



UNIVERSITÉ  
DE MONTPELLIER



## MASTER 2 MIND/SIAD

---

Etude de l'efficacité d'un traitement de  
biocontrôle contre l'oïdium de la tomate :

---

Fait par : **Ibrahim Gaizi**

Sous la direction de : **Xavier BR Y**

Établissement d'accueil : **INRAE**

Sous l'encadrement de : **Thomas Pressecq**

Et de : **Marc Bardin**



## Remerciements

En premier lieu je tiens à remercier M Thomas Pressecq pour la grande aide qu'il m'a apporté tout au long de ma période de stage, pour son accueil chaleureux, ses précieux conseils, ainsi que la qualité de son encadrement et le temps qu'il m'a accordé tout au long de mon passage à l'INRAE.

Un grand merci à M Marc Bardin, M Phillipe Nicot, M Marc Tchamitchian , et Mme Aurélie Rousselin, pour l'écoute et l'accompagnement dont j'ai bénéficié, ils m'ont beaucoup appris et ont facilité mon appréhension du sujet.

Je tiens aussi à remercier Mr Xavier Bry pour ces précieux conseils et orientations tout au long de la formation mais aussi du stage, il m'a toujours apporté un grand appui.

Je désire aussi remercier les professeurs de la FDS et de l'IAE Montpellier, qui m'ont fourni les outils nécessaires au bon déroulement de mon stage.

Je saisis également cette occasion pour adresser mes profonds remerciements aux responsables et au personnel de l'INRAE.

Pour finir, un grand merci pour mon défunt père, ma mère, mes frères, et ma bien aimée et mes amis pour leurs soutiens inconditionnels et sans faille.

## Table des matières

Introduction.....	7
La protection des plantes.....	7
Le biocontrôle, outil pour la protection des plantes .....	7
Recherches menées sur les facteurs d'efficacité du biocontrôle .....	8
Ampelomyces quisqualis, un agent de biocontrôle contre l'oïdium de la tomate.....	10
Essai sur l'influence des facteurs climatiques sur l'efficacité du traitement à base AQ10 .....	10
Objectifs du stage.....	12
Chapitre 0 : Préparation des données aux futures analyses.....	14
Présentation des données.....	14
Recodage de variable .....	15
Gestions des données manquantes .....	16
Calcul du pourcentage de protection, indicateur du succès d'un traitement .....	18
Chapitre 1 : Caractérisation de l'efficacité du traitement .....	19
Le développement de l'oïdium .....	19
L'effet du traitement sur le développement de la maladie .....	22
Le pourcentage de protection pour quantifier l'efficacité du traitement .....	24
Chapitre 2 : Quantification de l'impact des facteurs climatiques sur l'efficacité du biocontrôle .....	27
Introduction et méthodes .....	27
Analyse factorielle sur les variables de l'étude .....	28
Analyse de l'inertie des axes .....	30
Analyse des dimensions .....	31
Représentation des facteurs climatiques via l'ACP.....	33
Calcul des corrélations .....	35
Chapitre 3 : Classification des données, et comparaison des outils de modélisation pour déterminer des règles de décision les plus adéquates dans le cas rencontré .....	39
Introduction de la partie .....	39
Classification par seuil d'efficacité sous Power BI .....	40
Classification Hiérarchique sur Composantes Principales : HCPC .....	44
Dendrogramme: .....	45
Description des Clusters : .....	47
Clustering par Random Forest. ....	49
Comparaison des 3 méthodes :.....	53
Discussion & Conclusion.....	54

## Tableaux des figures

Figure 1 Représentation des différents facteurs influençant l'efficacité des biocontrôles microbiens contre les maladies des plantes .....	9
Figure 2 Feuille de tomate représentant environ 17 colonies d'oïdium (taches blanches) .....	11
Figure 3 Plan du tunnel d'expérimentation .....	11
Figure 4 Composition de la micro-parcelle .....	12
Figure 5 Schéma technique d'une cellule de Malassez.....	14
Figure 6 Courbe représentant la comparaison de l'évolution de la maladie selon le nombre de colonies par cm <sup>2</sup> de feuille entre juin et septembre.....	19
Figure 7 Courbe représentant la comparaison de l'évolution de la maladie selon le nombre de spores par cm <sup>2</sup> de feuille entre juin et septembre.....	20
Figure 8 Histogramme de l'évolution de la maladie en juin en fonction du jour de l'expérience .....	21
Figure 9 Histogramme de l'évolution de la maladie en juin en fonction du moment du traitement ...	21
Figure 10 Histogramme de l'évolution du nombre de colonie en fonction des traitements en juin ....	22
Figure 11 Histogramme de l'évolution du nombre de colonie en fonction des traitements en septembre .....	23
Figure 12 Histogrammes comparant les efficacités préventives en J14 et J21 en juin et septembre ..	24
Figure 13 Histogrammes comparant les efficacités pour les données de juin (matin, midi et soir) ....	25
Figure 14 Histogramme avec courbes représentant l'efficacité et le nombre de colonie du témoin en fonction des moments du traitement.....	25
Figure 15 Histogrammes représentant la comparaison entre traitements préventifs et curatifs et préventifs uniquement pour les notations de juin et septembre lors du deuxième comptage à 21 jours .....	26
Figure 16 Analyse factorielle avec toutes les variables prises en compte.....	28
Figure 17 AFDM avec variables d'intérêts contribuant à la construction des axes.....	29
Figure 18 Décomposition de l'inertie total par rapport aux nombres de dimension .....	30
Figure 19 Projection des individus sur l'ACP avec variables d'intérêts contribuant à la construction des axes par rapport à la variable traitement .....	32
Figure 20 Projection des individus sur l'ACP avec variables d'intérêts contribuant à la construction des axes par rapport à la variable traitement en retenant les données relatives à J28 .....	33
Figure 21 Projection des individus sur l'ACP avec différenciation selon la température lors du traitement dit « Moy_temp_inst » .....	34
Figure 22 Matrice de corrélation entre les efficacités et les variables de température .....	36
Figure 23 Matrice de corrélation entre les efficacités et les variables de VPD .....	37
Figure 24 Matrice de Corrélation entre les efficacités et les variables d'augmentation de la température .....	38
Figure 25 Graphique représentant les variables de moyenne de température par rapport aux groupes d'efficacité .....	41
Figure 26 Graphique représentant les variables de moyenne de VPD par rapport aux groupes d'efficacité .....	41
Figure 27 Graphique représentant les variables de minimum de température par rapport aux groupes d'efficacité .....	42
Figure 28 27 Graphique représentant les variables de maximum de température par rapport aux groupes d'efficacité .....	43
Figure 29. Dendrogramme de la CAH.....	45
Figure 30 Représentation des clusters sur le plan factoriel 1:2.....	46
Figure 31 Dendrogramme de la classification hiérarchique sur le plan factoriel 1:2 .....	47



Figure 32 Process suivi pour le Random Forest .....	50
Figure 33 Matrice de confusion .....	52
Figure 34 Représentation des individus suivant la variable heure .....	63

## Tableau des sigles et abréviations :

**ACP** : Analyse en composantes principales

**ACM** : Analyse des Correspondances Multiples

**AFDM** : Analyse factorielle de données mixtes

**HCPC** Classification Hiérarchique sur Composantes Principales

**CAH** : classification ascendante hiérarchique

**MCO** : Méthode des moindres carrés, elle consiste à minimiser la somme des carrés des écarts, c'est-à-dire entre chaque point du nuage de régression et son projeté

**missMDA** : Package R qui permet de gérer les données manquantes pour les méthodes d'analyses factorielles

**OAD** : outil d'aide à la décision

**Validation croisée** : une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage

**AQ10** : un bio-fongicide spécialiste des oïdiums à base d'une souche naturelle d'*Ampelomyces quisqualis*

**CTR** : Un coefficient qui exprime la contribution de chaque variable lors de la création des axes.

# Introduction

## La protection des plantes

L'apparition de l'agriculture fut un cap majeur pour l'évolution et la sédentarisation des sociétés humaines 9000 ans avant notre ère (Mercader 2009). Cette dernière se définit comme la production de ressources utiles pour les sociétés humaines dont par exemple la production alimentaire. (Mazoyer et Roudart 2002). Afin de pérenniser la production agricole soit l'apport de nourriture, la protection des plantes est devenue très rapidement l'enjeu majeur de l'agriculture. (FAO 2019) Les actions humaines liées à la protection des plantes ont évolué au cours de l'histoire : Pendant la première période préindustrielle, seule l'expérience empirique des agriculteurs ainsi que quelques décoctions minérales comme le cuivre et le soufre permettaient de lutter contre les fléaux inconnus qui constituaient les maladies des plantes. Ce fonctionnement a perduré tant que les outils nécessaires à la compréhension de la nature microscopique des maladies n'étaient pas disponibles. C'est avec la révolution chimique et microscopique du 19<sup>ème</sup> et surtout du 20<sup>ème</sup> siècle que la protection des plantes et par extension l'agriculture rentrent dans la période dite « productiviste », par la rationalisation de la main d'œuvre, par la mécanisation et l'augmentation des rendements et surtout par la systématisation de la protection par des produits chimiques. (Bernard 2007) Cette phase productiviste va s'accélérer tout au long du 20<sup>ème</sup> siècle avec comme catalyseur le deuxième conflit mondial et les bouleversements qui en découlent. (Fourastié 2020) On peut citer la profonde transformation de la société rurale en société urbaine et péri-urbaine, le recours massifs à des produits chimiques comme les benzodiazépines, les phthalimides ou encore les organo-azotés. La troisième phase est celle que nous vivons aujourd'hui, elle est toujours en cours de transition. Elle est caractérisée par des préoccupations croissantes pour l'environnement, la santé (du consommateur et du producteur) qui influencent les choix et les possibilités concernant la production agricole. (Ecophyto 2021) En effet, après plusieurs dizaines d'années du « tout chimique », la « prise de conscience écologique » [REF expression] engendre l'apparition ou réapparition de méthodes alternatives à la lutte chimique dans le paysage de la protection des plantes. Dans cette perspective on peut citer les variétés résistantes, les itinéraires techniques adaptés et le biocontrôle

## Le biocontrôle, outil pour la protection des plantes

Le biocontrôle fait partie de ces méthodes alternatives qui est au cœur de ce stage. Ce terme possède un sens spécifique en France qui est le suivant : « Agents et produits utilisant des mécanismes naturels dans le cadre de la lutte intégrée contre les ennemis des cultures » (Nationale, 2014). Cette définition n'a pas d'équivalent direct en anglais puisque les termes de « Biocontrol » ou de « biological control » (Eilenberg et al., 2001) nous donne : « utilisation d'un organisme vivant pour restreindre la densité de population ou l'impact d'un bio agresseur spécifique, le rendant moins abondant ou moins dommageable ». Pour donner plus de substance à cette définition de ce qu'est le biocontrôle, l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (Anses 2021) l'organisme régulateur, responsable de la mise sur le marché de ces produits le définit comme « Moyen de lutte contre les pathogènes et les ravageurs des plantes en utilisant des produits à base ou issus de sources naturelles ». Cette différence sémantique entraîne un flou dans la définition française puisque des produits inertes (non vivants) sont considérés comme du biocontrôle alors que dans la définition internationale il est précisé l'aspect vivant des produits utilisés. Les termes de lutte biologique, protection biologique ou biopesticide sont donc plus proches pour parler des

produits dit de biocontrôle à base par exemple d'extraits de levures, de sels minéraux ou d'extraits de plantes plutôt que biocontrôle.

En guise de synthèse des différentes définitions, le biocontrôle est donc un outil de lutte contre les ravageurs et les maladies des plantes utilisant des mécanismes dit naturels issus du monde du vivant, la synthèse chimique étant autorisée pour copier des molécules naturelles uniquement comme pour créer des phéromones d'insectes artificiellement.

Le biocontrôle se divise en plusieurs branches bien distinctes de par leur nature, leur cible et leur utilisation.

Le premier critère correspond à la nature des produits :

- Les macro-organismes (invertébrés, insectes, acariens ou nématodes), non soumis à autorisation de mise sur le marché. (AMN)
- Les produits phytopharmaceutiques.

Au sein des produits phytopharmaceutiques, on trouve les sous-catégories suivantes :

- Les micro-organismes (champignons, bactéries, virus) ;
- Les médiateurs chimiques comme les phéromones sexuelles (substances chimiques produites par les insectes jouant un rôle dans l'attraction sexuelle) ;
- Les substances naturelles (substances d'origine végétale, animale ou minérale).

Parmi les substances de biocontrôle, plusieurs sont susceptibles de satisfaire aux critères « faible risque environnemental » et donc compatibles avec la plupart des labels notamment le cahier des charges agriculture biologique ou celui moins connu du grand public comme « zéro résidus ». En conséquence, les agents de biocontrôle sont un des produits capables de protéger les plantes des maladies et des ravageurs en limitant le risque pour les consommateurs et les producteurs. Pour autant, bien que très prometteurs, leurs usages globaux restent limités par rapport à la protection chimique (Monus 2019). Ceci peut s'expliquer par leur complexité d'utilisation qui nécessitent la prise en compte de nombreux facteurs, ce qui engendre une efficacité variable et donc un résultat incertain pour des frais engagés souvent importants. Cette variabilité de l'efficacité est au cœur de nombreuses recherches sur le biocontrôle et constitue un verrou supposé comme majeur pour une adoption plus importante (Punja et al. 2019; Bardin et al. 2015).

## **Recherches menées sur les facteurs d'efficacité du biocontrôle**

Les agents de biocontrôle sont le plus souvent utilisés dans la pratique de la même façon que les pesticides chimiques, avec l'attente (souvent non satisfaite) qu'ils se comportent aussi bien en dépit de leurs caractéristiques spécifiques, en particulier pour les microorganismes. Quel que soit le système de culture, il est donc nécessaire de déterminer les conditions optimales d'utilisation de ces produits.

Pour comprendre et expliquer ces différents facteurs influençant le succès (ou l'échec) d'un traitement à base de biocontrôle, de nombreuses études ont été réalisées par différentes équipes avec des approches similaires (Fedele, Brischetto, et Rossi 2020). (Magan 2020) (Nicot 2011). Il ressort de ces études que les agents de biocontrôle microbiens sont souvent sous l'influence de facteurs d'efficacité multiple, peu ou pas encore réellement identifiés et enfin qu'il est difficile d'obtenir des résultats fiables

et concluant. La recherche n'est encore qu'au premier stade sur l'identification des facteurs d'efficacité, en effet la majorité des publications s'intéresse davantage à évaluer l'effet des biocontrôles plutôt qu'à expliquer pourquoi il fonctionne plus ou moins bien.

A titre d'exemple, le moment et la fréquence de leur application sur la culture sont reconnus comme des paramètres cruciaux pour assurer une efficacité de protection optimale. La détermination de ces paramètres nécessite de considérer de multiples facteurs. Les facteurs d'efficacité se rassemblent dans quatre catégories connectées entre elles, présentées dans le schéma ci-dessous (Figure 1) :

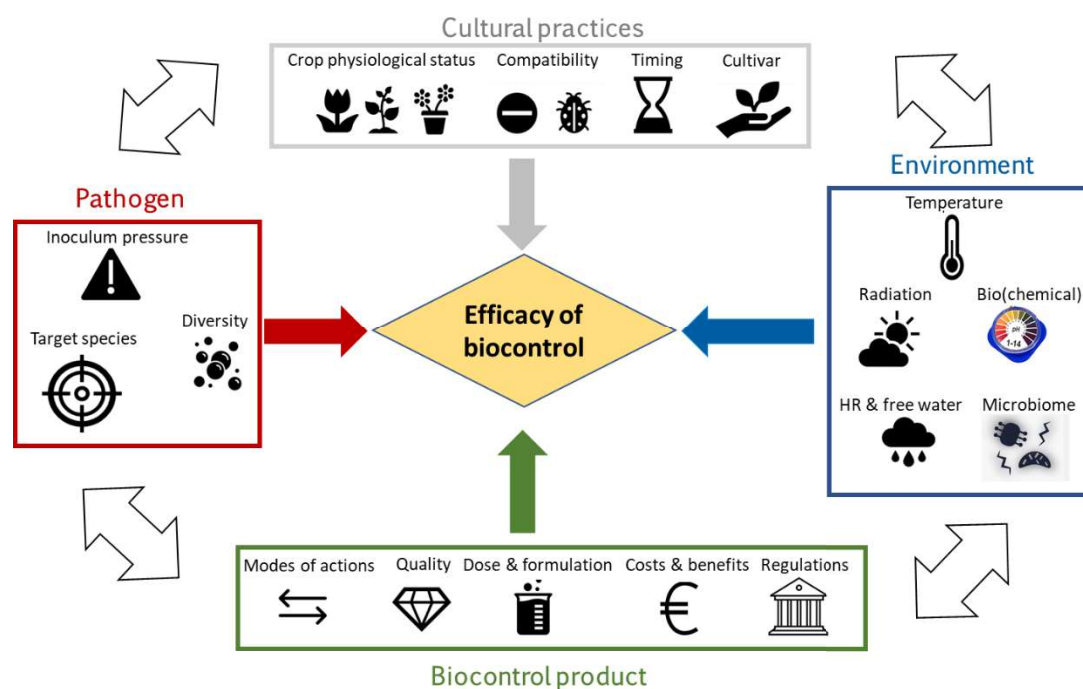


Figure 1 Représentation des différents facteurs influençant l'efficacité des biocontrôles microbiens contre les maladies des plantes

L'ensemble de ces facteurs peuvent impacter l'efficacité des agents de biocontrôle microbiens. Le premier ensemble de facteurs pouvant influencer l'efficacité des agents de biocontrôle rassemble les facteurs environnementaux (température, humidité, pH, oxydo-réduction, ...). Ces facteurs concentrent la majorité des recherches actuelles (Fedele, González-Domínguez, et Rossi 2020) puisque l'hypothèse est qu'ils sont aussi les plus impactant pour l'efficacité et aussi les plus facilement contrôlables par les utilisateurs.

On note également de très nombreuses études sur les modes d'actions des agents de biocontrôle (majoritairement effectué en condition *in vitro*). Les autres facteurs sont aussi étudiés mais dans une moindre mesure car les méthodologies d'analyse sont plus incertaines et difficiles à mettre en œuvre. De rares études ont été réalisées sur les facteurs liés au pathogène cible du biocontrôle car c'est un domaine très large (autant qu'il existe de pathogène des plantes soit des centaines). Enfin un nombre croissant d'études se concentrent sur les facteurs en relation avec les pratiques culturales. Comme par exemple les facteurs liés aux interactions des biocontrôles avec la plante et les actions et choix de l'agriculteur tout au long de la culture comme l'irrigation, la fertilisation ou le choix variétale. Les facteurs de ces dernières catégories sont cruciaux pour la réussite d'un traitement à base de biocontrôle car on imagine très vite qu'un biocontrôle à base de champignons ne pourra être efficace

s'il est associé à un antifongique lors d'un traitement. L'étude des facteurs d'efficacité reste donc au centre des préoccupations pour la recherche concernant les biocontrôles.

## **Ampelomyces quisqualis, un agent de biocontrôle contre l'oïdium de la tomate**

L'oïdium est une maladie fongique dont l'agent responsable sur la tomate est le champignon ascomycète *Oidium neolycopersici*. Cette maladie touche le monde entier depuis plus de 30 ans. Elle se caractérise par des taches blanchâtres sur la face supérieure des feuilles de tomate. Son importance économique est forte puisque mal gérée elle peut engendrer des pertes de rendement importantes dues notamment à la réduction de la photosynthèse sur les plantes touchées. Différents moyens de lutte existent, mais de nos jours c'est le soufre sous toutes ses formes qui reste le plus utilisé. Pour autant des questionnements apparaissent concernant la surutilisation de ce produit (pas réputé « bon pour la santé ») poussant la filière à trouver de nouvelles solutions plus adaptées. Dans la sphère du biocontrôle microbien, peu de solutions existent, en dehors des produits à base de *Bacillus* comme Serenade® ou encore *Ampelomyces quisqualis* sous la forme du produit commercial AQ10®, l'objet de l'étude. Cet agent de biocontrôle distribué par CBC Biogard, est un champignon de la famille des Phaeosphaeriaceae appartenant au groupe des ascomycètes. Il est reconnu comme étant un hyperparasite de *Oidium neolycopersici*. Il est au cœur de travaux de recherche menés précédemment dans le laboratoire d'accueil. Ces travaux ont permis d'évaluer les conditions optimales de développement en situation *in vitro* d'*A. quisqualis* en fonction de la température. Ils ont aussi étudié son efficacité protectrice en conditions contrôlées et sous tunnel contre l'oïdium de la tomate dans plusieurs expériences.

## **Essai sur l'influence des facteurs climatiques sur l'efficacité du traitement à base AQ10**

En amont du stage, deux expériences ont été menées afin de fournir des données pour évaluer l'effet du microclimat de la serre sur l'efficacité de la protection de l'agent de biocontrôle microbien *A. quisqualis* contre l'oïdium de la tomate (*Oidium neolycopersici*). Le concept de l'expérience est d'inoculer les tomates avec l'agent pathogène puis d'effectuer un traitement avec l'agent de biocontrôle dans différentes conditions pour pouvoir le comparer au témoin non traité. Le développement de l'épidémie d'oïdium dans la parcelle a été effectué par le comptage du nombre de colonies d'oïdium sur les feuilles de tomate à 14, 21 et 28 jours après inoculation. Ces dates ont été choisies pour obtenir un développement suffisant de l'oïdium. Ce comptage se fait à la main avec prise de photo pour calculer la surface foliaire afin de ramener le nombre de tâche par cm<sup>2</sup> de feuille.



Figure 2 Feuille de tomate représentant environ 17 colonies d'oidium (taches blanches)

En parallèle, le nombre de spores produites par l'oidium sur les feuilles a aussi été mesuré grâce à une méthode d'estimation microscopique avec des cellules de Malassez. (Voir chapitre 0) Pendant l'expérience, la température et l'humidité ont été contrôlées à l'aide de 9 capteurs placés dans le tunnel (7 pendant la deuxième période de test en septembre). La température et l'humidité ont varié tout au long de l'expérience avec des valeurs de température allant de 15 à 35°C et des humidités de l'air allant de quasi 100% jusqu'à 25%.

Deux expériences ont été menées en juin et en septembre 2020 dans un tunnel de 128 m<sup>2</sup> avec 240 plants de tomates divisés en 4 rangées (A, B, C & D). Chaque rangée est subdivisée en 3 groupes de 4 plantes (P1, P2, P3 & P4) correspondant à différents moments d'inoculation avec l'oidium et de traitement avec le produit de biocontrôle AQ10 (soit matin, midi et soir) afin de multiplier les conditions climatiques rencontrées pendant l'expérience. Ce dispositif est répété pendant 3 jours consécutifs dans chacune des périodes. (Figure 3)

Bloc A	Bloc D	Bloc C	Bloc B	Bloc A
4				
3	3 4	3 4	3 4	
2	1 2	1 2	1 2	
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	
	1 2	1 2	1 2	4
4				3
3	3 4	3 4	3 4	2
2	1 2	1 2	1 2	1
1				
	3 4	3 4	3 4	

Sur chaque parcelle de 4 plantes, toutes les plantes sont inoculées avec une suspension de spores d'oïdium dosée à  $2.10^3$  sp/ml, mais diffèrent par le traitement reçu (Figure 4). La plante 1 correspond à la plante témoin non traité, la plante 2 correspond à la plante ayant reçu un traitement curatif avec AQ10 (14 jours après inoculation de l'agent pathogène), la plante 3 correspond à la plante ayant reçu un traitement préventif avec AQ10 (traitement réalisé le même jour que l'inoculation de l'agent pathogène), et la plante 4 correspond à la plante ayant reçu les traitements préventifs et curatifs.

La dose d'AQ10 utilisée correspond à la dose recommandée par le fabricant, à savoir 70g/ha. Un volume de traitement à 1000 l/ha est appliqué sur la culture dont la densité est de 1,9 plantes/m<sup>2</sup>. Pour le traitement préventif, le traitement est effectué en même temps que l'inoculation de l'oïdium ; c'est donc un traitement préventif car il est effectué avant l'apparition des premiers symptômes de la maladie. Le traitement curatif est effectué quand les symptômes sont déjà présents soit 14 jours après inoculation de l'oïdium.

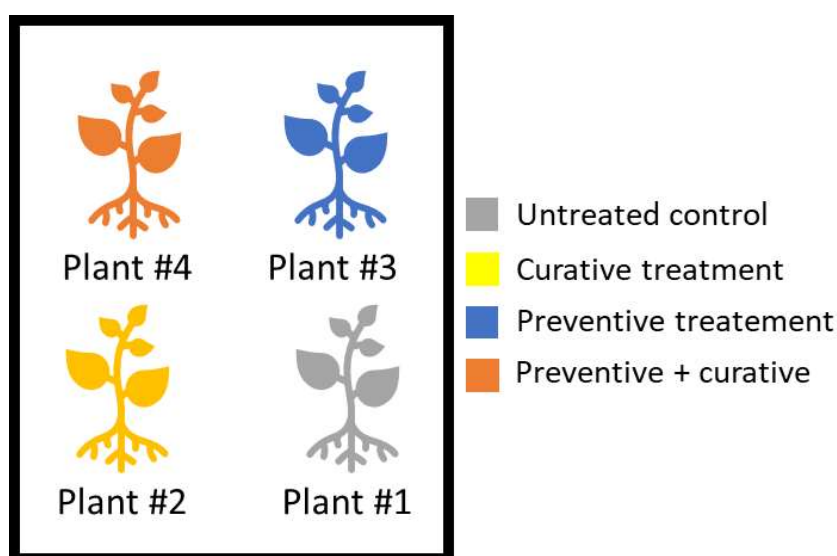


Figure 4 Composition de la micro-parcelle

## Objectifs du stage

Cette expérience s'inscrit dans le projet de thèse BIOAD (pour Biocontrôle et outil d'aide à la décision) qui cherche à prendre en compte la complexité des agents de biocontrôle microbiens afin de fiabiliser leur utilisation. Pour répondre à ce besoin et faciliter le conseil, la thèse vise au développement d'outils d'aide à la décision (OAD) intégrant les propriétés biologiques des agents de biocontrôle microbiens, celles des agents pathogènes, et les caractéristiques des systèmes de culture afin d'établir des règles de décision associés. Pour atteindre ces objectifs, de très nombreuses données sur l'efficacité sont nécessaire afin de les analyser en détail et de dégager les facteurs influençant le succès ou l'échec d'un traitement.

