert : pas d'évolution du rendement; Rouge : pas de valeur

Toutes années confondues, on voit un équilibre entre les augmentations et les diminutions du rendement à la parcelle. On a 504 augmentations et 534 diminutions. Dans 84 cas, l'exploitant a noté un rendement strictement identique à l'année antérieure, ce qui induit un doute sur la fiabilité des chiffres déclarés. De manière globale, on peut noter que les

rendements n'ont pas diminué chaque année et qu'en 24 ans, il y a seulement 30 diminutions de plus que le nombre d'augmentations de rendement.

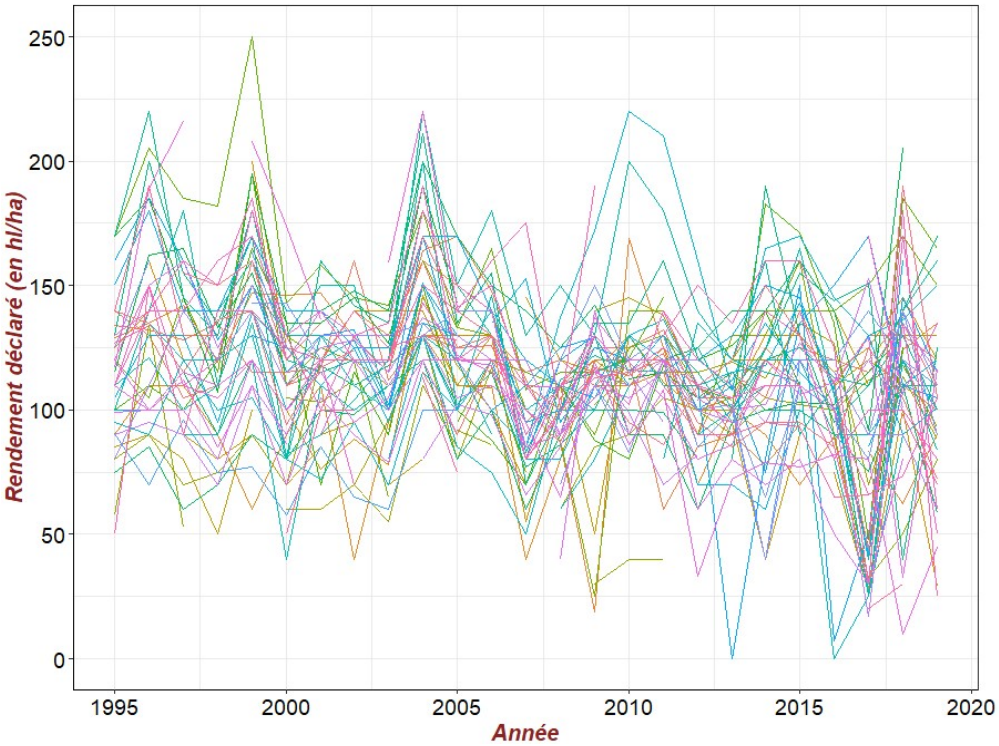


Figure 6 : Rendement déclaré par les exploitants selon l'année
Légende : Chaque courbe représente le rendement déclaré d'une parcelle au cours du temps

Sur la figure6sont représentés les rendements déclarés par l'exploitant des différentes parcelles du réseau selon l'année de production. On constate des similitudes entre parcelles, puisque de nombreuses parcelles ont le même comportement dans le temps, notamment en ce qui concerne les années "exceptionnelles", de fort ou faible rendement.

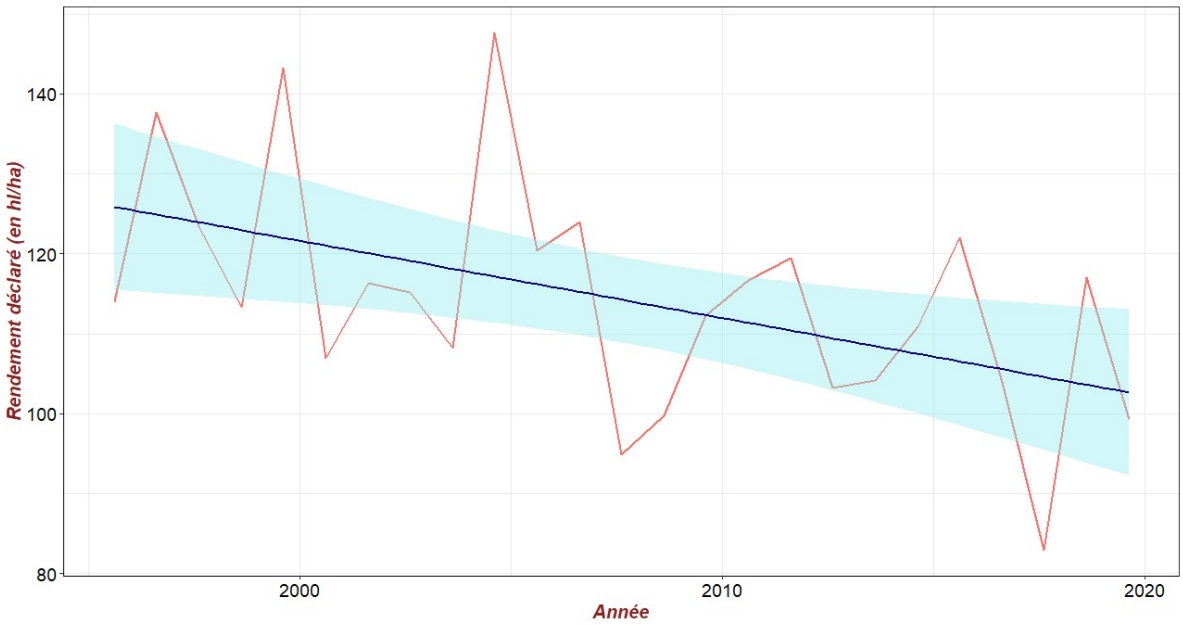
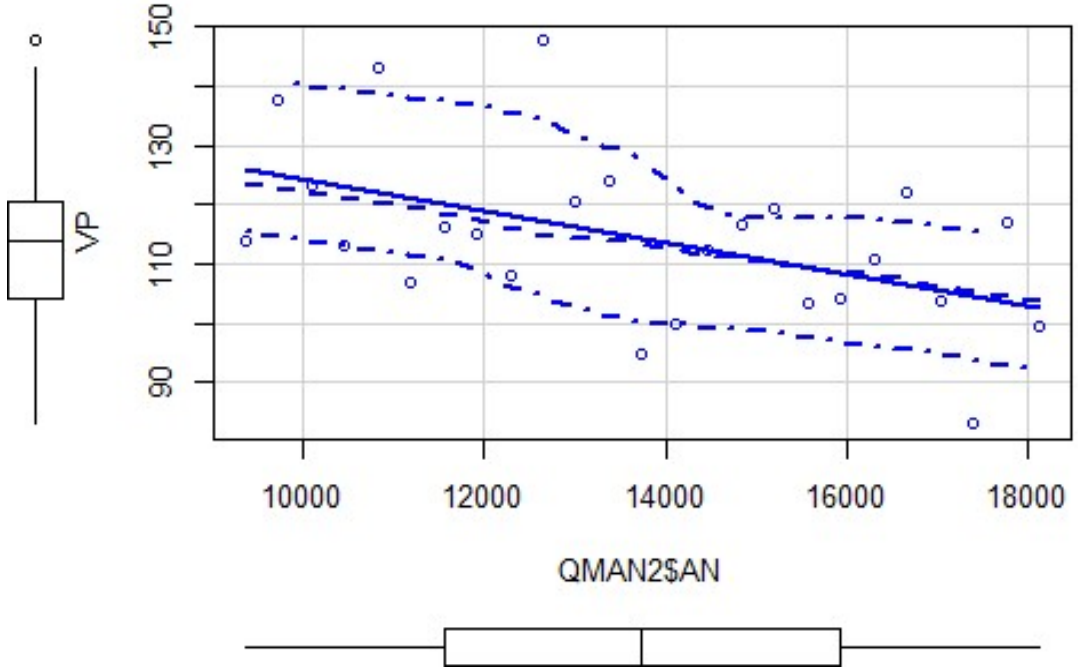


Figure 7 : Rendement moyen déclaré par les exploitants selon l'année
Légende : Rouge : rendement moyen déclaré ; Violet : droite de régression du rendement déclaré ; Bleu : intervalle de confiance de la droite de régression

Pour déterminer si le temps a une influence significative sur le rendement déclaré, on réalise une droite de régression sur le rendement moyen annuel (figure 7). Cette droite de régression indique si le rendement moyen déclaré augmente ou diminue significativement au cours du temps. Au premier abord, elle semble indiquer que ce rendement diminue au cours du temps puisque l'on passe de 126 hL/ha à 103 hL/ha.

Dans un premier temps les tests de validité d'hypothèse ont été réalisés. On réalise une évaluation visuelle de la linéarité car la régression linéaire simple suppose cette forme linéaire pour évaluer si deux variables ont un lien de significativité.

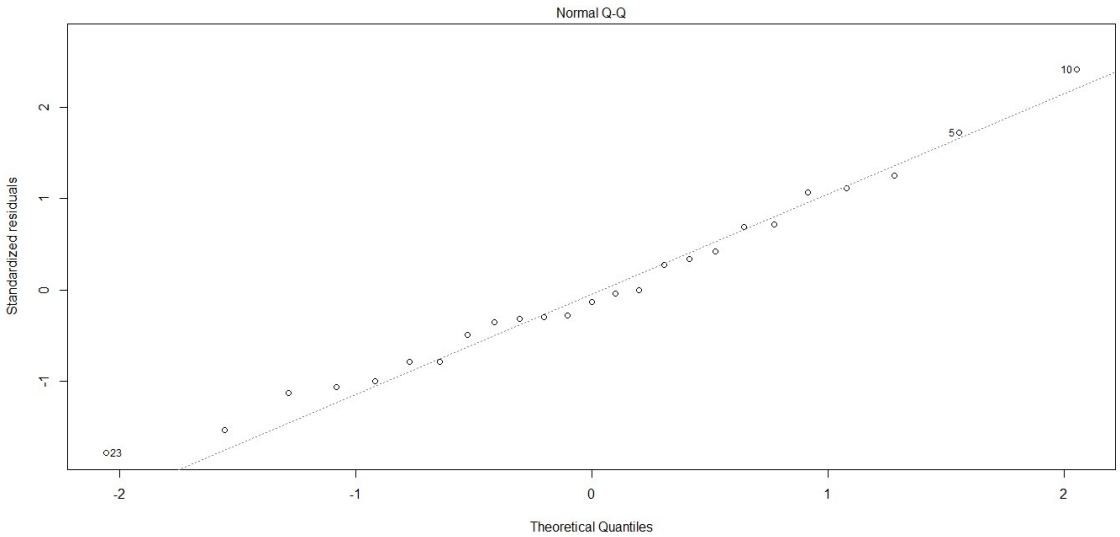


<u>Figure 8 :</u> Evaluation de l'hypothèse de linéarité du modèle linéaire du rendement déclaré moyen au cours du temps
<u>Légende :</u> VP : rendement déclaré moyen; QMAN2$AN : année; trait plein bleu : droite de régression linéaire définie par la méthode des moindres carrées; trait central en pointillés : courbe de régression locale de type lowess; traits extérieurs en pointillés : intervalle de confiance de la courbe lowess

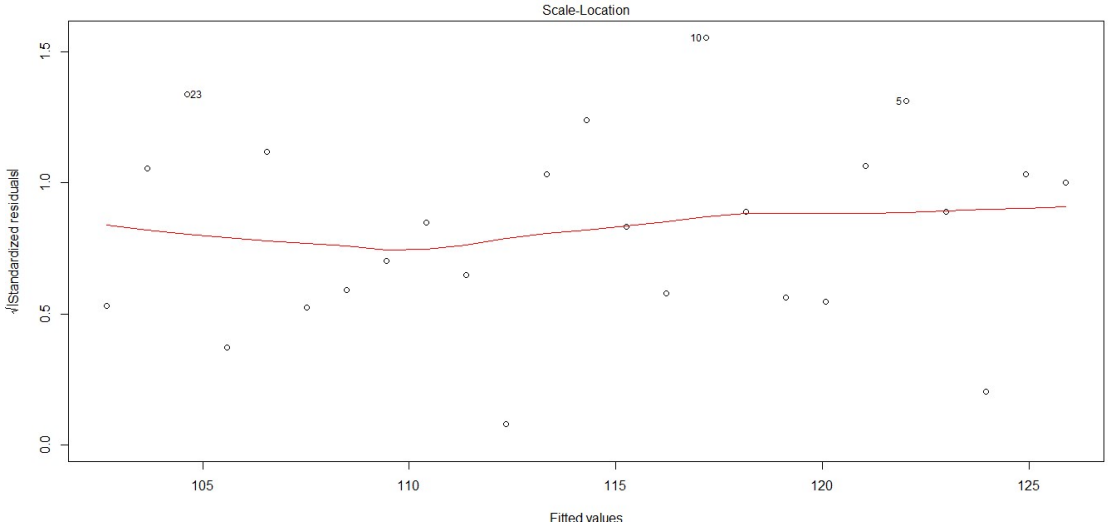
Dans la figure8, la droite de régression est dans l'intervalle de confiance donc l'hypothèse de linéarité est acceptable.

Pour le test d'indépendance des résidus du rendement déclaré moyen on a une dépendance puisque la variable prédictive est temporelle. Ainsi, les résidus sont auto-corrélés. Pour la réalisation du modèle on a donc pris en compte l'auto-corrélation par l'ajout d'une structure de corrélation comme indiqué paragraphe II/6.A.

Pour valider l'hypothèse de normalité des résidus, on regarde si les résidus sont bien distribués le long de la droite (figure 9). Ici, les points sont alignés, on peut valider l'hypothèse de normalité. Un test de Shapiro-Wilka également été fait pour confirmer l'évaluation visuelle. On considère un risque de première espèce alpha de 0,05. On a obtenu une p-value de 0,886, donc on ne rejette pas l'hypothèse de normalité.



<u>Figure 9 :</u> Evaluation de l'hypothèse de normalité des résidus du modèle linéaire du rendement déclaré moyen au cours du temps



<u>Figure 10 :</u> Evaluation de l'hypothèse d'homogénéité des résidus du modèle de rendement déclaré au cours du temps
<u>Légende :</u> Courbe rouge : courbe de régression locale

La figure 10 montre que les résidus ont tendance à être répartis de manière homogène tout le long du gradient des valeurs d'années. Ainsi, on ne peut pas rejeter l'hypothèse d'homogénéité des résidus. Pour confirmer ce résultat, on a réalisé un test de Breush-Pagan. L'hypothèse d'homogénéité est rejetée si la p-value est inférieure à 0,05, et nous avons une valeur de 0,69473 donc l'hypothèse d'homogénéité n'est pas rejetée.

Puisque les hypothèses de linéarité, de normalité, d'homogénéité et d'indépendance des résidus ont été validées, nous pouvons interpréter les résultats de la régression linéaire.

Dans le tableau 7, nous avons la partie coefficients de la sortie du modèle linéaire du rendement moyen déclaré. En première ligne, l'intercept est le paramètre qui représente l'ordonnée à l'origine. En seconde ligne se trouve la pente de la droite de régression.

Tableau 7 : Sortie du modèle linéaire issu de l'analyse du rendement moyen déclaré au cours du temps

	Estimation des paramètres	Estimation de leur erreur standard	Statistique T	p-value du test évaluant l'égalité à0 des coefficients
(Intercept)	2054.0942	719.2111	2.856	0.00894 **
Année	-0.9665	0.3583	-2.697	0.01286 *

Pour déterminer si une pente est significativement égale ou différente de zéro, il faut regarder la dispersion des points autour de la droite de régression donc la p-value. On prend un seuil alpha de 0,05. La p-value est égale à 0,0189, elle est donc inférieure à 0,05, ce qui écarte l'hypothèse de nullité de la pente et permet de conclure à l'existence d'un lien linéaire significatif entre le rendement déclaré moyen et l'année. Le coefficient est de -0,9665, il est donc négatif, la relation linéaire est donc décroissante. L'équation de la régression linéaire simple est: rendement déclaré = - 0,9665 * Année + 2054.Le rendement déclaré moyen baisse donc significativement au cours du temps, il baisse de 0,97 hL/ha par an.

Pour déterminer l'évolution au cours du temps des deux autres types de rendements, la même analyse que précédemment a été réalisée pour le rendement calculé à la placette et pour le rendement des ceps productifs. Les tests d'hypothèse sont présentés en annexe.

En annexe 4figure 1, on observe l'évolution du rendement calculé à la placette pour les différentes parcelles du réseau selon l'année de production. Comme pour le rendement déclaré, on constate des similitudes entre parcelles, puisque de nombreuses parcelles ont le même comportement dans le temps.

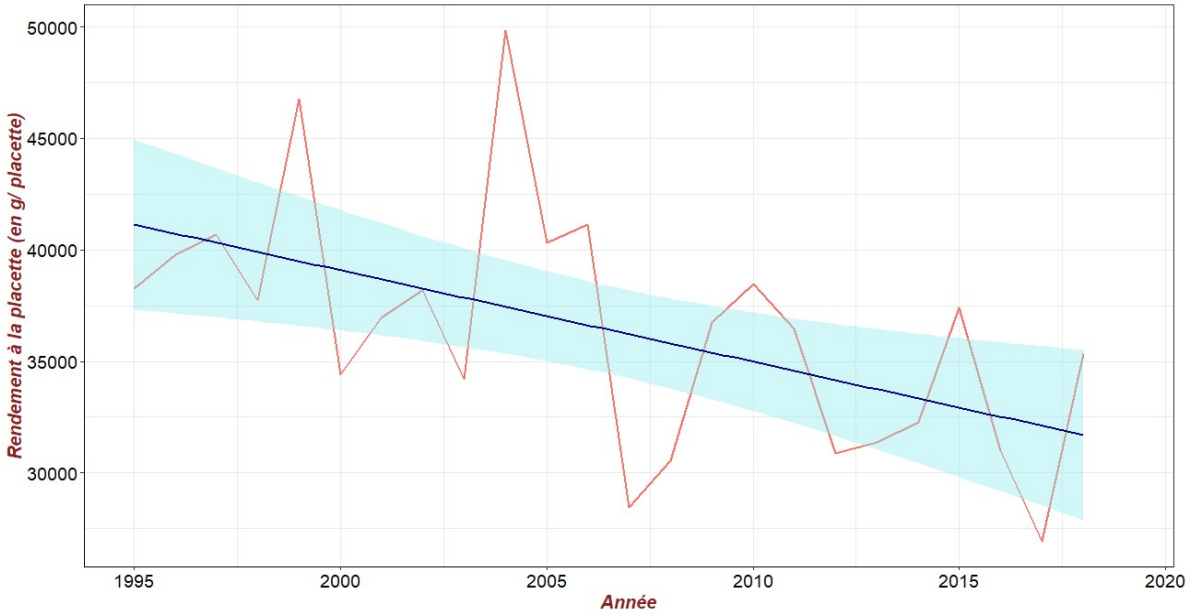


Figure 11 : Rendement moyen à la placette selon l'année
Légende : Rouge : rendement moyen à la placette ; Violet : droite de régression du rendement à la placette ; Bleu : intervalle de confiance de la droite de régression

Nous avons déterminé le rendement moyen calculé à la placette selon les années (figure 11). Puis nous avons réalisé une droite de régression qui permet de visualiser si ce rendement augmente ou diminue significativement au cours du temps. Visuellement, la droite de régression décroit au cours du temps, on passe de 41,25 kg par placette en 1995 à 31,7 kg par placette en 2018, ce qui indique une diminution du rendement moyen calculé à la placette au cours du temps. Comme les tests d'hypothèses sont satisfaisants on peut regarder les résultats issus du modèle de régression (annexe 5).

Tableau 8: Sortie du modèle linéaire issu de l'analyse du rendement calculé moyen à la placette au cours du temps

	Estimation des paramètres	Estimation de leur erreur standard	Statistique T	p-value du test évaluant l'égalité à0 des coefficients
(Intercept)	860193.6	274932.6	3.129	0.00489**
Année	-410.6	137	-2.996	0.00665**

Pour tester si la pente est significativement égale ou différente de zéro, on étudie la dispersion des points autour de la droite de régression grâce à la p-value. On prend un alpha de 0,05. La p-value est de 0,00665, elle est donc inférieure à 0,05, donc on peut dire qu'il y a une relation linéaire significative entre le rendement calculé moyen à la placette et l'année (tableau 8). La droite de régression a comme équation -410,6 * Année + 860193,6. Le rendement calculé moyen à la placette baisse significativement au cours du temps, il baisse de 411 g/placette/an.

De la même manière, on a étudié l'évolution du rendement des ceps productifs des différentes parcelles au cours du temps (annexe 4 figure 2). Après observation visuelle, il semble y avoir des similitudes entres les parcelles pour ce type de rendement également.

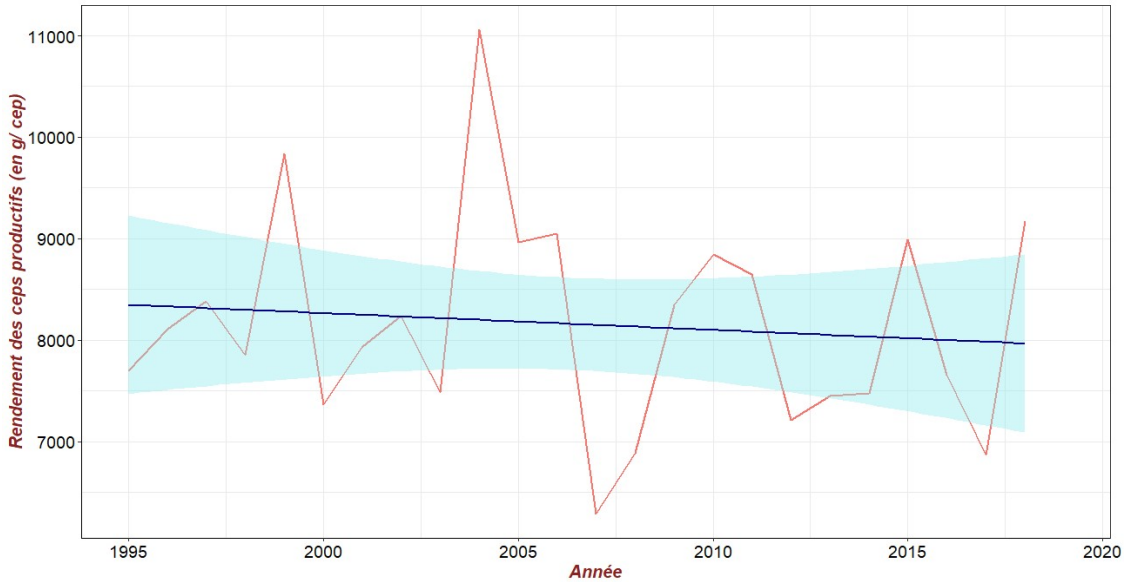


Figure 12: Rendement moyen des ceps productifs selon l'année
Légende : Rouge : rendement moyen des ceps productifs ; Violet : droite de régression du rendement des ceps productifs ; Bleu : intervalle de confiance de la droite de régression

La figure 12 présente le rendement moyen des ceps productifs au cours du temps et la droite de régression ajustée. Visuellement la droite de régression semble décroitre très légèrement. La droite de régression passe de 8,4 kg par cep productif en 1995 à 8,0 kg en 2018. Comme les tests d'hypothèses sont satisfaisants on peut regarder les résultats issus du modèle de régression (annexe 6).

