

# Caractérisation des génotypes à travers la modélisation de cinétique de NDVI en fin de cycle chez le blé

Auteur : Xavier Martin

Septembre, 2024

Jury : Jérôme DANTAN (Enseignant-Chercheur, UniLaSalle)





## Remerciements :

Ce projet de fin d'études, réalisé dans le cadre de mon alternance chez Limagrain, m'a offert l'opportunité de participer à une aventure scientifique passionnante. L'étude des plantes et la biologie appliquée à l'amélioration variétale sont des domaines en constante évolution, ouvrant sans cesse de nouvelles perspectives de recherche. Je suis particulièrement reconnaissant envers mon tuteur, Clément Buet, pour son encadrement et son soutien tout au long de ce stage. Ses conseils m'ont été précieux pour maintenir le cap malgré ma tendance à la dispersion, et son aide dans la révision du rapport a été importante pour en améliorer la clarté. Mon objectif a été de présenter des concepts parfois complexes de manière accessible, en espérant communiquer efficacement leur intérêt et leur portée.

## Table des matières

Remerciements : .....	4
Résumé .....	1
Abstract .....	1
Abréviations et termes techniques .....	2
1 : Introduction .....	2
1 – 1 : Contexte .....	2
1 – 1 – A : Le blé et ses variétés .....	3
1 – 1 – B : La sélection variétale .....	4
1 – 1 – C : Le Phénotypage .....	5
1 – 1 – D : Le phénotypage digitale .....	5
1 – 1 – E : L’acquisition phénotypique par drone .....	6
1 – 2 : La sénescence des plantes .....	7
1 – 2 – A : La sénescence étude exploratoire avec le phénotypage digitale .....	7
1 – 2 – B : Du développement de l’inflorescence à la floraison .....	8
1 – 2 – C : De la fécondation à la mort de la plante .....	10
1 – 3 : Problématique .....	10
1 – 3 – A : Le NDVI et la sénescence .....	11
1 – 3 – B : La régression non linéaire .....	11
1 – 3 – C : Objectifs poursuivis .....	12
2 : Matériel et Méthodes .....	12
2 – 1 : Choix des fonctions .....	14
2- 2 : Méthodologie de régressions .....	16
2 – 3 : Méthode 1 : récupération des données depuis les régressions .....	18
2 – 4 : Méthode 2 : récupération des données depuis les régressions sur un même EnvXVar .....	18
Voir annexes .....	18
2 – 5 : Méthode 3 : combinaison des deux méthodes .....	18
2 – 6 : Nettoyage des données et création des nouvelles variables .....	18
2 – 7 : Méthodes d’analyses statistiques .....	19
3 : Résultats .....	20
3 – 1 : Modélisation .....	20
3 – 2 : Analyse univariée .....	26
3 – 3 : Classement inter-environnement sur les variables d’intérêt .....	30
3 – 4 : Analyse bivariée des comportements des variables d’intérêts ajustées .....	36
3 – 5 : Classement des variétés sur un score de performance global .....	40
4 : Discussion .....	43

4 – 1 : Modélisation et récupération des données issues des régressions .....	43
4 – 1 – A : Standardisation.....	43
4 – 1 – B : Choix des modèles .....	43
4 – 1 – C : Augmentation de la puissance statistique .....	44
4 – 2 : Annovas .....	45
4 – 2 – A : Approches globales.....	45
4 – 2 – B : Approche par environnement .....	45
4 – 3 : Classement des génotypes .....	45
4 – 3 – A : Apport de la méthode 2 dans ce classement .....	45
4 – 3 – B : Apport sur la variable de pente .....	46
4 – 3 – C : Analyse bivariée des scores .....	46
4 – 3 – D : Classement sur le score de performance .....	47
4 – 4 : Note sur les génotypes et explications des comportements.....	48
4 – 4 – A : Anti-corrélation entre le Stay-Green et la force de sénescence .....	48
4 – 4 – B : Qualités des variétés, Stay-Green et sénescence .....	48
4 – 5 : Perspectives supplémentaires .....	48
5 : Conclusion .....	49
6 : Bibliographie.....	50
7 : Annexes .....	52
7 – 1 : Autres équations testées .....	52
7 – 2 : Formules des dérivées de Gompertz .....	54
7 – 3 : Méthode numéro 2 de récupération des données.....	55
7 – 3 – A : Récupération des valeurs moyennes .....	57
7 – 3 – B : Méthode de récupération des bornes .....	62
7 – 3 – C : Méthode de récupération des valeurs de seuils pour chaque variable.....	65
7 – 3 – D : Transformation de la moyenne et des bornes pour la création de répliquas de synthèse .....	70
7 - 4 : Limites de la méthode 2 .....	70
7 – 5 : Classement des génotypes par cluster .....	71

## Résumé

Le blé tendre est une espèce cruciale pour l'alimentation humaine. Cette étude s'est concentrée spécifiquement sur l'examen de la sénescence des variétés de blé d'hiver. Grâce à la caractérisation des cinétiques de sénescence, deux traits clés ont été identifiés : la force de sénescence et le Stay-Green. Ces traits présentent un intérêt particulier car ils servent de prédicteurs du rendement (Diaz-Mendoza et al. (2016)).