L'objectif du stage est donc d'analyser les données fournies par l'expérience afin de caractériser les facteurs d'efficacité du produit AQ10 contre l'oïdium de la tomate dans le but d'aider à l'établissement de règles de décision pour favoriser l'usage de cet agent de biocontrôle. Cet objectif global se divise en 4 objectifs particuliers :



- Préparer les données pour les futures analyses
- Caractériser le développement de la maladie ainsi que l'efficacité du traitement par des statistiques descriptives ;
- Quantifier l'impact des facteurs climatiques sur l'efficacité du traitement ;
- Classification et comparaison des outils de modélisation pour déterminer les règles de décision permettant d'obtenir la meilleure efficacité pour le produit en prenant en compte les facteurs climatiques.

# Chapitre 0 : Préparation des données aux futurs analyses.

## Présentation des données

L'expérience a généré de nombreuses données qualitatives et quantitatives, tout d'abord séparées en deux bases de données distinctes puis rassemblées à l'issue de la préparation des données.

La première base de données se concentre sur les individus (soit les plantes) indiquant pour chacune d'elles, leur positionnement dans les tunnels (A, B, C ou D), le moment d'inoculation (matin, midi et soir) et leur modalité de traitement. (0,1, 2 ou 3). En plus de ces variables descriptives, des variables quantitatives sont présentes : le nombre de colonies par cm<sup>2</sup> de feuille et le nombre de spores par cm<sup>2</sup> de feuille. Ces deux variables sont présentes sur plusieurs comptages sur les mêmes plantes mais sur des feuilles différentes car la méthode de comptage est destructive. Le comptage de colonie se fait manuellement par un membre de l'équipe qui observe et comptabilise le nombre de taches (soit les colonies d'oïdium). Pour le nombre de spores, leur comptage nécessite de recourir à une préparation plus complexe. On mélange les feuilles prélevées dans une solution d'eau osmosé de 20 ml pour obtenir une suspension de spores venant des colonies d'oïdium. Puis on prélève 17 µl (quantité fixée par cellule) qu'il faut déposer sur une lame de microscope spécifique : une cellule de malassez (voir figure 5). Une fois cette préparation effectuée, le comptage des spores peut avoir lieu. Seules les spores présentes sur les grilles et non détériorées sont prises en comptes. Les autres sont ignorées. Cette technique permet une estimation du nombre de spores présent sur la feuille. Tout comme la variable nombre de colonie, cette dernière est ramenée à la surface foliaire pour plus de précision.

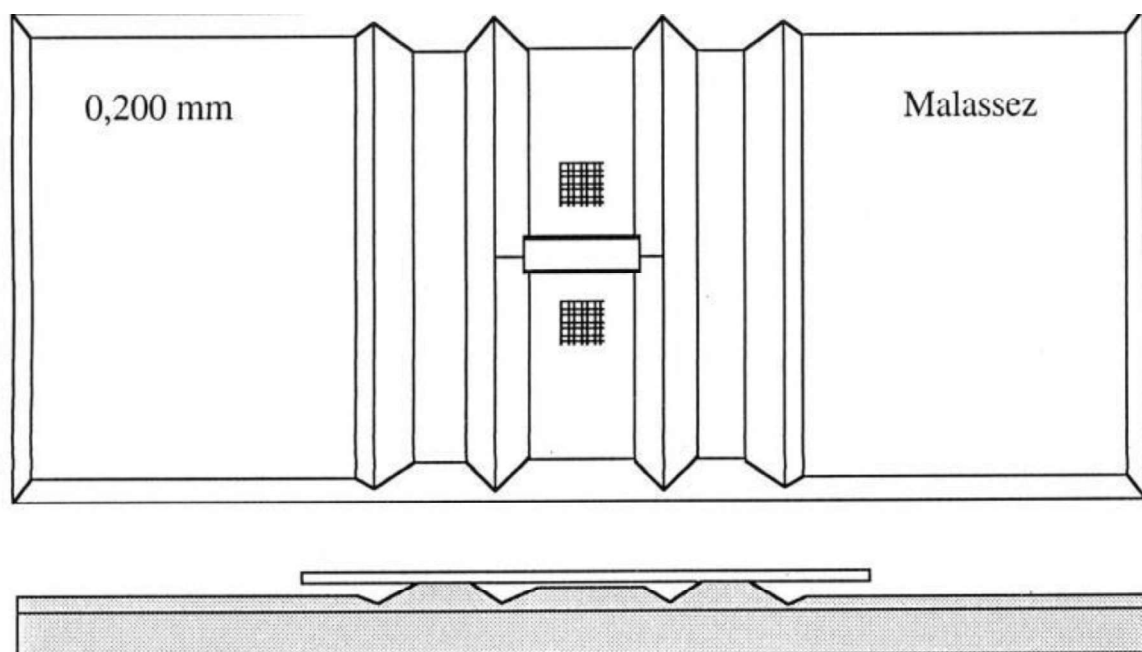


Figure 5 Schéma technique d'une cellule de Malassez

En plus de ces données quantitatives, l'expérience a fourni de très nombreuses données climatiques, rassemblées sur une deuxième base de données, issues des capteurs climatiques des tunnels

enregistrant la température (en °C) et l'humidité relative (en %) toutes les 10 minutes pendant toute la durée d'expérimentation. Pour l'humidité le choix fut de calculer le déficit de pression de vapeur saturante (VPD) plus juste et représentant mieux la réalité macroscopique rencontrée par les plantes et les microorganismes (Grossiord et al. 2020)

L'humidité relative de l'air est le rapport de la pression partielle  $e$  à la pression de saturation  $e^*$  à la température de l'air humide :

$$Hr = \frac{e}{e^*}$$

Hr l'humidité relative de l'air en (100 %).

$e^*$  Tension de la vapeur d'eau saturante est donnée en fonction de la température T.

Diverses formules ont été proposées qui permettent de calculer  $e^*$  en fonction de la température T.

Dans ce rapport, la formule de Tétens a été utilisée :

$$e^* = 0.611 \exp\left(\frac{17.27 T}{T + 237.3}\right)$$

où la température T est en degrés Celsius (° C) et la pression de vapeur saturante  $e^*$  est en kPa

## Recodage de variable

Le but de l'étude étant d'étudier l'impact du climat sur l'efficacité il est crucial de rassembler les deux bases de données dans une seule pour pouvoir les étudier.

La première étape consiste à la fusion de la première base de données (informations relatives aux plantes) avec une deuxième base de données qui nous informe sur la température et l'humidité dans chacun des emplacements du tunnel où a été faite l'expérience. On a donc relié chacune des plantes à sa température et humidité au moment de l'inoculation et de son traitement. Puis de nouvelles variables ont été établies servant à mieux exprimer l'impact de la température et de l'humidité (VPD) sur le développement de la maladie. Dans cette optique les variables de moyenne de température et de VPD dans les heures qui font suite au traitement ont été créées. Ces variables permettent de connaître les tendances climatiques jusqu'à 8h après le traitement et l'inoculation. Elle se compose des moyennes et des extrêmes heure par heure pour la température et la VPD. De plus, afin de mieux étudier l'aspect de l'évolution des températures, la variable d'accélération de la température a été calculée avec un intervalle d'une demi-heure pour plus de précision. Ces variables climatiques calculées sont cruciales pour l'étude des facteurs d'efficacité. En effet l'expérience est construite pour que les conditions climatiques soit différentes au moment du traitement et pendant les huit heures suivantes en fonction des modalités (matin, midi et soir) et selon les répétitions (jour 1, 2 et 3). Ceci implique que nos analyses se concentrent sur cette période afin de pouvoir faire émerger un profil des conditions climatiques aboutissant aux meilleures efficacités de protection de la plante par AQ10 contre l'oïdium.

## Gestions des données manquantes

Une bonne gestion des données manquantes nous permet de gagner en informations : Car laisser des cases vides dans le jeu de données nous emmènerait pour certains cas à ne pas prendre en considération la plante à qui il manque une case dans le jeu de données, et donc ça causerait la perte de toutes les informations apportées par cette plante concernant les autres variables. Néanmoins, il faut utiliser la gestion des données manquantes avec des pincettes, car une mauvaise utilisation risque de fausser toute l'analyse.

Les données manquantes ne sont pas des données nominales mais plutôt des données liées au comptage du nombre de colonies et de spores sur les feuilles, le volume des données manquantes est faible. Les données manquantes sont causées généralement par deux facteurs :

- L'attaque de ravageurs qui sont des petits insectes ou arachnoïdes qui détruisent la plante avant qu'on puisse récupérer le résultat du traitement.
- Des feuilles trop abîmées par la maladie pour que le nombre de colonies et de spores soit compté correctement.

Quatre méthodes sont proposées pour résoudre cette problématique des données manquantes :

- En remplaçant les données manquantes dans chaque variable par la moyenne de tous les individus.
- En remplissant les cases manquantes grâce à une ACP 2 dimension
- En remplissant les cases manquantes grâce à une ACP k dimension, le choix de K est fait par validation croisée en enlevant au hasard des données disponibles et en les "prédisant" avec la méthode, puis en calculant l'Erreur Quadratique Moyenne de Prédiction.
- En remplissant les cases manquantes en utilisant la méthode maximisation de vraisemblance (EM)(FactomineR 2008)

Jours	Heure	Place	Ef_Cur_2/1_NC/cm2_j21_09	Ef_Cur_2/1_NS/cm2_j21_09	Ef_Cur_4/1_NC/cm2_j21_09	Ef_Cur_4/1_NS/cm2_j21_09	Ef_Cur_4/3_NC/cm2_j21_09
1	1	1 A	-36.430921	58.6450772	-91.63669575	48.685420	-37.080667
2	1	1 B	-14.478118	-24.2380496	-144.46805068	-235.429978	-58.886330
3	1	1 C	30.832354	-62.8321658		-89.514596	-24.489664
4	1	2 A	31.169577	13.1882986	51.14450040	18.610743	0.610920
5	1	2 B	-51.230628	60.0552562	22.27061794	-32.002529	9.328515
6	1	2 C	39.747400	54.8115447	45.49972978		-57.244220
7	1	3 A	-40.388180	-26.7731299	-2.80157518	-22.532876	5.813950
8	1	3 B	12.324499	-48.5370742	22.11671149	1.622916	33.382757
9	1	3 C	-4.847285	-122.6567757	28.74507830	-139.424213	3.683040
10	2	1 A	-86.581713	33.5409943	-22.81482916	84.020131	-28.100683
11	2	1 B	-17.024799	58.8520985	47.13554615	53.287875	34.112882
12	2	1 C	-17.403405	12.3120529	-9.45623077	-4.758815	30.113640
13	2	2 A	-15.937319	-37.4563673	40.42076231	43.156692	-49.353449
14	2	2 B	12.456544	34.9971910	28.91999937	43.466364	-23.235700
15	2	2 C	38.849656	45.5788073	14.18084132	37.095905	-6.641101
16	2	3 A	31.333239	9.7304582	19.95572944	76.290177	-59.917779
17	2	3 B	-79.383274	-91.6901007	-3.50686619	3.685011	-54.851429
18	2	3 C	29.715788	3.1464708	36.36860909	-79.876093	-40.711258
19	3	1 A	-14.704790	0.6574589	-44.39016840	15.817603	-16.936425
20	3	1 B	-45.197061	42.8779637	-11.14515538	50.894865	-15.875955
21	3	1 C	-26.686113	-12.0289545	35.44454805	23.230986	27.886946
22	3	2 A	-50.090430	-41.7097457	25.94239934	2.799399	73.794237
23	3	2 B	-58.225442	-401.1887633	4.04005102	-134.082995	40.712067

### Remplissage des données manquantes par EM

Jours	Heure	Place	Ef_Cur_2/1_NC/cm2_j21_09	Ef_Cur_2/1_NS/cm2_j21_09	Ef_Cur_4/1_NC/cm2_j21_09	Ef_Cur_4/1_NS/cm2_j21_09	Ef_Cur_4/3_NC/cm2_j21_09
1	1	1 A	-36.430921	58.6450772	-91.63669575	48.685420	-37.080667
2	1	1 B	-14.478118	-24.2380496	-144.46805068	-235.429978	-58.886330
3	1	1 C	30.832354	-62.8321658	28.08320421	-89.514596	-24.489664
4	1	2 A	31.169577	13.1882986	51.14450040	18.610743	0.610920
5	1	2 B	-51.230628	60.0552562	22.27061794	-32.002529	9.328515
6	1	2 C	39.747400	54.8115447	45.49972978	31.5270227	-57.244220
7	1	3 A	-40.388180	-26.7731299	-2.80157518	-22.532876	5.813950
8	1	3 B	12.324499	-48.5370742	22.11671149	1.622916	33.382757
9	1	3 C	-4.847285	-122.6567757	28.74507830	-139.424213	3.683040
10	2	1 A	-86.581713	33.5409943	-22.81482916	84.020131	-28.100683
11	2	1 B	-17.024799	58.8520985	47.13554615	53.287875	34.112882
12	2	1 C	-17.403405	12.3120529	-9.45623077	-4.758815	30.113640
13	2	2 A	-15.937319	-37.4563673	40.42076231	43.156692	-49.353449
14	2	2 B	12.456544	34.9971910	28.91999937	43.466364	-23.235700
15	2	2 C	38.849656	45.5788073	14.18084132	37.095905	-6.641101
16	2	3 A	31.333239	9.7304582	19.95572944	76.290177	-59.917779
17	2	3 B	-79.383274	-91.6901007	-3.50686619	3.685011	-54.851429
18	2	3 C	29.715788	3.1464708	36.36860909	-79.876093	-40.711258
19	3	1 A	-14.704790	0.6574589	-44.39016840	15.817603	-16.936425
20	3	1 B	-45.197061	42.8779637	-11.14515538	50.894865	-15.875955
21	3	1 C	-26.686113	-12.0289545	35.44454805	23.230986	27.886946
22	3	2 A	-50.090430	-41.7097457	25.94239934	2.799399	73.794237
23	3	2 B	-58.225442	-401.1887633	4.04005102	-134.082995	40.712067

Figure 6 Gestion des données manquantes avec la maximisation de vraisemblance

La première méthode est la moins efficace, néanmoins elle sert à mesurer le gain de qualité apporté par les 3 autres méthodes. En les comparant c'est la 4<sup>ème</sup> méthode qui donne les meilleurs résultats et qui est utilisé dans les analyses.

Pour autant, les résultats sont proches avec les 3 méthodes, ce qui semble logique en vue de la dimension du jeu de données qui n'est pas très imposante. Si on avait un jeu de données plus grand, on aurait eu une différence plus importante sur les résultats obtenus avec une ACP à 2 dimensions et celle avec k dimension.

## Calcul du pourcentage de protection, indicateur du succès d'un traitement

Pour pouvoir comparer plus facilement les résultats du traitement, il est nécessaire d'établir des variables d'efficacité basées sur le niveau de protection de la plante contre l'oïdium. Ces variables sont calculées par parcelle de 4 plantes (témoin, curatif, préventif et préventif +curatif) et non pas par plante. En effet pour calculer un pourcentage de protection il faut comparer les plantes traitées et les plantes témoins au sein d'une parcelle ayant subi les mêmes conditions. Pour cela, on utilise la formule la suivante :

$$(1 - (\text{Nombre de colonies d'oïdium/cm}^2 \text{ d'une plante traitée avec AQ10}) / (\text{Nombre de colonies d'oïdium/cm}^2 \text{ d'une plante non traitée})) * 100$$

Cette formule est utilisée pour calculer :

- **L'efficacité préventive moyenne à 14 jour et 21 jour calculée sur le nombre de colonies d'oïdium** : cette variable représente l'efficacité préventive par rapport au nombre de colonies comptabilisé sur les plantes 3 et 4 par rapport aux plantes 1 et 2 au 14ème jour et 21ème jour après traitement et inoculation.
- **L'efficacité préventive à 21 jours calculée sur le nombre de colonies d'oïdium** : cette variable représente l'efficacité préventif par rapport au nombre de colonies comptabilisé sur la plante 3 par rapport à la plante 1 au 21ème jour ; on rappelle que les plantes 2 & 4 ont reçu un traitement curatif au 14ème jour. – Cette variable fut créée pour ne prendre en compte que les traitements préventifs.
- **L'efficacité préventive à 21 jours calculée sur le nombre de spores d'oïdium** : cette variable représente l'efficacité préventif par rapport au nombre de spores de la plante 3 par rapport à la plante 1 au 21ème jour.
- **Efficacité curative à 21 jours en nombre de colonies et le nombre de spores** : Cette variable représente l'efficacité curatif par rapport au nombre de colonies de la plante 2 par rapport à la plante 1 au 21ème jour.
- **Efficacité préventive et curative à 21 jours calculée sur le nombre de colonies et de spores** : cette variable représente l'efficacité préventive et curative par rapport au nombre de colonies de la plante 4 par rapport à la plante 1 au 21ème jour ; on rappelle que la plante 4 a subi les deux traitements : préventif et curatif.

Ces variables construites sont utilisées lors de l'analyse dans un deuxième temps puisqu'elle représente non plus le développement de la maladie mais uniquement l'efficacité du traitement, c'est-à-dire un rapport de performance entre des plantes traitées et non traitées. Cependant le calcul de l'efficacité entraîne une réduction du nombre d'individus dans l'analyse. En effet cela engendre un jeu de données avec trop de variables par rapport aux nombres d'individus. Ce point reste une des principales limites de l'étude, elle sera abordée en discussion.

# Chapitre 1 : Caractérisation de l'efficacité du traitement

Après avoir préparé les données, des analyses peuvent être effectuées sur ces dernières afin de dégager des hypothèses sur l'efficacité du traitement à base de AQ10.

Dans un premier temps une étude descriptive du jeu de données pour caractériser l'influence du traitement AQ10 sur le développement de la maladie va être présentée en vérifiant notamment l'absence d'influence spatiale. Puis dans un deuxième temps, l'hypothèse généralement admise sur la variabilité de l'efficacité du traitement de biocontrôle est vérifiée selon les différentes modalités de l'expérience. En effet selon les travaux précédents, les produits de biocontrôle peuvent être efficace mais aussi très variable voire parfois totalement inefficace dans certaines conditions

Pour répondre à ces objectifs, différentes méthodes sont utilisées : majoritairement des comparaisons de moyennes, de variances et d'écart type, synthétisées par des représentations graphiques. Le recours à des analyses de variance (ANOVA) ainsi que des corrélations entre variables permet par exemple de vérifier les hypothèses sur l'efficacité du traitement. On note que l'ANOVA est un modèle statistique qui nous permet de voir s'il existe ou non des similitudes sur certains aspects dans la population étudiée.

## Le développement de l'oïdium

Une première analyse descriptive des données permet d'observer le développement de la maladie lors des 3 comptages à 14, 21 et 28 jours (uniquement pour juin) en nombre de colonies par  $\text{cm}^2$  de feuille (figure 7) ainsi qu'en nombre de spores par  $\text{cm}^2$  (figure 8). L'évolution du nombre de colonies nous indique que la maladie se développe au cours des comptages successifs sur les témoins non traités.

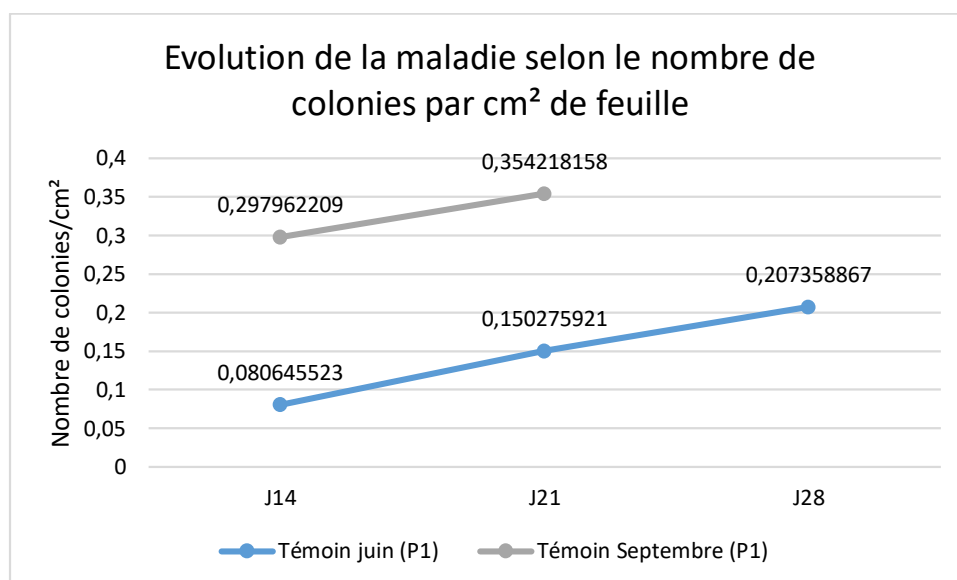


Figure 7 Courbe représentant la comparaison de l'évolution de la maladie selon le nombre de colonies par  $\text{cm}^2$  de feuille entre juin et septembre

Entre les données de juin et de septembre on note une différence très importante en termes de nombre de colonies, il y a plus de 3 fois et demi plus de colonie en septembre qu'en juin. De plus le taux variation entre le comptage à J14 et J21 est de 89% pour juin et 19 % pour septembre. Cela correspond au pourcentage d'augmentation entre ces comptages, soit à la croissance de l'oïdium pendant à 7 jours. Les pourcentages sont respectivement de 89% et 19 % pour juin et septembre.

Des conditions d'expérimentations différentes peuvent être la bonne explication pour expliquer les divergences entre les deux périodes : En effet l'âge et type des plantes ont été différents sur les périodes de Juin et de Septembre. Ce point rend les résultats difficilement exploitables en même temps, les analyses suivantes présenteront donc les données le plus souvent séparés.

Concernant l'évolution du nombre de spores peu d'interprétations peuvent être effectuées car le comptage du nombre de spores n'a pas été effectué en Juin lors du comptage à 14 jours après inoculation. Cependant on peut observer que le nombre de spores a tendance à diminuer entre le premier comptage (14 jours) et le deuxième (21 jours) en septembre. De plus le nombre de spores double entre le comptage à 21 jours et celui à 28 jours en juin ce qui correspond bien à l'explosion des symptômes observés à la fin de l'expérience soit en J28. Par contre on observe aussi le phénomène inverse en septembre où le nombre de spore d'oïdium diminue entre 14 et 21 jours.

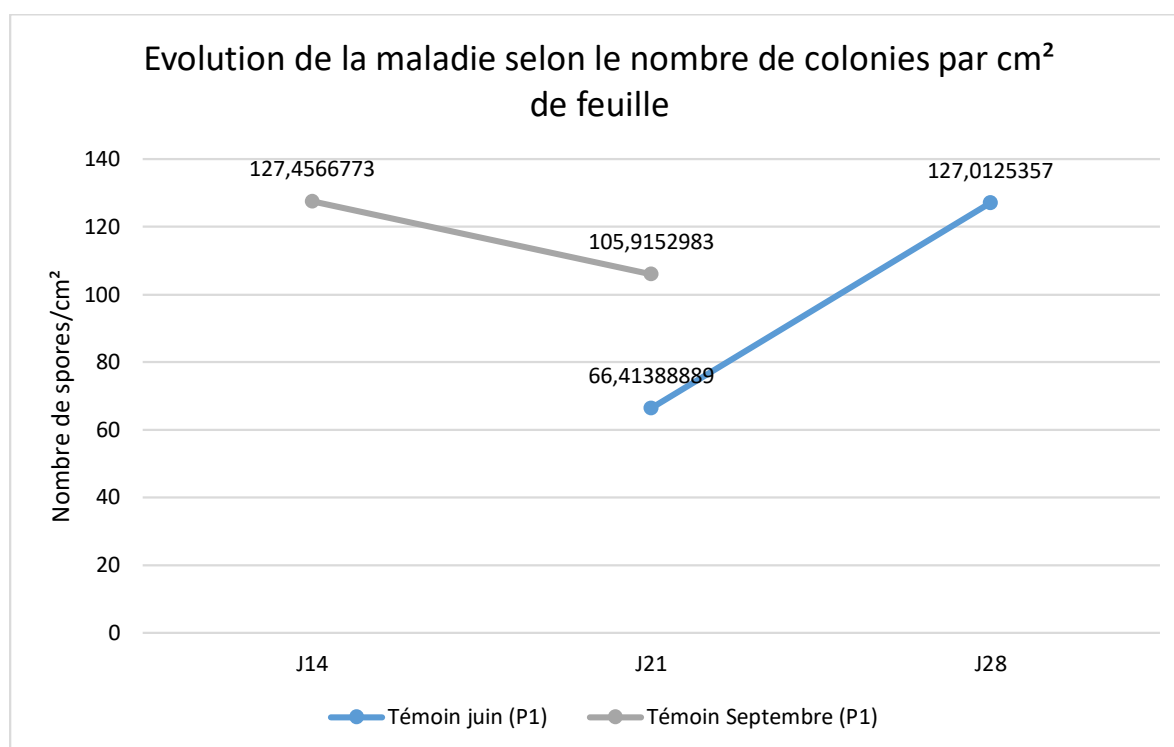


Figure 8 Courbe représentant la comparaison de l'évolution de la maladie selon le nombre de spores par cm² de feuille entre juin et septembre

Le développement de la maladie est donc différent entre juin et septembre mais identique au sein des répétitions dans une période données. En étudiant les jours (1, 2 et 3) de l'expérience, on peut observer que le développement de la maladie n'est pas significativement impacté par le jour de l'inoculation. (Figure 9) Donc les répétitions sont valides et peuvent être comparé entre elle.