Tableau 9: Sortie du modèle linéaire issu de l'analyse du rendement calculé moyen à la placette au cours du temps

	Estimation des paramètres	Estimation de leur erreur standard	Statistique T	p-value du test évaluant l'égalité à0 des coefficients
(Intercept)	41407.55	63305.67	0.654	0.520
Année	-16.57	31.55	-0.525	0.605

On prend un alpha de 0,05. La p-value est de 0,604, elle est donc supérieure à 0,05 donc on ne peut pas dire qu'il y a une relation linéaire significative entre le rendement moyen des ceps productifs et l'année (tableau 9). La pente de la droite de régression n'est pas significativement différente de zéro.

B. Les indicateurs de la mortalité

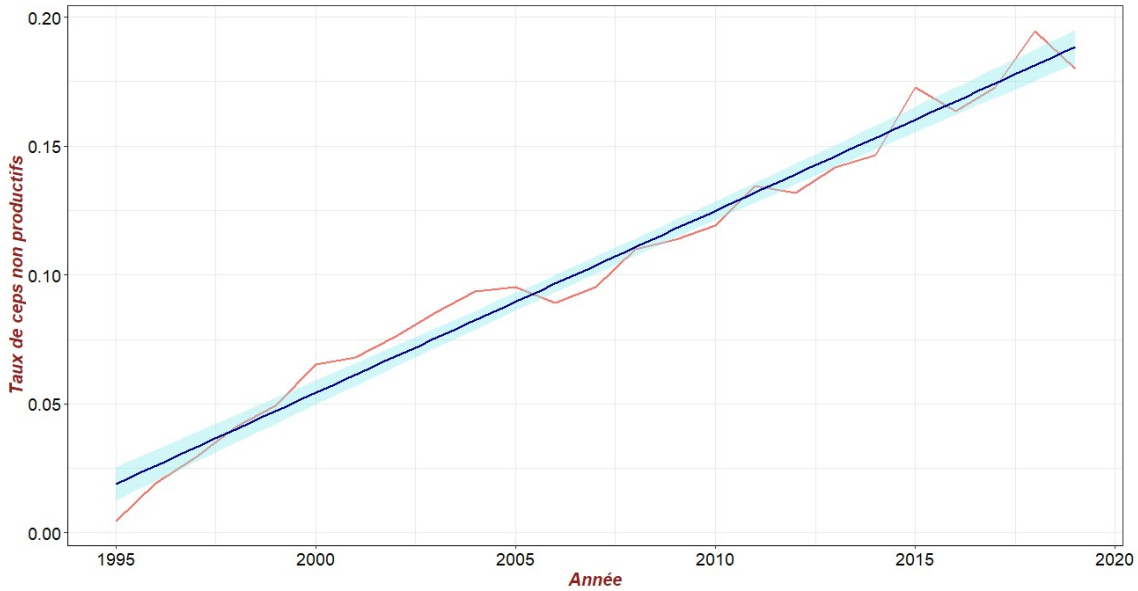


Figure 13 : Taux de ceps non productifs au cours du temps
Légende : Rouge : taux de ceps non productifs ; Violet : droite de régression du taux de ceps non productifs ; Bleu : intervalle de confiance de la droite de régression

Le rendement à la parcelle prend en compte les pieds de vignes productifs et non productifs. De ce fait, on a regardé l'évolution du taux des ceps non productifs du réseau au cours du temps dans la figure 13. On remarque une augmentation du nombre de ceps non

productifs donc une diminution du nombre de ceps de vigne productifs au cours du temps. Ainsi, on passe de 0,46 ceps non productifs sur 100 emplacements en 1995 à 18 ceps non productifs sur 100 emplacements en 2019.

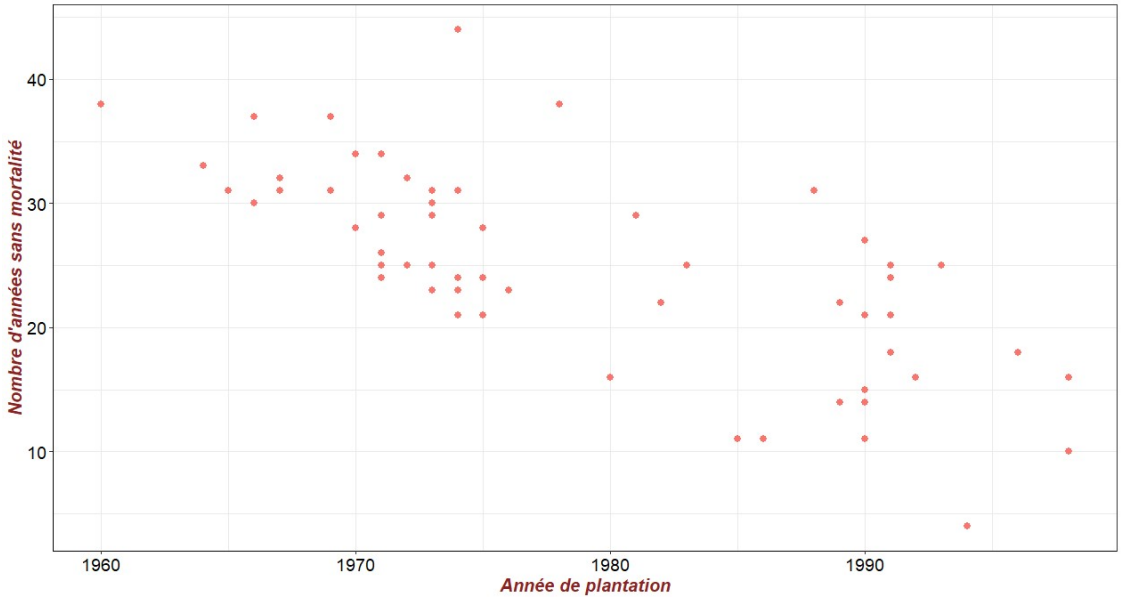


Figure 14 : Nombre d'années avant la mort du premier cep selon l'année de plantation
Légende : Chaque point représente une parcelle

Afin de vérifier si l'apparition de la mortalité dans les parcelles du réseau a évolué depuis sa création, on s'intéresse, pour chaque parcelle, au nombre d'années avant l'apparition d'un premier cep mort. Sur la figure 14, ce nombre d'années sans mortalité est représenté en fonction de l'année de plantation de la parcelle. On constate que les parcelles plantées après 1980 présentent pour certaines des ceps morts avant leurs 20 ans, ce qui n'est pas le cas des parcelles plantées avant 1980.

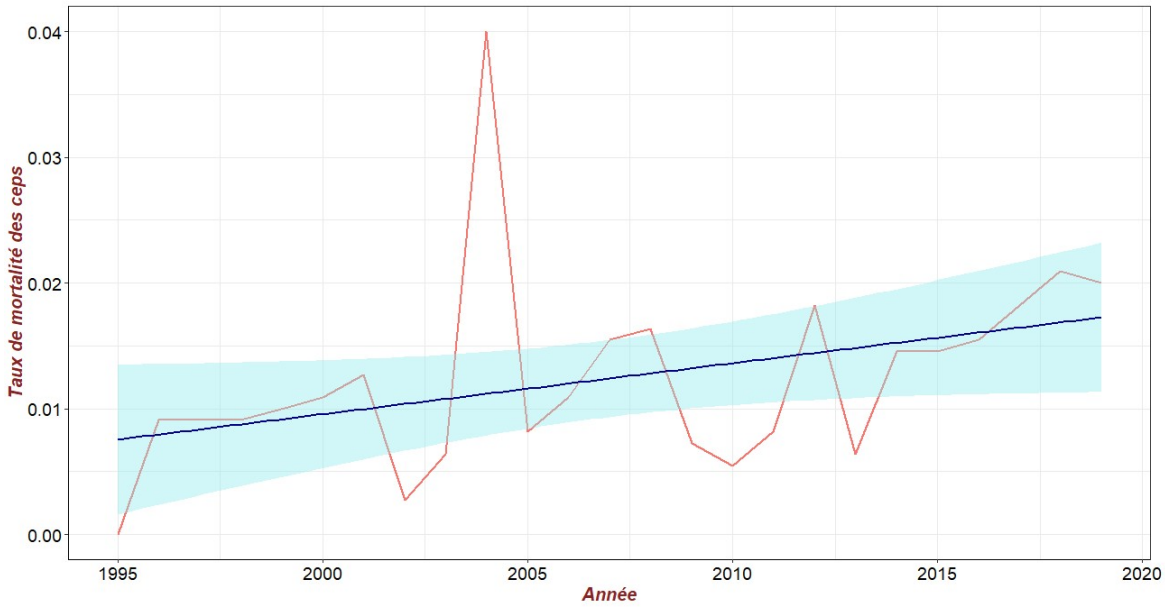


Figure 15 : Taux de mortalité des ceps selon les années
Légende : Rouge : taux de mortalité ; Violet : droite de régression du taux de mortalité ; Bleu : intervalle de confiance de la droite de régression

Globalement, l'évolution de la mortalité sur l'ensemble des parcelles du réseau est présentée sur la figure 15. Le taux de mortalité moyen correspond au nombre total de morts de l'année sur le nombre total d'emplacements de ceps présents dans le réseau. La droite de régression permet de tester si ce taux de mortalité augmente ou diminue au cours du temps. Le taux de mortalité des ceps était de 0 % en 1995, de 0,9 % en 1996 et de 2 % en 2019. Donc en 2019, deux ceps sur 100 emplacements sont morts en plus par rapport à l'année 1995. En 2004, on remarque une forte mortalité, on y trouve quatre ceps morts pour 100 emplacements. Cette augmentation de mortalité peut s'expliquer par un effet des fortes chaleurs subies en 2003.Visuellement, la droite de régression augmente au cours du temps, ce qui indique une augmentation du taux de ceps morts au cours du temps.

Les tests d'hypothèses nous indiquent que nous pouvons utiliser les résultats du modèle de régression mais que toutefois nous devons abandonner la linéarité (annexe7).

Tableau 10: Sortie du modèle linéaire issu de l'analyse du taux de mortalité au cours du temps

		Estimation des paramètres	Estimation de leur erreur standard	Statistique T	p-value du test évaluant l'égalité à0 des coefficients
Toutes les années	(Intercept)	-3.1302	8.665	-0.36	0.7212
	Année	0.001219	0.0006189	0.0609	0.0609.
Sans 2004	(Intercept)	-7.3488	4.97	-1.478	0.153450
	Année	0.001426	0.003537	4.042	0.000544**

Comme précédemment, on prend un seuil alpha de 0,05. La p-value est de 0,0609 donc supérieure à 0,05 donc on ne peut pas dire que la pente de régression est statistiquement différente de zéro (tableau 10). Le taux de mortalité observé en 2004 est exceptionnel et influence fortement la régression. Une nouvelle régression a donc été réalisée en supprimant ce point. La p-value résultante est 0,0005 donc inférieure à 0,05 donc nous pouvons dire qu'il existe un lien significatif entre le taux de mortalité et l'année sans l'année 2004 (tableau 10). L'équation du modèle de régression sans 2004 est : 0,0014 * Année -7,349. Donc on peut affirmer que pour les parcelles du réseau, hormis pour l'année 2004, le taux de mortalité augmente au cours du temps avec une progression relative de 0,0014. Chaque année, il meurt 0,14 cep de plus par emplacement que l'année précédente.

C. Sélection des années particulières

Au vu des analyses présentées ci-dessus nous avons sélectionné les années 1999, 2004 et 2007.

- L'année 1999 présente un rendement élevé avec un grand nombre de parcelles présentant une hausse du rendement et un taux de mortalité annuel faible.

- L'année 2004 est très similaire à l'année 1999. Elle présente un rendement élevé avec un grand nombre de parcelles présentant une hausse du rendement et on retrouve un fort taux de mortalité.

- L'année 2007 se caractérise par la présence d'une forte diminution de rendement avec un grand nombre de parcelles dont le rendement diminue et par un taux de mortalité faible.

5. Etudes annuelles des indicateurs de rendement et de mortalité

Sur les trois années sélectionnées précédemment, des analyses par arbres de décision de type random forest ont été réalisées pour les variables indicatrices du rendement et de la mortalité afin de mettre en évidence les facteurs déterminant les rendements et les mortalités observés.

A. L'année 1999

En 1999, un rendement fort et une mortalité faible ont été observés sur les parcelles du réseau. Pour la réalisation de la méthode random forest, on a pris un ntree de 500 et un mtry de 39 pour les indicateurs de rendement.

L'analyse par random forest permet d'expliquer 35,49 % de la variance du rendement déclaré. Parmi les principales variables impactant le rendement déclaré (annexe 8 figure 1A), trois variables se détachent. Il s'agit de la fumure azotée mise en place à la plantation, de la taille de la surface externe du couvert végétal de l'année précédente et du titre alcoométrique volumique probable (TAVP). Les autres variables dominantes sont des variables de nombre de grappes, d'amendement, de traitements phytosanitaires et de porte-greffe.

Pour le rendement calculé à la placette, on a 59,8 % d'explication de la variance. Les variables explicatives dominantes (annexe 8 figure 1B) sont des variables de composantes du rendement (le nombre de grappes à la placette et sur les ceps productifs des années 1998 et 1999, le poids moyen des grappes et le rendement de l'année antérieure), des variables d'amendement (azote, potassium et phosphore), des variables de traitements phytosanitaires, des variables fixes (le type de porte-greffe et la densité de plantation) et une variable de production, le TAVP.

La variabilité du rendement des ceps productifs est expliquée à 53,35 %. Les variables ayant une forte influence sur le rendement des ceps productifs (annexe 8 figure 1C) sont essentiellement des variables composantes du rendement, mais on y trouve également des variables d'amendement (azote et phosphate de fond), des variables phytosanitaires (nombre de cep présentant des signes d'Eutypiose, nombre de traitements de soufre de l'année antérieure et le nombre de traitements anti-oïdium de l'année 1998), une variable de production (le TAVP) et la densité de plantation.

Ainsi de manière globale, on voit que les variables dominantes des trois indicateurs du rendement en 1999, sont relativement proches puisque l'on retrouve les mêmes familles de variables. Parmi les quinze variables ayant le plus d'importance dans les random forest des trois variables indicatrices du rendement, on retrouve principalement des variables composantes du rendement, des variables d'amendement, des variables phytosanitaires et des variables de production.

L'analyse par random forest permet d'expliquer 49,3% de la variance du taux de mortalité de l'année 1999. Les variables contribuant le plus à expliquer cette variabilité (annexe 8 figure 2A) sont des variables d'amendements en azote, phosphate et potasse réalisés lors de la plantation, des variables phytosanitaires et des variables climatiques.

La variance du taux de mortalité moyen par an sur la période de 1995 à 1999 est expliquée à 47.48 % par le modèle. La taille de la surface externe du couvert végétal de l'année 1998 et les fumures en phosphore et en potasse à la plantation sont les informations permettant d'expliquer au mieux la variabilité de ce taux de mortalité (annexe 8 figure 2B).

Parmi les quinze premières variables de chaque analyse sur ces variables de mortalité, on retrouve sept variables identiques. Il s'agit de cinq variables d'amendement, d'une variable de rendement (le rendement déclaré) et d'une variable phytosanitaire (le nombre de traitement anti-mildiou).

B. L'année 2004

L'année 2004 présente un rendement fort avec une mortalité forte. Après optimisation des paramètres, on a ainsi pris un ntree de 500 et un mtry de 41 pour les indicateurs de rendement.

Pour le rendement déclaré, on obtient une explication de la variance de 15%. Parmi les principales variables impactant le rendement déclaré (annexe 9 figure 1A) cinq variables se détachent. Il s'agit du titre alcoométrique volumique probable, de la taille de la surface externe du couvert végétal des années 2003 et 2004, de l'enherbement sur l'inter-rang à la véraison et du nombre de traitement réalisé contre l'oïdium.

Pour le rendement calculé à la placette, nous obtenons une explication de 43 % de la variance de la variable à expliquer. Les principales variables qui ressortent sont des variables de composantes du rendement, des variables de végétation (SECVC, nombre de rognages, importance des entre cœurs et enherbement sur l'inter-rang à la nouaison) et des variables de plantation (type de porte-greffe et densité de plantation) (annexe 9 figure 1B).

La variabilité du rendement calculé par cep productif, est expliquée à hauteur de 43,67 % par la random forest. Les trois variables dominantes sont des variables de composantes du rendement (poids des grappes, nombre de grappes sur la placette et sur les ceps productifs) (annexe 9figure 1C). Les autres variables importantes dans cette analyse sont essentiellement des variables fixes comme le type de porte-greffe, la densité de plantation et la fumure potassique de fond.

Parmi les premières variables ayant le plus d'importance dans les random forest des trois variables indicatrices du rendement en 2004 on trouve deux variables présentes dans les trois analyses, il s'agit du type de porte-greffe et de la taille de la surface externe du couvert végétal de l'année précédente, 2003.

Pour les indicateurs de mortalité, nous avons une explication de 48,46 % de la variance du taux de mortalité annuel et une explication de 66,98 % de la variance du taux de mortalité global de la période de 1995 à 2004. Nos deux variables à expliquer sont principalement expliquées par le même type de variables (annexe 9figure 2). Il s'agit de variables d'amendement de fond. Parmi les quinze variables ayant le plus d'importance dans les random forest des deux variables indicatrices de la mortalité, on trouve douze variables présentes dans les deux analyses. Six de ces variables sont des variables d'amendements de fond, puis il y a le type de porte-greffe, l'âge des parcelles, la taille de la surface externe du couvert végétal de l'année 2003 et les rendements déclarés de l'année 2003 et 2004.

C. L'année 2007

L'année 2007 est marquée par un rendement faible et un taux de mortalité faible. L'optimisation des paramètres nous a fait prendre un ntree de 500 et un mtry de 41 pour le rendement déclaré et 42 pour les rendements calculés à la placette et aux ceps productifs.

Pour le rendement déclaré, on obtient une explication de la variance du rendement déclaré de 2 %. Au vu du faible pourcentage d'explication de la variable à expliquer pour l'année 2007 nous n'avons pas conservé les variables dominantes de ce modèle (annexe 10 figure 1A).

Pour le rendement calculé à la placette, nous obtenons une explication de 60,33 % de la variance. Les variables dominantes de ce modèle sont des variables de composantes du rendement, des variables d'amendement et des variables fixes (la sensibilité au gel, le type de porte-greffe, l'exposition et la densité de plantation) (annexe 10 figure 1B).

Au niveau du rendement calculé par cep productif, il y a 62,7 % de la variance qui a été expliqué par la random forest. Les variables importantes dominantes du modèle sont des variables de composantes du rendement, des variables d'amendement et des variables de plantation (type de porte greffe, densité de plantation et exposition de la parcelle) (annexe 10 figure 1C).

Parmi les quinze variables ayant le plus d'importance dans les random forest des deux variables indicatrices du rendement en 2007, on trouve dix variables présentes dans les deux analyses. Il s'agit de cinq variables de composantes du rendement, de deux variables d'amendement et de trois variables fixes (type de porte-greffe, exposition de la parcelle et densité de plantation).

Pour les indicateurs de mortalité, nous avons une explication de 3 % de la variance du taux de mortalité annuel et de 25,33 % de la variance du taux de mortalité global de la période de 1995 à 2004. Pour le taux de morts de l'année 2007, au vu du faible de pourcentage de

variance expliqué, nous ne conservons pas les variables dominantes du modèle (annexe 10 figure 2A).

Pour le modèle décrivant le taux de mortalité de l'année, les variables importantes sont le type de porte-greffe, la taille de la surface externe du couvert végétal, l'âge des parcelles, des variables de rendement et des variables climatiques (annexe 10 figure 2B).

D. Comparaison des caractéristiques

De manière plus globale, dans le tableau 11, on retrouve les variables expliquant le mieux les indicateurs de rendement selon l'année. Les variables composantes du rendement, le type de porte-greffe et la densité de plantation sont des variables expliquant le rendement. Pour des années à fort rendement, on retrouve aussi la variable du titre alcoométrique volumique probable.

Tableau 11 : Variables explicatrices principales des indicateurs de rendement

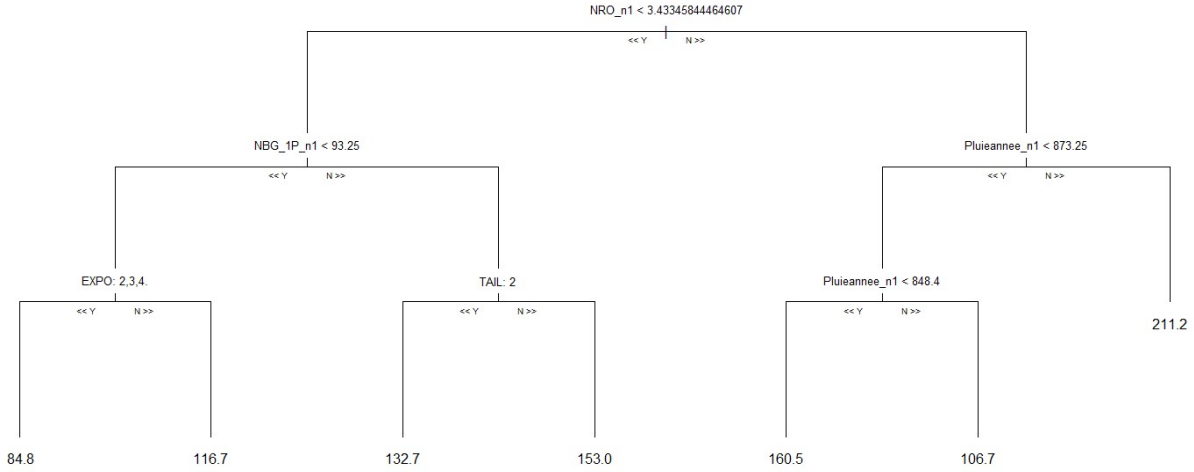
1999	2004	2007
- variables de composantes du rendement - amendement de fond (N, P et K) - titre alcoométrique volumique probable - densité de plantation - type de porte-greffe - nombre de traitements phytosanitaires	- variables de composantes du rendement - surface externe du couvert végétal de 2003 et de 2004 - titre alcoométrique volumique probable - végétation (enherbement, rognage et entre-cœur) - type de porte-greffe - densité de plantation - amendement annuel (N et K)	- variables de composantes du rendement - amendements annuels - porte greffe - densité de plantation - exposition de la parcelle

Le tableau 12 présente les variables permettant d'expliquer au mieux les indicateurs de mortalité. Toutes années confondus nous retrouvons des variables de rendement et la variable surface externe du couvert végétal. Les années plus proches entre elles présente des similitudes. Les années 1999 et 2004 ont comme variables explicatives importantes les amendements de fond. Les années 2004 et 2007 ont en variables explicatives similaires l'âge des parcelles et le type de porte-greffe.