La modélisation des cinétiques a été effectuée à l'aide de la fonction de Gompertz (Gompertz (1825)), qui s'est avérée la plus performante pour l'ajustement des séries temporelles. L'extraction des caractéristiques de ces cinétiques a été réalisée par deux méthodes indépendantes, toutes deux efficaces. La combinaison de ces méthodes a permis d'améliorer la lisibilité des tendances observées. Six variétés de blé ont été étudiées : Obiwan, Chevignon, Advisor, Arkeos, Extase et Rubisko.

Les résultats suggèrent des différences notables entre les variétés étudiées. Obiwan semble présenter un long Stay-Green mais une faible force de sénescence, tandis qu'Advisor montre apparemment des caractéristiques opposées. Arkeos paraît avoir un Stay-Green et une force de pente moins prononcés que Chevignon. Extase, quant à elle, semble démontrer la plus forte pente de sénescence et le Stay-Green le plus important. Il est important de noter que ces observations préliminaires nécessitent une confirmation par des études supplémentaires, impliquant une puissance statistique accrue et un plus grand nombre d'environnements.

## Abstract

Bread wheat is an essential species for human nutrition. This research concentrated on examining the senescence of winter wheat varieties. By characterizing senescence kinetics, two significant traits were isolated: senescence strength and Stay-Green. These traits are particularly valuable as they act as yield predictors, as shown by (Diaz-Mendoza et al. (2016)). For monitoring purposes, we utilized a non-destructive method: the kinetic tracking of a telemetry index (NDVI).

The results suggest notable differences among the studied varieties. Obiwan appears to exhibit a long Stay-Green period but a weak senescence strength, while Advisor seemingly displays opposite characteristics. Arkeos appears to have both a Stay-Green period and a slope strength less pronounced than Chevignon. Extase, on the other hand, seems to demonstrate the strongest senescence slope and the most significant Stay-Green effect. It is important to note that these preliminary observations require confirmation through additional studies involving increased statistical power and a greater number of environments.

The results suggest notable differences among the studied varieties. Obiwan appears to exhibit a long Stay-Green period but a weak senescence strength, while Advisor seemingly displays opposite characteristics. Arkeos appears to have both a Stay-Green period and a slope strength less pronounced than Chevignon. Extase, on the other hand, seems to demonstrate the strongest senescence slope and the most significant Stay-Green effect. It is important to note that these preliminary observations require confirmation through additional studies involving increased statistical power and a greater number of environments.

## Abréviations et termes techniques

Microparcelle : unité expérimentale au champ qui représente un individu dans les analyses

Cinétique-réplica : Ensemble de mesures réalisées sur une même microparcelle dans le temps.

EnvXVar : Combinaison Environnement X Variété composé de 3 à 4 cinétiques-réplica

Z55 : stade phénologique sur l'échelle de Zadocks correspondant à la demi-épiason du blé

VLB : Villiers-Saint-Christophe (Aisne, 02590)

Mons : Estrées-Mons (Somme, 80200)

Thizay : Thizay (Indre-et-Loire, 37500)

IC : intervalle de confiance

## 1 : Introduction

### 1 – 1 : Contexte

Les rendements du blé ont augmenté continuellement depuis les années 1950 jusqu'aux années 1990 dans pratiquement toutes les régions du monde (Calderini et Slafer (1998)), on remarque cependant une stagnation en France depuis la moitié des années 1990. Cet arrêt de l'augmentation du rendement ne semble pas être causée par l'atteinte du potentiel de la plante (Schauberger et al. (2018)). Les producteurs de blé subissent, en effet, des risques liés au réchauffement climatique et à l'apparition de nouvelles maladies créant des stress biotiques et abiotiques importants sur leurs cultures, en particulier sur les deux dernières décennies. De nombreux agents pathogènes comme la septoriose ont gagné des mutations de multirésistances dont la première apparition a eu lieu en 2008 (Omrane et al. (2015)). En conséquence les rendements n'ont pas augmenté depuis plus de deux décennies et ceux malgré les nombreuses innovations ayant eu lieu dans tous les secteurs de l'agriculture.

Parmi ces secteurs, l'industrie agrochimique par la découverte de nouvelles molécules actives pour le désherbage, la lutte contre certaines maladies des plantes et les ravageurs, reste essentielle dans le fonctionnement de l'agriculture moderne mais ces pratiques sont de plus en plus questionnées socialement et régulées par les autorités. Il est ainsi probable que les industries créant les nouveaux pesticides jouent un rôle moins important qu'autrefois pour répondre aux défis de l'agriculture de demain. Dans les nombreux domaines en amont de la production agricole, pour lequel les prochaines innovations joueront un rôle probablement majeur, on peut noter la sélection variétale, car, entre autres, les conséquences de la sélection de nouvelles variétés sur l'environnement sont mieux connues et impactent peu ou moins les milieux écologiques que l'industrie agrochimique.