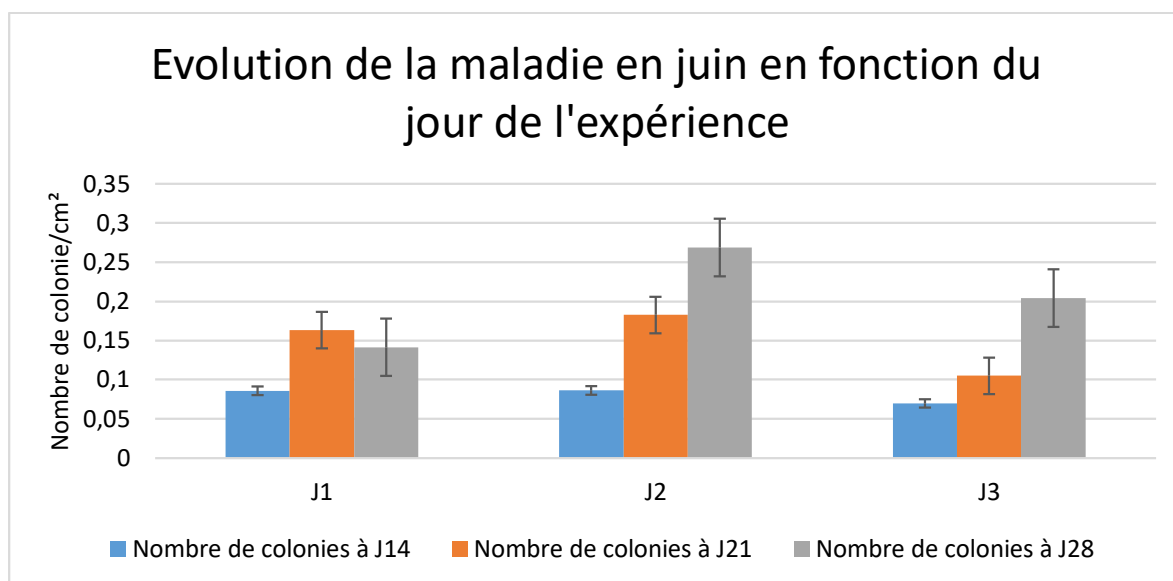


Figure 9 Histogramme de l'évolution de la maladie en juin en fonction du jour de l'expérience

De plus on peut observer que la maladie se développe plus quand elle est inoculée le matin plutôt que le midi ou le soir (figure 10). La littérature nous apprend que l'oïdium est sensible à des trop fortes chaleurs équivalent à celle rencontrée lors des inoculations le midi. (Sygenta 2020)

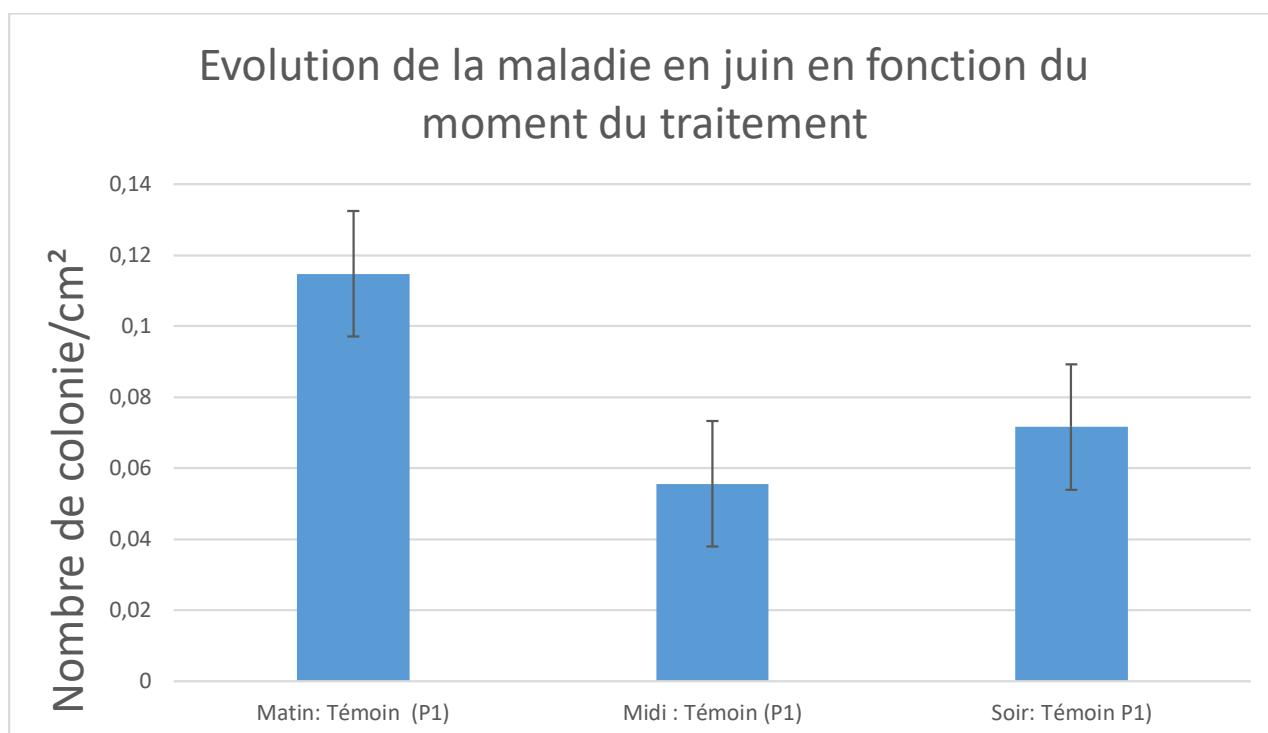


Figure 10 Histogramme de l'évolution de la maladie en juin en fonction du moment du traitement

Enfin une autre analyse des données permet de valider l'absence d'influence de la localisation des plantes sur le développement de la maladie (nombre de colonies d'oïdium/cm<sup>2</sup>). La position de la

plante dans le tunnel n'influence pas le développement de la maladie. Cette vérification est toujours nécessaire, en effet l'oïdium étant un champignon sensible à des variations des conditions climatiques (Blancard 2016), il aurait pu se développer différemment dans les tunnels et donc fausser les résultats

On a effectué une ANOVA pour justifier cela : On a relié les plantes à chacun des blocs auquel les plantes appartiennent, puis on a testé si le développement de la maladie est le même pour chaque groupe ou pas.

Notre hypothèse  $H_0$  ici est : Le développement de la maladie est le même dans tous les blocs. On obtient une p value de 0.12 ( $>0.05$ ) on ne rejette donc pas  $H_0$ , il n'y a pas d'influence spatiale.

L'étude des autres modalités en septembre donne des résultats similaires à ceux de juin, on a une p-value de 0.19 ( $>0.05$ ) pour la distribution spatiale.

On peut donc conclure à un développement différent de l'oïdium entre juin et septembre en termes de quantité mais avec la même dynamique de croissance. De plus, le développement de l'oïdium n'est pas influencé par des facteurs spatiaux ni pas les répétitions (jour 1, 2 et 3).

## L'effet du traitement sur le développement de la maladie

La deuxième partie de l'étude descriptive se focalise sur les données relatives au traitement des plantes avec l'agent de biocontrôle AQ10. En effet, selon toute vraisemblance le nombre de colonies sur les feuilles de tomate varie en fonction du traitement. Une rapide analyse valide cette hypothèse en juin et en septembre pour les notations à J14 et J21 (figure 11 et figure 12) Cependant il ressort aussi clairement qu'à J28, le traitement n'a plus aucun effet

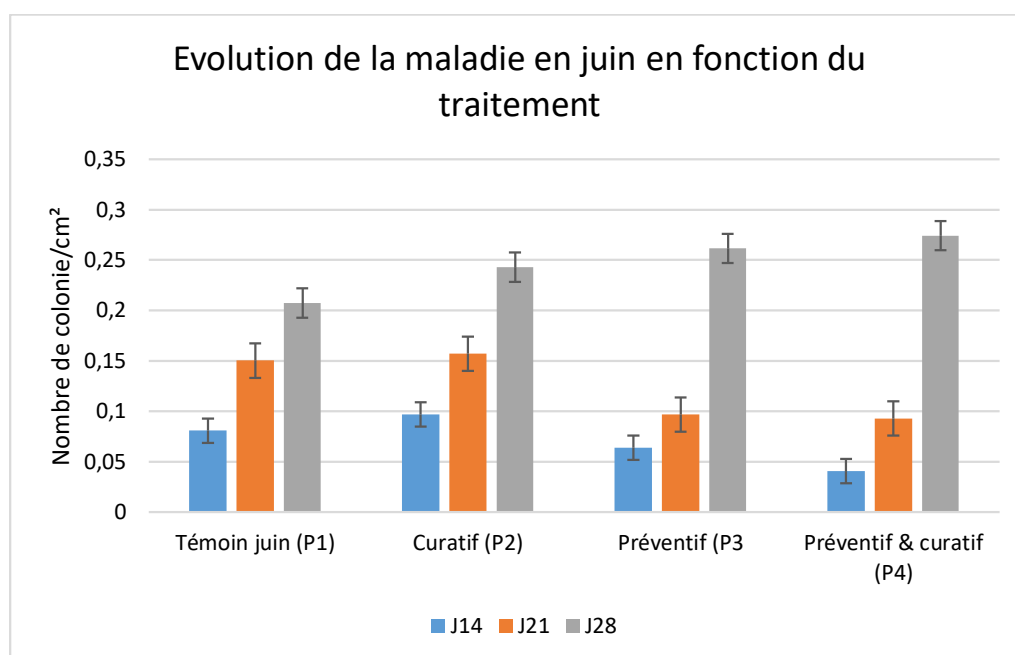


Figure 11 Histogramme de l'évolution du nombre de colonie en fonction des traitements en juin

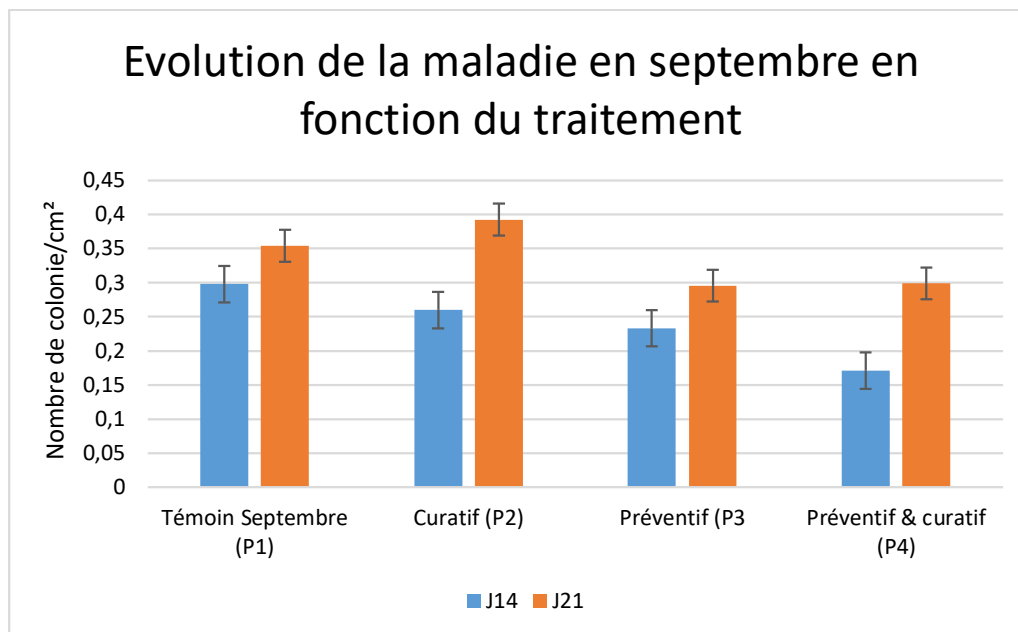


Figure 12 Histogramme de l'évolution du nombre de colonie en fonction des traitements en septembre

Une analyse plus approfondie à l'aide d'une ANOVA permet de savoir si ces différences sont significative

En premier temps nous n'allons pas faire de distinction entre les plantes recevant le traitement curatif et le traitement préventif.

Notre hypothèse nulle indique que les distributions sont semblables, quant à l'hypothèse H1, elle indique que les moyennes sont différentes.

Nous obtenons une p-value de 0.009 (inférieur à 0.05), la différence des moyennes chez les plantes traitées et les plantes non traitées est statistiquement significative.

Ces résultats suggèrent que le traitement préventif à base de AQ10 a un effet significatif sur le développement de la maladie lors des comptages à 14 jours et à 21 jours. Aucune différence entre les plantes traitées et celles non traitées n'est observée pour la notation réalisée à 28 jours. Ceci suggère que le traitement bien que réduisant la maladie ne l'éradique pas.

On note qu'ici les tests ont été fait vis-à-vis de la variable nombre de colonies, mais on obtient des résultats similaires avec la variable nombre de spores.

	Plantes 4 et 1	Plantes 4 et 2	Plantes 3 et 1	Plantes 3 et 2	Plantes 2 et 1	Plantes 3 et 4
P value J14	Différent	Différent	Différent	Différent	Similaire	Limite
P value J21	0.015	0.019	0.023	0.035	0.07	0.048

Avec la confirmation de l'effet du traitement sur les plantes, le but est maintenant de faire une distinction entre les traitements curatifs et préventifs.

Tout d'abord, les plantes non traitées et les plantes traitées avec le traitement curatif ne sont pas significativement différentes (p-value de 0.07). Le traitement curatif seul n'a pas d'efficacité.

Pour l'ANOVA entre les plantes traitées avec un traitement préventif, et les plantes non traitées nous avons obtenues une p values de 0.023. Ce qui signifie donc que les différences entres les moyennes sont statistiquement significatives.

Enfin pour la comparaison entre les plantes traitées avec le traitement curatif + Préventif, on se trouve dans limite de significativité, car on a une p value de 0.048 et qui est très proche de 0.05.

On peut synthétiser les résultats en 3 points majeurs :

1. Le traitement préventif réduit la présence de l'oïdium pendant 21 jours uniquement
2. Le traitement curatif seul ne réduit pas significative la présence de l'oïdium
3. L'addition d'un traitement curatif après un traitement préventif ne réduit que très peu la présence de la maladie limite de significativité

## Le pourcentage de protection pour quantifier l'efficacité du traitement

En utilisant le pourcentage de protection d'autres observations peuvent être effectuées. En effet le pourcentage de protection du traitement préventif moyen à 14 jours est de 41 % en juin et uniquement de 29% en septembre lors du premier comptage. À 21 jours le pourcentage de protection du traitement préventif moyen n'est plus que de 35% en juin et 16% en septembre. Le pourcentage de protection est même nul lors du comptage de 28 jours. Ces premières observations démontrent que bien le traitement est un effet sur le nombre de colonies et de spores, il reste modéré. En effet il n'est pas acceptable pour un agriculteur d'avoir 50 % de ses cultures qui restent contaminé par l'oïdium après un traitement

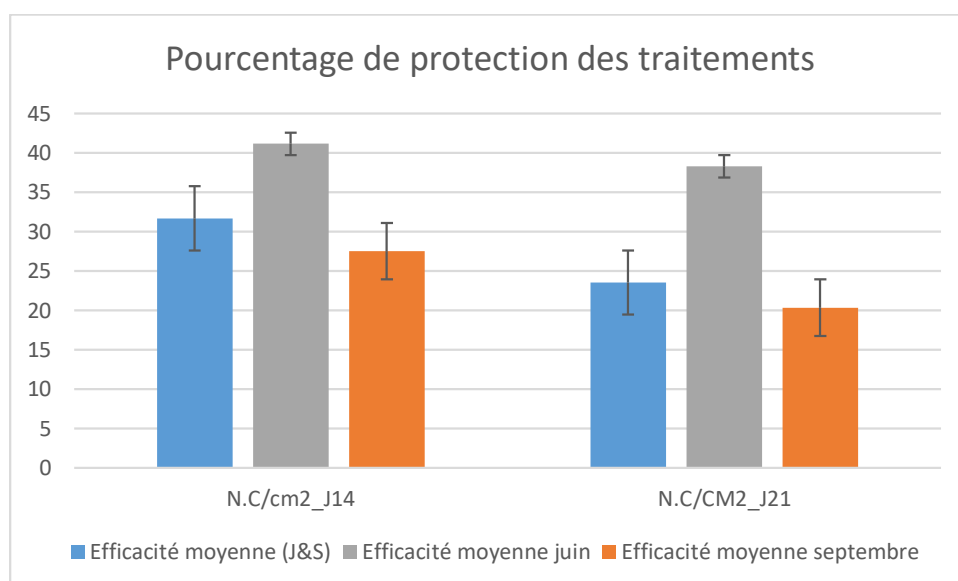


Figure 13 Histogrammes comparant les efficacités préventives en J14 et J21 en juin et septembre

En comparant les efficacités lors de trois modalités de traitement (matin, midi et soir), on observe que le traitement est significativement plus efficace lorsqu'il est réalisé le midi plutôt que le matin ou le soir, on en a fait aussi le test avec une ANOVA et nous avons obtenu une p-value=0.041 (figure 14).

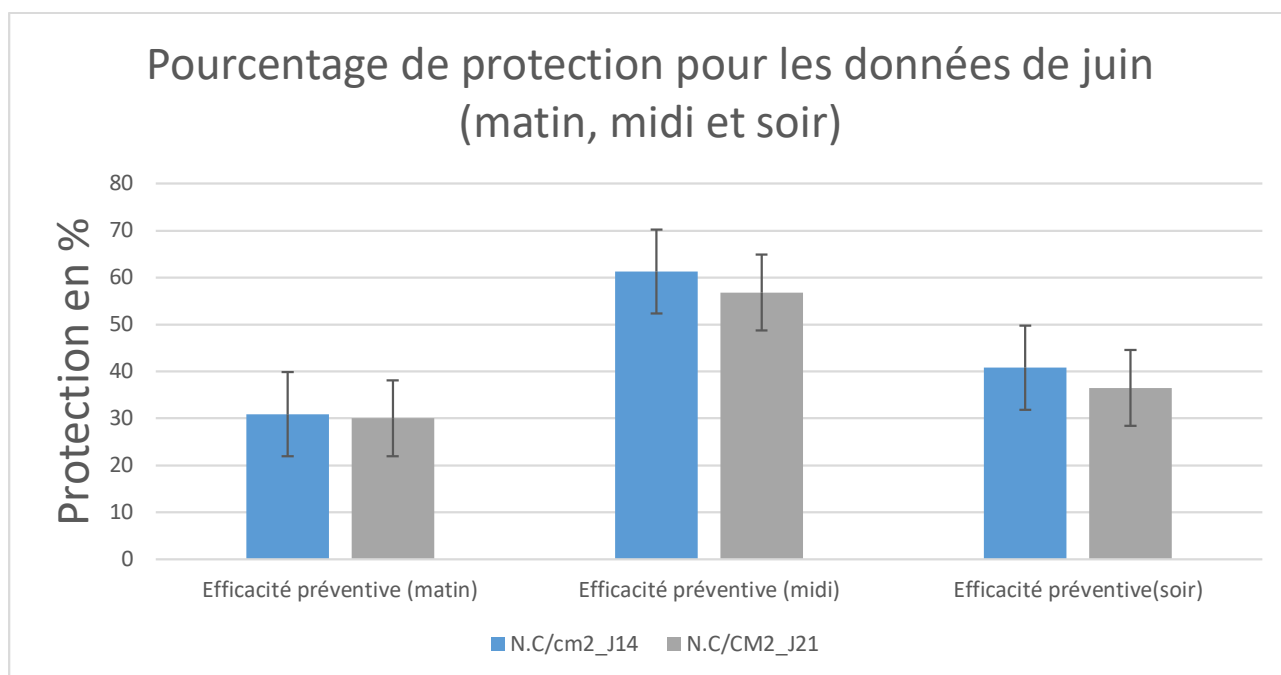


Figure 14 Histogrammes comparant les efficacités pour les données de juin (matin, midi et soir)

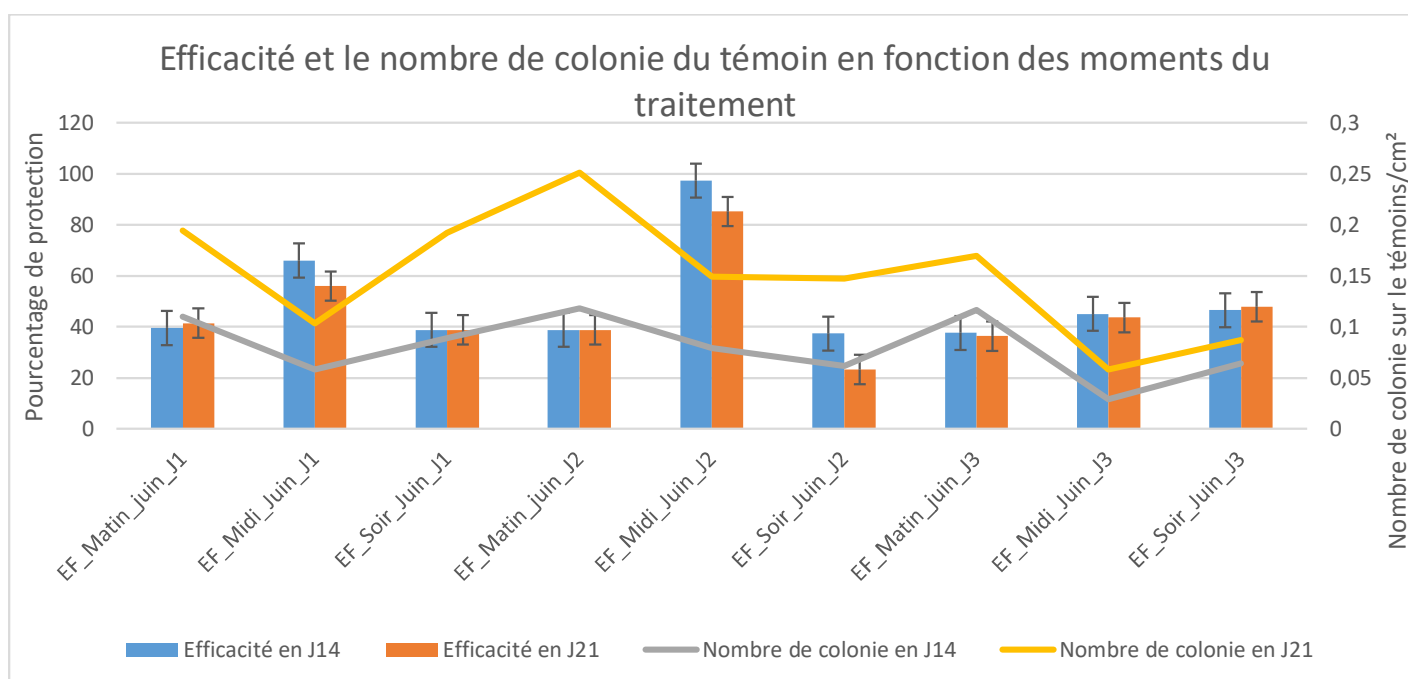


Figure 15 Histogramme avec courbes représentant l'efficacité et le nombre de colonie du témoin en fonction des moments du traitement

Comme indiqué lors de la comparaison des plantes, le pourcentage de protection du traitement curatif on s'aperçoit qu'il est nul en juin comme en septembre. Si on s'intéresse au taux de protection du traitement préventif et curatif, 38% (juin) et 15% (septembre), on remarque que l'ajout d'un traitement curatif n'améliore que très faiblement la protection (peu significatif). Enfin le taux de protection est nul en J28 (voir négatif) peu importe le type de traitement.

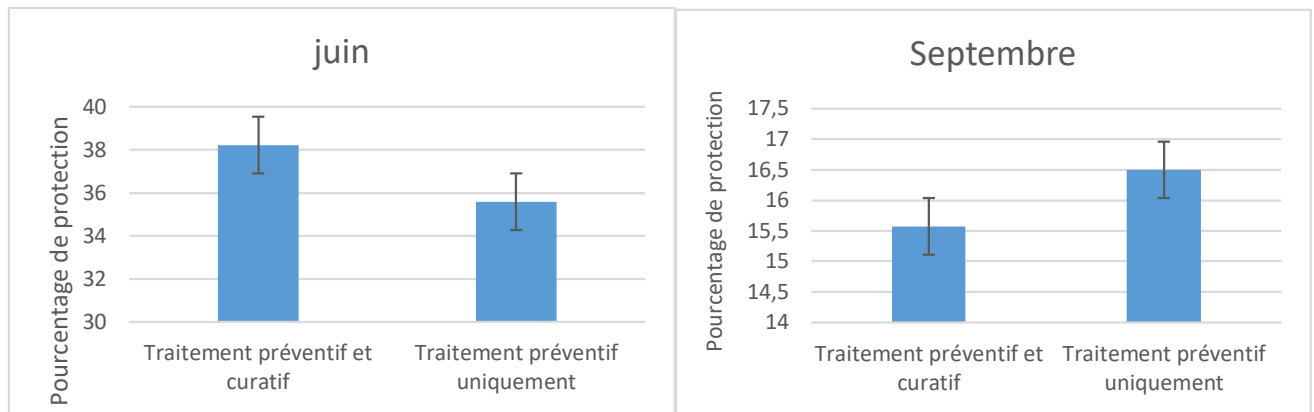


Figure 16 Histogrammes représentant la comparaison entre traitements préventifs et curatifs et préventifs uniquement pour les notations de juin et septembre lors du deuxième comptage à 21 jours

Plusieurs conclusions ressortent de ces analyses de l'efficacité du traitement : tout d'abord le traitement préventif diminue le développement de la maladie de manière certaine (40%), mais sur une durée limitée. Son impact paraît extrêmement variable selon les modalités d'application (matin, midi et soir) et reste plus important pour les traitements effectués le midi. D'autre part, l'efficacité du traitement curatif seul avec AQ10 est nulle voire négative. De ce fait, ce traitement n'a pas été considéré dans la suite des analyses. Enfin l'ajout d'un traitement curatif 14 jours après inoculation n'influence que très peu le niveau de protection contre la maladie.