Tableau 12 : Variables explicatrices principales des indicateurs de mortalité

1999	2004	2007
- amendements de fond - nombre de traitements phytosanitaires - variables de rendement - surface externe du couvert végétal	- amendements de fond - surface externe du couvert végétal - type de porte greffe - âge des parcelles - variables de rendement	- type de porte greffe - surface externe du couvert végétal - variables de rendement - variables climatique - âge des parcelles

De manière plus visuelle, l'arbre de régression moyen des différents indicateurs du rendement et de la mortalité a été réalisé pour les différentes années.



<u>Figure 16 :</u> Arbre de régression représentant l'arbre moyen des random forest du rendement déclaré de 1999

Dans la figure 16, on a l'arbre de régression représentant la moyenne des arbres réalisés dans la random forest pour le rendement déclaré de 1999. Ainsi, on peut déterminer quel chemin d'action engendre un rendement ou une mortalité selon l'arbre de décision. Par exemple, dans le réseau de Cognac, en 1999, pour avoir un fort rendement déclaré de 211,2 hL/ha, il fallait avoir réalisé plus de trois rognages l'année 1998 et que le cumul de précipitation sur l'année calendaire de 1998 soit supérieur à 873,25 mm. Le cumul de précipitation en 1998 a varié sur les parcelles du réseau de Cognac entre 823 mm et 904 mm. La valeur de 873.25 mm se situe donc dans la fourchette médiane haute de cette gamme de variation.

Pour les années 1999 et 2004, les rendements sont élevés tandis que pour l'année 2007, le rendement est faible. Lorsque l'on examine les différents chemins d'action pour avoir un rendement élevé lors d'années à rendement important on trouve un chemin impliquant beaucoup de grappes lourdes, des rendements élevés les années précédentes, les porte-greffes ruggeri 140 et gravesac sont mis en relief, des concentrations seuil en amendement, l'implication du cumul de précipitation de l'année antérieure et de l'influence de la surface externe du couvert végétal de l'année n-1 (annexe 8 figure 3 et annexe 9 figure 3). Pour l'année 2007, pour avoir un rendement élevé à la placette, il fallait avoir de grosses grappes, mis moins de 85 unité de fumure de potassium par hectare l'année 2006 et qu'il pleuve beaucoup l'année 2006 (>376 mm) (annexe 10 figure 3).

Les années 1999 et 2007 sont marquées par un taux de mortalité annuel faible. L'année 2004 présente un taux de mortalité important. En 1999, pour avoir un taux de mortalité élevé pour l'année, il fallait qu'il gèle légèrement l'année antérieure et apporter moins de 2,67 unité d'azote par hectare sur les parcelles (annexe 8 figure 3). Les parcelles présentant un taux de mortalité important en 2007 sont des parcelles à densité de plantation inférieure à 2721 ceps/ ha et ayant un porte-greffe de type 333 EM, 41 B, 161-49 c ou SO4 (annexe 10 figure 3). Pour l'année 2004, les parcelles avec un taux de mortalité élevé ont eu beaucoup de fertilisant azoté

(>137 28 unité/ha) et possèdent des porte-greffes de type 333 EM, 41 B, 161-49 c, SO4 ou rupestris (annexe 9 figure 3).

IV/ Discussion

Sur la base des informations que nous avons extraites de la base de données collectée sur les parcelles de Cognac, nous avons pu mettre en évidence une baisse progressive du rendement associée à une hausse de la mortalité des ceps. Cela confirme les constatations ayant conduit à la création du plan dépérissement (Baldassini and Despey, 2015; CNIV, 2018). La comparaison, année par année, du nombre de parcelles dont le rendement est croissant ou décroissant par rapport à l'année précédente montre un équilibre entre les augmentations et les diminutions de rendement. Cela semble confirmer que la baisse de rendement observée s'est installée sur le long terme et n'est pas un phénomène ponctuel.

Le rendement déclaré par le viticulteur et le rendement moyen calculé à la placette montrent une diminution significative au cours du temps. Le rendement moyen déclaré baisse de 0,97 hL/ha/an et le rendement moyen calculé de 410,6 g/placette sur la période où cette donnée est présente. Cette diminution de rendement déclaré semble importante mais reste bien inférieure à la perte de rendement moyen de 4,6 hL/ha/an mentionnée par le plan de dépérissement(CNIV and FranceAgriMer, 2018). Sur la même période, le rendement ramené au cep productif ne présente pas de diminution ou d'augmentation significative au cours du temps. Ainsi, un pied de vigne productif n'a pas tendance à produire plus ou moins de raisin sur le long terme. Comme les rendements à la parcelle baissent au cours du temps, mais que les rendements des ceps productifs ne changent pas, on peut considérer que la baisse de rendement observée à l'échelle de la parcelle ou de la placette sont dues à l'augmentation du nombre de ceps manquants ou non productifs.

Les ceps non productifs correspondent à des ceps morts, très jeunes, secs ou mourants. Le taux de ceps non productifs augmente au cours du temps dans les parcelles du réseau. On constate également que l'âge moyen des parcelles augmente dans le réseau au cours du temps. On peut ainsi mettre en évidence une augmentation du taux de mortalité des ceps concomitamment au vieillissement des parcelles, ce qui est cohérent avec les conclusions de Cassagnes (2011). Le vieillissement des parcelles peut donc apparaître comme un facteur favorisant le dépérissement. L'étude de l'âge des parcelles du réseau montre que les parcelles entrantes sont globalement jeunes. Entre 1987 et 2017, la moyenne d'âge des parcelles entrantes est de 7 ans. Pourtant, l'âge moyen des parcelles du réseau ne cesse pas d'augmenter au cours du temps. On peut donc conclure que la baisse de rendement au cours du temps est probablement liée à une hausse de la mortalité, favorisée par le vieillissement des parcelles du réseau. Toutefois, il serait intéressant d'élucider si l'augmentation de l'âge moyen des parcelles du réseau de Cognac est une particularité du réseau ou si cela indique que les vignobles produisant du raisin pour le Cognac sont de plus en plus vieux. Dans le premier cas, on pourrait proposer de renouveler plus vite les parcelles, l'arrivée de nouvelles parcelles jeunes limiterait le vieillissement moyen observé. En revanche, si le vieillissement des

parcelles est généralisé dans l'appellation, il y a probablement là un élément d'explication du dépérissement.

Par ailleurs, le plan de dépérissement de la vigne a constaté que le dépérissement viticole apparait de plus en plus tôt dans la vie de la vigne (CNIV, 2017). Pour vérifier si ceci peut être caractérisé via la base de données du BNIC, nous avons estimé, pour chaque parcelle, sa durée de vie sans mortalité de ceps à partir de l'année de plantation, et suivi l'évolution de cette variable au cours du temps. Les parcelles plantées récemment présentent des morts plus rapidement que les parcelles plantées il y a plus de 40 ans. Ceci permet de mettre en évidence sur les parcelles du réseau de Cognac un élément supplémentaire en lieu avec le dépérissement viticole. Toutefois, on constate que la hausse du taux de mortalité annuel n'est pas significative. La pente de la droite de régression du taux de mortalité annuelle devient significative lorsque l'on retire l'année 2004, qui se caractérise par une très forte mortalité explicable par un évènement ponctuel. Ainsi, à part pour l'année 2004, on peut affirmer qu'il y a une augmentation significative du taux de mortalité annuelle au cours du temps. Comme les parcelles du réseau sont de plus en plus âgées, il est ainsi difficile de déterminer si la mortalité augmente parce que les vignes sont de plus en plus vieilles ou pour d'autres raisons liées au temps comme par exemple l'évolution du climat, des changements de pratiques ou des pressions liées aux pathogènes ou aux ravageurs.

Pour identifier les facteurs déterminants du rendement et de la mortalité, nous avons réalisé des analyses par arbres de décision sur trois années particulières, choisies parce qu'elles présentaient des évolutions marquées des variables indicatrices du dépérissement. Cette analyse spécifique a été sélectionnée car la base de données fournie par le BNIC est riche en informations de types variés et qu'il s'agit d'une méthode d'analyse permettant une visualisation directe du chemin d'action pour obtenir une valeur donnée de la variable de rendement ou de mortalité étudiée (Genuer et al., 2017). Ainsi, on peut identifier l'organisation des variables et apporter des hypothèses d'explication sur le déterminisme de certains évènements.

Les années utilisées pour la suite des analyses ont été 1999, 2004 et 2007. Les années 1999 et 2004 ont été choisies pour leur fort rendement et l'année 2007 pour sa forte mortalité. Il est ressorti que, pour ces années précises, les facteurs explicatifs étaient plus similaires entre les modèles de rendement que pour les modèles de mortalité. Certains modèles permettent d'expliquer un très faible pourcentage de la variance des variables à expliquer. Cela peut être dû à des rendements déclarés par l'exploitant mal estimés ou encore à cause d'un manque de variables explicatives indispensables pour déterminer la variation de ces variables. Ceci signifierait que les causes de la variabilité observée ne se trouvent pas parmi les variables explicatives.

Les différents modèles de rendement et de mortalité présentent comme variables explicatives principales des variables de production, des variables d'amendement et des variables de végétation. Dans les variables de production, on retrouve les variables composantes du rendement, ce qui est logique puisqu'il s'agit de variables intervenant dans les rendements calculés. Le titre alcoométrique volumique probable est fortement lié au

rendement des années examinées. Cette variable est estimée à partir de la concentration en sucres à la vendange. Les amendements annuels, les fertilisants, permettent la satisfaction des besoins en minéraux des cultures. Ainsi, une insuffisance de fertilisation peut engendrer des effets indésirables et ainsi faire diminuer le rendement (Duchêne et al., 2003) voire affaiblir la vigne et concourir à des mortalités de ceps. L'impact et la diversité des variables d'amendement pour les modèles de rendement et de mortalité est telles qu'il devrait faire l'objet d'une étude spécifique pour la caractérisation du dépérissement. La présence de la variable de végétation explicatrice comme « surface externe du couvert végétal » semble également concorder avec les acquis du domaine, puisqu'elle caractérise le développement végétatif de la vigne et donc sa capacité à intercepter le rayonnement incident. Or il est démontré que la quantité de rayonnement intercepté impacte la mise en place du rendement avec la croissance des baies de raisin (Palma and Jackson, 1981).

Les modèles de rendement présentent spécifiquement également des variables phytosanitaires en variables explicatives. Les variables caractérisant les traitements phytosanitaires constituent des indicateurs indirects de la pression des bio-agresseurs sur les vignes. Plus une parcelle est impactée par la présence de pathogènes ou de ravageurs, plus le viticulteur apportera de traitements, mais la plante disposera également de moins de nutriments à distribuer aux organes pour constituer le rendement ou favoriser la longévité de la vigne. Les modèles de mortalité présentent également en variables explicatives principales des variables de plantation et des variables climatiques. Dans les variables de plantation, on identifie la densité de plantation, elle peut être considérée comme un indicateur de la concurrence des ceps de vigne entre eux sur la parcelle. Une trop forte densité engendre une concurrence entre ceps pour les nutriments, donc un risque d'appauvrissement sur le long terme et la baisse de production des ceps de vigne. Elle peut ainsi rendre la productivité et la longévité des ceps plus sensibles aux amendements apportés et à d'autres interventions techniques. L'âge des parcelles apparait également contre une variable importante au niveau des modèles de mortalité, cela semble logique puisqu'on a montré une hausse de la mortalité selon l'âge de la parcelle. Une différenciation de rendement a été observée selon le type de porte-greffe : certains porte-greffes favorisent les forts rendements. Les variables climatiques sont des facteurs impliqués dans la mortalité des ceps. Un cumul de précipitation minimal est nécessaire pour éviter un stress hydrique trop important qui engendrerait une réduction de l'activité photosynthétique et donc de la productivité (Epron, 2018). Le changement climatique entraine une augmentation de la précocité. De ce fait la maturation des raisins se déroule en période de plus en plus chaude, ce qui augmente le risque de sécheresse et donc de perte de rendement. L'évolution attendue des conditions climatiques pour les prochaines décennies entraînera un changement de la production viticole.

Les résultats obtenus permettent d'affirmer qu'il y a bien un dépérissement viticole sur le réseau de Cognac et des variables explicatives ont été mises en évidence selon les indicateurs de dépérissement et les années étudiées. Les résultats de cette analyse sont soumis à des limites et ne peuvent pas être extrapolés à d'autres réseaux viticoles. Dans ce mémoire, on travaille sur un réseau particulier dont les parcelles sont majoritairement de l'Ugni blanc et en moyennes vieillissantes. Cela limite grandement les conditions pédoclimatiques et génétiques

de l'étude. Les mesures de mortalité débutent en 1995 et de nombreuses variables explicatives courantes dans le domaine n'ont pu être utilisées, par exemple les analyses de sol puisque les mesures ont été très rarement effectuées.

Pour une caractérisation du dépérissement plus profonde du vignoble de Cognac, il serait profitable de poursuivre l'étude sur d'autres d'années afin d'identifier si les variables impliquées pour le rendement et pour la mortalité changent drastiquement sur certaines années ou évoluent progressivement au cours du temps. Une analyse Window Pane sur les différentes années, avec uniquement les parcelles présentes en continu sur une large période pourrait être proposée, permettant de réaliser un découpage temporel pour mettre en relief d'autres variables impactantes sur un plus long terme.

Conclusion

Lors de ce stage, j'ai mis en forme et analysé la base de données fournie par le BNIC. Mes objectifs étaient de voir s'il était possible de mettre en évidence des liens entres les variables caractéristiques du dépérissement, que sont le rendement et la mortalité, et les autres variables mesurées sur les parcelles du réseau.

J'ai pu mettre en évidence une baisse du rendement et une hausse de la mortalité des ceps de vigne significative au cours du temps comme signalé par la filière viticole. Toutefois, en examinant le rendement individuel des ceps productifs, on a remarqué une évolution non significative au cours du temps. De ce fait, la baisse de rendement serait plutôt liée à l'augmentation de la mortalité des ceps.

En sélectionnant des années particulières à rendements ou mortalités remarquables, j'ai pu mettre en évidence un certain nombre de variables qui sont significativement liées aux variables du dépérissement. Des variables de production, d'amendement, de végétation, de plantation, phytosanitaire et climatique ressortent principalement lors de l'étude de rendement et de mortalité sur des années spécifiques.

Les analyses réalisées ont permis de mettre en évidence le dépérissement présent dans le réseau de Cognac, en caractérisant le rendement et la mortalité. Pour une approche de caractérisation plus dynamique du rendement et de la mortalité, il serait intéressant d'effectuer des analyses Window Pane pour les différentes années, avec uniquement les parcelles présentes en continu de 1995 à 2019 et en prenant en compte les variables de l'année antérieure comme une nouvelle variable.

Bibliographie

I. Livre

Coussié, J.-V., 1996. Le cognac et les aléas de l'Histoire.

Tufféry, S., 2012. Data mining et statistique décisionnelle. L'intelligence des données 4e édition.

II. Article

Aubert, I., Mahé, H., Prêtet-Lataste, C., 2017. Le plan dépérissement en 31 fiches d'actions [WWW Document]. URL (accessed 8.20.20).

Baldassini, M., Despey, J., 2015. Les dépérissements du vignoble (Séminaire du 22/07/15 à la Maison de la Chimie).

Banerjee, M., Ding, Y., Noone, A.-M., 2012. Identifying representative trees from ensembles. Stat Med 31, 1601–1616. https://doi.org/10.1002/sim.4492

Barnier, M., Woerth, E., 2009. Arrêté du 11 février 2009 relatif aux conditions d'attribution de la prime à l'arrachage de vignes.

BIPE, 2015. Etat de l'art des connaissances sur le sujet du dépérissement du vignoble.

Carbonneau, A., Leclair, P., Dumartin, P., Cordeau, J., Roussel, C., 1977. Etude de l'influence chez la vigne du rapport « partie végétative / partie productrice » sur la production et la qualité des raisins. OENO One 11, 105–130. https://doi.org/10.20870/oeno-one.1977.11.2.1437

Clingeleffer, P.R., Martin, S.R., Dunn, G.M., Krstic, M.P., 2001. Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties : a national approach : final report to Grape and Wine Research & Development Corporation / principal investigator, Peter Clingeleffer ; [prepared and edited by Steve Martin and Gregory Dunn].

CNIV, FranceAgriMer, 2018. Dépérissement viticole : Evaluer l'exposition de votre exploitation viticole.

Comité national des appellations d'origine relatives aux, vins et aux boissons alcoolisées, et des boissons spiritueuses, 2018. CAHIER DES CHARGES DE L'APPELLATION D'ORIGINE CONTROLEE « COGNAC » ou « EAU-DE-VIE DE COGNAC » ou « EAUDE- VIE DE CHARENTES ».

des ROBERT, M.-L., JOBARD, E., RADUREAU, S., TONIN, P., 2016. BIPE : Présentation du Plan national de lutte contre les dépérissements du vignoble.

Desoutter, L., Vaillant, G., Zacharko, C., Ritter, M., 2020. Projet tutoré en agroforesterie viticole méditerrenéenne avec l'UMR System.

Duchêne, E., Jaegli, N., Salber, R., Gaudillère, J.-P., 2003. Effects of ripening conditions on the following season's growth and yield components for Pinot noir and Gewurztraminer grapevines (Vitis vinifera L.) in a controlled environment. OENO One 37, 39–49. https://doi.org/10.20870/oeno-one.2003.37.1.1682

Epron, D., 2018. Effet des déficits hydriques et des forts éclairements sur la photosynthèse de jeunes semis de chênes en conditions contrôlées et de chênes adultes en conditions naturelles 207.

FranceAgriMer, 2017. La consommation de vin en France (Enquête de 2015).

Gandin, 1966. Objective analysis of meteorological fields. Translated from the Russian. Jerusalem (Israel Program for Scientific Translations). Quarterly Journal of the Royal Meteorological Society 92, 447–447. https://doi.org/10.1002/qj.49709239320

Gary, C., 2017. Cadre de réponse technique : Déterminants, indicateurs et gestion technique des trajectoires de dépérissement des vignobles (TraDeVi).
Genuer, R., Poggi, J.-M., Tuleau-Malot, C., Villa-Vialaneix, N., 2017. Random Forests for Big Data. Big Data Research 9, 28–46. https://doi.org/10.1016/j.bdr.2017.07.003
Grimaldi, J., 2018. Impacts of agroforestry on microclimate for grappe and wine production: assessment in Southern France.
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
Howell, G.S., 2001. Sustainable Grape Productivity and the Growth-Yield Relationship: A Review. Am J Enol Vitic. 52, 165–174.
Jeong, J.H., Woo, J.H., Park, J., 2020. Machine Learning Methodology for Management of Shipbuilding Master Data. International Journal of Naval Architecture and Ocean Engineering 12, 428–439. https://doi.org/10.1016/j.ijnaoe.2020.03.005
Keller, M., Tarara, J.M., Mills, L.J., 2010. Spring temperatures alter reproductive development in grapevines. https://doi.org/10.1111/J.1755-0238.2010.00105.X
Montgomery, D.C., Peck, E.A., 1982. Introduction to Linear Regression Analysis. Wiley.
Palma, B.A., Jackson, D.I., 1981. Effect of Temperature on Flower Initiation in Grapes. Botanical Gazette 142, 490–493.
Préfecture de Charente, 2019. Viticuture.
Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., Morel, S., 2008. Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France. J. Appl. Meteor. Climatol. 47, 92–107. https://doi.org/10.1175/2007JAMC1636.1
Vasconcelos, M. Carmo, Greven, M., Winefield, C.S., Trought, M.C.T., Raw, V., 2009. The Flowering Process of Vitis vinifera: A Review. Am J Enol Vitic. 60, 411–434.
Weymark, J.A., 1981. Generalized gini inequality indices. Mathematical Social Sciences 1, 409–430. https://doi.org/10.1016/0165-4896(81)90018-4

III. Plateforme internet

AFP, 2020. La vigne et le vin en France: Cinq choses à savoir [WWW Document]. Ouest-France.fr. URL https://www.ouest-france.fr/economie/la-vigne-et-le-vin-en-france-cinq-choses-savoir-6726220 (accessed 8.20.20).
Cassagnes, J., 2016. Les raisons de la baisse des rendements.
CNIV, 2017. Plateforme collaborative du plan dépérissement: “https://www.plan-deperissement-vigne.fr/.”
CNRM, 2014. SAFRAN - National Centre for Meteorological Research [WWW Document]. URL http://www.umr-cnrm.fr/spip.php?article788&lang=en (accessed 8.25.20).
Vin-Vigne.com, 2018. Vignoble de France - Liste des vignobles français - [WWW Document]. URL http://www.vin-vigne.com/vignoble/ (accessed 8.25.20).

IV. Logiciel

Dasgupta, A., 2014. reprtree: Representative trees from ensembles. R package version 0.6.
Fox, J., Weisberg, S., 2019. car: Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Liaw, A., Wiener, M., 2002. randomForest: Classification and Regression by randomForest. R News 2(3), 18--22.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2018. nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-137,

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Therneau, T., Atkinson, B., 2018. rpart: Recursive Partitioning and Regression Trees. R package version 4.1-13. https://CRAN.R-project.org/package=rpart.