## 1 – 1 – A : Le blé et ses variétés

Le blé occupe une place centrale dans l'alimentation humaine mondiale, constituant l'une des principales sources de calories et de protéines végétales. Sa polyvalence remarquable se manifeste à travers une multitude de produits alimentaires, allant du pain et des pâtes aux céréales pour petit-déjeuner. Cette diversité d'utilisations est rendue possible grâce à la variété des génotypes de blé, chacun possédant des caractéristiques spécifiques adaptées à différentes applications culinaires et industrielles. Ainsi, la richesse génétique du blé permet de répondre à une large gamme de besoins nutritionnels et technologiques dans l'industrie agroalimentaire.

Cette richesse génétique du blé se manifeste concrètement à travers une multitude de variétés cultivées, chacune présentant des caractéristiques uniques adaptées à des conditions de culture spécifiques et à des usages variés. Pour illustrer cette diversité, examinons quelques variétés de blé qui seront utilisées dans cette étude et qui sont utilisées dans l'agriculture actuellement :

Chevignon :

- Excellente productivité et régularité de rendement
- Très bon comportement face aux maladies, notamment la septoriose (*Zymoseptoria tritici*)
- Qualité BPS (Blé Panifiable Supérieur) valorisant la récolte

Advisor :

- Bonne tolérance aux maladies foliaires (septoriose, rouille jaune (*Puccinia striiformis*), oïdium (*Blumeria graminis*))
- Adapté aux semis précoces et tardifs
- Bon poids spécifique (densité du grain)

Arkeos :

- Référence des blés biscuitiers (BB)
- Très bonne résistance à la germination sur pied
- Précocité intéressante pour les régions à risque d'échaudage (sensibilité au stress chaud du blé en fin de cycle, arrêtant prématurément le remplissage du grain)

Rubisko :

- Résistant à la cécidomyie orange ; insecte ravageur du blé (*Sitodiplosis mosellana*)
- Bon comportement face à la fusariose (maladie fongique) et aux DON (déoxynivalénol, une mycotoxine produite par certaines espèces de *Fusarium*)
- Recommandé par l'ANMF (Association Nationale de la Meunerie Française) dans les mélanges BPMF (Blés Pour la Meunerie Française)

Obiwan :

- Très bonne résistance à la rouille jaune
- Tolérant au chlortoluron (herbicide)
- Bon potentiel de rendement en conduite intensive

KWS Extase :

- Excellent profil sanitaire, notamment contre la septoriose
- Très bon potentiel de rendement
- Qualité BPS avec un bon taux de protéines

## 1 – 1 – B : La sélection variétale

Il est vraisemblable que pour parvenir à atteindre les objectifs de rendement, de résilience, et les exigences de durabilité environnementale fixés par les divers organismes supranationaux, la sélection variétale jouera un rôle central. Continuer d'innover dans de nombreux domaines ou technologies concernant ce secteur sera alors probablement décisif. Ces domaines sont en effet nombreux. Il y'a dans le développement de nouvelles variétés par les industries semencières comme Limagrain de nombreux sujets de travaux. Cela touche aussi bien les études sur les aspects physiques des farines produites par les variétés réservées que le dépôt légal de brevets protégeant l'innovation (généralement sur des gènes de résistance aux maladies).

Ces domaines d'études nombreux font chacun appel à de nombreuses savoir-faire. Ceux-ci vont des méthodes de séquençage de génome d'espèce entière, au développement d'algorithmes de prédiction du rendement pour différents types d'environnement. En définitive, ces technologies ne sont pas seulement liées à la biologie mais elles font appel à d'autres domaines comme la physique, l'informatique et les mathématiques.

On peut en outre, nommer parmi ces domaines multimodaux, un qui est particulièrement célèbre : le génie génétique. Le génie génétique permet de comprendre et prédire le comportement d'un panel de variétés ainsi que d'établir selon les positions des gènes d'intérêts sur le génome, des stratégies de croisements pour retrouver ces caractères dans les variétés qui seront commercialisés. La génétique donne accès aux sélectionneurs à une très grande quantité d'information dont le traitement, la caractérisation et les applications concrètes pour la sélection variétale sont pour elle, un espoir et des promesses d'innovations importantes.

Ces avancées permises par le génie génétique ont été dans les décennies passées très importantes. Toutefois si ce domaine génétique est encore plein d'avenir pour la sélection variétale, la nécessité de caractériser les plantes demeure. En effet, les agriculteurs souhaiteront constater dans les variétés qu'ils achèteront demain, un comportement des plantes obtenus sur le terrain. Si les gènes sont fortement corrélés à l'obtention de ces comportements, il reste toutefois des variables environnementales qui peuvent impacter la présence d'un caractère et il convient de vérifier par l'expérimentation en plein champ l'obtention des caractères désirés. Constater un caractère désiré sur le terrain est un domaine entier dans la sélection variétale et il fait appel à de nombreux systèmes digitaux à intégrer dans des approches larges.