Ces conclusions nous amènent naturellement vers l'analyse approfondie des variables différenciant les modalités matin, midi et soir, c'est-à-dire les variables climatiques mesurées lors du traitement. Ces derniers vont faire l'objet du chapitre suivant.

# Chapitre 2 : Quantification de l'impact des facteurs climatiques sur l'efficacité du biocontrôle

## Introduction et méthodes

Il a été démontré dans la partie précédente que le traitement bien qu'ayant un effet sur le développement de la maladie a des résultats très variables. Cette partie a pour but d'analyser et de comprendre les facteurs expliquant cette versatilité. Selon les hypothèses admises ces facteurs sont principalement d'ordre climatiques. L'étude va donc analyser l'impact des variables climatiques températures et humidités sur les variables du développement de la maladie (nombre de colonies et de spores lors des 3 comptages) ainsi que sur l'efficacité globale du traitement.

Le jeu de données contient 252 individus (soit 252 feuilles de plantes) et 61 variables, 50 variables quantitatives sont illustratives, 2 variables qualitatives sont illustratives (variable traitement). Les autres variables qualitatives permettent de localiser les individus dans les serres, et ne sont pas utilisées pour ces analyses, elles nous servent principalement à localiser les valeurs corrompues, et comprendre la raison de leurs différences, et aussi permettent de relier chacune des plantes avec les capteurs de températures et d'humidités correspondants.

La méthode utilisée est une étude factorielle. C'est une étude qui aide à simplifier la compréhension et le traitement de données assez complexes telles les nôtres. Les deux bases de données étudiées dans cette section comportent plusieurs variables, c'est pour ça que l'analyse factorielle semble appropriée. Des méthodes statistiques vont être utilisées pour regrouper les plantes et les variables ayant un comportement similaire pour obtenir les analyses les plus pertinentes. En conséquence du type de données, la méthode utilisée sera celle des ACP. Elles permettent de transformer les variables qui sont très corrélées entre elles en des axes décorrélés, ce qui va réduire le nombre de variables à étudier, et simplifiera l'interprétation. Des AFDM et des ACM ont été utilisées pendant les analyses mais non présentées dans le corps du rapport.

Les sorties de l'ACP vont être utiles aussi lors du 3ème chapitre. On va effectuer une CAH sur le plan factoriel que nous avons construit avec l'ACP, ce qui reviendra donc à se baser sur les plans factoriels et non seulement les variables pour interpréter les clusters, et puis déterminer un profil climatique.

Une fois l'étude des axes terminés, l'étude a été complétée avec des matrices de corrélation représentant les différents liens existants entre les variables (et non les axes comme sur une ACP) en utilisant des tests statistiques pour exprimer la corrélation et autre.

## Analyse factorielle sur les variables de l'étude

La première étape dans la démarche consiste à réaliser une ACP rassemblant toutes les variables afin d'établir les interactions entre les variables climatiques et celle représentant le développement de la maladie.

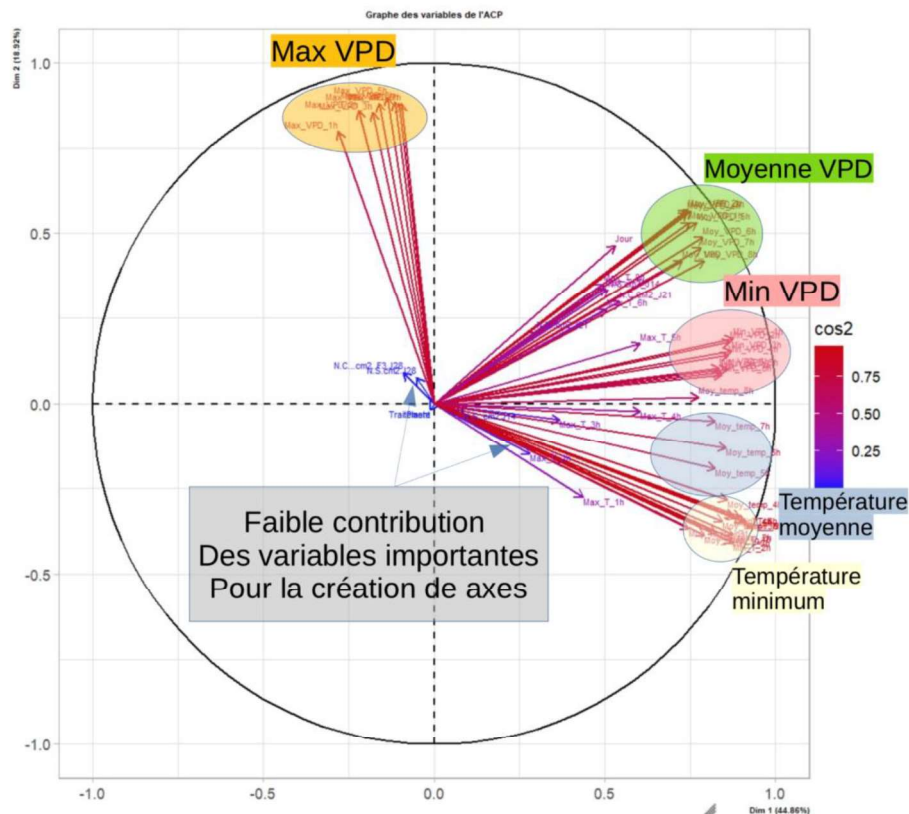


Figure 17 Analyse factorielle avec toutes les variables prises en compte

On remarque plusieurs faisceaux de variables regroupés. D'un côté un premier faisceau qui contient les variables qui décrivent les valeurs moyennes de températures au cours des heures qui suivent l'inoculation de l'oïdium et le traitement. Un deuxième faisceau regroupant les variables représentant les valeurs minimums de températures lors des heures qui suivent l'inoculation. Puis trois autres, un pour les moyennes de VPD, le deuxième pour les valeurs minimums de VPD, et le troisième pour les valeurs maximums de VPD.

On note aussi que les variables qu'on souhaite expliquer (Nombre de colonies, nombre de spores, et traitement) grâce aux variables explicatives (variables climatiques) ne sont pas bien représentées dans le plan 1 et 2 de l'ACP. Afin de remédier à ce problème, nous allons modifier le coefficient CTR des variables climatiques en sorte de construire nos axes en fonctions des variables les plus pertinentes à notre étude. Car le graphique des variables ne correspond pas à l'étude qu'on souhaite faire, et ne peut pas fournir les informations utiles pour l'étude : les variables climatiques absorbent l'effet des autres variables.



On effectue maintenant une ACP en prenant les variables climatiques comme variables supplémentaires, ces variables vont donc être visibles dans notre ACP, par contre elles ne vont pas intervenir lors de la création des axes.

On note que dans le graphique des variables suivant, les variables nombre de colonies ainsi que les variables du traitement ont les meilleures qualités de la présentation sur cette dimension parmi nos variables.

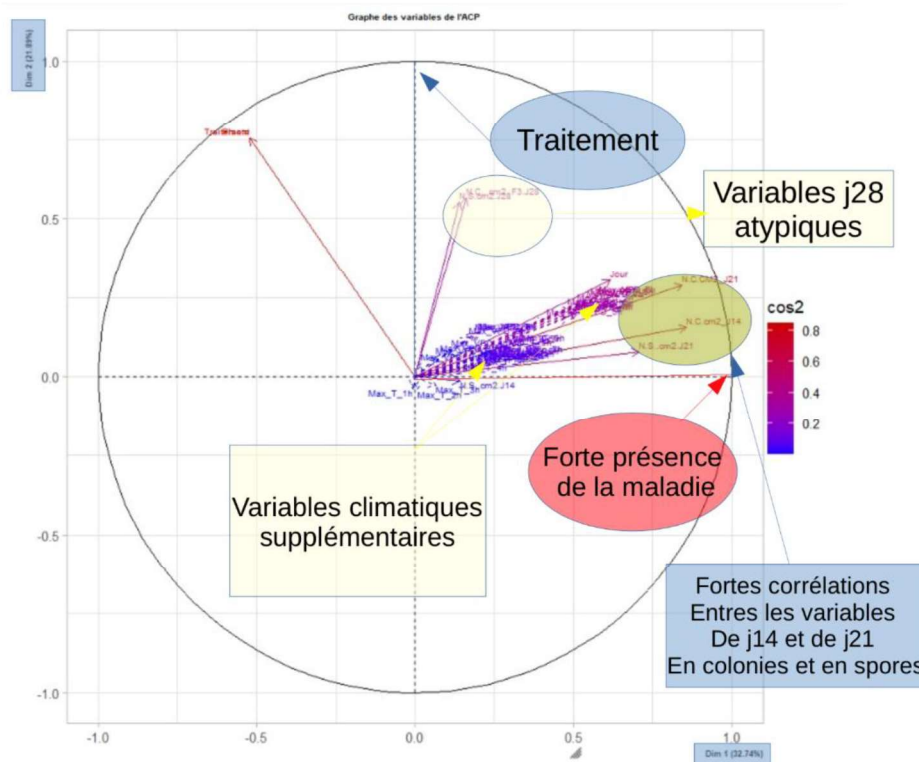


Figure 18 AFDM avec variables d'intérêts contribuant à la construction des axes

On rappelle que les variables en traits complet sont les variables actives, celles en traits pointillés sont les variables supplémentaires avec un coefficient CTR=0. Les variables libellées sont celles les mieux représentées sur le plan.

Les variables actives sont représentées selon le coefficient  $\cos^2$ . On note que ce coefficient représente la qualité de présentation des variables sur la dimension présentée, plus la couleur est rouge, plus le  $\cos^2$  est grand et plus on a une meilleure qualité de la présentation des variables sur cette dimension, et plus c'est bleu, plus la valeur du  $\cos^2$  se rapproche de 0 et donc plus la qualité de la présentation des variables sur cette dimension est mauvaise.

Ce graphique nous permet d'avoir un premier aperçu sur les corrélations entre les différentes variables contenues dans notre jeu de données. On remarque un comportement très proche entre les données de J14 et de J21 concernant le nombre de colonies ainsi que le nombre de spores/cm<sup>2</sup>. On peut donc en déduire que le nombre de colonies ou de spores chez une plante au J21 est proportionnel à celui

de J14. Cela confirme aussi le comportement atypique pour les données de J28 : on a une faible corrélation entre les variables symptomatiques de l'oïdium au J14 & J21 avec celles du J28, ce qui conforte la conclusion du chapitre précédent sur l'état des plantes au J14, J21 et J28. Les plantes contaminées par l'oïdium vont finir par atteindre le même stade au J28 : elles vont devenir toutes très contaminées. On observe donc par la suite qu'à partir d'une certaine date, les plantes traitées et les plantes témoins vont avoir le même stade d'avancement de la maladie. Le produit du traitement, AQ10 a donc plutôt un effet retardateur que réellement curatif selon les résultats de l'expérience. De plus, des traitements de rappels sont peut-être nécessaires pour assurer une continuité dans la protection.

Ces informations viennent confirmer les observations précédentes sur la durabilité du traitement au-delà de 21 jours. Ceci pourrait être dû à une inoculation secondaire ; c'est-à-dire une contamination issue des plantes témoins. En effet après un certain stade de l'infection, l'oïdium est capable de former des spores pouvant aller contaminer d'autres plantes et ainsi diminuant l'efficacité du traitement.

Le choix des dimensions de l'ACP se justifie par une analyse de l'inertie afin de savoir le nombre de dimensions choisies pour l'ACP.

#### Analyse de l'inertie des axes

On rappelle que l'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'analyse expriment **67.93%** de l'inertie totale du jeu de données ; cela signifie que 67.93% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage assez important, et le premier plan représente donc convenablement la variabilité contenue dans une grande part du jeu de données actif, la variabilité expliquée par ce plan est donc hautement significative.

On peut observer cela avec le graphique de décomposition de l'inertie :

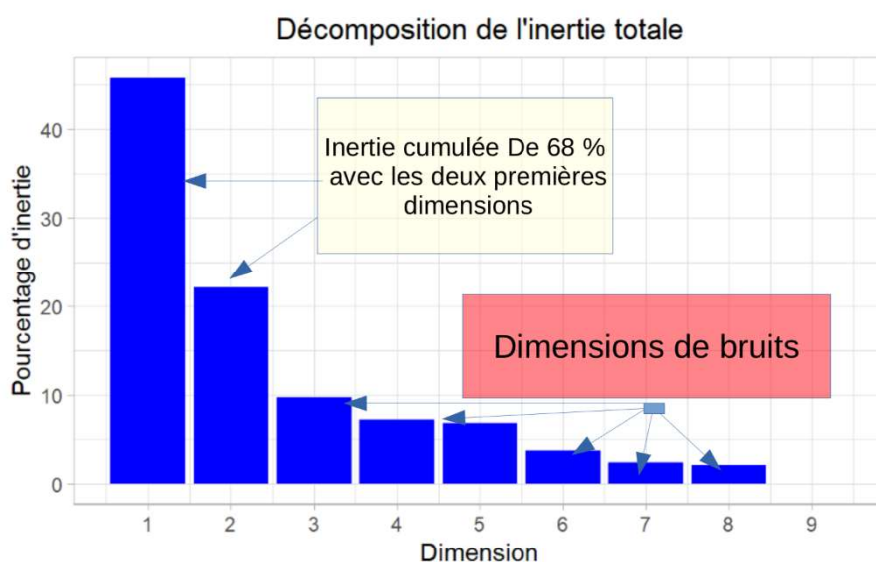


Figure 19 Décomposition de l'inertie total par rapport aux nombres de dimension

On observe que le plus grand de l'inertie est cumulé avec les deux premières dimensions, l'étude se limite donc à 2 dimensions.

Maintenant que nous avons déterminé le nombre de dimension que nous allons traiter, on va les analyser.

### Analyse des dimensions

Dimension 1 :

La **dimension 1** oppose des individus tels que 241 et 181 (à droite du graphique, caractérisés par une coordonnée fortement positive sur l'axe) à des individus caractérisés par une coordonnée fortement négative sur l'axe (à gauche du graphe 19 et 20 en dessous).

Cette dimension se caractérise par :

- Plus les plantes ont une grande valeur vis-à-vis de l'axe 1, plus la maladie est présente chez elles, c'est à dire que les plantes ont de grandes valeurs concernant le nombre de colonies ainsi que le nombre de spores, et ceci pour le J14 et le J21. On remarque aussi que les plantes qui ont de très grandes valeurs de colonies et spores sont principalement celles qui n'ont subi aucun traitement, ou celles qui ont eu que le traitement curatif.
- Les plantes ayant de fortes valeurs sur l'axe, ont de fortes valeurs de VPD à partir de 7h après le traitement des plantes, on note que le plus la VPD est grande, plus l'humidité est faible.

La **dimension 2** oppose des individus caractérisés par une coordonnée fortement positive sur l'axe 2 (en haut du graphe) à des individus comme 241 et 181 (en bas du graphique, caractérisés par une coordonnée fortement négative sur l'axe).

Cette dimension se caractérise par :

- Plus on a une grande valeur par rapport à cet axe, le plus la plante a subi de traitement : Les plantes qui ont les plus grandes valeurs pour cet axe ont toutes eu au moins le traitement préventif, si ce n'est pas le traitement préventif + curatif (une vérification a été faite pour ce point).

On observe maintenant la distribution des plantes sur le plan construit par l'axe 1 et 2 par rapport aux différentes modalités de la variable traitement :

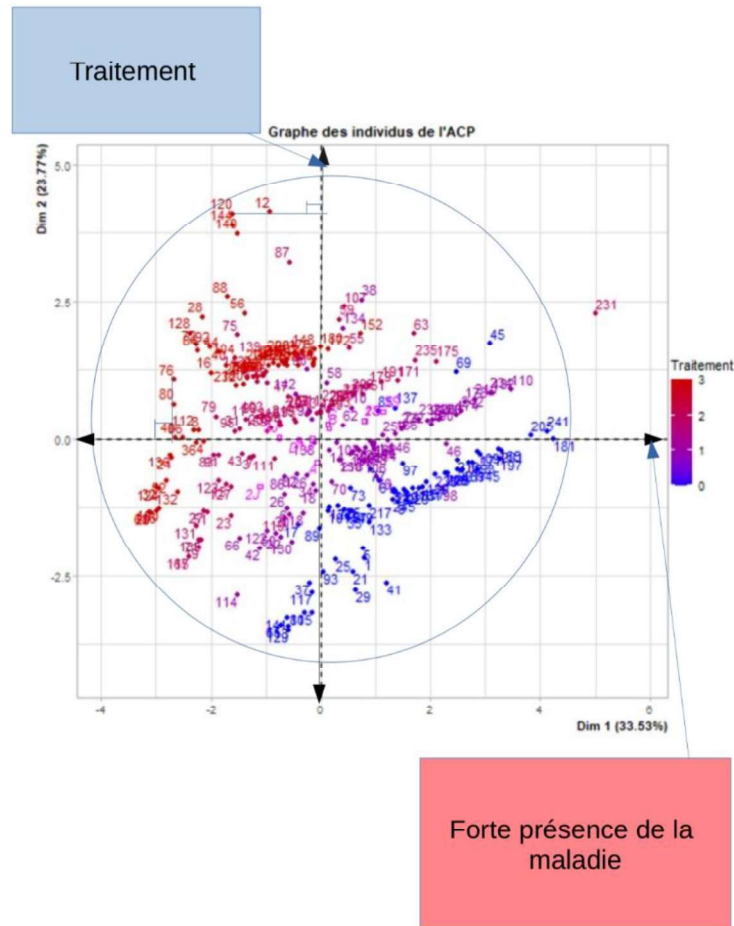


Figure 20 Projection des individus sur l'ACP avec variables d'intérêts contribuant à la construction des axes par rapport à la variable traitement

On remarque un certain regroupement des individus par modalité de la variable traitement.

On rappelle que dans ce contexte, la modalité 0 (en bleu) correspond à l'absence de traitement, la modalité 1 correspond au traitement curatif à 14 jours, la modalité 2 correspond à un traitement préventif à 0 jour, et la modalité 3 aux deux traitements curatifs et préventifs à 0 et 14 jours.

En projetant les groupes des individus sur les axes de l'ACP, on retrouve les résultats suivants :

- Il existe un effet du traitement : les plantes qui n'ont pas subi de traitement présentent plus de symptômes de la maladie (colonies et spores) que celles qui ont eu un traitement.
- Le traitement curatif n'a que peu d'efficacité par rapport au témoin
- Les plantes qui ont eu les deux traitements représentent légèrement moins de symptômes de présence de maladie que celles qui n'ont eu qu'un seul.

On rappelle aussi que le comptage au J28 ne distingue pas les individus qui ont été traités ou pas. Les variables liées aux comptages de J28 rajoutent un certain bruit si on souhaite étudier toutes les données ensemble, donc pour la suite on va ajouter les variables concernant le nombre de colonies et de spores du J28 comme variables supplémentaires :

### Etude des données en excluant les valeurs de J28

En retirant J28 :

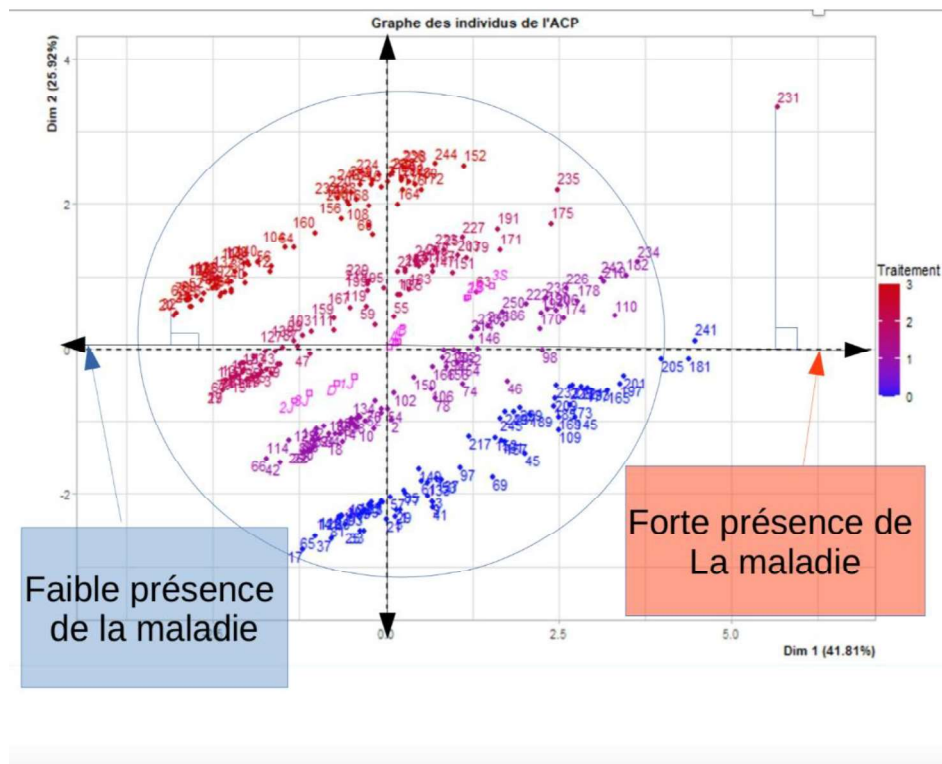


Figure 21 Projection des individus sur l'ACP avec variables d'intérêts contribuant à la construction des axes par rapport à la variable traitement en retriçant les données relatives à J28

En ajoutant les variables de J28 comme variables supplémentaires, on arrive à bien distinguer les plantes selon leurs traitements, on valide les résultats précédant.

On retrouve une distinction très claire entre les plantes qui ont subi les deux traitements, celles qui ont subi que le traitement curatif, ou que le traitement préventif, et celles qui n'ont eu aucun traitement. On peut projeter très facilement les individus sur les axes, et voir le réel intérêt du traitement préventif, et le faible intérêt du traitement curatif seul.

On observe donc que le traitement curatif seul n'est pas efficace, mais associé au traitement préventif les résultats deviennent meilleurs.

## Représentation des facteurs climatiques via l'ACP

Après avoir préparé le jeu de données et vérifié les conclusions des chapitres précédents, on cherche à établir les liens entre les variables de la maladie et les variables climatiques

On affiche maintenant la distribution des plantes sur le plan construit par l'axe 1 et 2 par rapport aux différentes modalités de la variable température instantanée :

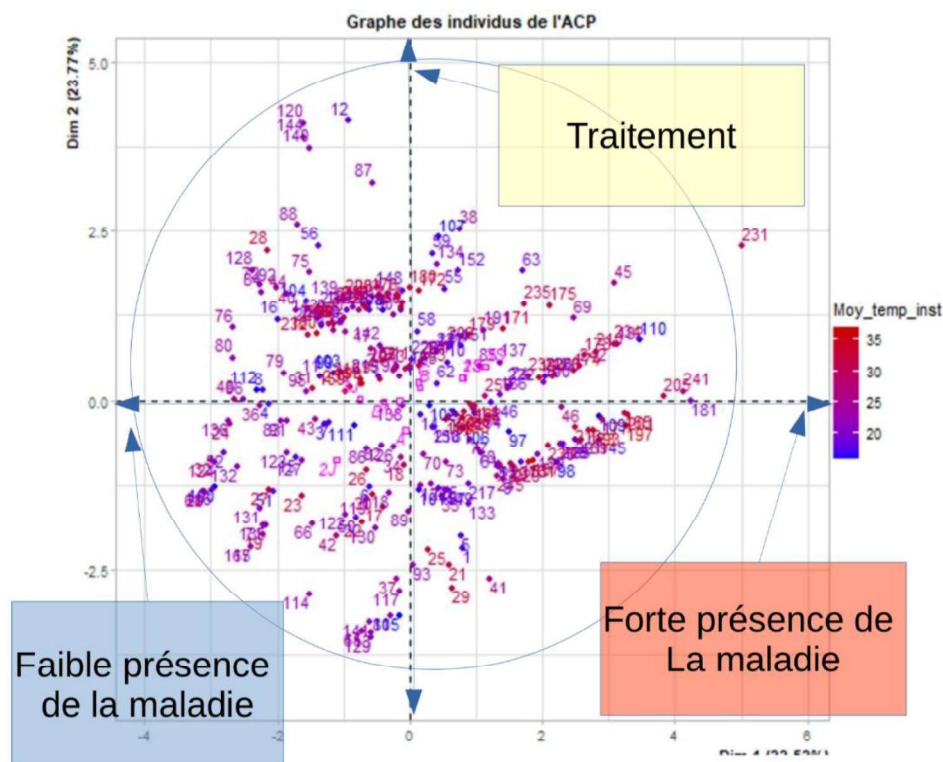


Figure 22 Projection des individus sur l'ACP avec différenciation selon la température lors du traitement dit « Moy\_temp\_inst »

La distribution des plantes selon la température ou la VPD au moment du traitement sur les deux axes que nous avons sélectionné ne permet pas de dégager de conclusion. Il en va de même avec toutes les autres variables climatiques créées (moyenne heure par heure, extremum et augmentation. En effet les variables climatiques ne semblent pas suivre de schéma de corrélation simple du genre « plus il fait chaud au moment du traitement plus le traitement diminue le nombre de colonie d'oïdium ». Notre hypothèse est donc que l'influence des facteurs climatiques est modérée voire faible et surtout dilué dans les nombreuses variables calculées qui n'arrivent pas à être synthétisées au sein des dimensions choisies.