Sommaire des annexes

Annexe 1 : L'indice héliothermique de Huglin 2
Annexe 2 : Lexique des variables 3
Annexe 3 : Liste des variables sélectionnées 6
Annexe 4 : Rendement des différentes parcelles au cours du temps 10
Annexe 5 : Test de validité des hypothèses pour la régression linéaire du rendement placette calculé 11
Annexe 6 : Test de validité des hypothèses pour la régression linéaire du rendement des ceps productifs 13
Annexe 7 : Test de validité des hypothèses pour la régression linéaire du taux de mortalité annuel 15
Annexe 8 : Sortie des random forest pour les indicateurs du rendement et de mortalité de 1999 18
Annexe 9 : Sortie des random forest pour les indicateurs du rendement et de mortalité de 2004 22
Annexe 10 : Sortie des random forest pour les indicateurs du rendement et de mortalité de 2007 27

Annexe 1 : L'indice héliothermique de Huglin

Tableau : Classification des climats viticoles d'après l'indice héliothermique de Huglin

Très chaud	IH+3	3000 < IH
Chaud	IH+2	2400 < IH ≤ 3000
Tempéré chaud	IH+1	2100 < IH ≤ 2400
Tempéré	IH-1	1800 < IH ≤ 2100
Frais	IH-2	1500 < IH ≤ 1800
Très frais	IH-3	IH ≤ 1500

Annexe 2 : Lexique des variables

Comme le lexique des variables fait 53 pages Word, je mets uniquement les 3 premières pages du dictionnaire en annexe. Les autres pages sont disponibles sous format numérique.

Tableau : Lexique d'identification des variables
Légende : n° : numéro de la variable; sélection : o : variable gardée, n : variable mise de cotée; var : abréviation présent dans les fichiers Excel; var_id : identification de la variable; mod : modalité présente par variable; mod_id : identification des modalités par variable; fichier : nom du fichier source de la variable; Type_données : indication du type de donnée de la variable; Type_information : indication de la thématique de la variable, 1 : caractéristiques de la parcelle fixes dans le temps, 2 : variables agronomiques, 3 : pratiques du viticulteur sur la parcelle, 4 : caractéristiques commerciales; var_fixe : indication de la dynamique de la variable; d_début et d_fin : indication de l'année de première et de dernière année de mesure de la variable

n°	sélection	var	var_id	mod	mod_id	fichier	Type_données	Type_information	var_fixe	d_debut	d_fin
1	o	AN	Année			AF				1990	2011
2	n	B	Bore			AF	Quantitatif	2	evo	1993	2002
3	n	Ca	Calcium			AF	Quantitatif	2	evo	1997	2011
4	o	CODE	Code d'identification de la parcelle			AF	Qualitatif		fixe		
5	n	Cu	Cuivre			AF	Quantitatif	2	evo	1990	1997
6	n	DATE	Date des prélèvements des feuilles			AF	Qualitatif		evo	1999	2005
7	o	EPOQ	Période de mesure	1	Floraison	AF	Qualitatif				

7	o	EPOQ	Période de mesure	2	Véraison	AF	Qualitatif				
8	n	Fe	Fer			AF	Quantitatif	2	evo	1990	1997
9	o	K	Potassium			AF	Quantitatif	2	evo	1990	2011
10	n	Mg	Magnésium			AF	Quantitatif	2	evo	1990	2011
11	n	Mn	Manganèse			AF	Quantitatif	2	evo	1990	1997
12	o	N	Azote			AF	Quantitatif	2	evo	1990	2011
13	n	Na	Sodium			AF	Quantitatif	2	evo	1997	2011
14	n	NBO	Nombre de feuille observée			AF	Quantitatif	2	evo	1999	2010
15	o	ORGA	Organe analysé	1	Limbe	AF	Qualitatif				
15	o	ORGA	Organe analysé	2	Pétiol	AF	Qualitatif				
15	o	ORGA	Organe analysé	3	Feuille entière	AF	Qualitatif				
16	o	P	Phosphore			AF	Quantitatif	2	evo	1990	2011
17	n	POID	Poids des feuilles prélévées			AF	Quantitatif	2	evo	1990	2010
18	n	S	Soufre			AF	Quantitatif	2	evo	1997	1997
19	n	Zn	Zinc			AF	Quantitatif	2	evo	1990	1997
20	n	AN	Année			AF1					
21	n	B	Bore			AF1	Quantitatif	2	evo	1993	1997

22	n	Ca	Calcium			AF1	Quantitatif	2	evo	1997	1998
23	n	Cu	Cuivre			AF1	Quantitatif	2	evo	1990	1997
24	n	DATE	Date des prélèvements des feuilles			AF1	Qualitatif				
25	n	EPOQ	Période de mesure	1	Floraison	AF1	Qualitatif				
25	n	EPOQ	Période de mesure	2	Véraison	AF1	Qualitatif				
26	n	Fe	Fer			AF1	Quantitatif	2	evo	1990	1997
27	n	GEN	Numéro de génération			AF1					
28	n	K	Potassium			AF1	Quantitatif	2	evo	1990	1998
29	n	Mg	Magnésium			AF1	Quantitatif	2	evo	1990	1998
30	n	Mn	Manganèse			AF1	Quantitatif	2	evo	1990	1997
31	n	N	Azote			AF1	Quantitatif	2	evo	1990	1998
32	n	N°LABO	Identification du laboratoire			AF1	Qualitatif		evo		
33	n	Na	Sodium			AF1	Quantitatif	2	evo	1997	1998
34	n	ORGA	Organe analysé	1	Limbe	AF1	Qualitatif				
34	n	ORGA	Organe analysé	2	Pétiol	AF1	Qualitatif				
34	n	ORGA	Organe analysé	3	Feuille entière	AF1	Qualitatif				

Annexe 3 : Liste des variables sélectionnées

Tableau : Identification des familles de déterminants pour les variables sélectionnées
Légende : Famille de déterminant : catégorie de la variable ; 1 : variables de production ; 2 : variables de végétation ; 3 : variables d'amendement ; 4 : variables de plantation ; 5 : variable phytosanitaire ; 6 : variables climatiques ; 7 : sol ; var : abréviation de la variable; var_id : identification de la variable; Type_données : quantitative ou qualitative

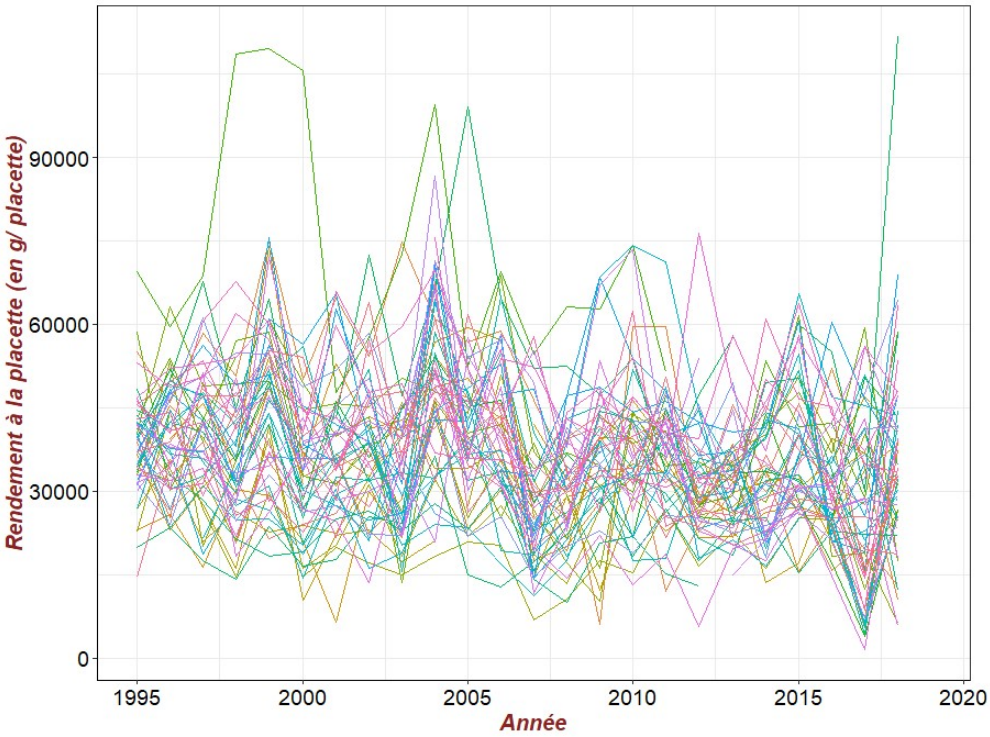
Famille de déterminant	var	var_id	Type_données
1	AGE_1er_MORT	âge des parcelles lorsque le premier cep est mort	Quantitatif
1	CRU	cru de la parcelle	Qualitatif
1	Nb_etablie	nombre de cep productif sur les quatre placettes	Quantitatif
1	Nb_MORT	Nombre de mort (calculé à partir d'ESCAC)	Quantitatif
1	NBG_1C	nombre de grappes moyen par cep	Quantitatif
1	NBG_1P	nombre de grappes moyen par placette	Quantitatif
1	NEW_MORT	nombre de mort de l'année	Quantitatif
1	Rdt_cep	rendement moyen par cep productif calculé	Quantitatif
1	Rdt_placette	rendement moyen par placette calculé	Quantitatif
1	taux_mort_annuel	nombre de cep mort de l'année divisé par le nombre de cep regardé	Quantitatif
1	TAVP	degré de la parcelle (% vol.)	Quantitatif
1	Tx_mort	pourcentage de cep mort dans une parcelle lors de l'année n	Quantitatif
1	tx_mort_annuel_global04	taux de ceps morts par an entre 1995 de mortalité et 2004	Quantitatif
1	tx_mort_annuel_global07	taux de ceps morts par an entre 1995 de mortalité et 2007	Quantitatif
1	tx_mort_annuel_global99	taux de ceps morts par an entre 1995 de mortalité et 1999	Quantitatif
1	TX_pas_etablie	pourcentage de cep pas productif dans une parcelle lors de l'année n	Quantitatif
1	VP	rendement de la parcelle (hl/ha)	Quantitatif
2	CIR1V	concurrence de l'enherbement à la véraison IR1	Qualitatif
2	CIR2V	concurrence de l'enherbement à la véraison IR2	Qualitatif
2	COUP	nombre de coupes	Quantitatif
2	CSRV	concurrence de l'enherbement à la véraison sous le rang	Qualitatif

2	EC	importance des entre cœurs	Qualitatif
2	HERB	importance de l'herbe	Qualitatif
2	HIR1N	importance de l'herbe sur l'interrang à la nouaison	Qualitatif
2	HIR1V	importance de l'herbe sur l'interrang à la véraison	Qualitatif
2	HIR2N	importance de l'herbe sur l'interrang à la nouaison	Qualitatif
2	HIR2V	importance de l'herbe sur l'interrang à la véraison	Qualitatif
2	HIRN	importance de l'herbe sur l'interrang à la nouaison	Qualitatif
2	HIRV	importance de l'herbe sur l'interrang à la véraison	Qualitatif
2	HSRN	importance de l'herbe sous le rang à la nouaison	Qualitatif
2	HSRV	importance de l'herbe sous le rang à la véraison	Qualitatif
2	MOD	mode de conduite (hauteur et forme de palissage)	Qualitatif
2	NRO	nombre de rognages	Quantitatif
2	PGP	Poids grappe prélevé	Quantitatif
2	SECH	sensibilité à la sécheresse	Qualitatif
2	SECVC	Surface Externe du Couvert Végétal	Quantitatif
2	TAIL	mode de conduite (longueur et attachage)	Qualitatif
2	VIG	vigueur selon viticulteur	Qualitatif
2	VIGO	vigueur observée	Qualitatif
3	FOL	apports de fertilisants foliaires	Qualitatif
3	K	Potassium	Quantitatif
3	K_fum	fumure potassique minérale (unités K2O/ha)	Quantitatif
3	KF	fumure potassique minérale de fond (unités K2O/ha)	Quantitatif
3	KO	fumure potassique organique (unités K2O/ha)	Quantitatif
3	KOF	fumure potassique organique de fond (unités K2O/ha)	Quantitatif
3	N	Azote	Quantitatif
3	N_fum	fumure azotée minérale (unités N/ha)	Quantitatif
3	NF	fumure azotée minérale de fond (unités N/ha)	Quantitatif
3	NO	fumure azotée organique (unités N/ha)	Quantitatif
3	NOF	fumure azotée organique de fond (unités N/ha)	Quantitatif

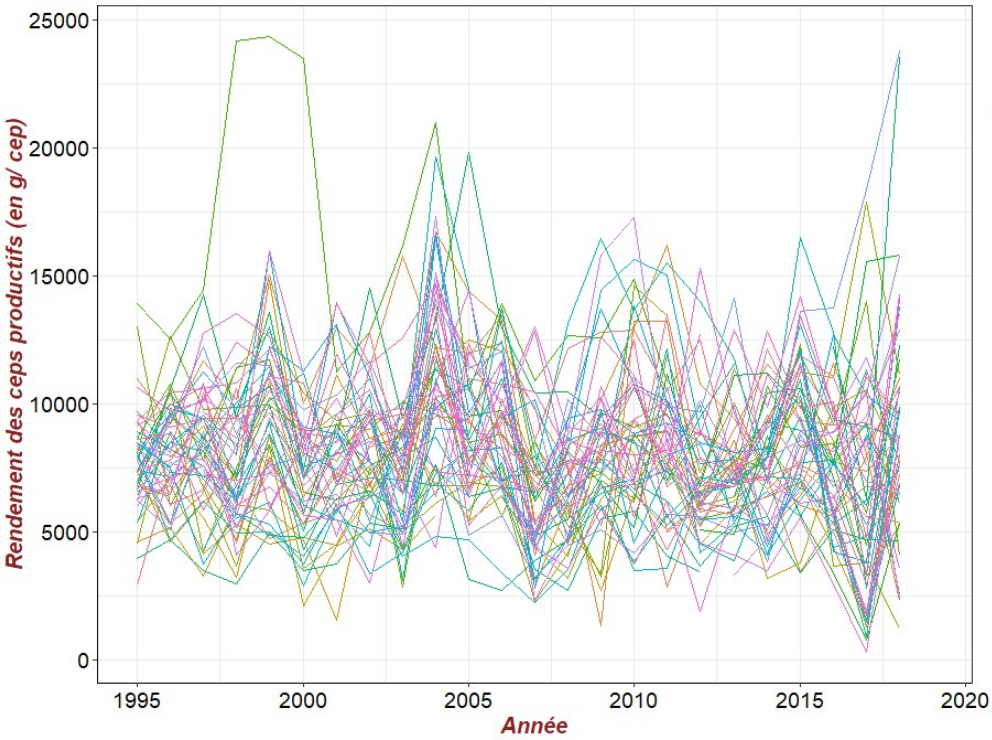
3	ORG	apports organiques	Qualitatif
3	ORGF	apports organiques de fond	Qualitatif
3	P	Phosphore	Quantitatif
3	P_fum	fumure phosphatée minérale (unités P2O5/ha)	Quantitatif
3	PF	fumure phosphatée minérale de fond (unités P2O5/ha)	Quantitatif
3	PO	fumure phosphatée organique (unités P2O5/ha)	Quantitatif
3	POF	fumure phosphatée organique de fond (unités P2O5/ha)	Quantitatif
4	AGE	Age de la parcelle au moment de la mesure	Quantitatif
4	densO	densité de plantation observée (nb ceps/ha)	Quantitatif
4	EXPO	exposition de la parcelle	Qualitatif
4	PG	porte greffe	Qualitatif
4	SITU	situation topographique de la parcelle	Qualitatif
5	BOUI	volume de bouillie /ha en pleine végétation	Quantitatif
5	CHAF	chaulage avant plantation	Qualitatif
5	CHLO	symptômes de chlorose observés par le viticulteur	Qualitatif
5	DES	désinfection du sol avant plantation	Qualitatif
5	HUMI	sensibilité à l'humidité	Qualitatif
5	Nb_MALADE_ESCA	Nombre de malade de l'ESCA	Quantitatif
5	Nb_MALADE_EUTYC	Nombre de malade de l'Eutypiose	Quantitatif
5	SCHL	sensibilité à la chlorose	Qualitatif
5	SPOU	sensibilité à la pourriture	Qualitatif
5	TB	nombre de traitements anti-botrytis	Quantitatif
5	THIV	traitement d'hiver	Qualitatif
5	TI	nombre de traitements insecticides ou acaricides	Quantitatif
5	TIBS	nombre de traitements IBS	Quantitatif
5	TM	nombre de traitements anti-mildiou	Quantitatif
5	TOI	nombre de traitements anti-oïdium	Quantitatif
5	TP	nombre de poudrages de soufre	Quantitatif
5	TQOI	nombre de traitements QOI	Quantitatif

5	TV	nombre total de traitements en végétation	Quantitatif
6	DJsup25	cumul du surplus de températures moyennes journalières supérieures à 25°C	Quantitatif
6	Froid_cum	cumul de la différence de température entre la température moyenne et 5°C pour les jours ayants une température inférieure à 5°C	Quantitatif
6	GEL	dégâts de gel estimés par le viticulteur	Qualitatif
6	Huglin	indice héliothermique de Huglin	Quantitatif
6	Jchauds	nombre de jours dans l'année où la température a dépassé 30°C	Quantitatif
6	Jgel	nombre de jours où la température est inférieure à 0°C pour la période allant de septembre de l'année n-1 à août de l'année n	Quantitatif
6	JTchauds	nombre de jours dans l'année où la température a dépassé 35°C	Quantitatif
6	Pluie_annee	cumul de précipitations sur l'année calendaire	Quantitatif
6	Pluie_saison	cumul de précipitations sur la saison (du 10/04 au 30/09)	Quantitatif
6	SGEL	sensibilité au gel	Qualitatif
7	ESUY	ressuyage du sol après pluies	Qualitatif
7	PREC	précédent cultural	Qualitatif
7	SOL	type de sol	Qualitatif
	AN	Année	
	CODE	Code d'identification de la parcelle	Qualitatif

Annexe 4 : Rendement des différentes parcelles au cours du temps



<u>Figure 1 :</u> Rendement à la placette selon l'année
<u>Légende :</u> Chaque courbe représente le rendement à la placette d'une parcelle au cours du temps



<u>Figure 2 :</u> Rendement des ceps productifs selon l'année
<u>Légende :</u> Chaque courbe représente le rendement des ceps productifs d'une parcelle au cours du temps

Annexe 5 : Test de validité des hypothèses pour la régression linéaire du rendement placette calculé

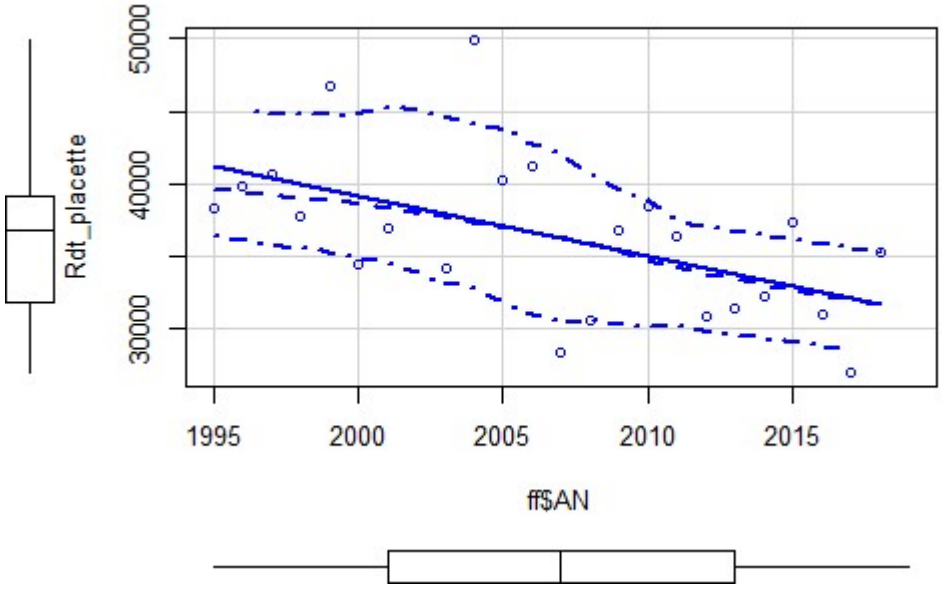


Figure 1 : Evaluation de l'hypothèse de linéarité
Légende : Rdt_placette : rendement calculé moyen à la placette; ff$AN : année; trait plein bleu : droite de régression linéaire définie par la méthode des moindres carrées; trait central en pointillée : courbe de régression locale de type lowess; traits extérieurs en pointillée : intervalle de confiance de la courbe lowess

Dans la figure, la droite de régression est dans l'intervalle de confiance donc l'hypothèse de linéarité est acceptable.

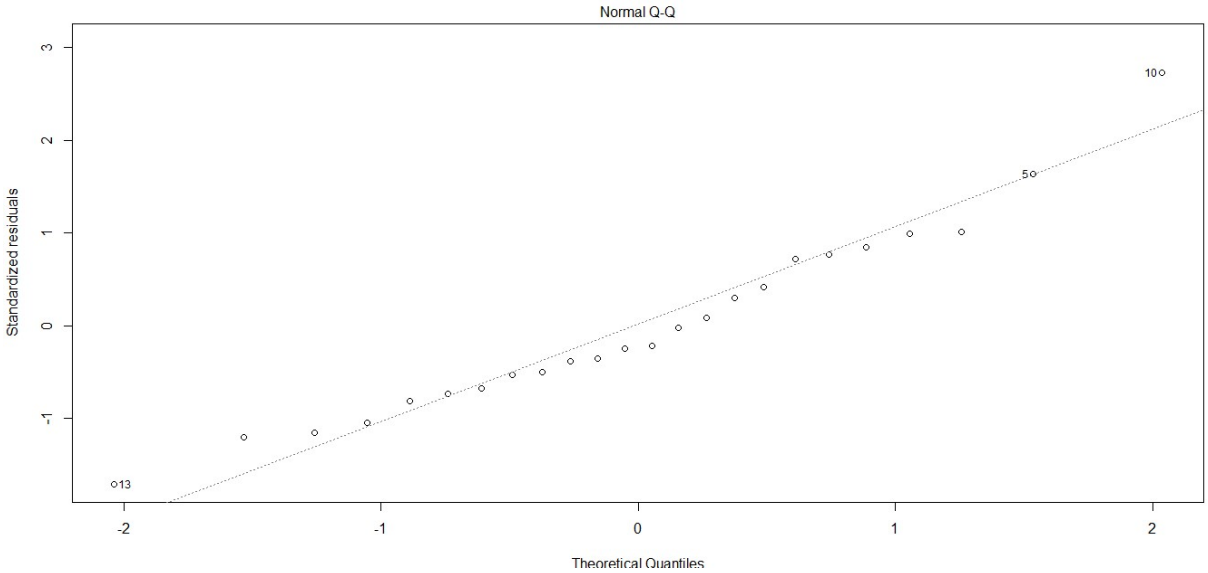


Figure 2 : Evaluation de l'hypothèse de normalité des résidus

```
shapiro.test(residuals(prest.lm1))

        Shapiro-Wilk normality test

data:  residuals(prest.lm1)
W = 0.95797, p-value = 0.399
```

Figure 3 : Test de Shapiro-Wilk

Les points sont alignés donc on peut valider l'hypothèse de normalité. Un test de Shapiro-Wilk à également été fait pour confirmer l'évaluation visuelle. On considère un alpha de 0,05. On a obtenu une p-value supérieure, donc on ne rejette pas l'hypothèse de normalité.