## 1 – 1 – C : Le Phénotypage

Cette caractérisation des plantes appelée phénotypage permet en effet de relier les données moléculaires obtenues par le génie génétique à des caractères des plantes. Le phénotype constitue l'ensemble des caractères que peut posséder une plante. La définition d'un caractère reste relativement difficile à donner mais son illustration par des exemples est suffisant pour comprendre sa portée. On peut citer parmi les caractères les plus communs : le rendement, la teneur en protéine du grain, la résistance à une maladie donnée, le nombre de feuilles à un stade phénologique donnée. Ces caractères peuvent être mesurés par :

- Des grandeurs physiques
- D'autres types de caractérisations non physique : allant du simple décompte d'organes à des calculs de concentration
- Des données ordinales comme l'évaluation de l'impact d'une maladie sur une plante par un indice standard.

Les caractères sont en somme autant de caractéristiques utiles pour décrire les plantes. Les caractères utilisés peuvent être différents selon l'espèce d'intérêt.

La diminution drastique de l'obtention de marqueurs et des coûts du séquençage, ont déjà donné de nombreuses technologies et avancées pour la sélection variétale. Le facteur limitant dans la sélection de nouvelles variétés ne se situe alors plus tant au niveau génotypique qu'au niveau de la description phénotypique des plantes. Si les industriels de la sélection variétale ont cherché jusqu'ici à relier leurs données moléculaires à quelques caractères précis comme le rendement (sous différentes conditions), la résistance à la verse, ou à la résistance à certaines maladies, de nombreuses autres caractéristiques pourraient être mesurées dans les années à venir.

En effet, en plus des stress croissants économiques et climatiques que nous avons déjà mentionnés, on peut également prendre en compte d'un point de vue de l'histoire récente de l'amélioration variétale l'atteinte par les semenciers d'un plafond de rendement des plantes et de résistance aux maladies en particulier pour les plantes de grandes cultures. Si le potentiel génétique réel de la plante n'est pas encore acquis on note cependant de nombreux caractères d'intérêts qui subissent des anti-corrélations génétiques persistantes (Roby (2023)). On peut citer la teneur en protéine par rapport au rendement (source interne) rendant plus difficile l'amélioration conjointe de ces deux caractéristiques. Dans ces conditions, il est intéressant de participer à toutes les possibilités d'innovations permises et notamment par l'étude exploratoire sur de nouvelles données qui ne sont encore pas, ou peu, exploitées jusqu'à présent.

## 1 – 1 – D : Le phénotypage digitale

Certains caractères d'intérêt sur les plantes peuvent être mesurés rapidement et avec une grande précision lors des différents passages d'engin agricoles du semis à la récolte. En effet, certains outils de captures d'informations embarquées sur les machines agricoles permettent d'obtenir pour la parcelle mesurée une cartographie précise du rendement et de diverses autres caractéristiques. Toutefois peu

de caractères peuvent être mesurés grâce à ces outils embarquées à l'heure actuelle. Ces caractères ne couvrent pas tout le spectre des caractéristiques descriptives des plantes pouvant intéresser les industriels semenciers. Ces technologies offrent néanmoins un gain majeur sur la notation humaine pour la vitesse, la précision et les coûts réduits qu'elles autorisent.

Les outils de captures d'informations embarquées sur les machines agricoles ne sont pas seuls. Si des caméras sont utilisées sur ces engins pour décrire certains caractères, les possibilités d'utiliser ces appareils de capture d'images sur d'autres supports sont très nombreuses. Ces appareils photo sont d'un point de vue physique, des capteurs d'ondes électromagnétiques permettent de mesurer en chaque point de l'espace couvert (pixel) un certain nombre de longueurs d'ondes données. Ces capteurs renverront en information, pour chacune de ces longueurs d'ondes et pour chacun des pixels, une valeur d'intensité. En utilisant des capteurs capables de récupérer des longueurs d'ondes différentes des longueurs d'onde classiquement utilisée dans les images (bleu, vert et rouge) correspondant aux 3 couleurs observés par l'œil l'humain, on obtient autant de nouvelles variables permettant d'explorer des aspects moins connus des plantes.

Les possibilités offertes par ces appareils de mesure se déclinent aussi par les vecteurs d'acquisition comme les satellites, le drone ou le smartphone et sont regroupées dans les technologies de l'imagerie. L'ensemble de ces techniques sont étudiées et utilisées à Limagrain par l'équipe de phénotypage digital permettant d'aider les sélectionneurs dans leurs prises de décisions.

#### 1 – 1 – E : L'acquisition phénotypique par drone

Dans cette équipe, l'acquisition par image drone reste l'élément central pour la partie phénotypage et il utilise un grand nombre de sites partout dans le monde. Sa précision, l'envergure des données qu'elle permet de collecter en fait un atout majeur pour Limagrain et les entreprises de sélection en général. En une heure de vol, des centaines de variétés sont photographiées plusieurs fois. Il est possible de reconstituer en trois dimensions ces images et d'établir de nombreuses caractéristiques des plantes à partir de ces reconstitutions ou simplement des images brutes.

Chaque nouvelle variété créée met en moyenne une dizaine d'année à être inscrit au catalogue officiel. En effet, un nombre d'étapes important nécessitent d'être réalisées de l'ébauche des caractéristiques voulues dans la nouvelle variété créée à sa commercialisation. L'édification des caractéristiques d'intérêt est la première étape de la création d'un nouveau produit commercial. Pour la sélection variétale, cette étape est complexe car les personnes responsables de l'étude du marché et de l'établissement des spécifications doivent prédire les besoins d'un marché dix ans à l'avance. Ces caractéristiques une fois établies serviront de guide tout au long du processus de sélection et de création de la variété. Ces caractéristiques seront regroupées dans un idéotype qui comptera ainsi l'ensemble des phénomènes attendus selon des conditions de cultures déterminées.