Pour résoudre ce problème et vérifier l'hypothèse, une nouvelle approche va être essayée. Elle consiste à calculer les corrélations entre les différentes variables qui expriment la présence de la maladie, et les variables climatiques deux à deux pour connaître l'influence de chacune.

## Calcul des corrélations

Des tests statistiques de corrélations ont été effectués entre chacune des variables afin d'étudier les relations entre elles. Par la suite nous avons sélectionné les corrélations qui nous semblent les plus informatives pour étudier les effets climatiques sur le traitement. Nous avons représenté les corrélations calculées sous forme de matrice afin de faciliter la visualisation et d'avoir à chaque fois une vue globale sur l'ensemble des corrélations. D'autres matrices de corrélations ont été ajoutées en annexe.

Selon les conditions, différents tests ont été appliqués :

Dans des conditions favorables : variables quantitatives continues, distribution gaussienne des valeurs, moyenne et écart type connus... des tests de Pearson ont été favorisés. Dans le cas contraire nous avons appliqué des tests de Kendall ou de Spearman, les résultats pour les deux derniers sont identiques.

On commence par visualiser la matrice qui contient les variables d'efficacités, ainsi que celles de températures (minimum, maximum, et moyenne) sur la figure 22 :

On retrouve une très grande corrélation entre l'efficacité en nombre de colonies par  $\text{cm}^2$  au J14 et au J21. Cette corrélation est d'ordre 65%. Donc quand on a une bonne efficacité en j14, elle reste maintenue en j21 soit 7 jours après.

On note aussi que l'efficacité en nombre de spores en J21 est corrélée avec l'efficacité en nombre de colonies du même jour, mais cette corrélation n'est pas très grande, car en effet on a pu observer dans les données que la liaison est faite dans un sens plus que dans l'autre : l'absence des spores implique l'absence des colonies, mais l'absence des colonies n'impliquent pas l'absence des spores : On a pu constater cela chez plusieurs feuilles qui n'ont pas colonies, mais ont tout de même des spores.

On observe aussi une corrélation faible entre l'efficacité en nombre de colonies et de spores, et le nombre de colonies chez les témoins, ce qui signifie donc, que la pression de l'inoculum impacte légèrement l'efficacité du traitement ; en effet un inoculum plus faible implique une maladie plus facilement jugulée par le traitement

On observe que les plus grandes corrélations sont celles qui sont entrées les variables de températures. Ce qui est logique car la température en à l'heure h est très corrélée à celle à l'heure h+1, h+2...

On note que les corrélations varient entre faibles et moyennes entre les températures et les efficacités en colonies et en spores, ainsi qu'avec le nombre de colonies chez les témoins. On parle d'une corrélation qui atteint pour certaines variables de température les 50%, mais pour une bonne moitié des variables elle ne dépasse pas les 35%. Ces résultats corroborent les précédents tests statistiques effectués avec les analyses factorielles.



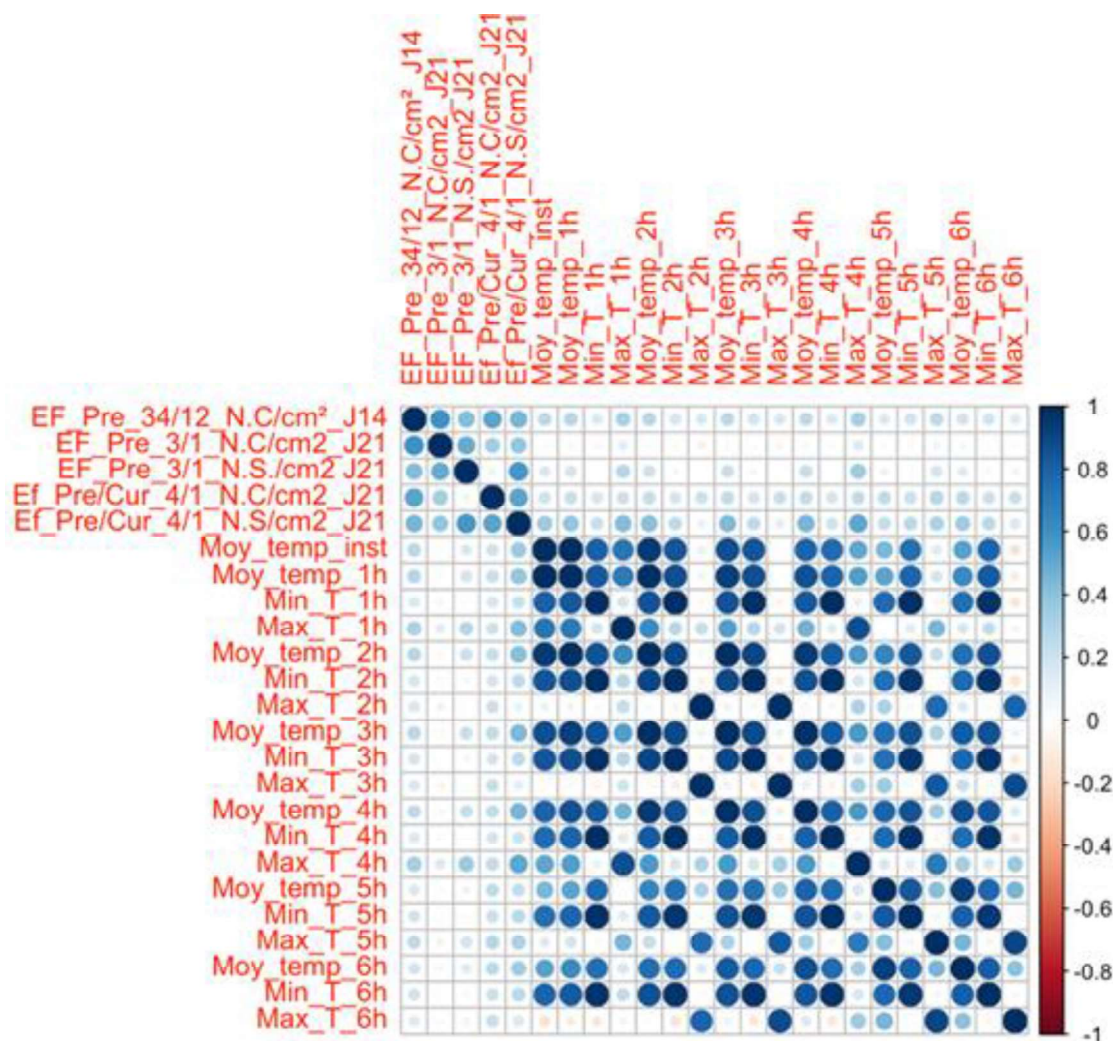


Figure 23 Matrice de corrélation entre les efficacités et les variables de température

On remarque aussi que les température moyennes et maximums ont de plus grandes corrélations avec les variables d'efficacités que les températures minimums. Les températures minimums ne varient pas autant que les températures moyennes et maximum lors de l'expérimentation, on note que l'expérience a été faite en mois de juin et de septembre, des mois ou les températures peuvent atteindre des seuils maximums élevés.

On note aussi que les variables de températures ont de plus grandes valeurs de corrélations avec la variable efficacité préventive + curative qu'avec les autres variables qui expriment que l'efficacité préventive.

On remarque aussi de très faibles valeurs de corrélations entre la variable qui exprime l'état du traitement préventif effectué au j21 avec les variables climatiques. On a quasiment une absence de corrélation. Ceci pourrait être expliquer par la période importante entre le traitement et l'inoculation d'un côté et le comptage à 21 jours. En effet même s les facteurs climatiques ont un impact comme on peut le voir en lors du comptage à J14, il reste faible et cet effet ne se retrouve presque plus dans les comptages suivants.



On traite maintenant la matrice qui contient les variables d'efficacités, ainsi que celles de VPD (minimum, maximum, et moyenne) :

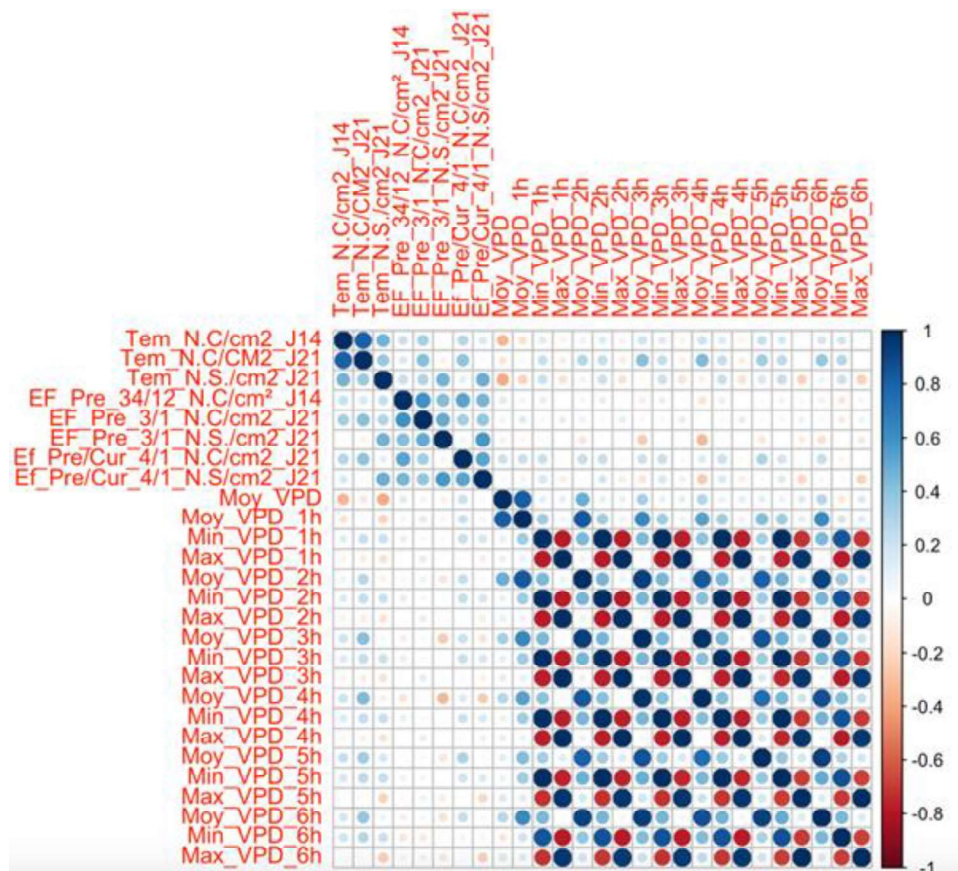


Figure 24 Matrice de corrélation entre les efficacités et les variables de VPD

On remarque que l'efficacité préventive en nombre de spores à J21 est corrélée négativement avec toutes les variables de VPD. Ces corrélations peuvent atteindre les -35%. Par contre l'efficacité préventive en nombre de colonies à j21 est corrélée positivement avec toutes les variables de VPD avec des corrélations qui varient autour de 20%. Ce résultat est donc assez paradoxal car l'efficacité en nombre de colonies et l'efficacité en nombre de spores sont très corrélées entre elles (corrélation autour de 70%).

On retrouve une corrélation de -57% entre la moyenne VPD et le nombre de spores/cm<sup>2</sup> chez les témoins au J21 ainsi que le nombre de colonies /cm<sup>2</sup> chez les témoins en j14. C'est une corrélation négative assez forte. Elle indique que si la VPD est élevée au moment de l'inoculation, le nombre de colonies ainsi que le nombre de spores diminue, en effet une VPD élevée signifie un environnement plus sec et donc moins favorable au développement de l'oïdium.

On note aussi que les plus grandes corrélations obtenues (hors celles qui sont entre les variables construites à partir de la VPD), sont celles qui sont entre le nombre de colonies chez les plantes témoins et les variables qui expriment la VPD moyennes : corrélations peuvent atteindre jusqu'à 60%.

On analyse maintenant la troisième matrice qui contient les variables d'efficacités, ainsi que celles de l'accélération de la température.

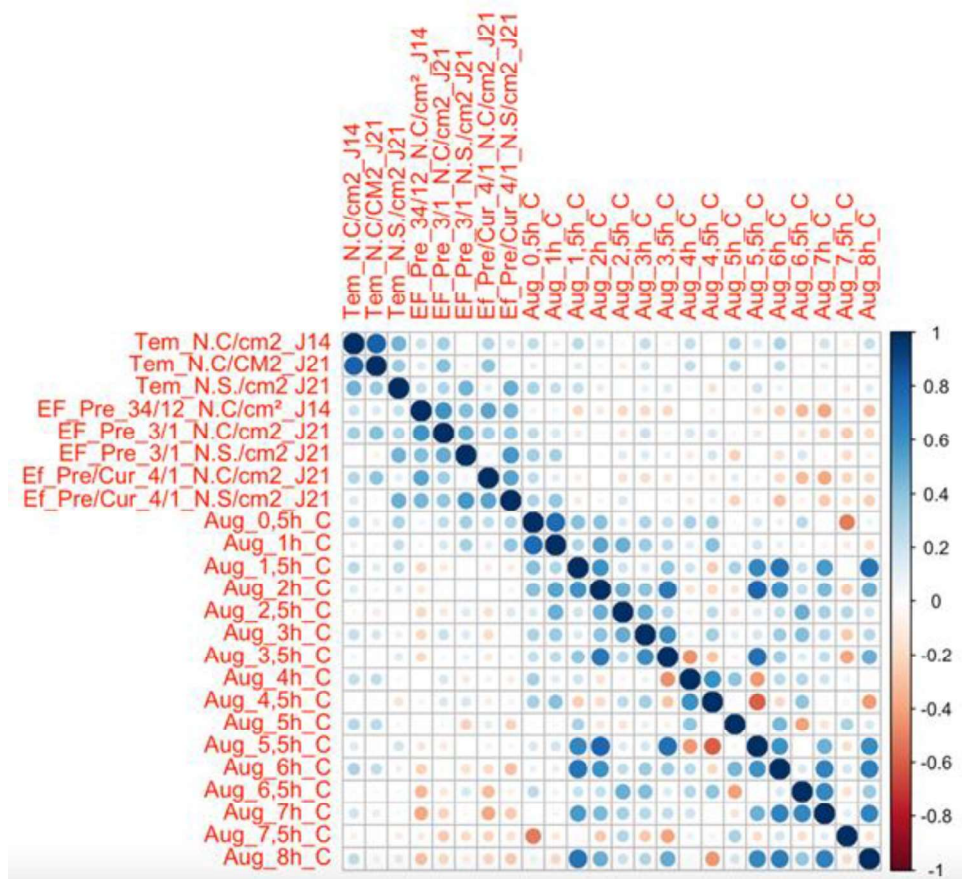


Figure 25 Matrice de Corrélation entre les efficacités et les variables d'augmentation de la température

Tout d'abord, on remarque que chez les plantes témoins, l'augmentation de la température est toujours corrélée positivement avec l'augmentation du nombre de colonies, on peut alors supposer que l'augmentation de température est propice à l'augmentation du nombre de colonies.

On remarque aussi que globalement les efficacités préventives sont corrélées négativement avec les augmentations de températures. On voit cet effet encore plus pour les augmentations de températures supérieures 5.5h après le traitement que pour les premières heures qui le suivent.

On note aussi que les plus grandes valeurs de corrélation négatives des efficacités sont obtenues pour l'augmentation de température 7h après les traitements.

On remarque que les corrélations entre les efficacités préventives et les augmentations de températures ne dépassent pas les 30%.

### Conclusion sur les variables impactant les efficacités du produit :

L'étude factorielle, et par la suite les tests statistiques nous ont permis de voir plusieurs points importants, tout d'abord la différence très claire entre les plantes qui ont eu un traitement préventif et celles qui ne l'ont pas eu. On observe une corrélation entre les variables climatiques et les efficacités d'environ 30 %. Ceci reste relativement faible et explique les difficultés rencontrées avec l'approche factorielle. Néanmoins la variation de température et de l'humidité impact bel et bien l'efficacité du produit mais de manière faible et diffuse environ 30% pour de nombreuses variables liées à l'humidité ou à la température. L'approche factorielle à elle seule, ne permet pas d'établir un profil d'efficacité, qui serait la base afin de pouvoir établir des règles de décision. C'est pour atteindre cet objectif que des approches plus poussées ont été testé et comparées dans le chapitre suivant.

## Chapitre 3 : Classification des données, et comparaison des outils de modélisation pour déterminer des règles de décision les plus adéquates dans le cas rencontré.

### Introduction de la partie

Dans cette partie l'objectif est d'établir un profil qui décrit les conditions climatiques qui permettent d'obtenir les meilleures efficacités.

Plusieurs méthodes empiriques plus ou moins complexes ont été réalisées pour atteindre cet objectif.

La première approche est la plus simple. Elle se réalise via PowerBI et permet une classification des individus selon leurs efficacité et leurs variables climatiques

On a commencé par faire diverses régressions entre variables pour étudier plus en détails leurs liens, et par la suite modéliser l'efficacité en fonction des variables descriptives par des PLS, Régression logistique à modalités ordonnées et en effectuant des régularisation ridge/lasso selon le besoin. Le résultat de ces modèles n'a pas été très concluant, soit à cause de faibles  $R^2$ , soit à cause de scores assez bas pour le côté prédictif du modèle.

La troisième approche consiste à une classification par forêt aléatoire qui a donné de meilleurs résultats. Deux autres classifications : Classification Hiérarchique ascendante, et une classification par seuil d'efficacité. Cette dernière sert principalement de référence de comparaison entre les deux autres méthodes.

Les trois approches de classifications sont présentées et analysées en présentant à chaque fois, leur mise en place, leur réalisation ainsi que les forces et faiblesse de chacune.

Leur méthodologie respective sera décrite dans chaque partie avant de présenter laquelle semble la plus adéquate au vu de notre jeu de données.

## Classification par seuil d'efficacité sous Power BI

Cette partie a été faite sous l'outil Power BI, c'est un outil qui facilite la visualisation des données, et qui permet aussi de traiter les sorties R et Python

Une fois les données récupérées, la visualisation de données avec des histogrammes groupées a été effectuée. Ce genre de méthode permet de visualiser pour chacun des groupes d'efficacité ; les différentes valeurs de températures et de VPD qui lui correspondent, et aussi pouvoir visualiser sur le même graphique l'évolution climatique dans chacun des groupes d'efficacité. Ces groupes sont construits de manière contrainte, c'est à dire à priori sur une échelle choisie. Les groupes sont constitué de cette manière :

Nombre d'individus	Groupe d'efficacité	Condition d'efficacité à J14
8	Très bonne	Supérieur à 70%
9	Bonne	Entre 70% et 50%
10	Moyenne	Entre 50% et 1%
8	Nulle	0% et moins

Maintenant que nous avons calculé les différentes corrélations, nous allons vérifier s'il existe une différence entre les efficacités selon les températures instantanées, en utilisant des ANOVA.

Pour ces tests nous avons pris en considération que les données des comptages fait en J14 et en J21. Les comptages de J28 ont été exclu car comme on a pu le voir ; à ce stade il n'existe plus une grande différence entre les plantes traitées et les plantes non traitées. On a aussi pris que les efficacités préventifs et non curatifs.

Donc ici notre hypothèse  $H_0$  est la suivante : la distribution de l'efficacité est la même dans chacun des groupes vis à vis de la température instantanée.

Pour toutes les ANOVA on a eu des p-value  $< 0.05$ , donc l'hypothèse  $H_0$  n'est pas vérifié pour tous ces cas-là. Et donc la température instantanée n'est pas distribuée de la même façon dans chacun des groupes, ce qui nous pousse alors à chercher la température qui donne les meilleures efficacités à chaque fois.

Pour résumer ces tests ont permis donc de déduire qu'il existe une meilleure température pour l'efficacité, maintenant il faut la déterminer en appliquant le test de Wilcoxon Pairwise.

L'approche utilisée dans cette partie n'est pas valable pour la VPD. En effet pour pouvoir appliquer le test de Wilcoxon Pairwise il faut préalablement vérifier l'hypothèse préliminaire : "Est-ce qu'on peut différencier les efficacités selon la VPD" avec une ANOVA.

Or en faisant l'ANOVA, on obtient une P Value de 0.08 qui est supérieur à 0.05, l'hypothèse  $H_0$  est alors maintenue, et donc on ne peut pas appliquer le test de Wilcoxon Pairwise.

Ces résultats sont illustrés par les représentations graphiques des figure suivantes. En effet pour la température, la différence en très bonnes efficacité et efficacité nulle est très nette ce qui n'est pas le cas pour la VPD.

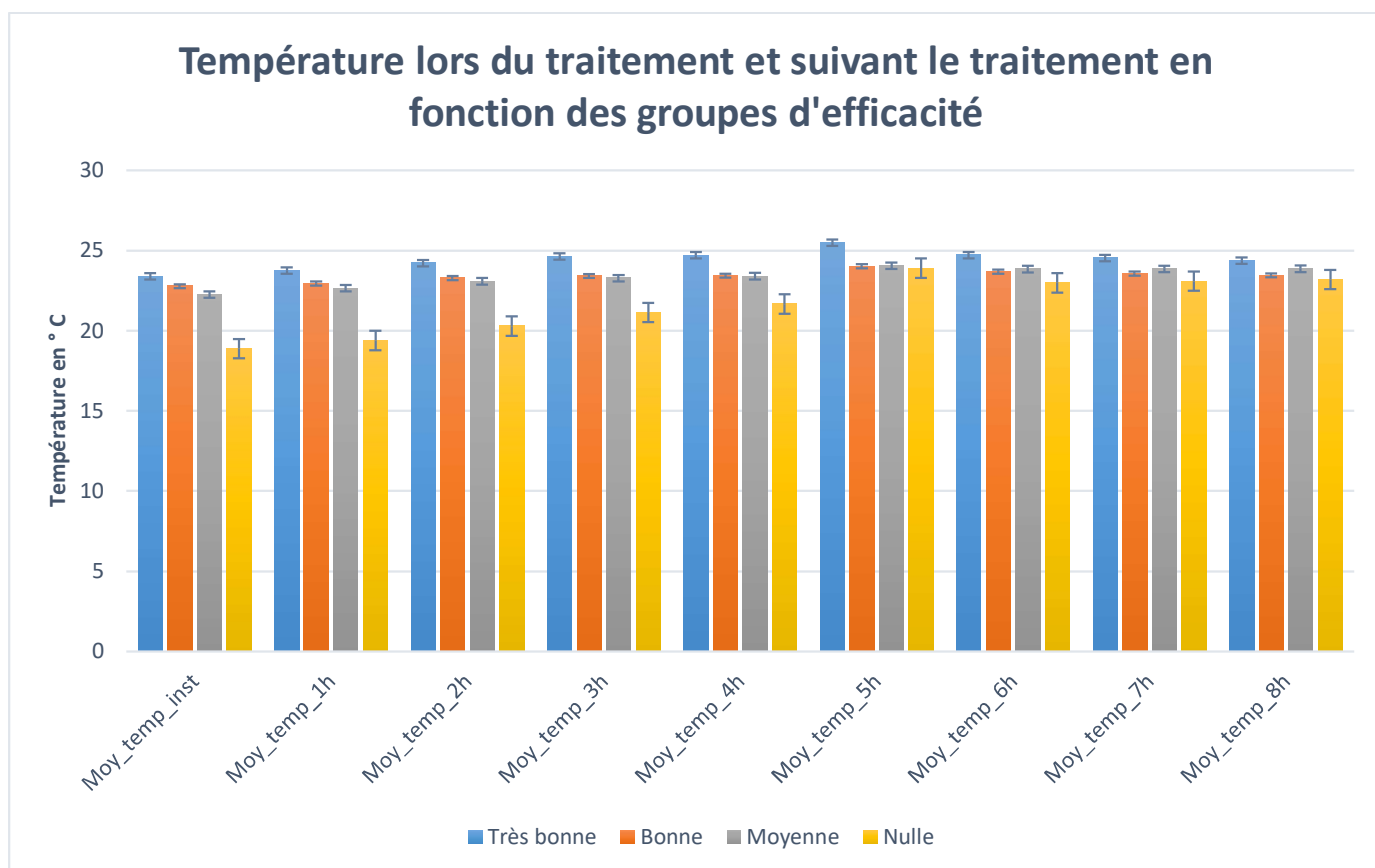


Figure 26 Graphique représentant les variables de moyenne de température par rapport aux groupes d'efficacité

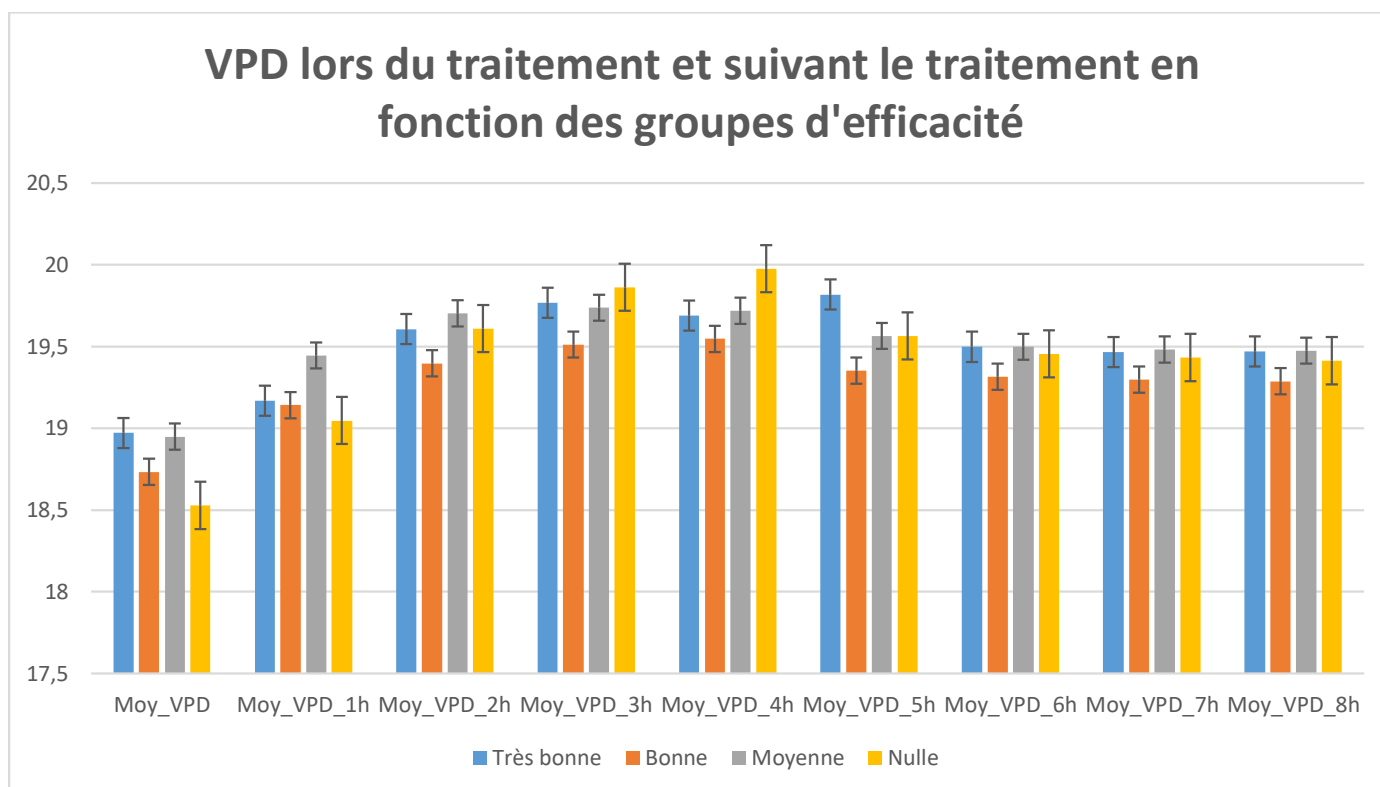


Figure 27 Graphique représentant les variables de moyenne de VPD par rapport aux groupes d'efficacité

En affichant les extremums pour mieux étudier les variations de température, d'autres interprétations sont possibles.