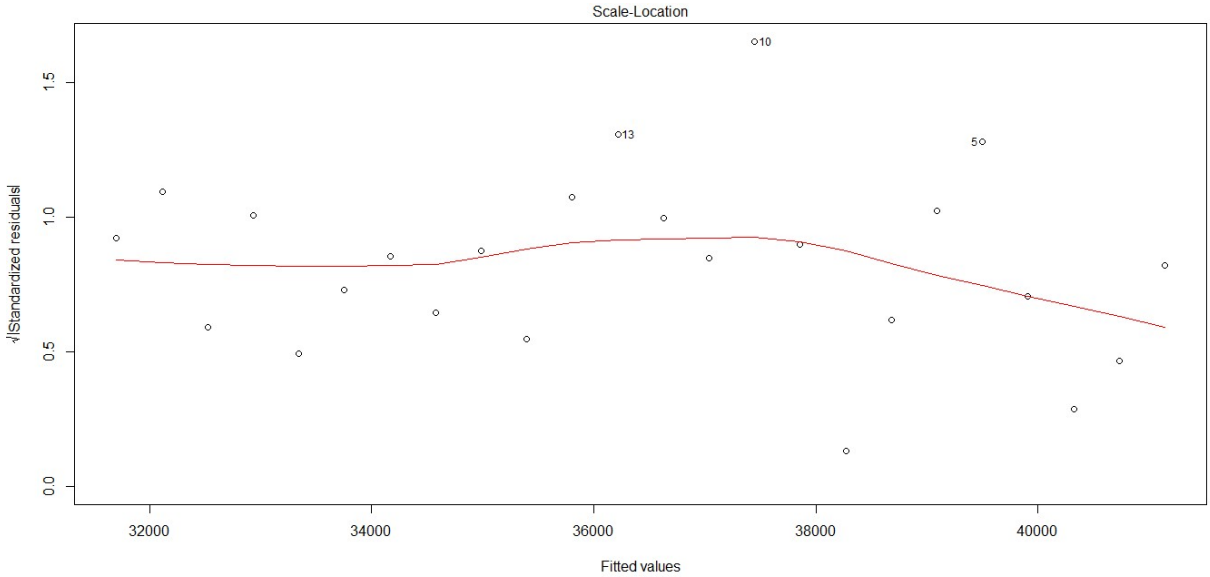


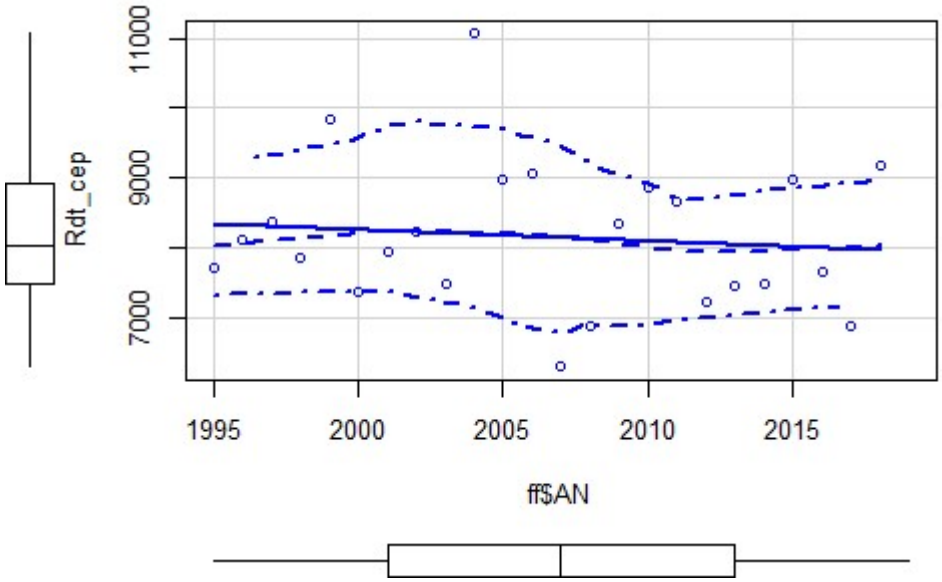
Figure 4 : Evaluation de l'hypothèse d'homogénéité des résidus
Légende : Courbe rouge : courbe de régression locale

```
ncvTest(prest.lm1)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.09690113, Df = 1, p = 0.75558
```

Figure 5 : Test de Breush-Pagan

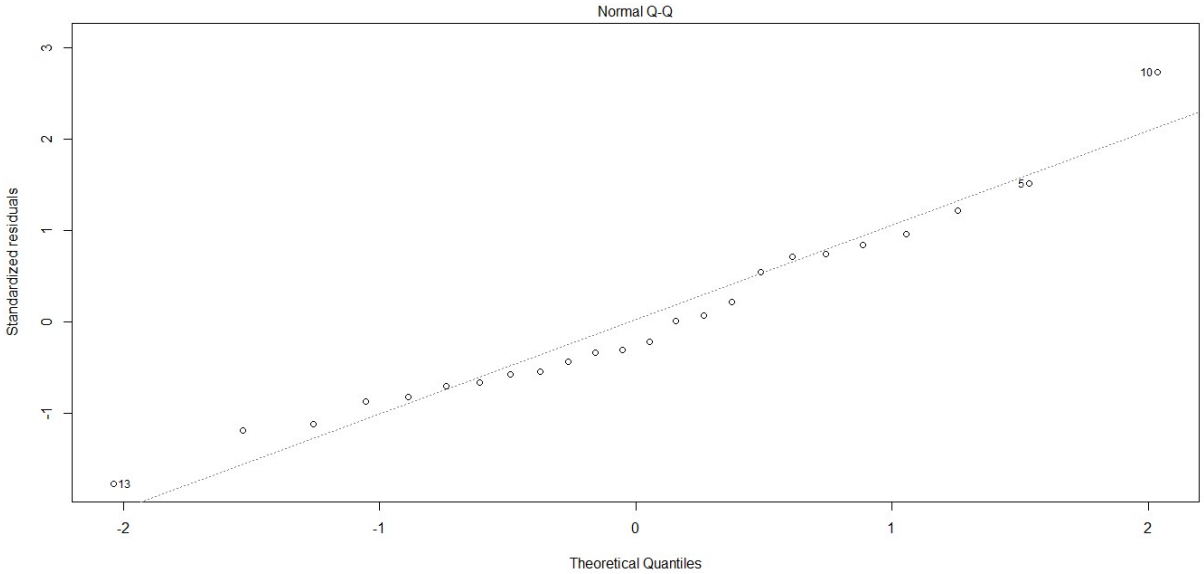
Les résidus ont tendance à être répartis de manière homogène tout le long du gradient des valeurs d'années. Ainsi, on ne peut pas rejeter l'hypothèse d'homogénéité des résidus. Pour confirmer ce résultat on a réalisé un test de Breush-Pagan. L'hypothèse d'homogénéité est rejetée si la p-value est inférieure à 0,05, et nous avons une valeur de supérieure donc l'hypothèse d'homogénéité n'est pas rejeté.

Annexe 6 : Test de validité des hypothèses pour la régression linéaire du rendement des ceps productifs



<u>Figure 1 :</u> Evaluation de l'hypothèse de linéarité
<u>Légende :</u> Rdt_cep : rendement moyen des ceps productifs; ff$AN : année; trait plein bleu : droite de régression linéaire définie par la méthode des moindres carrées; trait central en pointillée : courbe de régression locale de type lowess; traits extérieurs en pointillée : intervalle de confiance de la courbe lowess

Dans la figure, la droite de régression est dans l'intervalle de confiance donc l'hypothèse de linéarité est donc acceptable.



<u>Figure 2 :</u> Evaluation de l'hypothèse de normalité des résidus

```
shapiro.test(residuals(prest.lm1))

        Shapiro-Wilk normality test

data:  residuals(prest.lm1)
W = 0.95823, p-value = 0.4038
```

Figure 3 : Test de Shapiro-Wilk

Les points sont alignés donc on peut valider l'hypothèse de normalité. Un test de Shapiro-Wilk à également été fait pour confirmer l'évaluation visuelle. On considère un alpha de 0,05. On a obtenue une p-value supérieure, donc on ne rejette pas l'hypothèse de normalité.

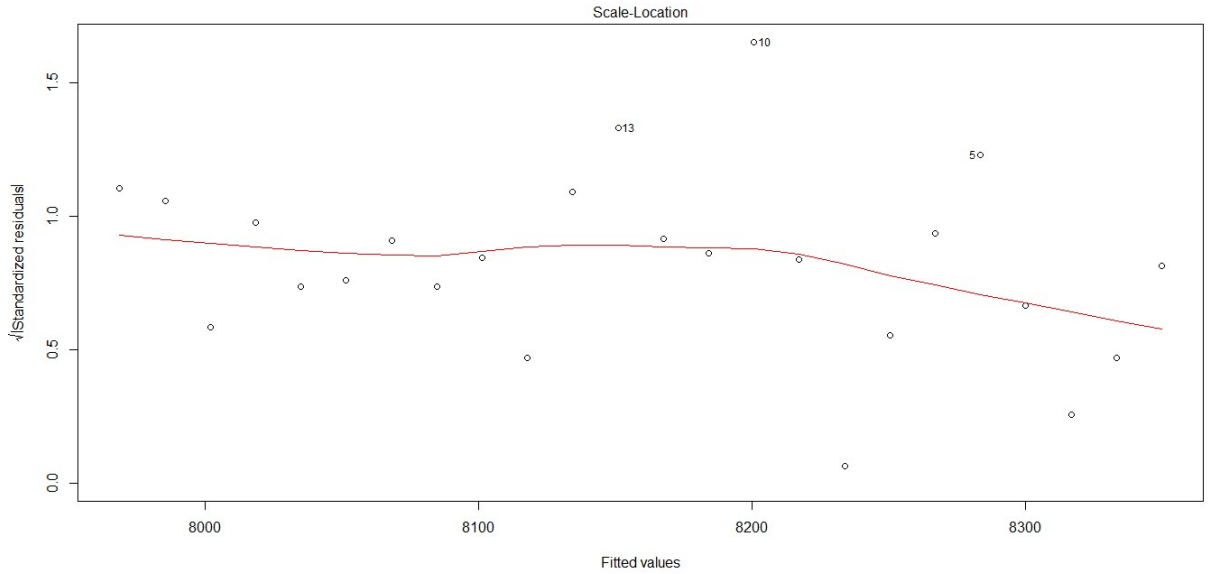


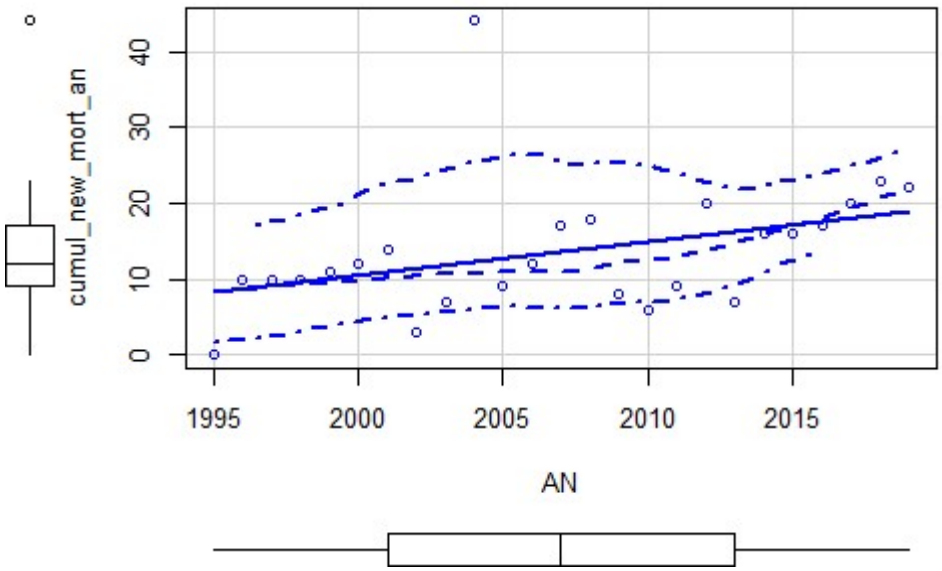
Figure 4 : Evaluation de l'hypothèse d'homogénéité des résidus
Légende : Courbe rouge : courbe de régression locale

```
ncvTest(prest.lm1)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 7.594996e-05, Df = 1, p = 0.99305
```

Figure 5 : Test de Breush-Pagan

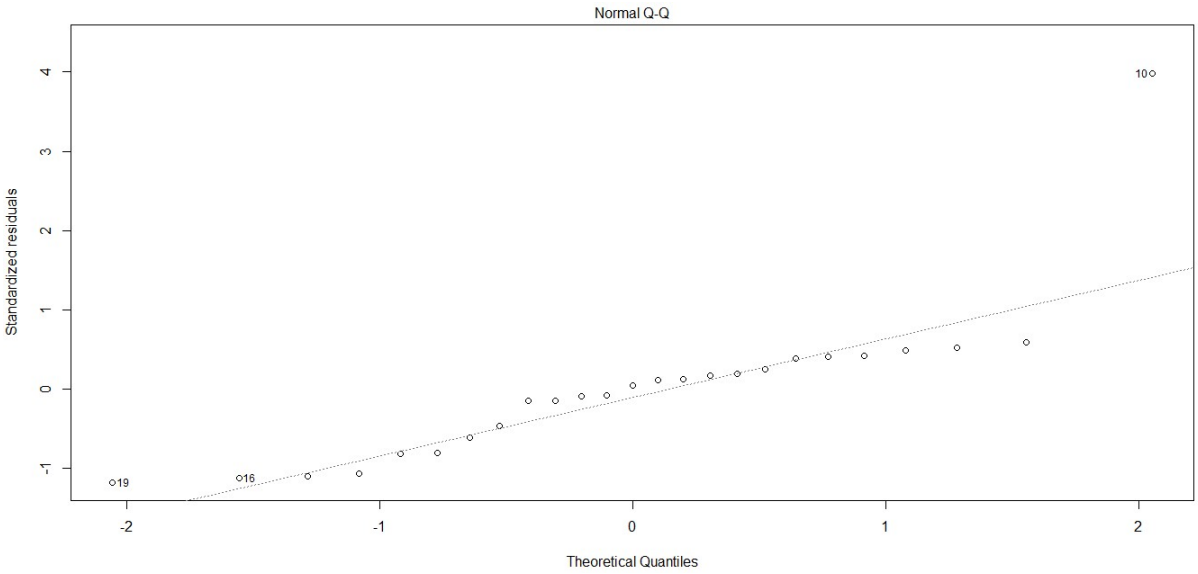
Les résidus ont tendance à être répartis de manière homogène tout le long du gradient des valeurs d'années. Ainsi, on ne peut pas rejeter l'hypothèse d'homogénéité des résidus. Pour confirmer ce résultat on a réalisé un test de Breush-Pagan. L'hypothèse d'homogénéité est rejetée si la p-value est inférieure à 0,05, et nous avons une valeur de supérieure donc l'hypothèse d'homogénéité n'est pas rejeté.

Annexe 7 : Test de validité des hypothèses pour la régression linéaire du taux de mortalité annuel



<u>Figure 1 :</u> Evaluation de l'hypothèse de linéarité
<u>Légende :</u> cumul_new_mort_an : taux de mort; AN : année; trait plein bleu : droite de régression linéaire définie par la méthode des moindres carrées; trait central en pointillée : courbe de régression locale de type lowess; traits extérieurs en pointillée : intervalle de confiance de la courbe lowess

Dans la figure, la droite de régression est dans l'intervalle de confiance donc l'hypothèse de linéarité est donc acceptable.



<u>Figure 2 :</u> Evaluation de l'hypothèse de normalité des résidus

```
shapiro.test(residuals(prest.lm1))

        Shapiro-Wilk normality test

data:  residuals(prest.lm1)
W = 0.73229, p-value = 2.046e-05
```

Figure 3 : Test de Shapiro-Wilk

Les points ne sont pas alignés donc on ne peut pas valider l'hypothèse de normalité. Un test de Shapiro-Wilk à également été fait pour confirmer l'évaluation visuelle. On considère un alpha de 0,05. On a obtenue une p-value inférieure, donc on rejette l'hypothèse de normalité.

Pour pouvoir tout de même regarder la relation entre le taux de mortalité et l'année on abandonne la linéarité. On réalise un test de Spearman. La p-value est largement inférieure au seuil de significativité généralement employé de 0,05. On conclut donc à la dépendance monotone significative entre les variables du taux de mortalité et l'année.

```
Spearman's rank correlation rho

data:  nb_new_mort_an$cumul_new_mort_an and nb_new_mort_an$AN
S = 1296.5, p-value = 0.01068
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.5013497
```

Figure 4 : Test de Spearman

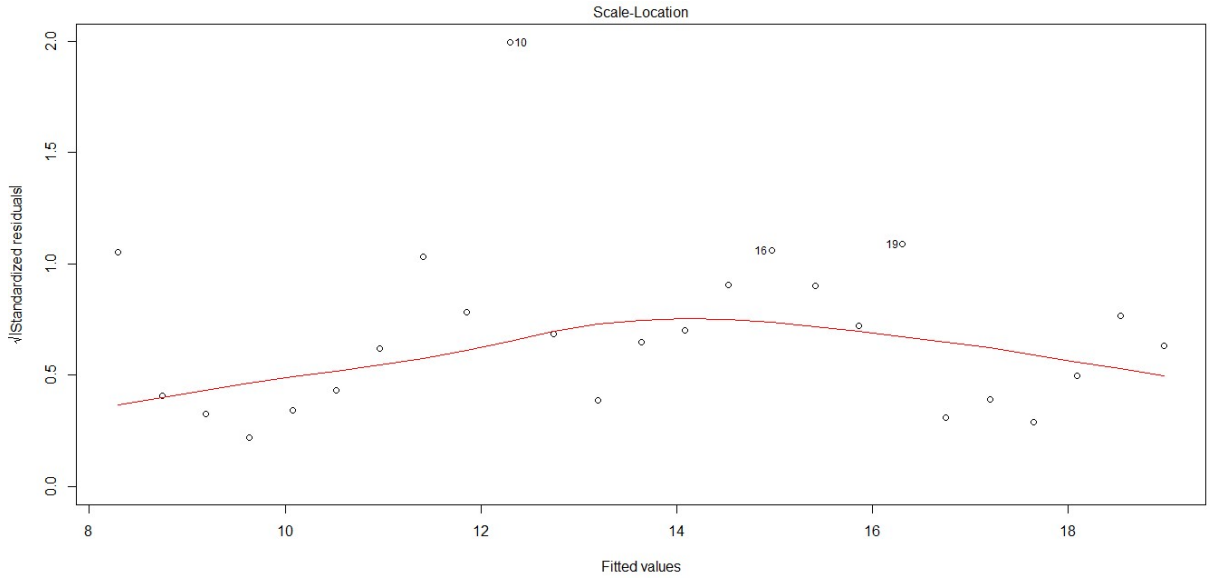


Figure 5 : Evaluation de l'hypothèse d'homogénéité des résidus
Légende : Courbe rouge : courbe de régression locale

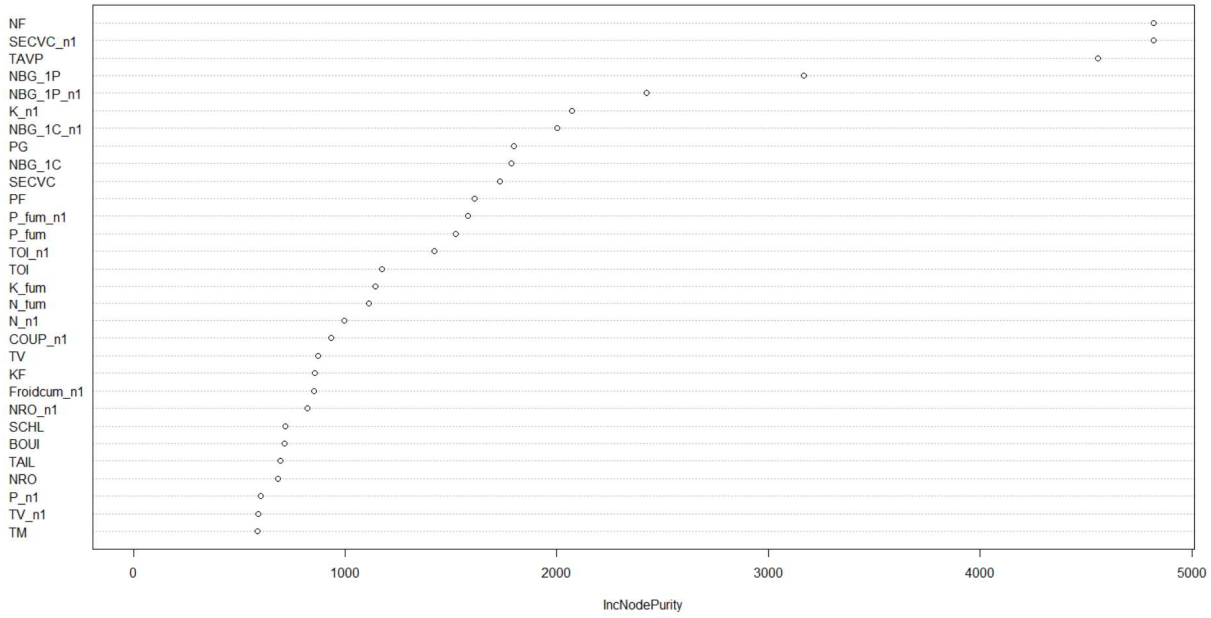
```
ncvTest(prest.lm1)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.8800391, Df = 1, p = 0.34819
```

Figure 6 : Test de Breush-Pagan

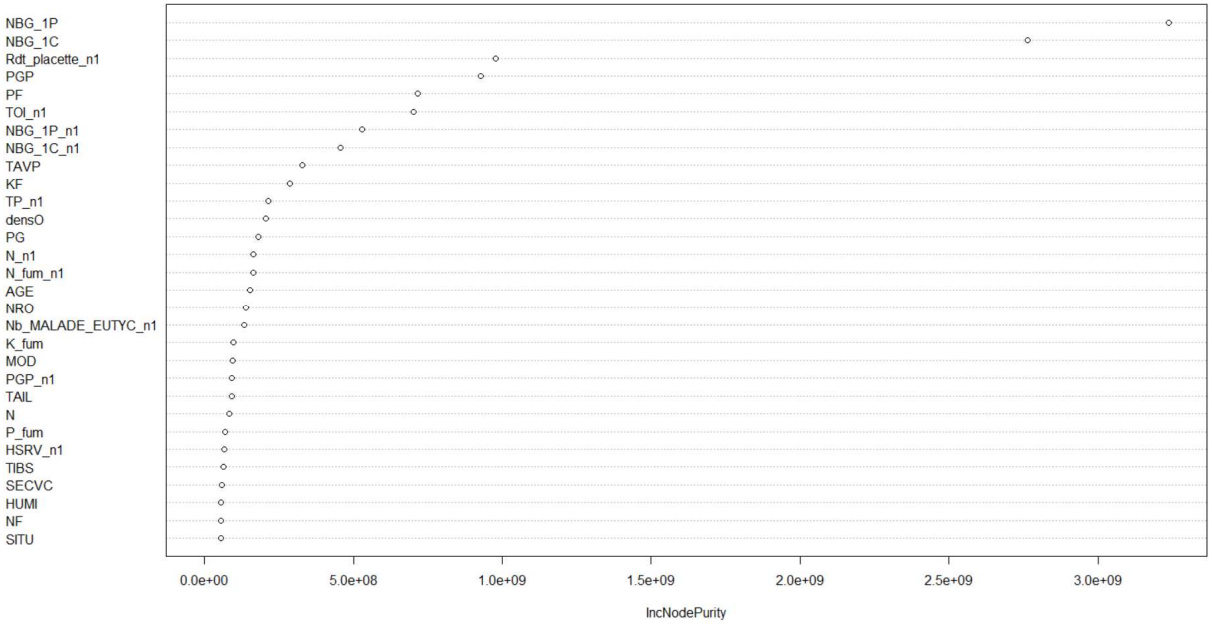
Les résidus ont tendance à être répartis de manière homogène tout le long du gradient des valeurs d'années. Ainsi, on ne peut pas rejeter l'hypothèse d'homogénéité des résidus. Pour confirmer ce résultat on a réalisé un test de Breush-Pagan. L'hypothèse d'homogénéité est rejetée si la p-value est inférieure à 0,05, et nous avons une valeur de supérieure donc l'hypothèse d'homogénéité n'est pas rejeté.