Les sélectionneurs doivent pouvoir s'assurer que les génotypes qu'ils croiseront pour créer les futures variétés expriment les caractéristiques d'intérêts dans les conditions pédoclimatiques de leurs clients. Ainsi des expérimentations en plein champ, de grande ampleur, pour réaliser un phénotypage global et précis sont essentielles dans le processus de sélection. Les caractères étudiés dans le cadre du phénotypages sont nombreux mais il est encore possible d'augmenter leurs nombres pour explorer de

nouveaux aspects des plantes permettant de différencier les génotypes entre eux et aider les sélectionneurs à faire des choix durant le processus de création des nouvelles variétés.

## 1 – 2 : La sénescence des plantes

Notre étude s’inscrit donc dans la continuité de l’implémentation de ces nouveaux caractères. Nous nous sommes ici intéressés à la fin du cycle de développement du blé tendre d’hiver.

Nous avons plus précisément étudié des valeurs d’un indice télémétrique indiquant un état général de vitalité de la plante (moyennée sur un micro-parcelle) pour différentes dates, différentes variétés et différents environnements de cultures.

Une variété avec une vitesse de sénescence élevée permet comparativement avec une variété avec une phase de remplissage du grain équivalente (Stay-Green), un rendement équivalent ou meilleur avec une récolte plus précoce. Cela limitera pour l’agriculteur, les risques liés aux aléas climatiques, biologiques et économiques (Alt et al. (2019)), (Saphir (2020)).

### 1 – 2 – A : La sénescence étude exploratoire avec le phénotypage digitale

La digitalisation récente du phénotypage est une possibilité supplémentaire de combler ce manque d’études exploratoires. Le phénotypage digitale, par imagerie drone, permettrait d’explorer de nouvelles possibilités d’obtenir un nouveau type de marqueur caractérisant les plantes, à même d’inférer plus tard, un caractère d’intérêt. Cela peut être simplement fait en utilisant un indice de végétation moyennée sur les pixels d’une même image de micro-parcelle (accession génétique cultivée sur une zone donnée) sur plusieurs dates. Cela donnerait donc une évolution de cet indice à même de donner la possibilité d’obtenir des caractères inférant et discriminants les génotypes.

Depuis une dizaine d’année, cette optique prospective de découverte et de mesure de caractères phénotypiques nouveaux est un sujet d’étude plein d’espoir. En effet, l’accumulation grandissante des informations phénotypiques sur les plantes permises par l’imagerie drone recèle de possibilité de mesurer de nouvelles caractéristiques comme le nombre de grains par épi, la verse, la hauteur de plantes, par le traitement des images à l’aune du machine learning, du deep learning et de l’ia. De nombreuses perspectives d’approches sont également possibles dans ce traitement d’images. En effet, les nombreux vols effectués sur la saison, permettent d’obtenir des indices de santé ou de vigueur des plantes, à partir de ratio de longueur d’ondes électromagnétiques mesurées et dont l’évolution, au cours du temps, peut référer en elle-même à une information d’intérêt pour le sélectionneur.

Afin de prédire le rendement et les caractères agronomiques d’intérêts en sélection il est parfois préférable de mesurer des caractères le plus tôt possible dans le cycle de développement des plantes pour adapter l’itinéraire technique si des corrélations avec certains comportements ultérieurs des plantes sont connus. Cependant, les caractères apparaissant dans les dernières étapes du développement de la plante sont les plus prédictifs du rendement. Ces dernières étapes sont généralement moins héritables car plus contrainte par l’environnement que les étapes de début et de milieu de cycle de développement de la plante (Zeng et al. (2017); De Meyer et al. (2023); Chang et al. (2021)). Elles ont cependant un intérêt agronomique majeur car elles illustrent le comportement

de la plante vis-à-vis de l'allocation de ses ressources et donc du potentiel rendement et la qualité du grain obtenu.

## 1 – 2 – B : Du développement de l'inflorescence à la floraison

Le blé est une graminée avec un cycle de vie monocarpique. C'est-à-dire que le blé ne fleurit qu'une fois. Les plantes de blés sont également des plantes annuelles. Ces plantes monocarpiques et annuelles utilisent les ressources acquises au cours de l'année pour créer des graines sèches ou orthodoxes, capable de survivre l'hiver tandis-ce que le plant mère dessèche et meurt. Les plantes monocarpiques comme le blé ont ainsi au cours de l'année des étapes de développement bien précises. On note 8 étapes dans le cycle de développement caractérisant la culture du blé qui sont : la germination, la levée, le tallage, la montaison, l'épiaison, la floraison et fécondation, le remplissage des grains et la maturation.