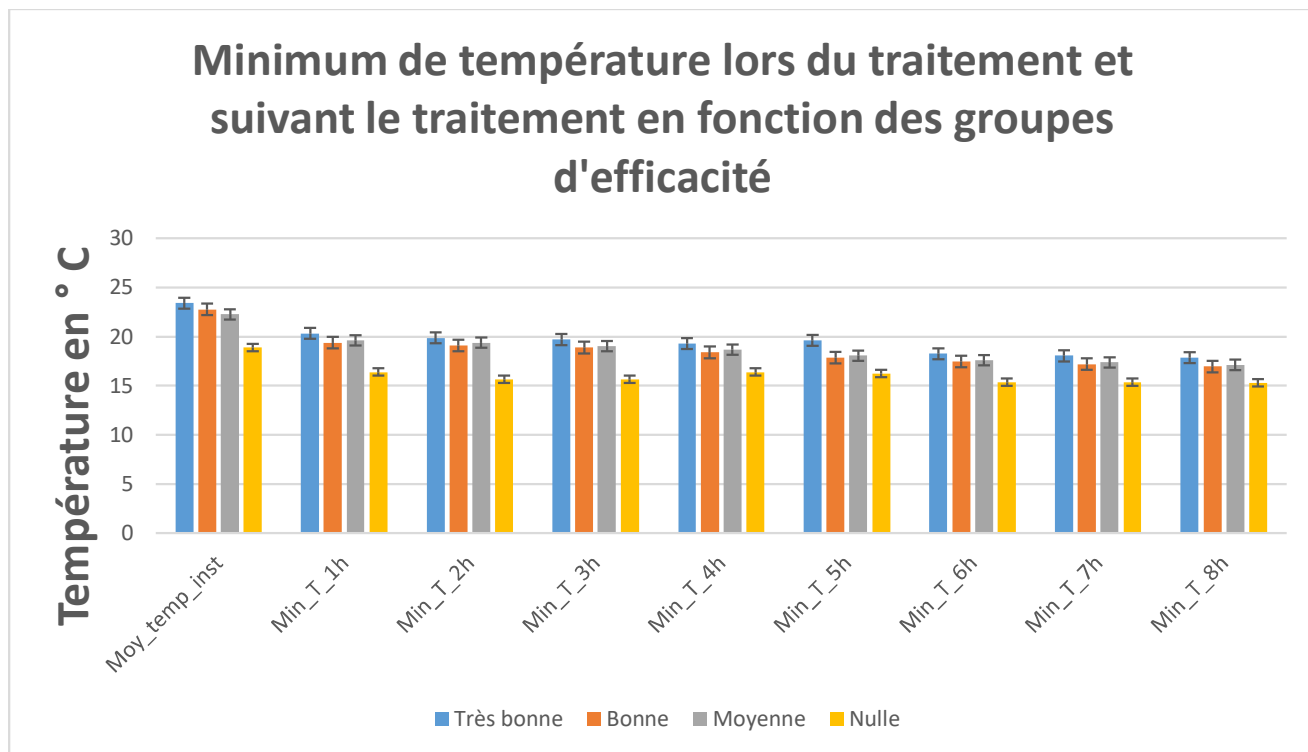


Figure 28 Graphique représentant les variables de minimum de temperature par rapport aux groupes d'efficacité

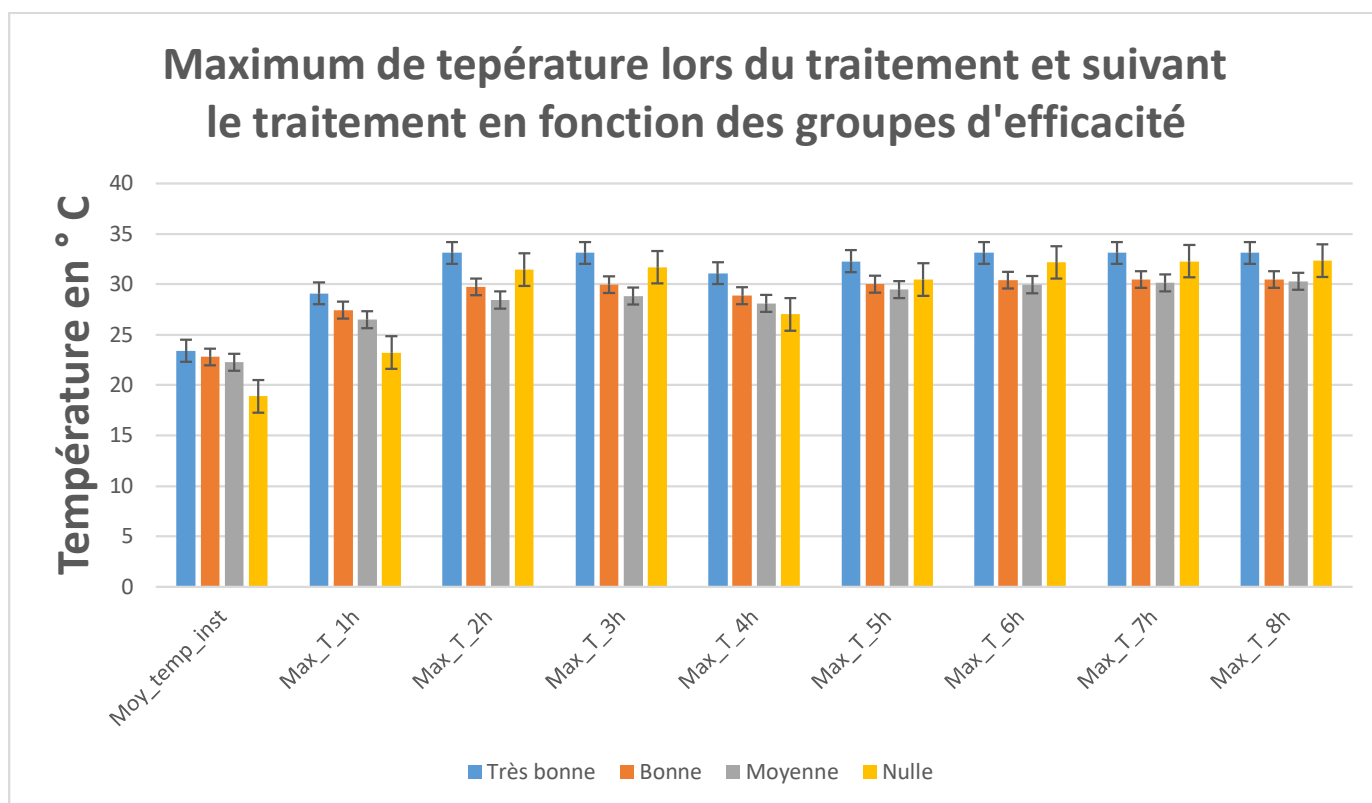


Figure 29 27 Graphique représentant les variables de maximum de température par rapport aux groupes d'efficacité

On peut ainsi observer que c'est avant tout le minimum de température atteint qui différencie les efficacités plus que les maximum qui sont mélangés entre les groupes.

Les mêmes graphiques pour les données de juin et de septembre ont été effectués ainsi que pour les comptages faits au 21<sup>ème</sup> jour, on a présenté les graphiques de j14 Juin car elles sont plus représentatives car les efficacités en J21 ont tendance à être moindres donc les différences moins prononcées.

En conclusion de cette approche :

Le profil climatique des très bonnes, bonnes et parfois moyennes efficacités sont proches, cependant le profil climatique des efficacités nulles est différent de par ses variables températures.

Cette approche permet donc d'établir le profil des efficacités nulles c'est-à-dire de période de traitement où il faut éviter de traiter sous peine de ne pas avoir d'effet du traitement.

Afin d'éviter une mauvaise efficacité il ne faut pas traiter en dessous de 20°C, où dans un moment où la température va atteindre 20°C ou moins dans les 3h qui suivent le traitement, il n'est donc pas judicieux de traiter le soir. De plus dans une moindre mesure une VPD trop faible, c'est-à-dire trop humide ne semble pas être trop profitable au traitement.

Cette première approche n'a donc pas permis de remplir l'objectif de profil d'une efficacité, une autre approche va donc être présentée et analysée.



# Classification Hiérarchique sur Composantes Principales : HCPC

Dans cette approche le HCPC va être utilisé. Il s'agit d'une combinaison de méthodes à composantes principales (ACP pour notre cas), classification ascendante hiérarchique, et partitionnement en k-moyennes. (STHDA 2017)

Le choix de cette méthode plutôt qu'un CAH simple, est dû au très grand nombre de variables comparé aux nombres d'individus. L'approche factorielle avant de faire les clusters permet de réduire le nombre de variables et de faire l'analyse sur les axes déjà créés avec l'ACP.

La HCPC est donc comme la CAH une méthode de classification hiérarchique itérative. Elle consiste à calculer les dis-similarités existantes entre l'ensemble des individus, qui sont ici les plantes/feuilles de l'expérience, puis à les regrouper deux à deux de manière à avoir un critère d'agrégation optimum, ce qui nous permet de créer une classe. On refait cette action jusqu'à avoir toutes les classes. Avec la même logique, on refait la même opération, mais cette fois-ci entre les classes que nous avons créées et non pas avec les individus. On continue à itérer cette opération jusqu'à ce que tous les individus soient regroupés. (XLSTAT 2020)

Une des sorties de la HCPC est le dendrogramme qui est un arbre binaire. Cet arbre nous permet d'obtenir une classification. La racine de l'arbre regroupe tous les individus, qui se découpe en plusieurs partitions

On note aussi que la méthode des K-means est utilisée ici pour diviser le jeu de données en plusieurs groupes. C'est une méthode de partitionnement directe et qui ne nécessite pas d'arbre hiérarchique. C'est une méthode d'agrégation autour des centres mobiles.

Dans cette partie, la base de données utilisée est une base de données des efficacités construite à partir des parcelles de 4 plantes plutôt que la base de données des individus à partir de l'ancienne : on a calculé les efficacités préventives et curatives à partir des nombres de colonies et de spores qu'on avait avant, et on a ajouté aussi l'accélération de la température. Le jeu de données contient 36 individus (le nombre d'individus a diminué par rapport à la première base de données car plusieurs plantes sont nécessaires pour calculer chacune des efficacités) et 96 variables, 3 variables sont qualitatives, et n'influencent pas la construction des composantes principales de l'analyse. Elles aident à l'interprétation des dimensions de variabilité.

On va donc appliquer la CAH sur les axes de l'ACP à qui nous avons préalablement appliqué la méthode des K-means. L'ACP correspondant à cette CAH est mise en Annexe.

La CAH vient donc après une analyse factorielle faite par ACP, et où les plantes à classer ont été repérées par leurs coordonnées sur les premiers axes factoriels

On note que pour ce jeu de données nous avons seulement 36 individus, le clustering risque d'être un peu juste vis-à-vis du nombre d'individus : nous avons un jeu de données qui contient beaucoup plus de variables que d'individus.

On va essayer de regrouper des plantes qui ont un profil similaire, puis on va définir un profil climatique pour chacun des clusters.



On commence par visualiser l'arbre hiérarchique construit avec la distance euclidienne et le critère de Ward. On rappelle que le critère de Ward consiste à choisir à chaque étape de la classification que le regroupement effectué soit de manière que l'augmentation de l'inertie intra-classes soit minimum.

Dendrogramme:

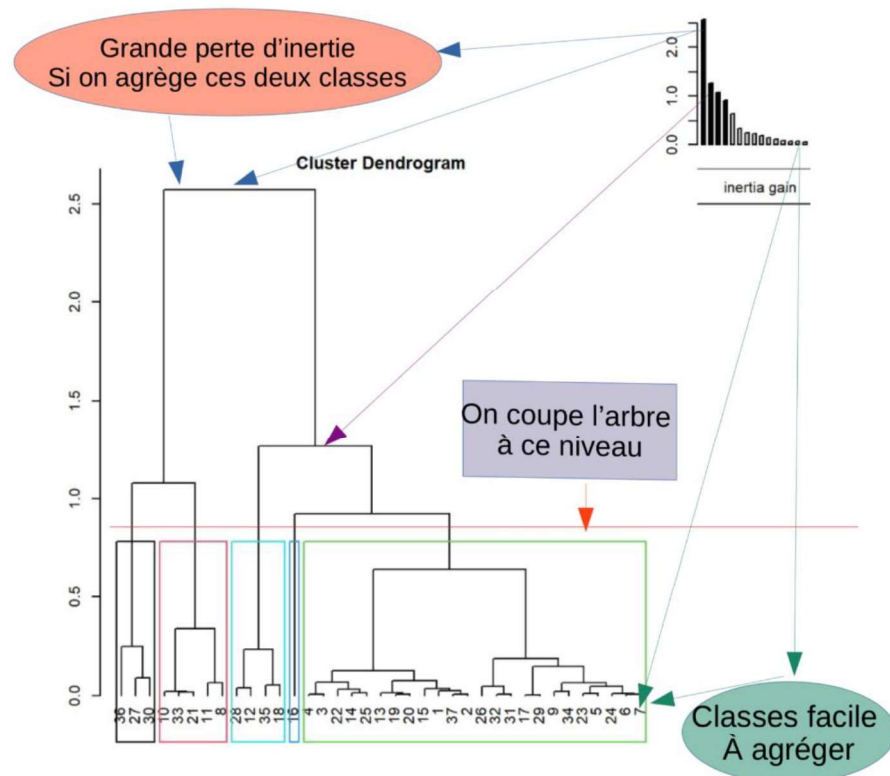


Figure 30. Dendrogramme de la CAH

On remarque que les individus 19 et 20 ont un comportement très proche, de même pour les individus 2 et 37 ou encore 6 et 7. Le niveau de présence de la maladie chez ces individus est proche.

On remarque aussi que les individus 5 et 6 se regroupent très bien avec l'individu 24 et construisent une seule classe, qui à son tour se combine très bien avec la classe construite avec les individus 5 et 23. La combinaison de ces deux classes crée à son tour une classe assez homogène.

Cette méthode nous permet alors de voir les proximités entre les individus ainsi qu'entre les groupes.

Le graphique en haut à droite exprime l'évolution de l'inertie par rapport pour différentes partitions, il exprime alors la perte d'inertie interclasses si on regroupe lors deux regroupements des classes.

Il faut alors trouver le bon équilibre, un compromis entre inertie perdue et nombre de classes obtenues.

Dans notre cas on voit que si on passe d'une à deux classes on a une perte 2.2 (la somme de perte de l'inerties est de 7.3) qui est assez grande comparée aux autres, ce qui signifie que ce regroupement agrège des individus qui sont différents, on a perdu beaucoup d'information en l'appliquant.

L'histogramme nous montre aussi que le regroupement des premières classes de l'arbre est naturel, il n'implique pas de grande perte d'inertie.

Avec le découpage que nous avons appliqué, on obtient 75.34% de l'inertie totale. On va comparer ce pourcentage avec le pourcentage de variabilité expliquée par le premier axe de l'ACP. (On rappelle que l'analyse de cet ACP se trouve en Annexe).

Le premier axe de l'ACP récupère 33.32 de l'information du jeu de données, alors qu'avec le découpage que nous avons établi on récupère 75.34%, la CAH communique alors une information plus fine.

Donc pour résumer La classification réalisée sur les individus fait apparaître 5 classes, ce choix a été fait par rapport aux inerties cumulées avec ce choix ainsi que la grande chute dans le dendrogramme.

Le CAH nous montre comment on est passé de 36 individus à 5 classes par une succession de regroupement.

On découpe l'arbre là où les branches sont un peu longues

Maintenant que nous avons choisi le nombre de cluster, nous allons visualiser la distribution des plantes et des clusters dans le plan factoriel.

Distribution des plantes et des clusters dans le plan factoriel :

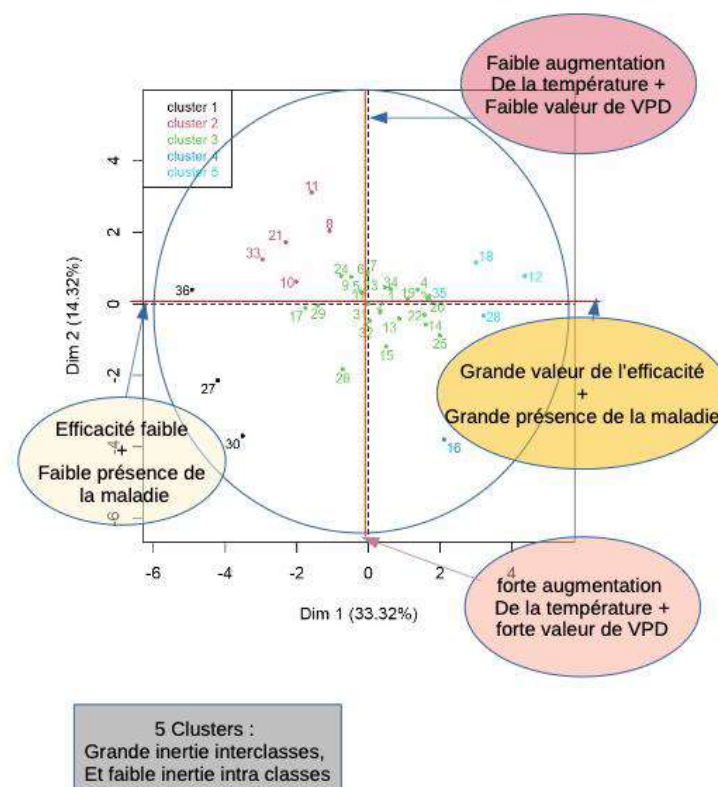


Figure 31 Représentation des clusters sur le plan factoriel 1:2

## Description des Clusters :

### Description des clusters :

La **classe 1** est composée d'individus tels que 27, 30 et 36. Ce groupe est caractérisé par de faible valeur d'efficacité du traitement, ainsi que par de fortes valeurs l'accélération de température.

La **classe 2** est composée d'individus tels que 8, 10, 11, 21 et 33. Ce groupe est caractérisé par de faibles valeurs d'efficacités, d'accélération de température ainsi que les températures instantanées lors du traitement.

La **classe 3** est composée d'individus tels que 25 et 26. Ce groupe est caractérisé par des valeurs moyennement fortes en efficacité et de température, ces plantes se trouvent au centre de l'ACP, elles n'ont pas de valeurs extrêmes sur les axes.

La **classe 4** est composée de la plante 16. Cet individu a un comportement très atypique, on a donc convenu de la laisser seul dans une classe, l'ajouter à une autre classe ajouterai de l'inertie infra-classes.

Ce groupe est alors caractérisé par une très forte valeur en nombre de colonies chez les témoins, mais contrairement aux autres individus qui ont ce caractère, l'efficacité est faible, et l'augmentation de la température est importante.

La **classe 5** est composée d'individus tels que 12, 18 et 28. Ce groupe est caractérisé par une très grande contamination chez les témoins, mais aussi une grande efficacité contre la maladie.

Maintenant que nous avons identifié et caractérisé les clusters, on va les visualiser en 3D.

On observe maintenant l'arbre hiérarchique en 3D avec les plans 1 et 2 :

### Visualisation en 3D des clusters :

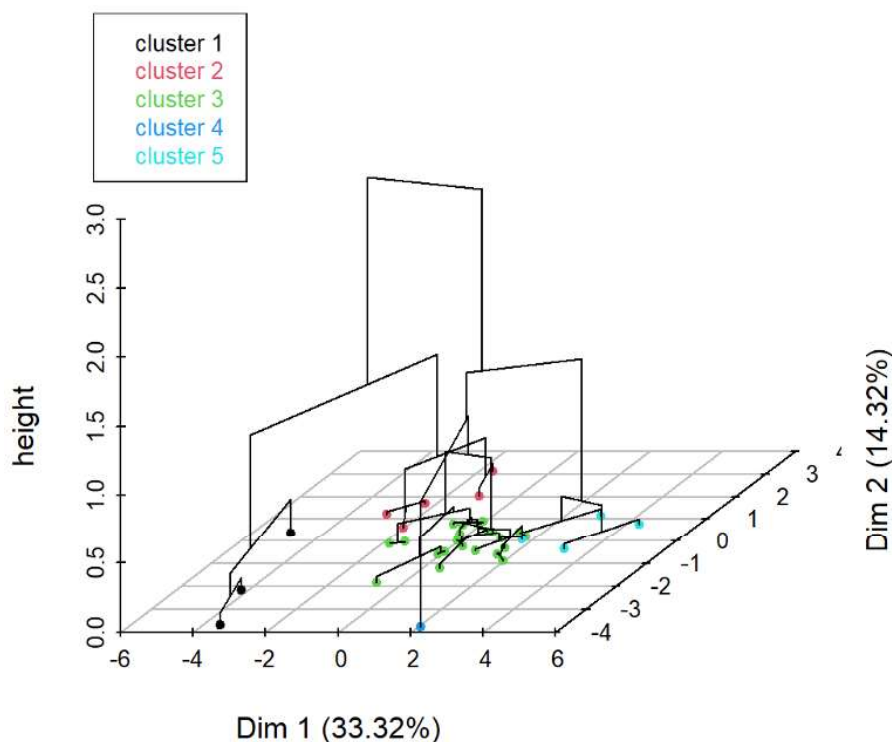


Figure 32 Dendrogramme de la classification hiérarchique sur le plan factoriel 1:2

Les 5 clusters sont bien représentés dans le plan factoriel, on voit qu'« ils ne sont pas mélangés les uns aux autres ». On a une inertie inter classe qui est grande et une inertie interclasses faibles (on peut l'optimiser encore plus en prenant plus de classes, mais ceci aura en contrepartie une perte d'informations, tout est donc question d'équilibre).

On note qu'à chaque fois que les regroupements sont fait, l'inertie intra-classe augmente et l'inertie interclasse diminue. On a ce résultat car la somme de l'inertie intra-classe et l'inertie interclasse est une constante.

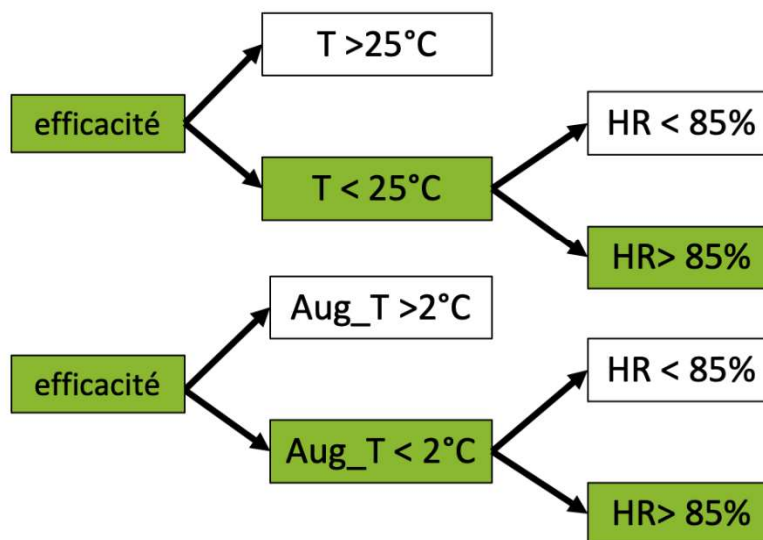
De point de de vue biologiques cette approche a permis de montrer que les facteurs climatiques ne peuvent différencier seule les efficacités en effet avec ce type de classification, les variables autres que climatique joue un rôle prépondérant dans la classification. Cependant cette approche à aussi mis en avant l'importance de la variable augmentation de température qui n'avait pas pu être étudiée dans l'approche précédente car non adapté. Enfin cette approche met aussi en évidence qu'une efficacité données peut être obtenue par plusieurs facteurs. En effet une mauvaise efficacité peut être lié à une mauvaise température de traitement, par une forte augmentation de température ou par les deux à la fois. Les raisons de succès comme d'échec sont donc multiples et imbriqués.