Les tests d'hypothèse sans l'année 2004 donnent les mêmes conclusions que les tests d'hypothèse de toutes les années.

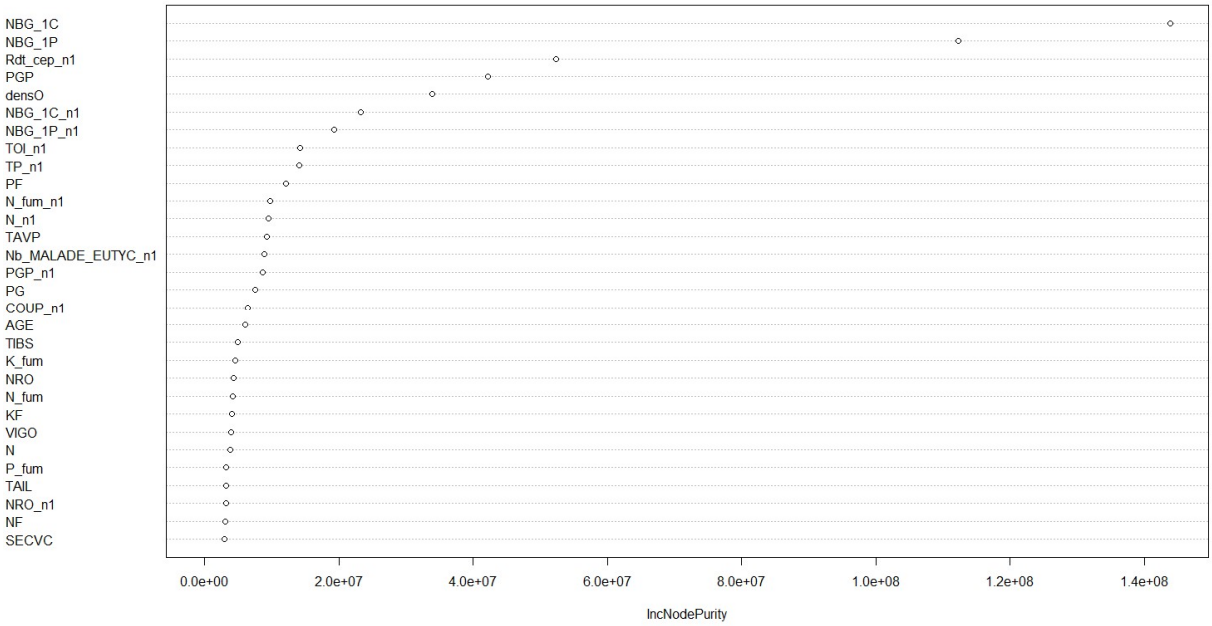
Annexe 8 : Sortie des random forest pour les indicateurs du rendement et de mortalité de 1999



A.



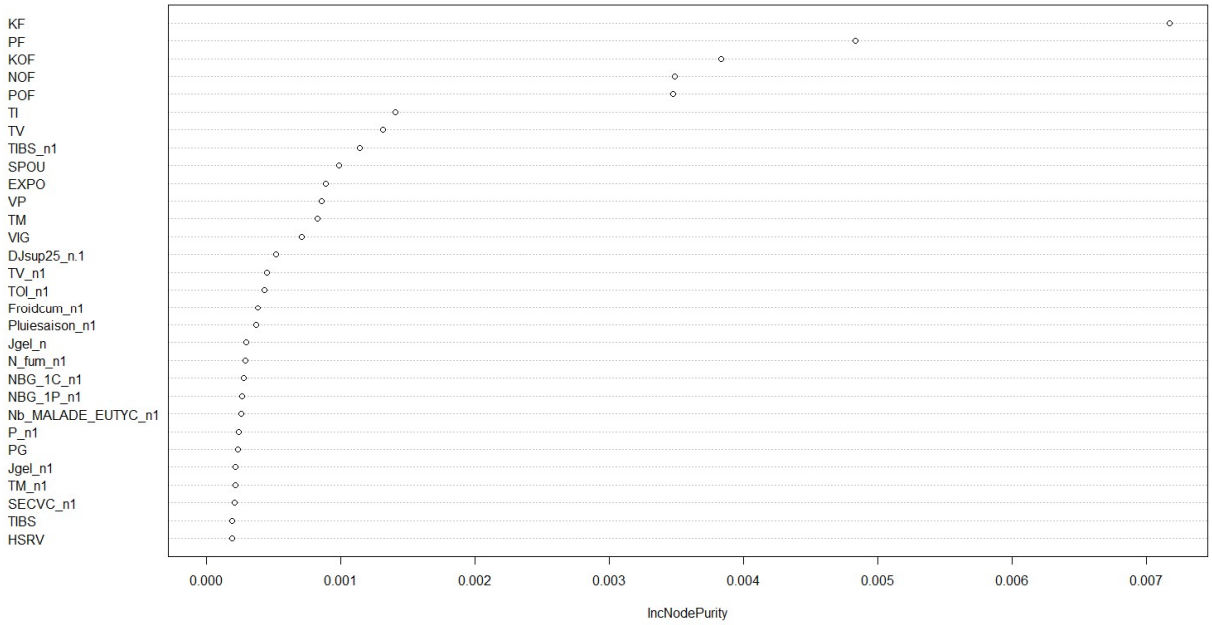
B.



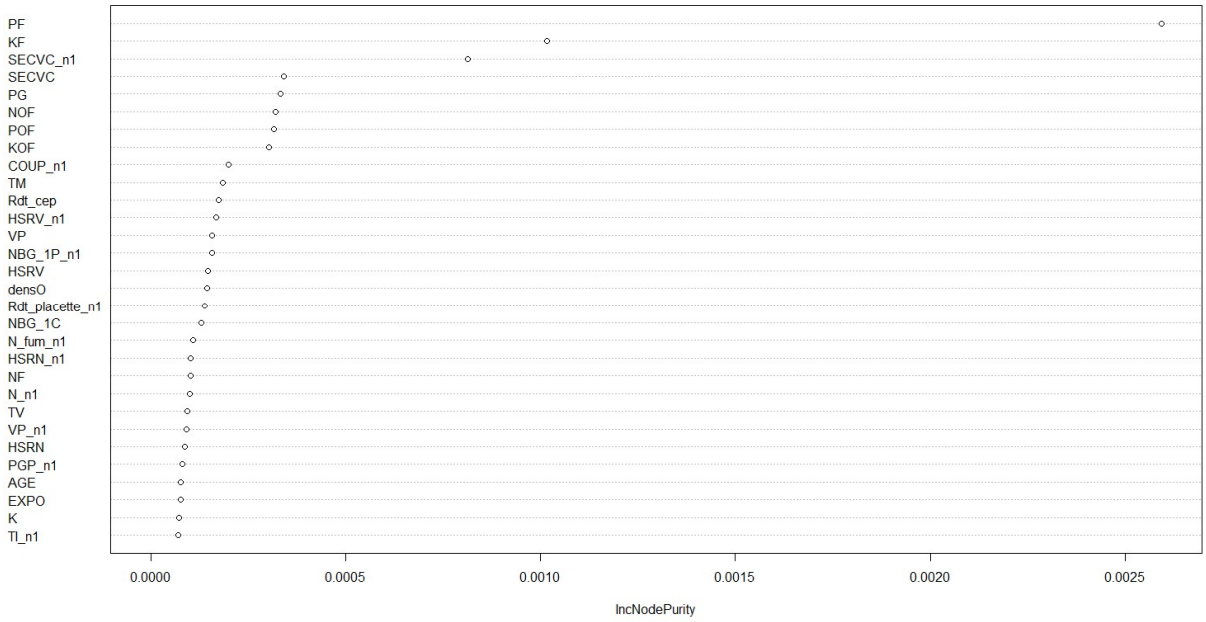
C.

<u>Figure 1 :</u> Importance des variables explicatives pour les différentes sorties des indicateurs de rendements en 1999

<u>Légende :</u> A : Rendement déclaré ; B : Rendement calculé à la placette ; C : Rendement calculé par cep productif ; Chaque point correspond à une variable explicative indiquée en ordonnée et à son implication dans l'analyse indiqué en abscisse. Les variables sont classées par poids d'implication décroissant.

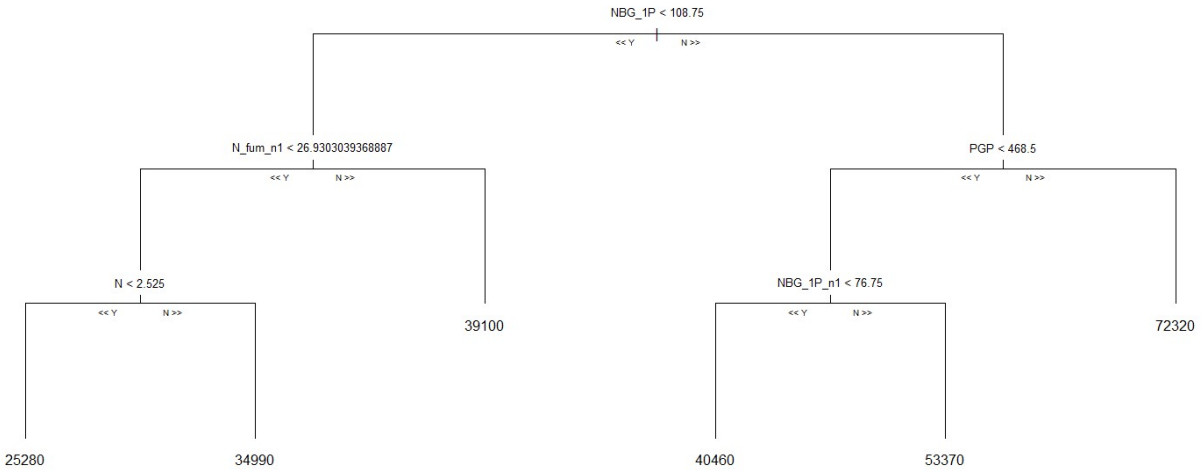


A.

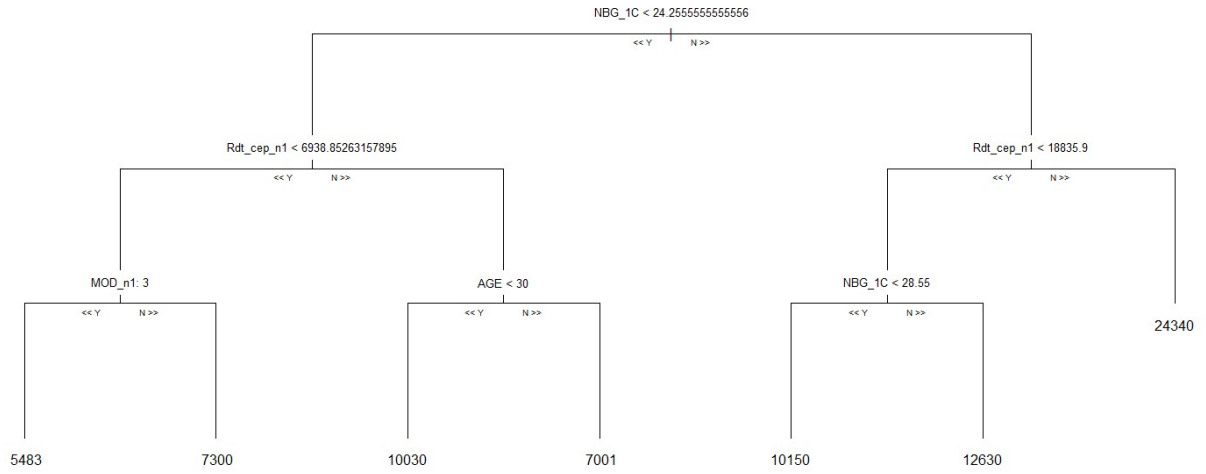


B.

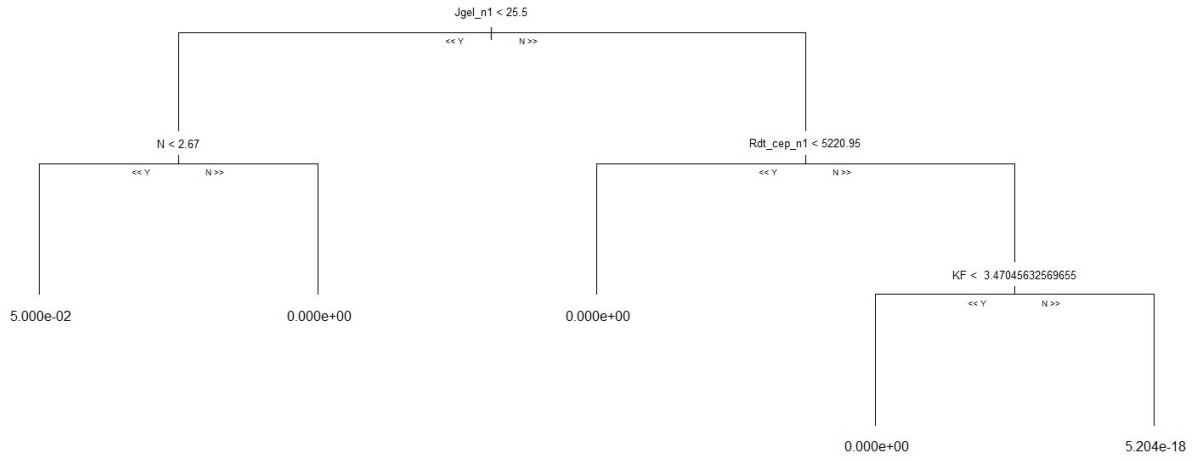
<u>Figure 2 :</u> Importance des variables explicatives pour les différentes sorties des indicateurs de mortalité en 1999
<u>Légende :</u> A : Taux de mortalité de l'année (ntree=500, mtry=38) / B : Taux de mortalité de 1995 à 1999 (ntree=500, mtry=38)



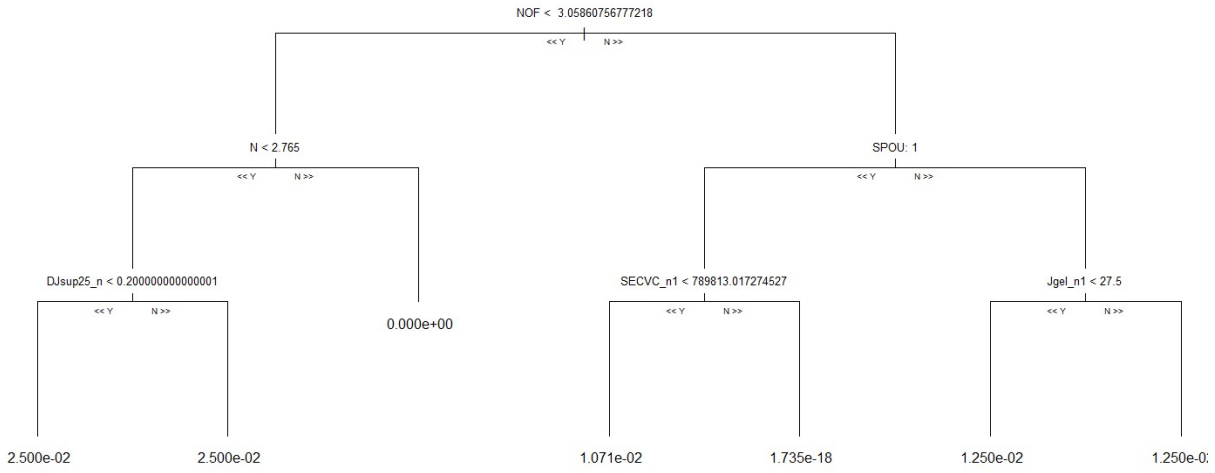
B.



C.



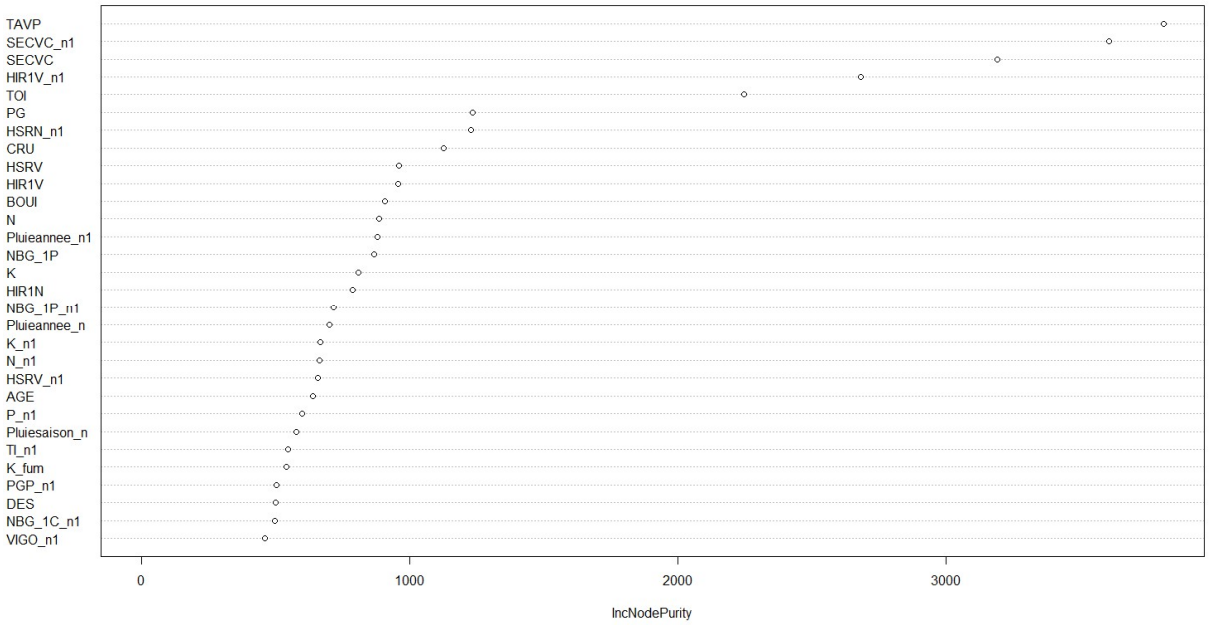
D.



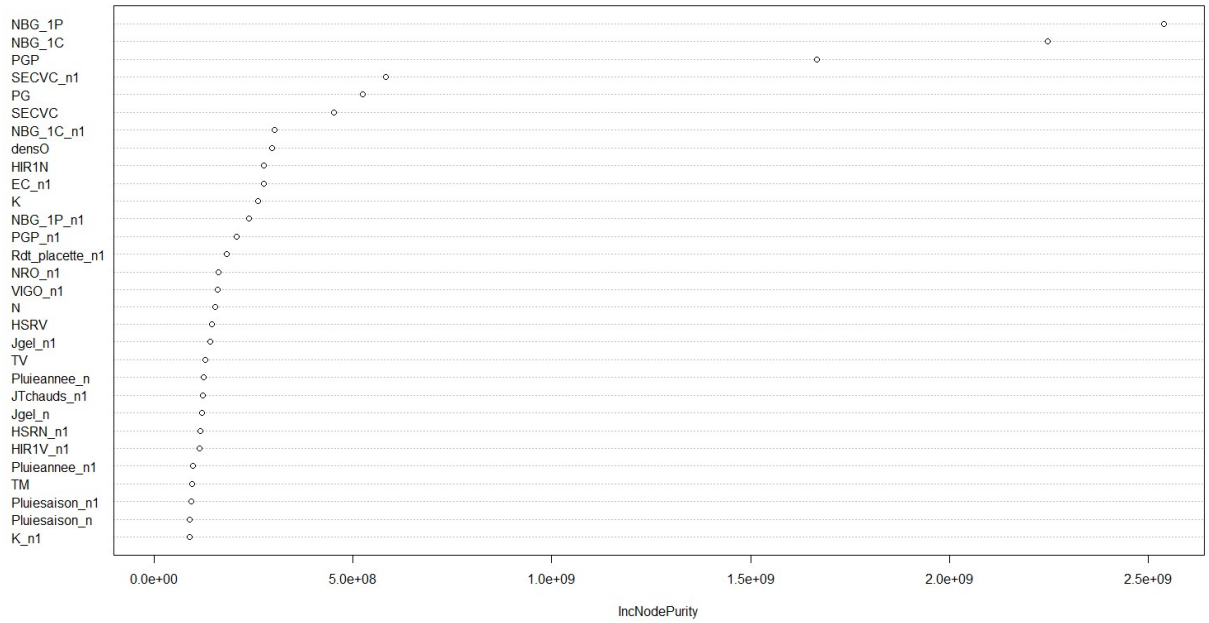
E.