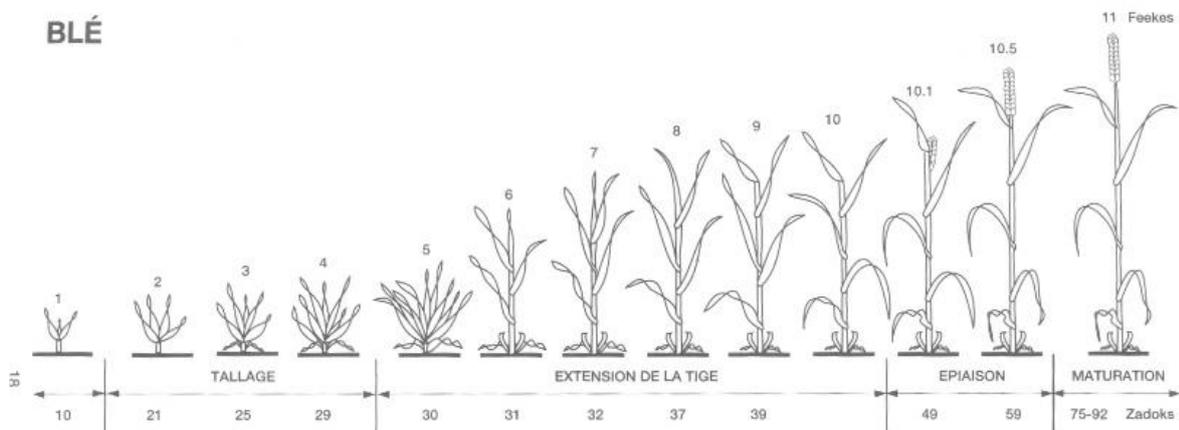


Figure 1 : Illustration des différents stades de développement du blé selon l'échelle BBCH (Ach (2006))

Illustration chez le blé pour différentes fourchettes de l'échelle BBCH d'une plante de blé. Indication écrite en dessous du stade phénologique correspondant.

L'inflorescence du blé émerge progressivement de la graine foliaire cela marque le début de l'épiaison A ce stade l'activité photosynthétique est maximale et la canopée est dense (Kirby (2002); Lobell et al. (2015)). Nous nous intéresserons plus précisément à l'étude de la dynamique de la plante de la fin de la formation de son épi, jusqu'à sa mort et son dessèchement complet. La *Figure 1* présente les différents stades phénologiques selon l'échelle BBCH (notons que l'échelle BBCH et Zadocks sont identiques pour le stade BBCH 55 ou Z 55 (Lindsey, Paul, et Lentz (2017)) ). L'étude porte sur une gamme assez large de temporalité



Figure 2: (Guo, Chen, et Schnurbusch (2018)) Les 7 étapes du développement et des stades d'avortements des fleurons durant la phase d'élongation de l'épi

TS : stade de l'épillet terminal, WA : l'anthere blanche, GA : stade de l'anthere verte, YA : stade de l'anthere jaune, TP : Stade de transition pré épiaison « tipping », HD : épiaison, AN : Anthèse

L'épi ressemble d'abord à un épillet minuscule marquant le début de la phase reproductive pour la plante de blé (figure 2 'terminal spikelet' ou stade d'épillet terminal TS). Cette phase est particulièrement sensible à la variation de température et à la disponibilité en eau, qui peuvent affecter le nombre de grains sur l'épi (Farooq et al. (2011)).

Le stade de l'anthere verte, (figure 2 'Green Anthere', GA) caractérisé par la présence d'anthers vertes et turgescents dans les fleurs centrales (Kirby et Appleyard (1984)). Ce stade est également la fin de l'initiation des fleurs comprises dans les épillets chez le blé (Guo, Chen, et Schnurbusch (2018)).

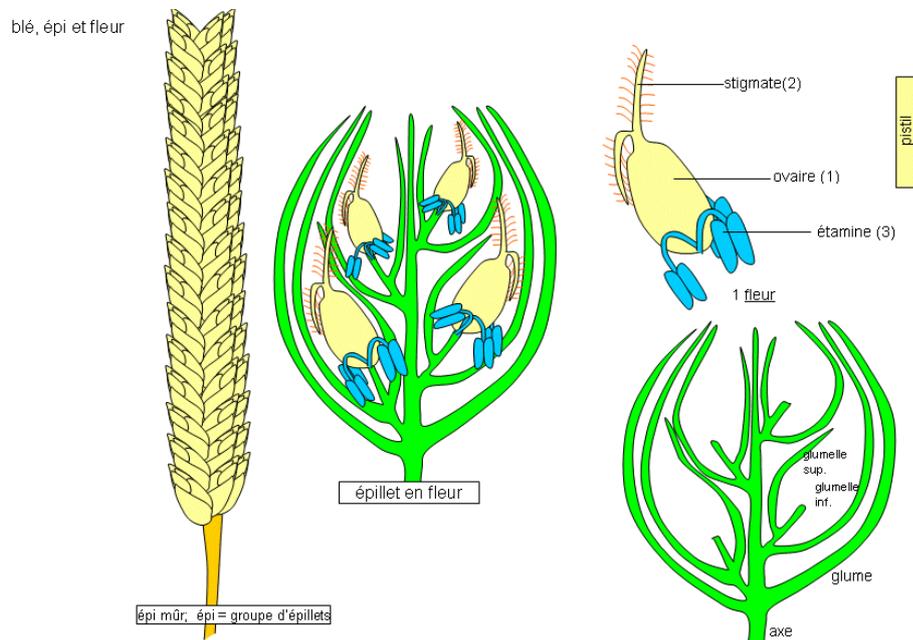


Figure 3: (Gallien (2008)) structures de l'épi de blé

4 dessins : à gauche, l'épi complet ; au centre, un épillet et ses 4 fleurs, en bas à droite : un épillet sans fleurs, en haut à droite une fleur.