## Clustering par Random Forest.

La classification avec les forêts aléatoires est une méthode basée sur le machine Learning, c'est donc une méthode qui se base sur l'apprentissage automatique à la machine, le concept est de faire apprendre à la machine d'appliquer une tâche (ici c'est classer les plantes selon les différents critères que nous avons établi (climatiques, état de la plante vis-à-vis de la maladie...), et cet apprentissage se fait en donnant à la machine des exemples sur lesquels se baser, et donc le plus on a d'exemples, le mieux l'apprentissage est. La random forest ou la forêt aléatoire se compose d'une multitude d'arbres de décision. Les arbres utilisés dans ce modèle sont des arbres de décision binaire. Le concept derrière cette méthode est qu'au lieu d'avoir un seul arbre de décision et qu'on essaye de l'optimiser, on a une multitude d'arbres de décision et qu'on essaye d'optimiser aussi, et chacun de ces arbres contient une ou plusieurs conditions à vérifier et se base sur une partie des données à analyser. (XLSTAT 2021)

Concrètement, on peut traduire cela comme suivant pour notre dataset : nous avons la variable efficacité qui est expliquée par une multitude de variables (climatiques et de présence de la maladie). On utilise alors la fonction de mapping qui représente le lien entre la variable à expliquer et les variables explicatives afin de faire notre classification et par la suite une prédiction. Pour notre cas on applique une classification, donc cette fonction est une fonction à valeurs discrètes (on note que pour les cas de régressions, la fonction est continue)

En voici deux exemples d'arbres qu'on a utilisé dans la forêt aléatoire :



Dans notre modèle nous avons donc une multitude d'arbres (n=92 : obtenu par validation croisée) ressemblant à ceux qu'on vient de présenter. Et chacun des arbres créer permet d'acquérir une vision parcelaire de la situation traitée, et ceci est donc causé par le double tirage aléatoire : Tirages pour les individus (tree bagging), et tirage pour les variables (feature sampling).

On note que chacun de ces arbres est un classificateur, et se compose de de racine, nœuds, branches, et puis finalement feuilles. Le procès fonctionne de manière à descendre dans l'arbre, et donc aller de la racine aux feuilles. La racine de l'arbre regroupe alors tous les individus, et puis des branches sont créés par rapports aux différentes modalités établies dans l'arbre. Dans notre exemple on a une branche qui est créé si la condition : la température instantanée est supérieure à 25°C est vérifiée, si

c'est le cas, vient alors la branche créer par la condition : si la VPD est inférieure à 85%, et vis vers ça.... Et avant que chaque branche ne se décompose à son tour en deux, nous avons les nœuds, et c'est au niveau des nœuds que les conditions sont testées. Finalement on a les feuilles de l'arbre qui correspondent aux sorties de l'algorithme, et donc aux différents clusters que nous avons établis.

On va donc combiner nos n=92 arbres de décisions avec une approche bagging, c'est-à-dire qu'on va calculer à chaque étape la moyenne des prédictions afin de réduire l'erreur de prédiction générale. L'apprentissage se fait aussi sur 3 répartitions de données d'apprentissages. Les répartitions sont faites de manière aléatoire, et le nombre de répartitions est choisi par validation croisée.

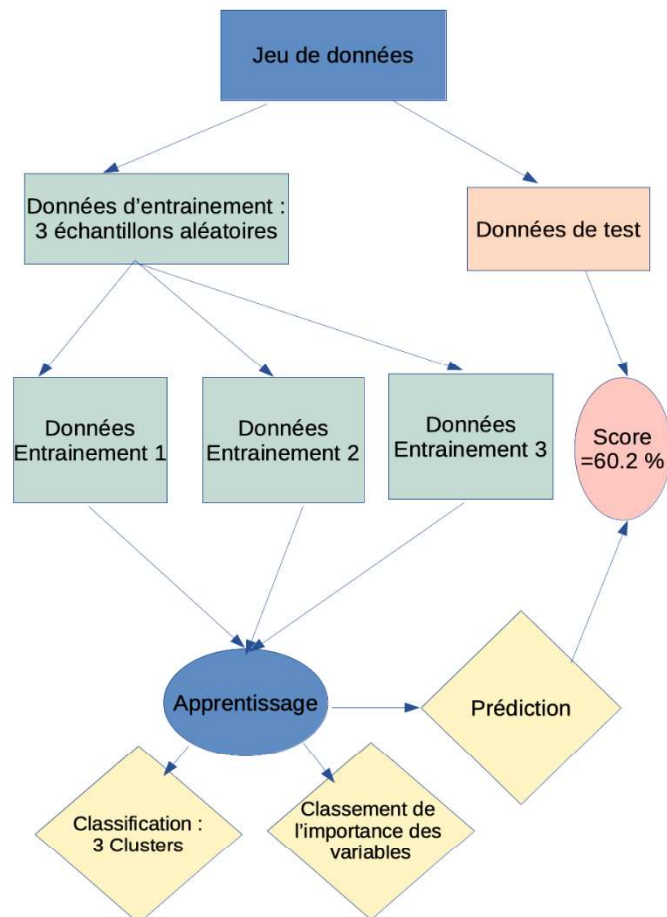


Figure 33 Process suivi pour le Random Forest

Le bagging nous permettra alors de réduire la variance par rapport à un unique arbre de classification, d'où l'intérêt de son utilisation.

L'avantage qu'on a avec le bagging est qu'on a en général un bon rapport de prédiction, et qu'on évite d'avoir un problème d'overfitting (surapprentissage), néanmoins le problème que nous avons ici pour nos données est que le nombre d'individus est très petit (36 plantes pour l'efficacité), et donc le

problème qui nous gêne le plus c'est l'underfitting. Le modèle sera bien meilleur avec plus d'individus, c'est à dire de refaire l'expérience avec plus de plantes. Chose qui sera faite dans un futur proche car pour des expérimentations complémentaires vont être établies à l'INRAE.

#### **Clusters Obtenus avec la Random Forest :**

Nous avons obtenu 3 Clusters avec la Random forest :

Cluster 1 : 36, 27, 30, 33, 10, 24, 29, 6, 17, 23, 24, 29.

Ces plantes sont caractérisées par de faibles valeurs d'efficacité (globalement moins 20%) et de températures, et des VPD qui atteignent des seuils Maximum.

Cluster 2 : 11, 8, 24, 6, 7, 34, 9, 5, 1, 2, 10, 13, 14, 15, 31, 32.

Ces plantes ont une bonne efficacité, mais aussi une présence moyenne de la maladie chez les témoins.

Cluster 3 : 12, 21, 18, 28, 35, 16, 4, 25, 26, 22, 3, 19, 20.

Ces plantes ont les meilleures efficacités, avec des valeurs de températures instantanées autour de 24°, et des VPD Faibles.

Pour cette partie le R<sup>2</sup> du modèle est de 0.41

Nous avons utilisé par la suite les random forest afin d'obtenir plus de renseignements sur nos données plus précisément pour classer les variables selon leurs importances à expliquer les variables d'efficacité. Le résultat qu'on a obtenu est le suivant :

La variable qui influence le plus l'efficacité est la VPD 7h après le traitement :

On remarque que la variable de VPD 7h après le traitement revient souvent comme résultat. Chose à laquelle on ne s'attendait pas. On a découvert par la suite qu'il existe une chute assez brutale de l'efficacité quand la VPD atteint le seuil de : 17.5 et globalement ce seuil n'est pas toujours atteint, et quand il l'est, c'est au bout de 7h après le traitement.

La deuxième variable a été la VPD 8h après le traitement : On peut justifier cela par la très grande corrélation positive qui existe entre la VPD 7h et la VPD 8h.

Puis ensuite vient les variables de températures, et là on n'a pas eu de grandes surprises, la première variable est celle de la température moyenne aux premières heures qui suivent le traitement.

Vient ensuite la variable Max et min température pour les premières heures suivant le traitement. Pour ces variables là on a pu constater que c'est surtout le changement très brutal qui fait un grand effet sur l'efficacité du produit.

On note aussi que la random forest utilisée pour cette partie a un R<sup>2</sup>=0.34

#### **Prédiction avec les Random Forest :**

Nous allons effectuer maintenant une évaluation de notre classification, pour cela nous allons créer une nouvelle variable à 3 modalités : Bonne efficacité, efficacité moyenne, et mauvaise efficacité.

La bonne efficacité regroupera les efficacités qui sont supérieures à 40%.

L'efficacité moyenne regroupera les efficacités qui ont des valeurs entre 20% et 40%.

La mauvaise efficacité regroupera les efficacités qui sont inférieures à 20%.

Maintenant on va essayer de faire une prédiction vis-à-vis de ce critère, puis calculer un score de prédiction, et pour cela nous allons avoir besoin d'établir la matrice de confusion.

On note que la matrice de confusion se présente de la manière suivante :

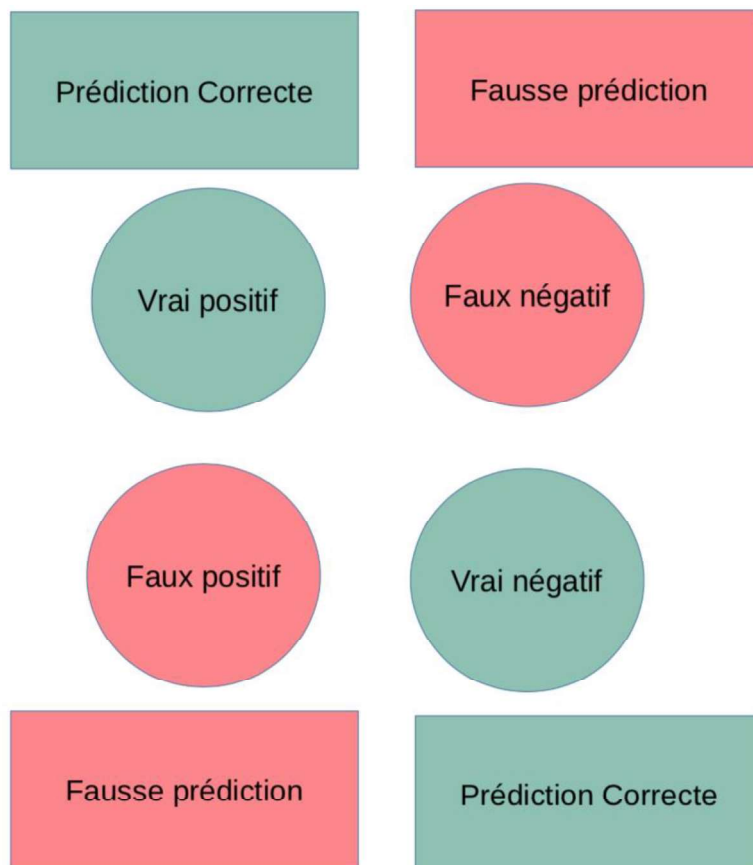


Figure 34 Matrice de confusion

- La case (1,1) correspond aux Vrais Positifs, c'est à dire aux résultats où le modèle prédit correctement et sans erreur la classe positive.
- La case (1,2) correspond aux Vrais négatifs, c'est à dire aux résultats où le modèle prédit correctement et sans erreur la classe négative.
- La case (2,1) correspond aux Faux positif, c'est à dire aux résultats où le modèle prédit avec erreur la classe positive.



- La case (2,2) correspond aux Faux négative, c'est à dire aux résultats où le modèle prédit avec erreur la classe négative.

Maintenant que nous avons la matrice de confusion, nous calculons le score, le score se calcule en divisant la somme des vrais positifs et des vrais négatifs sur la somme de toutes les cases de la matrice.

Pour notre random forest, nous avons un score de 60.34%, ce score est moyennement bien, on aura bien plus avec plus de données, mais il reste tout de même acceptable.

A titre informatif, le score obtenu avec un seul arbre est de 54.43%, ce résultat est assez faible, il est même très proche d'une solution aléatoire, l'intérêt de ce score est de nous donner un repère de comparaison, et qui nous servira même pour la suite afin de savoir le pourcentage de bonne prédiction gagné avec d'autres modèles.

Maintenant qu'on a une idée sur la pertinence de ce modèle, on va comparer les classes qu'on a eu avec le RF et la CAH.

## Comparaison des 3 méthodes :

Dans ce chapitre nous avons fait 3 types de classifications, la première est une classification assez simple, et nous a servi de point de comparaison pour les deux autres méthodes.

Pour la **HCPC** et la random forest, le résultat obtenu est moyennement proche, pour la **HCPC** nous avons 5 classes, et ces 5 classes sont construites sur le plan factoriel alors que pour la random forest nous en avons seulement 3. Les classes 4 et 5 du **HCPC** sont fusionnées en une seule classe (classe 3) du random forest : en effet l'individu 16 qui appartient à une classe à part dans la HCPC est dans la classe 4 de la random forest. On note aussi que la classe 2 en HCPC est partagée entre la classe 1 et 2 de Random Forest, puis une dernière différence est que la plante 35 dans le **HCPC** appartient à la troisième classe qui est caractérisée par une grande efficacité et une faible augmentation de la température, or dans la random forest cette plante appartient à la deuxième classe.

Pour résumer, les individus de la classe 1 et 5 du HCPC font toutes partie de la classe 1 et du random forest, de même pour là les classes sont composées des mêmes individus à quelques exceptions près, et ces individus ont été localisés au bord des classes dans le plan factoriel du CAH. On va les citer :

Les résultats des deux méthodes ne sont pas très éloignés, mais pour l'état actuel des données la Classification Hiérarchique sur Composantes Principales semble être plus adaptée, car on a un jeu de données très riches en variables, mais très pauvres en individus, cette méthode nous semble plus appropriée.

Des expériences similaires à celles que nous avons traitées durant mon stage vont continuer à être faites à l'INRAE au futur proche, et donc le jeu de données contiendra de plus en plus d'individus, et dans ce cas précis, la random forest a le potentiel de donner de meilleurs résultats que ce soit pour la classification ou la prédiction, et surmontera le problème actuel d'underfitting.

## Discussion & Conclusion

Ce stage a permis de dégager plusieurs conclusions concernant l'utilisation du produit AQ10 contre l'oïdium de la tomate.

Tout d'abord on a pu observer concrètement que le traitement a un effet : le nombre de colonies observé chez les plantes traitées de manière préventive est largement inférieur à celui observé chez les plantes témoin. Le niveau de protection du traitement est en moyenne de 31% à 14 jours après traitement mais il chute rapidement par la suite à environ 25% à 21 jours après traitement pour être totalement inefficace 28 jours après traitement.

Dans certaines conditions comme établies lors des Chapitres 2 et 3, l'efficacité peut être très bonne et se maintenir pendant 21 jours. On peut citer par exemple les traitements effectués le Midi en juin où l'efficacité est supérieure à 60 % lors du comptage à 14 jours et se maintient à plus de 50% lors du deuxième comptage à 21 jours.

En revanche, le traitement curatif n'a qu'un effet très limité. Celui faisant suite à un traitement préventif apporte une légère amélioration dans certains cas notamment pour la diminution du nombre de spores.

Ces « bonnes conditions » de traitement ont été rassemblées dans un profil climatique, on retrouve des températures et des VPD associés. Les plantes les mieux protégées par l'agent de biocontrôle partagent presque tous les conditions d'application suivantes :

- Température instantanée de traitement qui varie entre 23° et 25°.
- Température maximum dans la première heure qui suit le traitement qui varie entre 25° et 28° et une température minimum qui varie entre 21° et 23°.
- VPD qui varie très peu pendant les 7 premières heures, avec une valeur proche de 19.
- VPD qui ne chute pas au bout de la 7ème heure, et n'atteint pas le seuil de 17.5 dans les 10 h qui suivent le traitement.

Cependant, force est de constater que le profil climatique des bonnes efficacités et celui des moyennes efficacités paraissent assez proches. A contrario le profil des mauvaises efficacités de protection est très différent :

- Température moyenne instantanée qui varie entre 15° et 18°
- Température minimum inférieure à 16°C, pour presque toutes les plantes
- Température maximum inférieure à 24°C, pour presque toutes les plantes
- VPD qui varie de plusieurs unités dans les premières heures qui suivent le traitement (une variation qui peut valoir 4).
- VPD qui atteint le seuil de 17.5 dans les 10 premières heures qui suivent le traitement.

Concernant les mauvaises efficacités il existe trois critères qui ont le plus attiré notre attention :

### Température Non Linéaire Non monotone :

Le premier critère pour une mauvaise efficacité est que les variables de températures (température min, max, moyenne de température au moment du traitement) soient plus basses de 5°C en comparaison des variables de température des plantes bien protégées contre l'oïdium.

Le deuxième est que pour les mauvaises efficacités, il y a souvent des pics de température dans les heures qui suivent le traitement, et donc cette accélération soudaine de la température a peut-être un effet affectant l'efficacité d'AQ10. En étudiant le profil des mauvaises efficacités on remarque la présence d'une montée soudaine de température.

Le troisième point est dans le caractère linéaire de descente de température :

- Pour les bonnes et les moyennes efficacités cette pente a la même accélération en général, on a donc des courbes de températures en escaliers.
- Pour les mauvaises efficacités le caractère linéaire de la pente est perturbé, ce qui peut être lié au deuxième point : car un grand pic de température perturbe la linéarité. Il est donc possible d'avoir plusieurs petites variations de températures qui, une par une ne représentent pas un pic, mais ensemble permettent de faire varier la linéarité de la température.

La VPD est un peu plus délicate à interpréter, car cette dernière varie moins que la température. Nous avons tout de même constaté qu'à partir du seuil de VPD= 17.5, l'efficacité fait une grande chute.

- L'effet climatique sur les efficacités du traitement (efficacité curative et préventive) est présent : corrélations d'environ 30%. La réelle complexité qu'on a cependant est qu'on n'a pas des données concernant la présence et la propagation de la maladie avant J14. Il y a donc plusieurs événements qui se sont produits avant que le premier comptage de nombre de taches soit effectué, ce qui rend difficile de définir avec une grande précision lequel de ces événements a joué quel rôle exact lors des interactions. On note que les colonies d'oïdium ne sont pas visibles à l'œil nu durant les premiers jours, donc si on souhaite observer la présence de la maladie lors des premiers jours on est obligé de passer par un comptage de microscopique impossible sur le terrain.
- Une autre approche consistera à essayer de faire les mêmes conditions climatiques avec les mêmes quantités et qualités de l'inoculum sur des plantes jumelles, et donc étudier sur une partie des plantes l'évolution de la maladie, et sur la deuxième partie un comptage microscopique dans les premiers jours, une méthode de mesure qui serait, elle destructive. Ce genre d'étude bien que difficile à mettre en place permettrait un suivi plus détaillé de l'efficacité du traitement.

Le traitement avec AQ10 a un effet retard sur la propagation de la maladie, il retarde l'extension de plusieurs jours, ce qui peut avoir de très grands avantages :

- Les jours gagnés avant que les plantes n'atteignent des stades très avancés peuvent servir à isoler les plantes malades (isolation par bâche en plastique) et donc faire un bouclier face à la propagation de la maladie et donc offrir un moyen de lutte complémentaire avec AQ10 contre l'oïdium.
- A partir du moment où les plantes malades sont isolées, il faut rapidement appliquer un traitement curatif sur les plantes malades pour réduire le nombre de spores viables.
- Cela permettrait aux agriculteurs de cueillir leurs récoltes avant que la maladie n'affecte le potentiel de rendement lorsque le traitement est effectué proche de la récolte.

### **Limites de cette étude et propositions**

À partir de J21 (comptage de J21 & J28), à ce stade l'infection est trop avancée sur les plantes, il ne semble plus utile de traiter les plantes avec AQ10, car la maladie s'est bien propagée, on ne peut plus la stopper ou la retarder avec le produit. Il faudrait donc prévoir un traitement curatif plutôt pour vérifier s'il n'a pas une efficacité plus importante.

Les modèles de machine Learning ont besoin de plus de données pour affiner les résultats. (Le nombre de variables est beaucoup plus grand que le nombre d'individus).

Le « background génétique » n'est pas pris en considération (différence entre les plantes utilisées dans l'expérimentation de juin et de septembre) : L'état de la plante joue un rôle sur l'extension de la maladie, on a pu le constater en comparant les données de juin à celle de septembre : Les plantes de Juin proviennent de plantes « Gourmandes » qui sont donc plus vieilles et matures que les plantes utilisées en Septembre, et sont donc beaucoup plus solides face à l'oïdium.

Le traitement curatif après 14 jours de l'inoculation est peut-être réalisé trop tard, car la maladie a commencé son extension, et il est difficile de la contrôler. Il est possible que pour le traitement curatif le dosage suggéré ne soit pas bon et qu'il devrait être augmenté pour de meilleurs résultats.

Le troisième comptage de colonies se fait à J28, et à cette étape la maladie s'est largement propagée chez les plantes traitées et non traitées. Ce comptage ne nous rapporte pas beaucoup d'informations. Il serait éventuellement mieux d'avancer ce comptage de quelques jours, ou bien de faire un autre comptage entre le J21 et J28.

Les variables qui caractérisent l'extension de la maladie ne sont que 2 : Nombre de colonies et nombre de spores. La possibilité d'avoir plus de variables qui permettent de suivre l'avancée de la maladie serait un avantage certain pour la modélisation et la recherche de profil d'efficacité.

## Bibliographie

- Anses. 2021. « Anses - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail », 2021. <https://www.anses.fr/fr>.
- Bardin, Marc, Sakhr Ajouz, Morgane Comby, Miguel Lopez-Ferber, Benoît Graillot, Myriam Siegwart, et Philippe C. Nicot. 2015. « 176 Is the Efficacy of Biological Control against Plant Diseases Likely to Be More Durable than That of Chemical Pesticides? » *Frontiers in Plant Science* 6. <https://doi.org/10.3389/fpls.2015.00566>.
- Bernard, Jean-Louis. 2007. « Protection chimique des plantes cultivées et durabilité ». *Oléagineux, Corps gras, Lipides* 14 (6): 332-44. <https://doi.org/10.1051/ocl.2007.0166>.
- Blancard, Dominique. 2016. « Tomate - Pseudoidium neolycopersici (oidium externe) ». *Ephytia*. <http://ephytia.inra.fr/fr/C/4998/Tomate-Pseudoidium-neolycopersici-oidium-externe>.
- Ecophyto. 2021. « EcophytoPIC ». <http://www.acta.asso.fr/ressources/ecophytopic.html>.
- FactomineR. 2008. « missMDA ». 2008. [http://factominer.free.fr/missMDA/index\\_fr.html](http://factominer.free.fr/missMDA/index_fr.html).
- FAO. 2019. « Minimizing the impact of plant pests - a global surveillance system for crop diseases ». International Plant Protection Convention. 2019. <https://www.ippc.int/fr/news/minimizing-the-impact-of-plant-pests-a-global-surveillance-system-for-crop-diseases/>.
- Fedele, Giorgia, Chiara Brischetto, et Vittorio Rossi. 2020. « 426 Biocontrol of Botrytis Cinerea on Grape Berries as Influenced by Temperature and Humidity ». *Frontiers in Plant Science* 11. <https://doi.org/10.3389/fpls.2020.01232>.
- Fedele, Giorgia, Elisa González-Domínguez, et Vittorio Rossi. 2020. « 492 Influence of Environment on the Biocontrol of Botrytis Cinerea: A Systematic Literature Review ». In *How Research Can Stimulate the Development of Commercial Biological Control Against Plant Diseases*, édité par Antonieta De Cal, Paloma Melgarejo, et Naresh Magan, 61-82. Progress in Biological Control. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-53238-3\\_5](https://doi.org/10.1007/978-3-030-53238-3_5).
- Fourastié, Jean. 2020. « "Les Trente Glorieuses", » 24 décembre 2020. <https://www.fayard.fr/sciences-humaines/les-trente-glorieuses-9782213006833>.
- Grossiord, Charlotte, Thomas N. Buckley, Lucas A. Cernusak, Kimberly A. Novick, Benjamin Poulter, Rolf T. W. Siegwolf, John S. Sperry, et Nate G. McDowell. 2020. « Plant Responses to Rising Vapor Pressure Deficit ». *New Phytologist* 226 (6): 1550-66. <https://doi.org/10.1111/nph.16485>.
- Magan, Naresh. 2020. « 490 Importance of Ecological Windows for Efficacy of Biocontrol Agents ». In *How Research Can Stimulate the Development of Commercial Biological Control Against Plant Diseases*, édité par Antonieta De Cal, Paloma Melgarejo, et Naresh Magan, 1-14. Progress in Biological Control. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-53238-3\\_1](https://doi.org/10.1007/978-3-030-53238-3_1).
- Mazoyer, M, et L Roudart. 2002. *Histoire des agricultures du monde , Ma...* Seuil. <https://www.seuil.com/ouvrage/histoire-des-agricultures-du-monde-marcel-mazoyer/9782020530613>.
- Mercader, Julio. 2009. « Mozambican Grass Seed Consumption during the Middle Stone Age ». *Science (New York, N.Y.)* 326 (5960): 1680-83. <https://doi.org/10.1126/science.1173966>.
- Monus, LENA. 2019. « 470 Enquête de l'IBMA France, colloque-2019 ». 2019. <https://www.academie-agriculture.fr/sites/default/files/agenda/enquetelenamomusibmafrancecolloque-2019.pdf>.
- Nicot, Philippe C. 2011. *459 Classical and Augmentative Biological Control against Diseases and Pests: Critical Status Analysis and Review of Factors Influencing Their Success*. S.l.: IOBC/WPRS. [http://www.iobc-wprs.org/pub/biological\\_control\\_against\\_diseases\\_and\\_pests\\_2011.pdf](http://www.iobc-wprs.org/pub/biological_control_against_diseases_and_pests_2011.pdf).
- Punja, Zamir K., Ananchanok Tirajoh, Danielle Collyer, et Li Ni. 2019. « 41 Efficacy of Bacillus Subtilis Strain QST 713 (Rhapsody) against Four Major Diseases of Greenhouse Cucumbers ». *Crop Protection* 124 (octobre): UNSP 104845. <https://doi.org/10.1016/j.cropro.2019.104845>.
- STHDA. 2017. « Classification Hiérarchique sur Composantes Principales: L'Essentiel - Articles - STHDA », 2017. <http://www.sthda.com/french/articles/38-methodes-des-composantes-principales-dans-r-guide-pratique/78-classification-hierarchique-sur-composantes-principales-l-essentiel/>.