<u>Figure 3 :</u> Arbre de régression représentant l'arbre moyen des random forest de 1999 des indicateurs du rendement et de mortalité
<u>Légende :</u> B : Rendement calculé à la placette ; C : Rendement calculé par cep productif ; D : Taux de mortalité annuel ; E : Taux de mortalité pour la période allant de 1995 à 1999

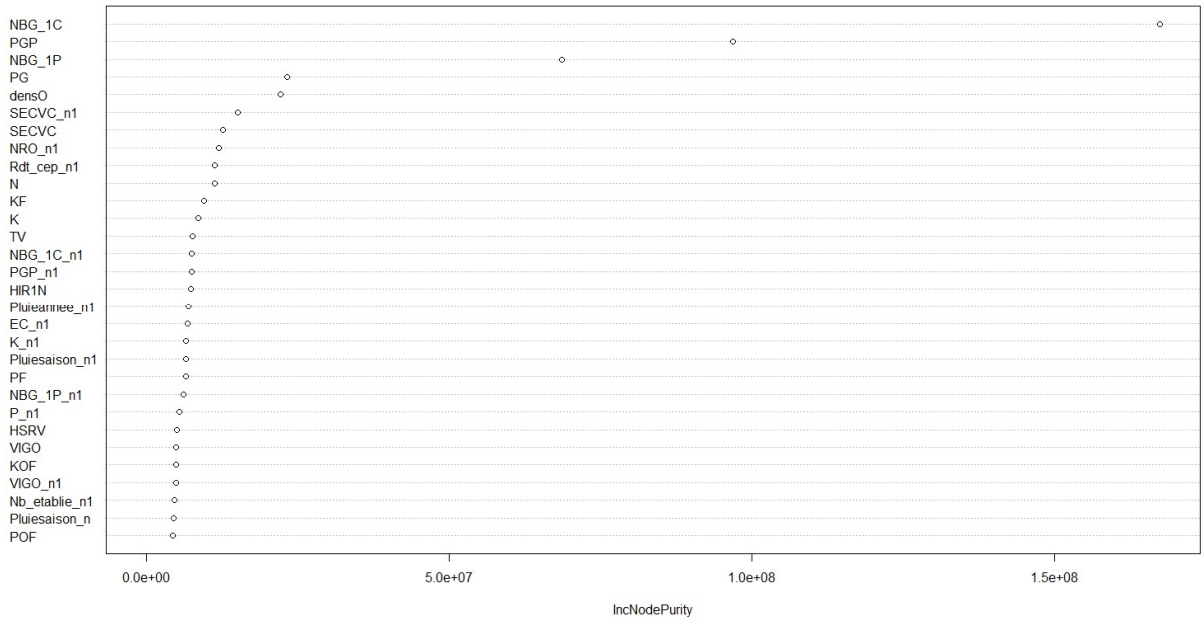
Annexe 9 : Sortie des random forest pour les indicateurs du rendement et de mortalité de 2004



A.

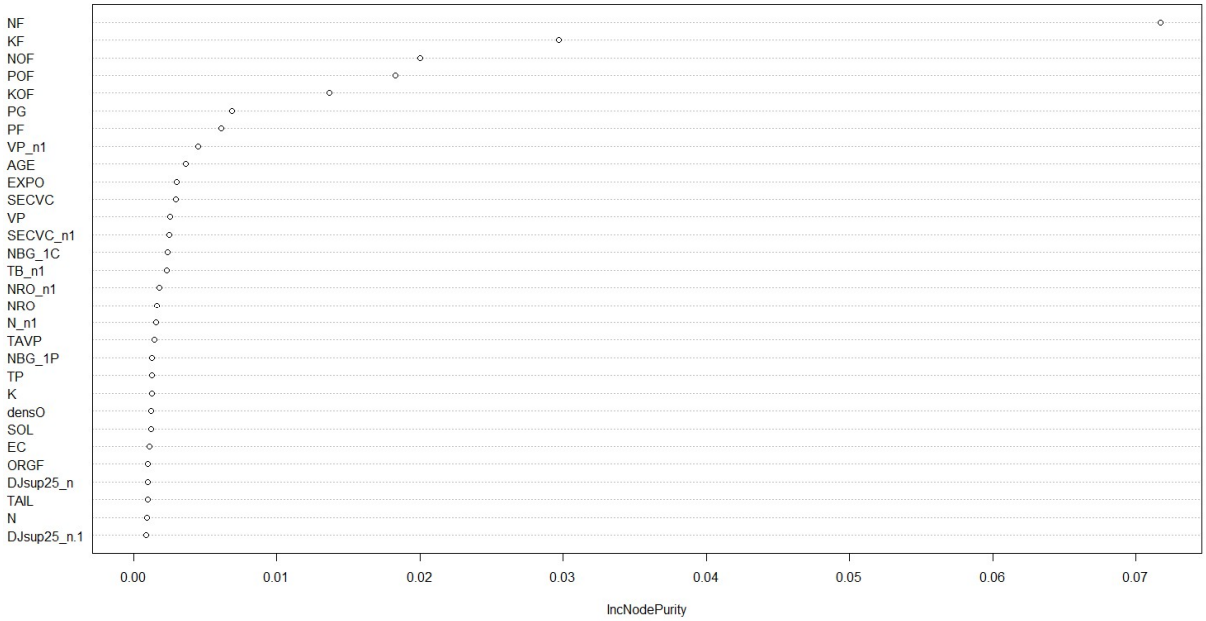


B.

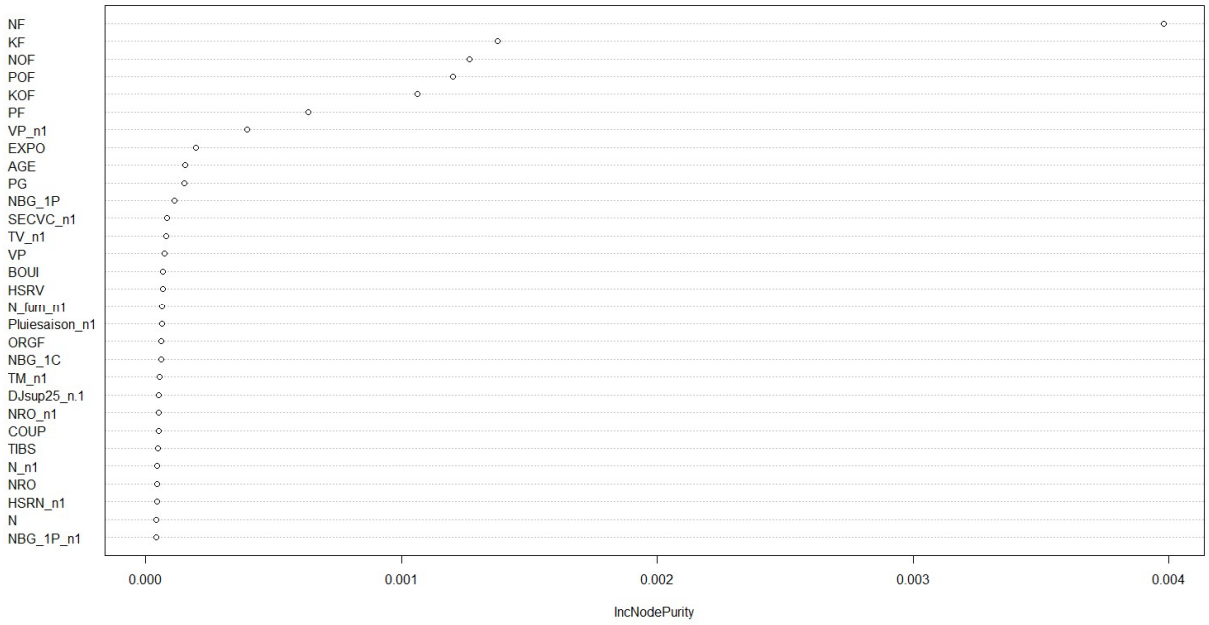


C.

<u>Figure 1 :</u> Importance des variables explicatives pour les différentes sorties des indicateurs de rendements en 2004
<u>Légende :</u> A : Rendement déclaré ; B : Rendement calculé à la placette ; C : Rendement calculé par cep productif



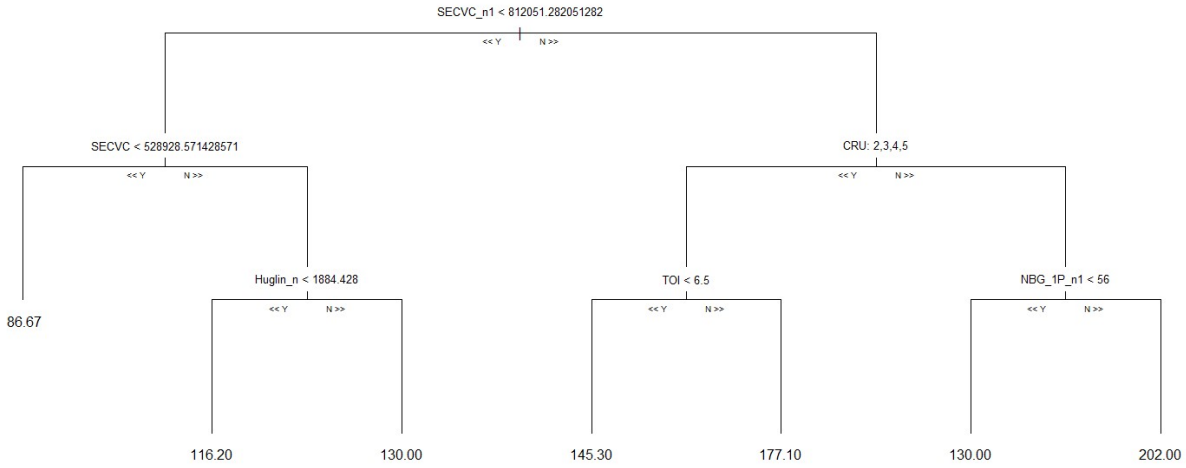
A.



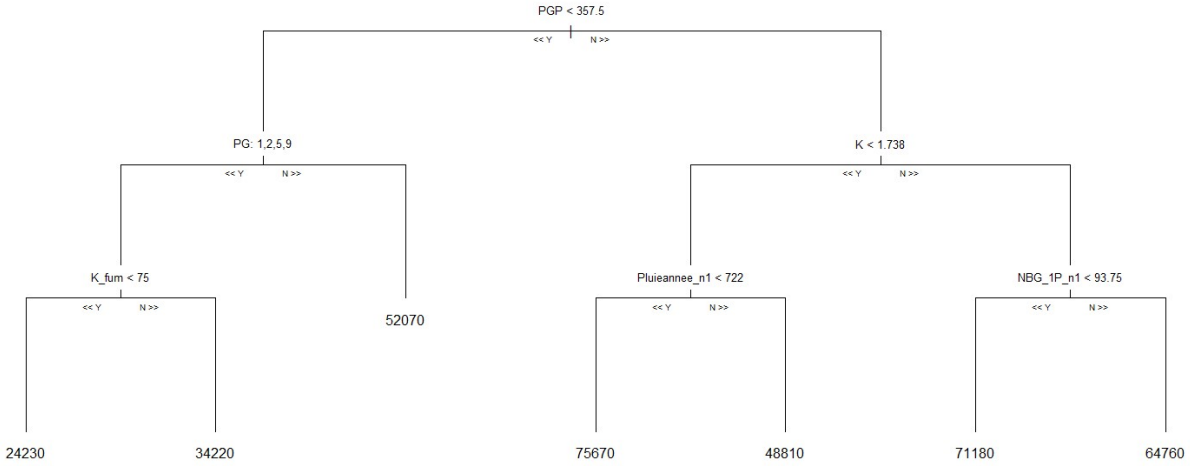
B.

<u>Figure 2 :</u> Importance des variables explicatives pour les différentes sorties des indicateurs de mortalité en 2004

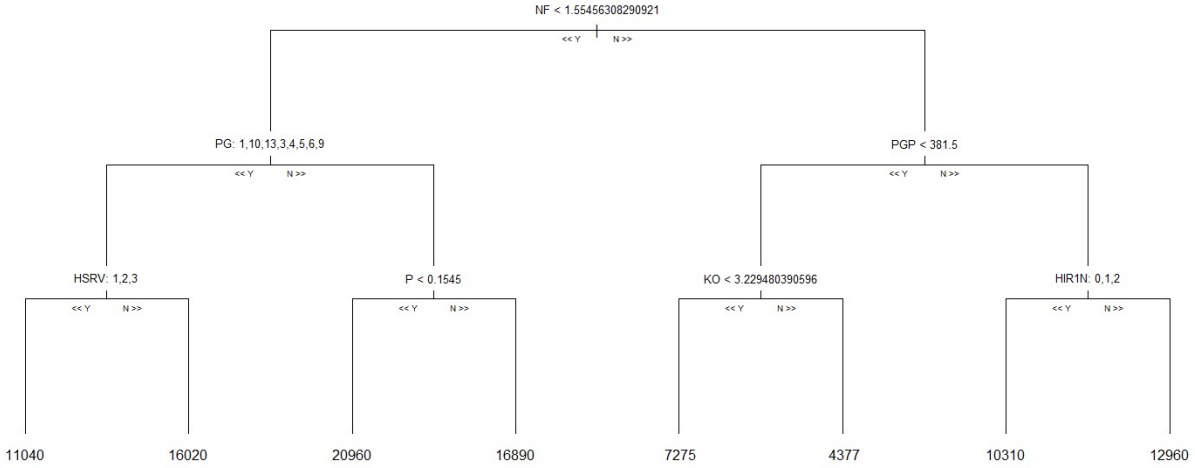
<u>Légende :</u> A : Taux de mortalité de l'année (ntree=500, mtry=40) ; B : Taux de mortalité de 1995 à 2004 (ntree=500, mtry=40)



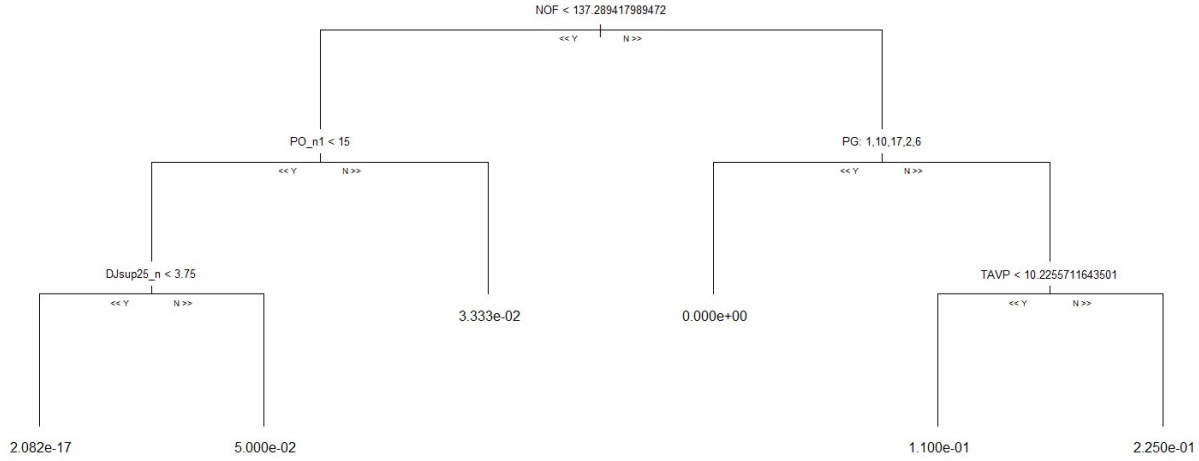
A.


PGP < 357.5
<< Y
N >>
PG: 1,2,5,9
K < 1.738
K_fum < 75
52070
Pluieannee_n1 < 722
NBG_1P_n1 < 93.75
24230
34220
75670
48810
71180
64760

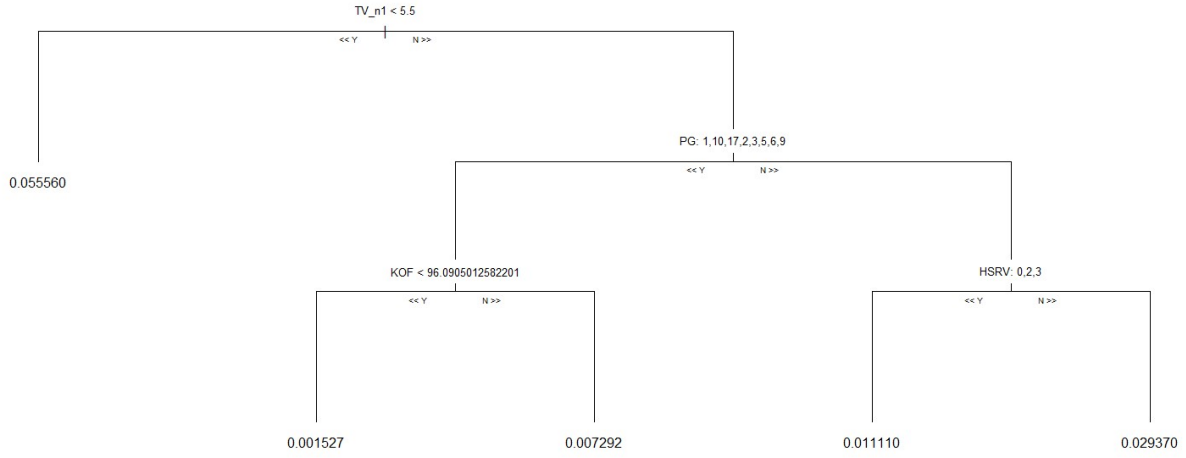
B.


NF < 1.55456308290921
<< Y
N >>
PG: 1,10,13,3,4,5,6,9
PGP < 381.5
HSRV: 1,2,3
P < 0.1545
KO < 3.229480390596
HIR1N: 0,1,2
11040
16020
20960
16890
7275
4377
10310
12960

C.


NOF < 137.289417989472
<< Y
N >>
PO_n1 < 15
PG: 1,10,17,2,6
DJsup25_n < 3.75
3.333e-02
0.000e+00
TAVP < 10.2255711643501
2.082e-17
5.000e-02
1.100e-01
2.250e-01

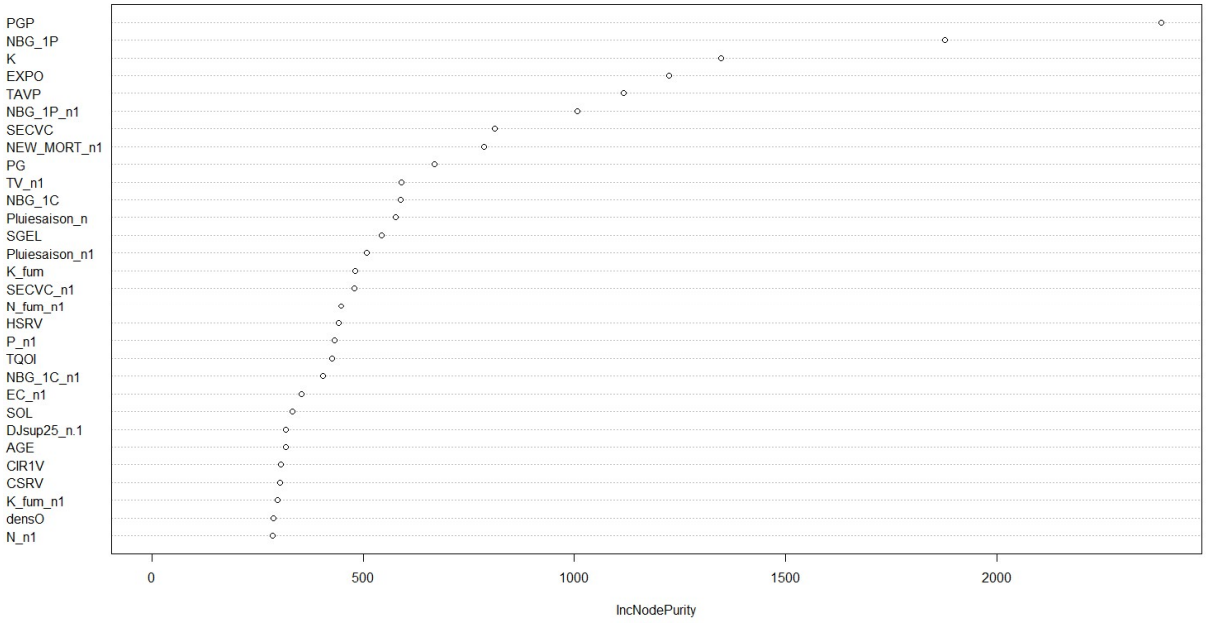
D.



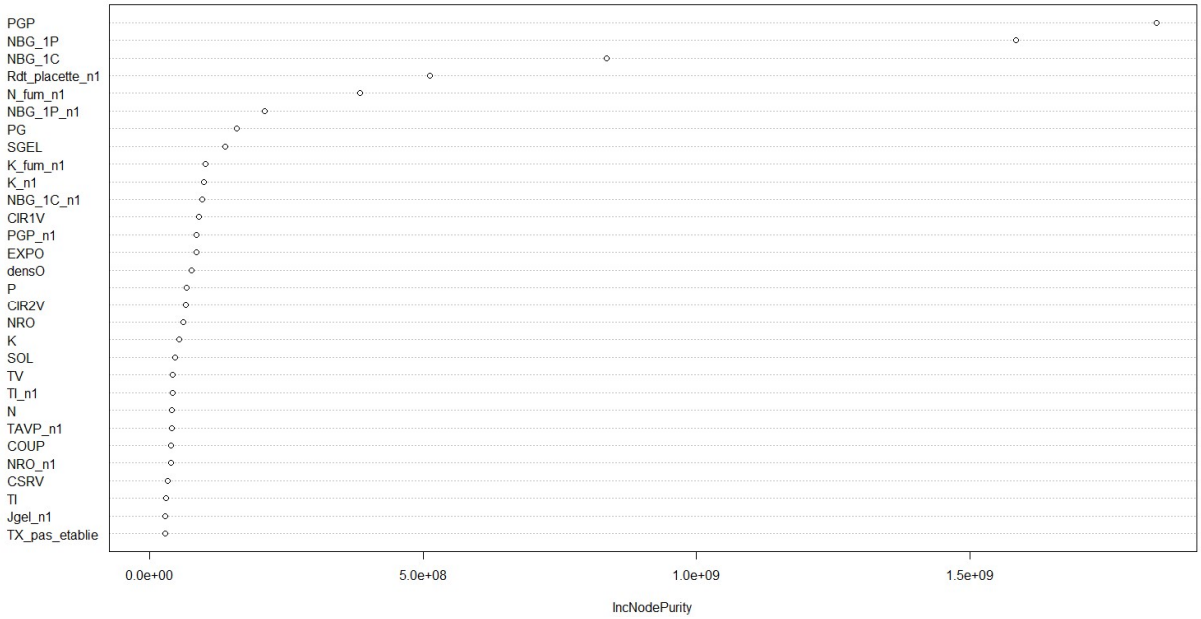
E.