Au stade suivant, l'anthere jaune (YA voir Figure 2), les glumes (Figure 3) sont totalement formées et une des feuilles stériles appelée 'lemme' devient visible. Au stade suivant TP, les barbes qui forment le bout des lemme commencent à apparaître. Enfin, à l'avant dernier stade présent sur la (Figure 2) l'épiaison, sera défini comme le moment où 50 % des épillets de l'épi auront des barbes visibles. Cela correspond au stade Z55 dans le développement du blé (Figure 1). Enfin, le dernier stade de la figure 2 montre l'anthere et ainsi la floraison.

## 1 – 2 – C : De la fécondation à la mort de la plante

Ces dernières étapes de la vie de la plante qui marquent également le développement de la graine sont particulièrement importantes. 48 heures après le début de la floraison, la germination marque en effet le début du développement de l'organe d'intérêt pour l'agriculteur : la graine. La plante mère, lors de la sénescence caractérisée par le dessèchement de la plante et sa mort, réalloue toutes ses ressources aux graines. Ce développement se déroule en 4 étapes :

- 1- La transformation de l'ovaire en grain
- 2- Le remplissage du grain
- 3- Le stade Pâteux
- 4- La Maturation

Les deux premières étapes se déroulent pendant la dernière phase photosynthétique de la plante (Shewry et al. (2012)). La troisième étape marque quand a-elle le début de ce dessèchement (Couturier (2014)). Enfin, la dernière étape s'effectue pendant le dessèchement de la plante et de la réallocation totale des ressources vers les graines (Baillot (2020)).

La capacité de la plante à rester verte longtemps (Stay-Green) et à engager une sénescence rapide dans l'étape qui suit (drying pre-harvest) semblent être des étapes essentielles pour la prédiction du rendement (Diaz-Mendoza et al. (2016)) et pour prédire la qualité, en particulier la quantité de protéine dans le grain (GPC) (Distelfeld, Avni, et Fischer (2014)). La sénescence est aussi le moment de la remobilisation des ressources de la plante. Durant cette période, la plante remobilise des ressources en carbone et en azote vers les graines.

Certains processus cellulaires de la sénescence, interviennent durant cette remobilisation des ressources. L'autophagie et la protéolyse sont des étapes clés permettant de récupérer dans le phloème les produits de dégradations des cellules des organes foliaires, qui seront ainsi amenés aux graines (Diaz-Mendoza et al. (2016)). Cette étape est particulièrement prononcée chez les espèces monocarpiques qui réemploient l'intégralité de leur ressource pour les graines entraînant leur mort (Distelfeld, Avni, et Fischer (2014)). Ces processus cellulaires de la sénescence font peu à peu baisser l'activité photosynthétique de la plante. Les longueurs d'ondes spécifiques de la chlorophylle et des feuilles habituellement renvoyées lorsque la plante est à son pic de maturité apparaissent de moins en moins (Sanger (1971)). Une surveillance de la diminution de cet indice à la fin de la vie de la plante, peut permettre de caractériser ce processus de sénescence.

## 1 – 3 : Problématique

Nous nous concentrerons donc ici sur l'analyse prospective de nouveaux caractères à même de différencier des cultivars sur une évolution cinétique d'index télémétrique en fin de cycle chez le blé d'hiver.

### 1 – 3 – A : Le NDVI et la sénescence

De l'inflorescence à sa mort, le blé peut-être suivi par un indice de végétation, le NDVI qui est un outil non invasif permettant d'évaluer l'état physiologique des cultures (Rouse et Haas (1974)). Le NDVI est un index basé sur la différence de longueur d'onde normalisé (Normalized Difference Vegetation Index).

$$NDVI = \frac{NIR_{850} - RED_{650}}{NIR_{850} + RED_{650}}$$

Les longueurs d'ondes utilisées correspondent à l'importance de la bande d'absorption de la chlorophylle  $RED_{650}$  par rapport à la longueur d'onde la plus émise par la structure foliaire  $NIR_{850}$ . Plus la chlorophylle absorbe, moins le second terme de la soustraction au numérateur diminue la valeur de NDVI. Le dénominateur permet de normaliser la mesure en fonction de la densité de la structure foliaire et de l'intensité lumineuse générale pour la longueur d'onde absorbé par la chlorophylle. Il est ainsi relativement indépendant de l'exposition lumineuse générale et de la structure du spectre ainsi que de l'importance du couvert végétal étudié. L'équation du NDVI devrait ainsi isoler relativement bien la vitalité de la plante des biais précédemment mentionnés.