- Syngenta. 2020. « Oïdium de la tomate ». *Syngenta France*, 2 novembre 2020. <https://www.syngenta.fr/traitements/oidium-de-la-tomate>.
- XLSTAT. 2020. « Classification Ascendante Hiérarchique (CAH) ». XLSTAT, Your data analysis solution. 2020. <https://www.xlstat.com/fr/solutions/fonctionnalites/classification-ascendante-hierarchique-cah>.
- . 2021. « Forêts aléatoires de classification et de régression ». XLSTAT, Your data analysis solution. 2021. <https://www.xlstat.com/fr/solutions/fonctionnalites/forets-aleatoires-de-classification-et-de-regression>.

## Annexe

### Code R utilisé

```
#Descriptive :
library(ggplot2)
plot(IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 1 - J14`,IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`)
boxplot(IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 1 - J14`,IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 2- J14`)
boxplot((IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 2- J21`,IG_AQ10_T2_06_2020$`Nombre de tache/ feuille 3- J21`)
boxplot(IG_AQ10_T2_06_2020$`Nombre de colonie/ cm2 - J14`,IG_AQ10_T2_06_2020$`Nombre de colonie/ cm2 - J21`)
boxplot(IG_AQ10_T2_06_2020$`Nombre e spores/cm2 J21`,IG_AQ10_T2_06_2020$`Nombre e spores/cm2 J28`)
cor(IG_AQ10_T2_06_2020)

boxplot(Plante1$`Nombre de colonie/ cm2 - J14`,Plante1$`Nombre de colonie/ cm2 - J21`)
boxplot(Plante1$`Nombre de colonie/ cm2 - J14`,Plante2$`Nombre de colonie/ cm2 - J14`,Plante1$`Nombre de colonie/ cm2 - J21`,Plante2$`Nombre de colonie/ cm2 - J21`,Plante1$`Nombre de colonie/ cm2 - feuille3 J28`,Plante2$`Nombre de colonie/ cm2 - feuille3 J28`)
boxplot(Plante1$`Nombre de colonie/ cm2 - J14`)
boxplot((IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 2- J21`,IG_AQ10_T2_06_2020$`Nombre de tache/ feuille 3- J21`)
boxplot(IG_AQ10_T2_06_2020$`Nombre de colonie/ cm2 - J14`,IG_AQ10_T2_06_2020$`Nombre de colonie/ cm2 - J21`)
boxplot(IG_AQ10_T2_06_2020$`Nombre e spores/cm2 J21`,IG_AQ10_T2_06_2020$`Nombre e spores/cm2 J28`)
boxplot(Plante1$`Nombre e spores/cm2 J21`,Plante2$`Nombre e spores/cm2 J21`,Plante1$`Nombre e spores/cm2 J28`,Plante2$`Nombre e spores/cm2 J28`)

plot(Plante1$`Nombre e spores/cm2 J21`,Plante2$`Nombre e spores/cm2 J21`,Plante1$`Nombre e spores/cm2 J28`,Plante2$`Nombre e spores/cm2 J28`)
##### Variables ajustées
#on regroupe les données pour plante = 1&2 puis celles pour plante 3&4

Plante1<-IG_AQ10_T2_06_2020[IG_AQ10_T2_06_2020$Plante== 1 |
IG_AQ10_T2_06_2020$Plante== 2,]
Plante2<-IG_AQ10_T2_06_2020[IG_AQ10_T2_06_2020$Plante== 3 |
IG_AQ10_T2_06_2020$Plante== 4,]
boxplot(Plante1$`Nombre de colonie/ cm2 - J14`)
boxplot(Plante2$`Nombre de colonie/ cm2 - J14`)

boxplot(Plante1$`Nombre de colonie/ cm2 - J21`)
boxplot(Plante2$`Nombre de colonie/ cm2 - J21`)
```

```

#View(Plante1)
#View(Plante2)

### quelques TEST
pwr.t2n.test(n1=30, d=0.5, sig.level = 0.05, power=0.8, alternative = "greater") #On trouve 147 donc
pas bien
kruskal.test(IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 1 -
J14`,IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`)
shapiro.test(IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`)

wilcox.test(IG_AQ10_T2_06_2020$NC_1_J14,IG_AQ10_T2_06_2020$SF_J14,alternative="two.sided
")
t.test(IG_AQ10_T2_06_2020$NC_1_J14,IG_AQ10_T2_06_2020$SF_J14,alternative="two.sided")

#IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 1 - J14`,IG_AQ10_T2_06_2020$`Surface
foliaire (mm2) feuille 1 - J14`
wilcox.test(IG_AQ10_T2_06_2020$NS_J14,IG_AQ10_T2_06_2020$SF_J14,alternative="two.sided")

pairwise.wilcox.test(IG_AQ10_T2_06_2020$NC_1J14,IG_AQ10_T2_06_2020$T_inst)

pwr.t2n.test(n1=30,d=0.5,power = 0.8,sig.level=0.05,alternative="greater") #FIIIIIRE 50 WARNING
#bartlett.test(IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`,c(rep(1,30),rep(2,30)))
#Marche pas, utiles?

Kendall(effi_8$`Ef_N.C/cm2_J14`,effi_8$`Tem_N.C/cm2_J14`)

pairwise.wilcox.test(Data_col,as.numeric(data_T),p.adjust.method="holm",alternative="two.sided")

#tau = 0.18, 2-sided pvalue =0.13223
cor.test(effi_8$`Ef_N.C/cm2_J14`,effi_8$`Tem_N.C/cm2_J14`,method="kendall",exact = FALSE)
cor.test(effi_8$`Ef_N.C/cm2_J14`,effi_8$`Tem_N.C/cm2_J14`,method="kendall",exact = FALSE)

#corrélation :
crrr_efiTemp <-
matrice[,c(5,6,7,8,9,10,11,12,13,15,16,17,21,22,23,27,28,29,33,34,35,39,40,41,45,46,47)]
crrr_efiVPD <-
matrice[,c(5,6,7,8,9,10,11,12,14,18,19,20,24,25,26,30,31,32,36,37,38,42,43,44,48,49,50)]
crrr_efiTemp_aug <- matrice[,c(5,6,7,8,9,10,11,12,63:78)]
corrplot(cor(crrr_efiTemp),method = "circle")
corrplot(cor(crrr_efiVPD),method = "circle")
corrplot(cor(crrr_efiTemp_aug),method = "circle")

corTot<-
effi_8[c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,3
4,35,36),c(55,60,56,57,61,62,58,59,63,64)]

```



```

corrplot(cor(cor28),method = "circle")
ggcorrplot(cor(cor28))
## régression
model1 <- lm(IG_AQ10_09$`N.C./ cm2`-F3
J21~IG_AQ10_09$Moy_temp_1h+IG_AQ10_09$Moy_temp_2h+IG_AQ10_09$Moy_temp_3h+IG_A
Q10_09$VPD_inst,data=IG_AQ10_09)

```

```

#fig
ggplot(mpg, aes( x=class,y=hwy, fill=class)) +
  geom_boxplot()+
  xlab(label = "Jours") +
  ylab(label = "Nombre de colonies ") +
  theme(axis.text.x = element_text(angle=30, hjust=1, vjust=1))+
  theme(legend.position="none")+
  ggtitle("Boxplot Nombre de colonies par jour")

```

```
View(mpg)
```

```

## Descriptif
mean(IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`)
sd(IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`)
length(IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`)
hist(IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 - J14`, col = blues9,probability = T)
# Normal??

```

```

boxplot(IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 1 - J14`,alternative="two.sided")
boxplot(IG_AQ10_T2_06_2020$`Nombre de colonie/ cm2 - J14`)
boxplot(IG_AQ10_T2_06_2020$`Nombre de colonie/ cm2 - J21`)
boxplot(IG_AQ10_T2_06_2020$`Nombre de spores feuille 2 J21-1`)
boxplot(IG_AQ10_T2_06_2020$`Nombre e spores/cm2 J21`)

```

```

lines(IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 1 - J14`, col="red", lwd=2) #Prob?
boxplot(IG_AQ10_T2_06_2020$`Surface foliaire (mm2) feuille 1 -
J14`,IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 1 - J14`)
boxplot(IG_AQ10_T2_06_2020$`Nombre de colonies / feuille 2- J14`)
####
toto<-effi_all2[,c(8,17:66)]

```

```

toto<-na.exclude(toto)
toto2<-as.matrix(toto)

```

```

#fonction du package randomForest pour trouver une valeur optimale pour mtry = nb de variables
pour les split
mtry_opt<-tuneRF(toto[,-8], toto[,8], trymtry = 10, stepFactor = 1, ntreeTry = 10000, improve = 0.01,
doBest = FALSE)
ring<- ranger( EF_Pre_34_12_NC_cm_J14 ~ ., data=toto2, num.trees=10, mtry=16, replace=FALSE,

```

```

importance="permutation",respect.unordered.factors=TRUE)
shit.ring.14<-ring$variable.importance
shit2.ring.14<-shit.ring.14[order(shit.ring.14, decreasing=TRUE)]
par(mar=c(4.1, 10.1, 2.1, 2.1))
barplot(rev(shit2.ring.14[1:20]), horiz=TRUE, las=1, xlab="",

      main="2014")

#calcul de l'importance de chaque variable et plot par Random Forest
library(iml)
library("ggplot2")
X <- toto[which(names(toto) != "lag")]
pfun<-function(object, newdata) predict(object, data =

      newdata)$predictions
predictor <- Predictor$new(model=ring, data = X, y = toto$lag,

      predict.fun = pfun )
impo <- FeatureImp$new(predictor, loss = "mse", n.repetitions=10)#ici
juste 10 rep pour aller vite
plot(impo)

#Accumulated Local Effect (ALE) plots = relation entre un predicteur
#et la variable réponse

ale <- FeatureEffect$new(predictor, feature = "Dim.3")

ale$plot()

```

```

install.packages("randomForest")
library(randomForest)
Data_ALL.RF<-
randomForest(Traitement~IG_AQ10_dataall$N.C.cm2_J14+IG_AQ10_dataall$N.C.CM2_J21,data =
IG_AQ10_dataall,importance=TRUE,proximity=TRUE)
print(Data_ALL.RF)

Data_ALL.RF$importance
View(IG_AQ10_dataall)

```

## Exemple d'analyse effectuée mais non intégrée au rapport

ACP 1

Description du plan 1 :2

On applique maintenant le test de Wilks afin de connaître la quelle des variables qualitatives sépare le mieux les individus sur le plan 1 :2 :

## Heure

## 4.921271e-43

La meilleure variable qualitative pour illustrer les distances entre individus sur le plan est la variable : *Heure*.

Ce résultat conforte le résultat précédant concernant le fort impact du moment du traitement et de l'inoculation dans le développement de la maladie.

On retrouve dans le graphique suivant la distribution des individus (plantes) par rapport à cette variable qualitative heure :

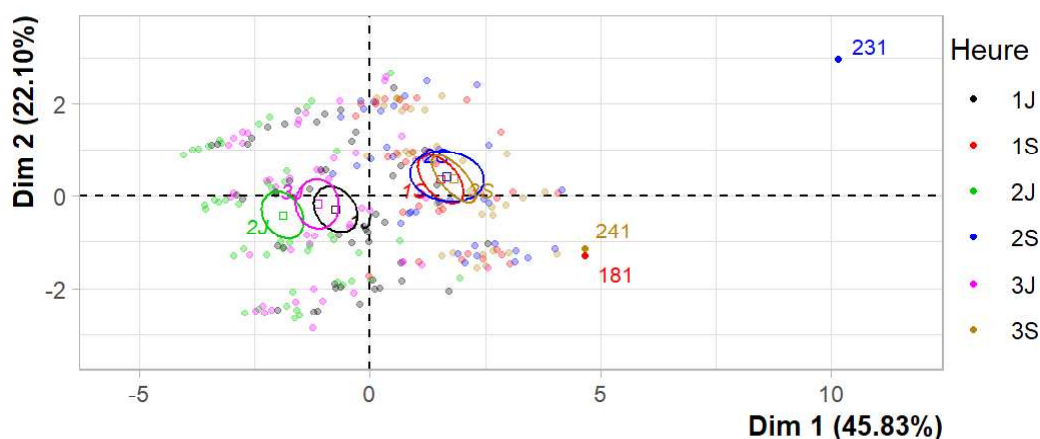


Figure 35 Représentation des individus suivant la variable heure

*On rappelle aussi que les individus libellés sont ceux ayant la plus grande contribution à la construction du plan. Les individus sont colorés selon leur appartenance aux modalités de la variable Heure.*

On a ajouté les ellipses de confiance aux individus afin d'avoir une aide graphique à localiser les individus. C'est des ellipses qui permettent de résumer les nuages de points avec comme objectif de voir si les modalités d'une variable qualitative sont significativement différentes les unes des autres. Elles peuvent aussi être paramétriser pour avoir l'intervalle de confiance.

On remarque que la maladie est beaucoup plus présente chez les plantes traitées en septembre que chez les plantes traitées en juin. En effet les 3 ellipses qui représentent les 3 modalités de la variable heures du mois de septembre sont très proches les unes des autres, et ont de très grandes valeurs sur l'axe comparées à celles de juin.

Maintenant que nous avons choisi de faire l'ACP avec deux dimensions grâce à l'analyse de l'inertie, on va analyser ces dimensions choisies.

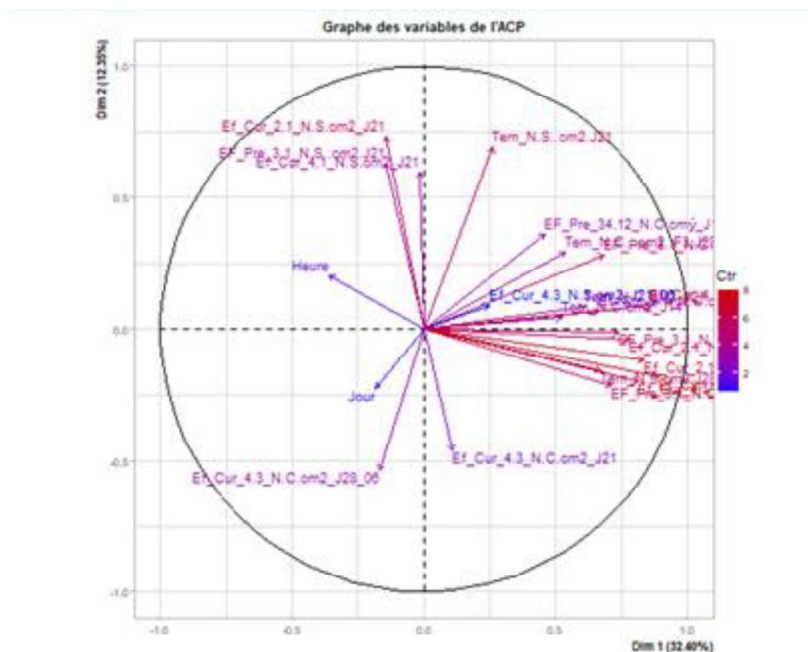


Il est tout de même difficile d'interpréter de faire l'étude des variables tels qu'elles sont, afin de remédier à ce problème, nous allons redéfinir les contributions des variables lors de la création des axes.

Les détails concernant cet ACP grossiers sont en Annexe.

Variables climatiques en variables supplémentaires :

On va maintenant faire une étude de l'ACP en prenant comme variables supplémentaires les variables climatiques :



## 2. Distribution de l'inertie

On rappelle que l'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'analyse expriment 45.85% de l'inertie totale du jeu de données ; donc 45.85% de la variabilité totale du nuage des individus et des variables sont représentées dans ce plan. C'est un pourcentage relativement moyen, et le premier plan représente donc seulement une part de la variabilité contenue dans l'ensemble du jeu de données actif.

Il est alors utile de considérer également les dimensions supérieures ou égales à la troisième dans l'analyse.

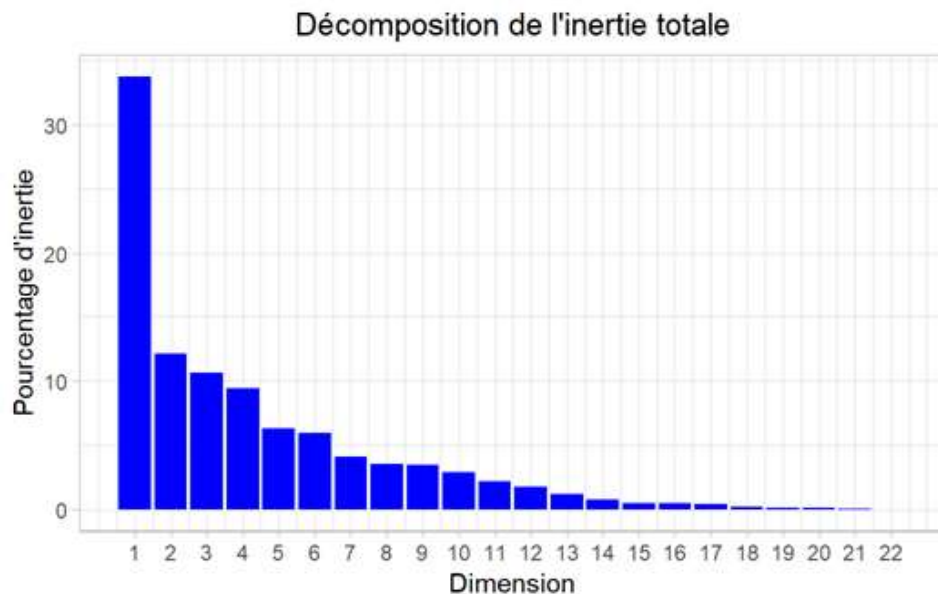


Figure 2 - Décomposition de l'inertie totale

On va donc restreindre l'analyse à la description des 4 premiers axes

### 3. Description du plan 1:2

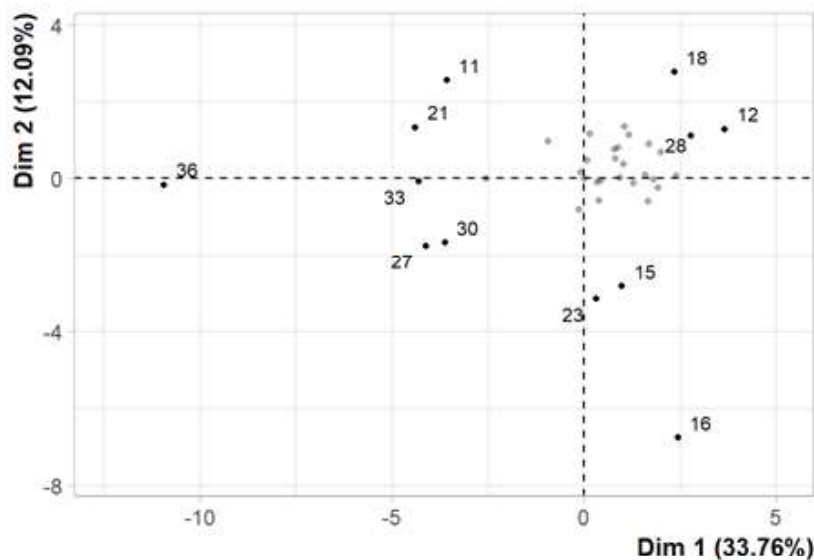


Figure 3.1 - Graphe des individus (ACP)

Comme pour les autres ACP, on va appliquer le test de Wilks pour voir la quelle des variables qualitatives séparent le mieux nos données.

## Place ## 0.9936398

La meilleure variable qualitative qui permet d'illustrer les distances entre individus sur le plan est la variable : *Place*, c'est la même variable qu'on a obtenue par le test de Wilks pour l'ACP grossier pour les mêmes dimensions. Par contre l'inertie cumulée par les deux dimensions ici est plus grande que pour l'ACP grossier.

#### Description des dimensions :

La dimension 1 oppose des individus tels que 12, 18 et 28 (à droite du graphique, caractérisés par une coordonnée fortement positive sur l'axe) à des individus comme 36, 21, 27, 11 et 30 (à gauche du graphique, caractérisés par une coordonnée fortement négative sur l'axe).

Le premier groupe contient les individus 12, 18 et 28, et qui ont des coordonnées positives par rapport aux deux axes. On peut caractériser ces plantes par une grande présence de la maladie chez les plantes traitées mais aussi les plantes témoins (principalement au j28) mais aussi avec une grande valeur de l'efficacité curative au J28.

Le groupe d'après contient un seul individu qui est la feuille n 36, *cette plante a un comportement atypique, elle a une coordonnée négative sur l'axe 2, et une valeur nulle pour l'axe 1.*

Le 3<sup>ème</sup> groupe qui inclut les individus : 21, 27, 11 et 30, et a comme caractéristiques une faible présence de la maladie chez les témoins, et avec de faibles efficacités.

La dimension 2 oppose des individus tels que 12, 18 et 28 (en haut du graphe, caractérisés par des coordonnées fortement positives sur l'axe) à des individus comme 16, 15 et 23 (en bas du graphe, caractérisés par des coordonnées fortement négatives sur l'axe).

Le groupe auquel les individus 16, 15 et 23 appartiennent (caractérisés par des coordonnées négatives sur l'axe) partage :

Pour le 4<sup>ème</sup> groupe, il regroupe les individus 12, 18 et 28 et qui ont des coordonnées positives sur l'axe partage une grande présence de la maladie chez les témoins, mais aussi de grande valeurs pour les efficacités préventives. On a pu observer que le traitement préventif a été meilleur quand les plantes sont assez contaminées

Le 5<sup>ème</sup> groupe regroupe les individus 16, 15 et 23, *et ces derniers ont en commun* de faibles valeurs pour les efficacités en nombre de spores.

#### 4. Description du plan 3:4

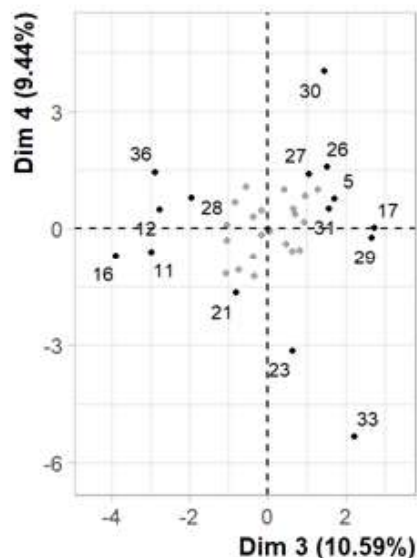


Figure 4.1 - Graphe des individus (ACP)

On remarque que les individus les plus représentatives et qui sont donc libellés sont celles qui sont décentralisées, c'est donc les individus avec « les plus grandes valeurs » si on effectue une projection des individus sur les axes de l'ACP (idem pour les variables avec un grand CTR).

La dimension 3 oppose des individus tels que 17, 29, 26, 30 et 31 (à droite du graphique, caractérisés par une coordonnée fortement positive sur l'axe) à des individus comme 12, 11, 16 et 28 (à gauche du graphique, caractérisés par une coordonnée fortement négative sur l'axe).

Le premier groupe se compose des individus 17, 29, 26, 30 et 31. Ces individus partagent des valeurs très grandes en température, et faibles en VPD (grande valeur de l'humidité)

Le deuxième groupe rassemble les individus 12, 11, 16 et 28. Ces individus sont caractérisés à la fois par une forte présence de la maladie chez les témoins, ainsi que de fortes valeurs de la VPD minimal, mais aussi de faibles valeurs de la VPD maximum. Principalement les VPD au bout de 7h. On peut alors se poser la question sur l'effet de la VPD à cet instant. Nous allons creuser un peu plus ce point dans la suite avec les Radom Forest.

La dimension 4 oppose des individus tels que 17, 29, 26, 30 et 31 (en haut du graphique, caractérisés par des coordonnées fortement positives sur l'axe) à des individus comme 33 et 23 (en bas du graphique, caractérisés par des coordonnées fortement négatives sur l'axe).

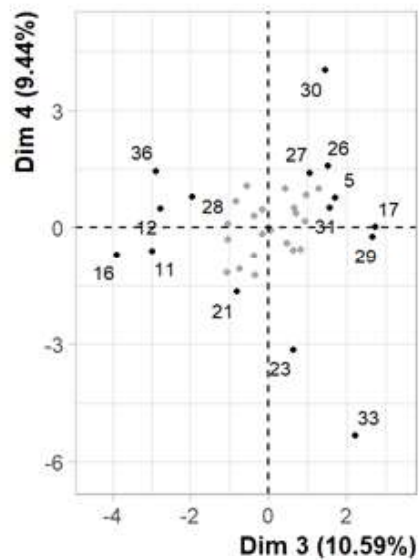
Le 3ème groupe de ce plan regroupe les individus 17, 29, 26, 30 et 31, ces individus ont donc de grandes valeurs pour la température, mais de faibles valeurs pour la min VPD au bout de 7h.

Le groupe auquel les individus 33 et 23 appartiennent (caractérisés par une coordonnée négative sur l'axe) partage :



Le 4<sup>ème</sup> groupe se compose que de deux individus. Ces deux individus ont de faibles valeurs d'efficacité curatives, ainsi que des températures moyennes faibles au bout de 6h après le traitement.

On va essayer maintenant de voir la répartition des plantes selon leurs informations climatiques. Pour y faire nous allons utiliser les deux ACP que nous avons fait : Le premier qui est fait avec les informations sur le nombre de colonies (par feuilles et par cm<sup>2</sup>), et le deuxième qui contient les efficacités.



La distribution des plantes selon la température instantanée de traitement ne nous semble pas très pertinente. En effet les températures instantanées c'est-à-dire au moment du traitement ne semble pas suivre de schéma de corrélation précis.

Afin de voir l'effet de la température plus en détails, on va procéder à tests statistiques suivants