<u>Figure 3 :</u> Arbre de régression représentant l'arbre moyen des random forest des indicateurs du rendement et de mortalité
<u>Légende :</u> A : Rendement déclaré; B : Rendement calculé à la placette; C : Rendement calculé par cep productif ; D : Taux de mortalité annuel ; E : Taux de mortalité pour la période allant de 1995 à 2004

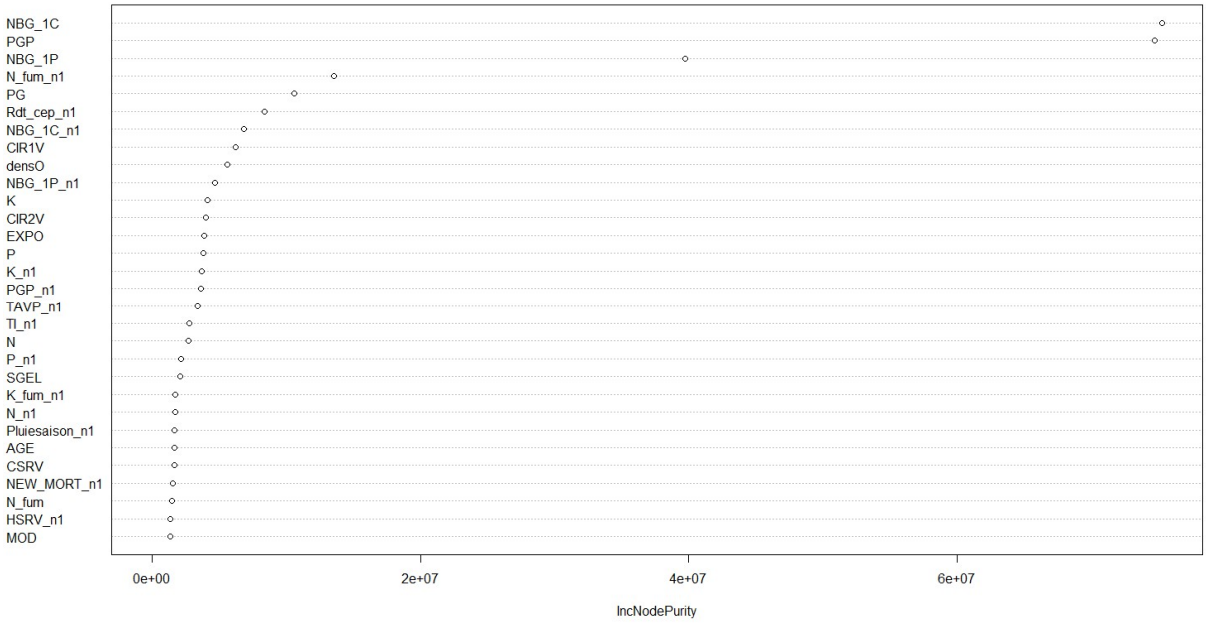
Annexe 10 : Sortie des random forest pour les indicateurs du rendement et de mortalité de 2007



A.



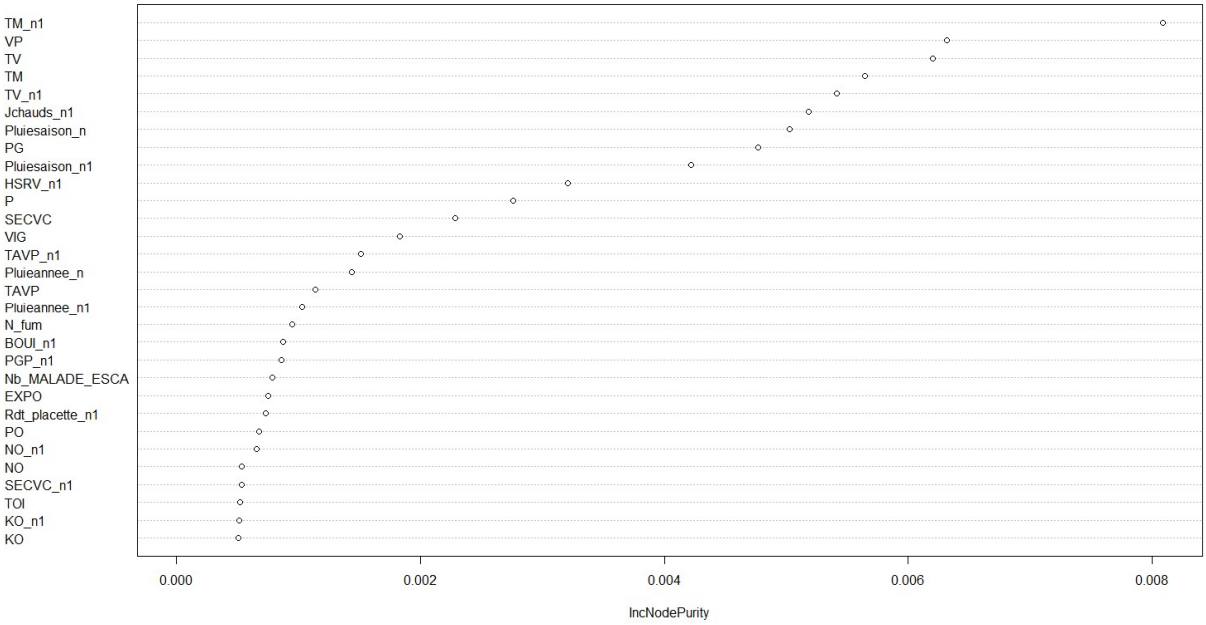
B.



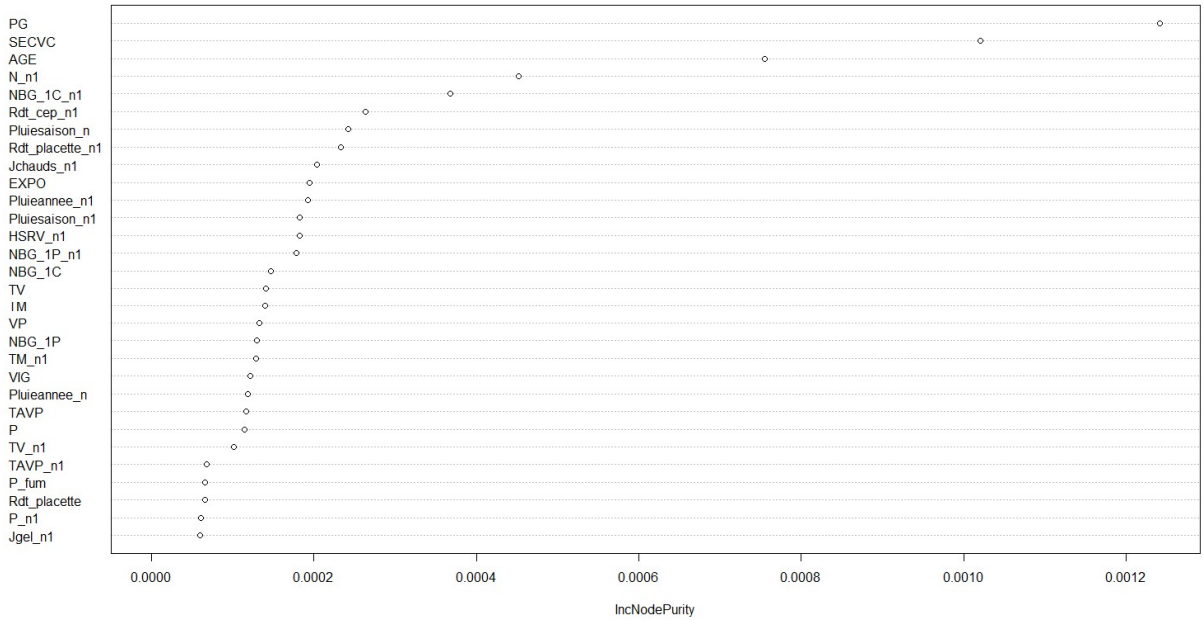
C.

<u>Figure 1 :</u> Importance des variables explicatives pour les différentes sorties des indicateurs de rendements en 2007

<u>Légende :</u> A : Rendement déclaré ; B : Rendement calculé à la placette ; C : Rendement calculé par cep productif



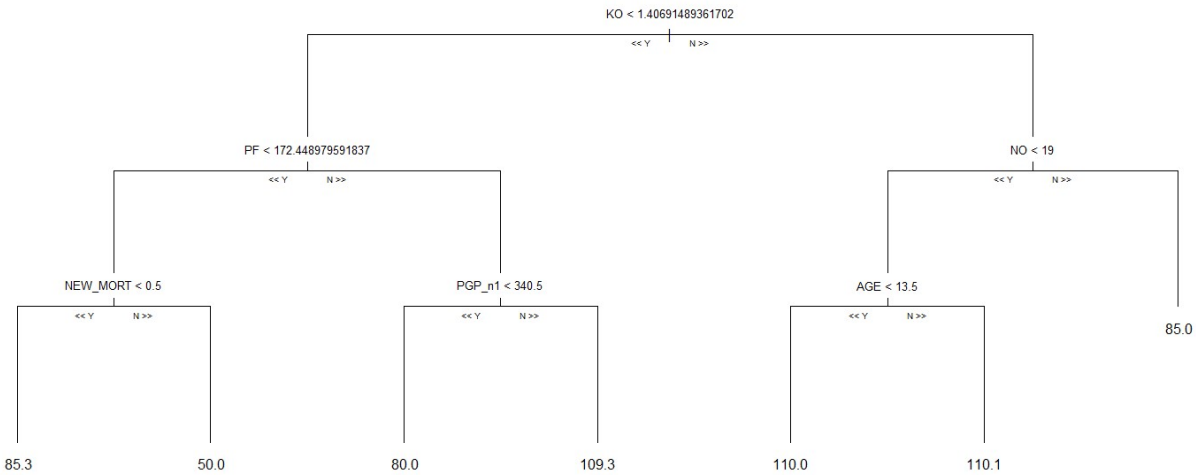
A.



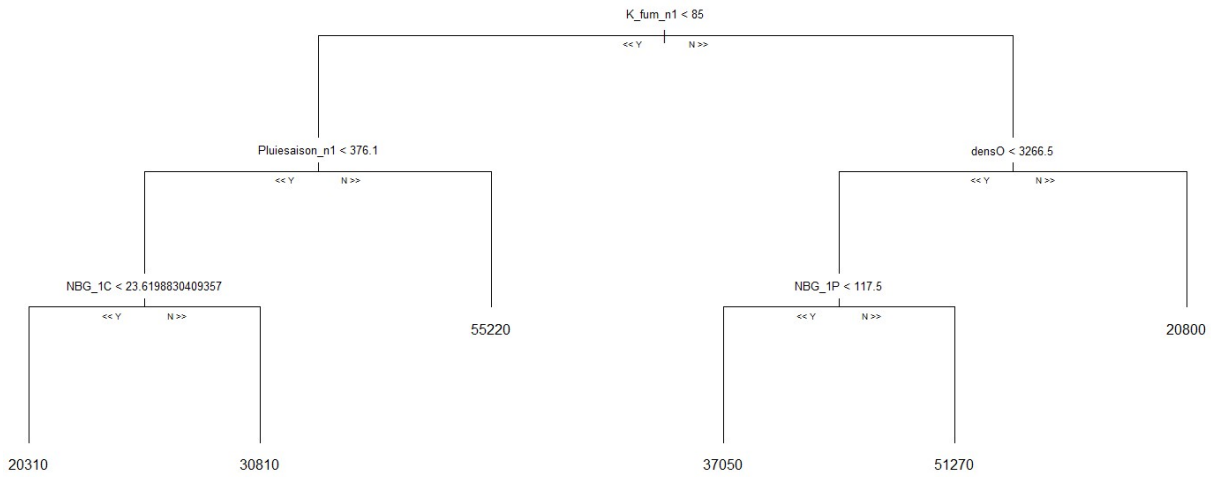
B.

<u>Figure 2 :</u> Importance des variables explicatives pour les différentes sorties des indicateurs de mortalité en 2007

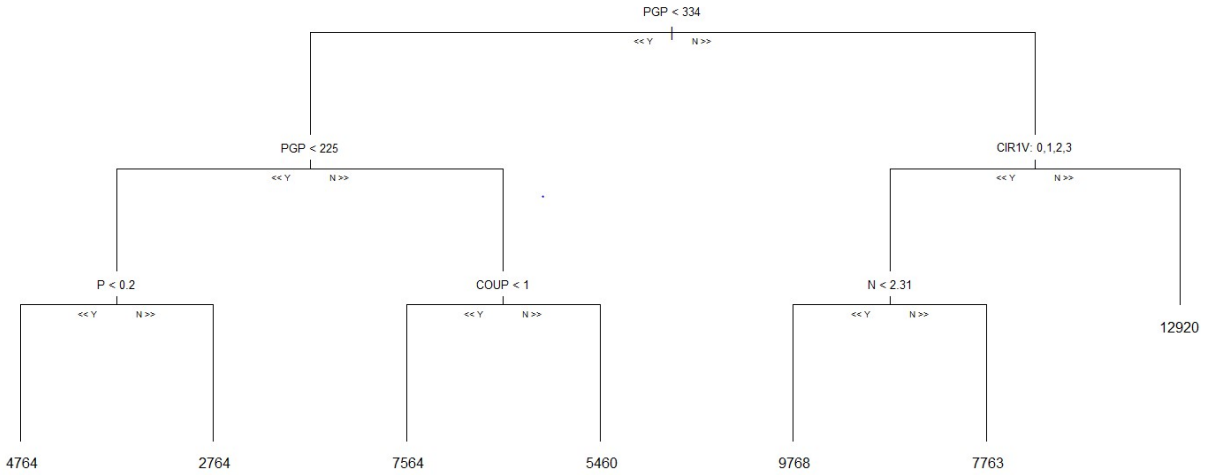
<u>Légende :</u> A : Taux de mortalité de l'année (ntree=500, mtry=41) ; B : Taux de mortalité de 1995 à 2004 (ntree=500, mtry=40)



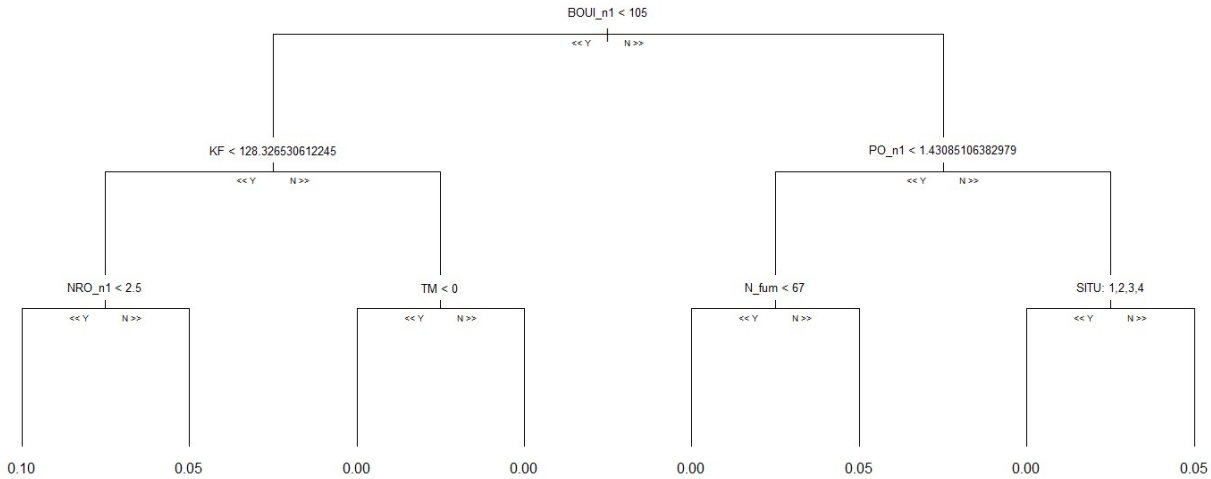
A.


K_fum_n1 < 85
<< Y
N >>
Pluiesaison_n1 < 376.1
<< Y
N >>
densO < 3266.5
<< Y
N >>
NBG_1C < 23.6198830409357
<< Y
N >>
55220
NBG_1P < 117.5
<< Y
N >>
20800
20310
30810
37050
51270

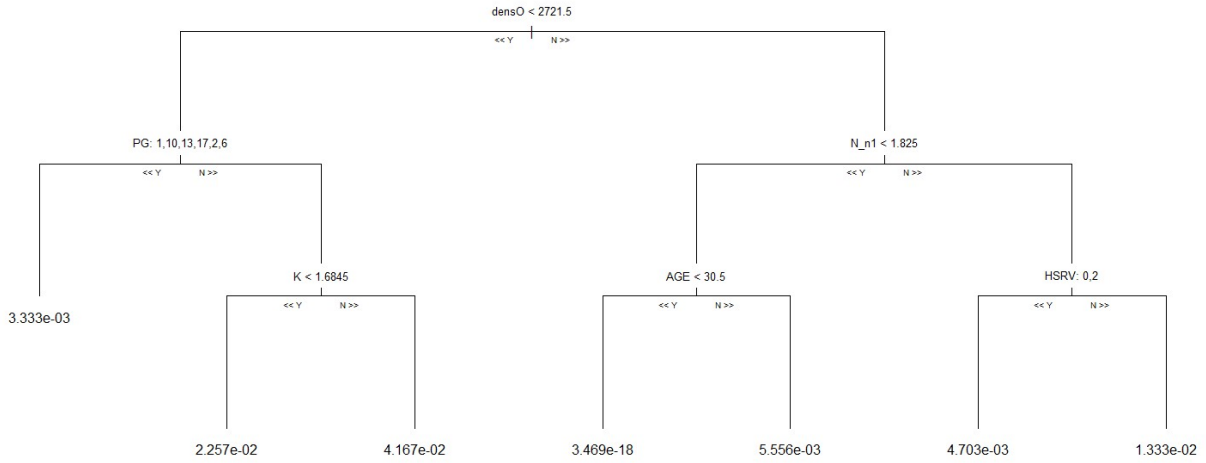
B.


PGP < 334
<< Y
N >>
PGP < 225
<< Y
N >>
CIR1V: 0,1,2,3
<< Y
N >>
P < 0.2
<< Y
N >>
COUP < 1
<< Y
N >>
N < 2.31
<< Y
N >>
12920
4764
2764
7564
5460
9768
7763

C.


BOUI_n1 < 105
<< Y
N >>
KF < 128.326530612245
<< Y
N >>
PO_n1 < 1.43085106382979
<< Y
N >>
NRO_n1 < 2.5
<< Y
N >>
TM < 0
<< Y
N >>
N_fum < 67
<< Y
N >>
SITU: 1,2,3,4
<< Y
N >>
0.10
0.05
0.00
0.00
0.00
0.05
0.00
0.05

D.



E.

<u>Figure 3 :</u> Arbre de régression représentant l'arbre moyen des random forest des indicateurs du rendement et de mortalité

<u>Légende :</u> A : Rendement déclaré; B : Rendement calculé à la placette; C : Rendement calculé par cep productif ; D : Taux de mortalité annuel ; E : Taux de mortalité pour la période allant de 1995 à 2007

Résumé

Le dépérissement viticole affecte la performance de production des exploitations en entrainant une diminution de la productivité des ceps et leur mort prématurée de manière progressive ou brutale. A l'heure actuelle, on estime la perte moyenne de rendement liée au dépérissement à 4.6 hL/ha/an en France. Le but de cette étude est d'étudier un jeu de données très diversifié, collecté par le BNIC, pour déterminer s'il est possible d'identifier et de mettre en évidence des facteurs du dépérissement viticole sur la période analysée, sur le réseau de parcelles de Cognac. Dans un premier temps, le jeu de données a été mis sous forme exploitable pour pouvoir être utilisé pour les analyses. Dans un second temps, des analyses descriptives ont été effectuées pour déterminer les variables explicatives pouvant être gardées. Enfin, des analyses de type régression ont été réalisées pour caractériser le dépérissement viticole sur le réseau au cours du temps. Il a été constaté une baisse de rendement et une augmentation de la mortalité significatives. Les rendements des ceps productifs ne montrent pas d'évolution croissante ou décroissante significative. Ainsi, la baisse de rendement serait plutôt liée à l'augmentation de la mortalité au cours du temps. L'analyse spécifique d'années particulières montre que la plupart des variables caractérisant le rendement caractérisent également la mortalité. La mortalité est liée notamment à des variables caractéristiques de l'âge des parcelles et du climat. Les résultats obtenus ouvrent des pistes à approfondir, puisque le jeu de données provient d'un terroir viticole particulier, l'analyse a une portée limitée, nous travaillons sur la période de 1995 à 2019 et plus spécifiquement sur les années 1999, 2004 et 2007. Augmenter cette plage d'étude améliorerait la généricité du travail.

Mots clés : dépérissement viticole, viticulture, approche systémique, régression linéaire, random forest, Cognac

Summary

Grapevine dieback affects French vineyard production through a decrease of vine's yield and gradual or sudden premature death of vines. Currently, loss on average yield caused by grapevine dieback is estimated at 4.6 hL/ha per year in France. The goal of the present work is to study a very diverse dataset, collected by the BNIC, to determine if it's possible to identify and highlight dieback factors on a plot network of the Cognac winegrowing region. Firstly, the dataset was put into usable form to be used for the analyzes. Secondly, descriptive analyzes were carried out in order to determine explanatory variables that can be kept. Finally, regression-type analyzes have been realised to characterize grapevine dieback on the network over time. We found a significant drop in yield and an increased mortality during the studied period. Productive vines showed no significant increasing or decreasing evolution of their yield. Thus the yield drop was rather linked to the increase in mortality over time. Specific analysis of particular years showed that most of the variables characterizing yield also characterized mortality. Mortality is notably linked to variables know to be indicators of the age of the plots and to the climate. Results obtained open avenues for further study: because the dataset comes from a particular wine region, the analysis is limited in scope, we work over the period from 1995 to 2019 and more specifically on the years 1999, 2004 and 2007, and this could be expanded to give more genericity to the study.

Keywords : Grapevine dieback, viticulture, systematic approach, linear regression, random forest, Cognac