Il devrait ainsi logiquement bien représenter la première étape de la sénescence qui consiste en un dessèchement et désactivation de la photosynthèse mais au maintien de la structure carbonée de la plantes (tiges et feuilles). Si l'on suggère que le 'plant senescence reflectance index' (PSRI) serait plus efficace pour mesurer la sénescence, le NDVI semble rester un bon indice de mesure de cette sénescence (Anderegg et al. (2020)).

### 1 – 3 – B : La régression non linéaire

La relation étudiée ici est entre une variable et le temps. Cette variable est le NDVI et le temps étudié correspond aux jours écoulés depuis le semis (ou un autre stade phénologique d'intérêt des plantes de blé utilisé pour standardiser les mesures).

Nous cherchons ici à modéliser cette relation. Nous utiliserons une régression non linéaire. Contrairement à la régression linéaire, la régression non linéaire permet d'ajuster des courbes et des formes plus sophistiquées aux données. Les modèles non linéaires peuvent prendre diverses formes, telles que polynomiales, exponentielles, logarithmiques ou sinusoïdales, offrant ainsi une grande flexibilité pour s'adapter à différents types de données. Plus difficile à mettre en œuvre et à interpréter, la régression non linéaire est souvent obligatoire lorsque la relation entre les variables étudiées est loin d'être linéaire.

Cependant, elle présente aussi des défis par rapport à la régression linéaire car elle met en jeu plus souvent plus qu'un ou deux paramètres et parfois on obtient plusieurs courbes différentes qui relient les points bruts avec un coefficient de détermination ( $R^2$ ) identique ou très semblable conduisant à des aberrations et à la nécessité d'interprétation humaine. Cette problématique intervient en particulier quand le nombre de paramètre de l'équation utilisée est quasiment le même que le nombre de points bruts utilisés dans l'algorithme d'obtention de ces paramètres. Malgré ces difficultés, la régression non linéaire reste un outil puissant et largement utilisé dans l'analyse de données et la modélisation

prédictive, permettant aux chercheurs et aux analystes, de visualiser ou de comprendre, des schémas complexes dans leurs données.

Une fois les points ajustés avec plusieurs équations, il convient de ne sélectionner qu'une seule de ces équations. Cela se fait sur trois points majeurs.

- La parcimonie : la capacité du modèle à prédire les points avec le moins de paramètres possibles
- Une faible erreur de prédiction : la capacité du modèle à prédire au mieux les points bruts
- La cohérence biologique : l'explication biologique derrière le modèle (Paine et al. (2012))

### 1 – 3 – C : Objectifs poursuivis

L'étude de la cinétique par plusieurs points de mesure de NDVI en fin de cycle des cultivars de blé d'hiver a été ici réalisé. L'objectif était de caractériser le Stay-Green (le temps ou la plante continue sa période photosynthétique après les derniers stades phénologique) et la sénescence (fin de la photosynthèse). Cette étude constitue une cartographie du comportement de l'évolution des indices de végétation en fin de cycle dans différentes conditions mais surtout la volonté de différencier et la tentative de marquer l'héritabilité sur 6 cultivars différents les caractéristiques extraites de ces cinétiques de NDVI.

De la maturité à la sénescence des plantes, dans différentes conditions, ces cinétiques offrent une cartographie du comportement de l'évolution du NDVI sur différentes variables.

Nous allons donc procéder par étapes pour tenter d'extraire des informations issues des modélisations des cinétiques. L'objectif est ainsi d'avoir des informations valorisables par l'équipe chargée de la mise des nouvelles technologies de phénotypage pour Limagrain.

## 2 : Matériel et Méthodes

L'analyse a porté sur un tableau de 1597 mesures réparties en 210 répétitions. Ce tableau comporte des informations concernant la modalité de culture, le bloc, la variété, la date de semis, date de Z55 de Z65, le numéro de mesure effectué dans la cinétique, la date de cette mesure, le ST\_mesure\_Z55 et la valeur de NDVI. Chaque cinétique fait entre 5 et 10 points uniques.

Essai	Annee	Modalite	Bloc	Var_minu	Semis	Date_Z55	Date_Z65	Num_mesure	Date_mesure	ST_mesure_Z55	Valeur_mesure_MDVI
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	1	24/05/2023	88.3	0.8
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	2	31/05/2023	205.3	0.74
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	3	07/06/2023	332.5	0.71
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	4	15/06/2023	506.75	0.71
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	5	21/06/2023	640.5	0.61
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	6	26/06/2023	747.5	0.3
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	7	30/06/2023	828.55	0.16
VLB	2023	D1	1	Advisor	19/10/2022	19/05/2023	27/05/2023	8	03/07/2023	887.05	0.16

Figure 4 : Tableau des données étudiées

Tableau représentant les premières lignes de données pour l'étude qui ont été utilisé pour réaliser les cinétiques avec différentes dates et informations de catégories disponibles.

La Figure 4, représente l'ensemble des types d'informations et les premières lignes sont un exemple visuel des informations disponibles pour les analyses ultérieures, de la régression à la caractérisation.

Variétés	Modalités	Explication modalité	Année	Lieu	Bloc
Advisor	Normal	-	2022	Mons	1
Extase	D2	Semis précoce	2023	VLB	2
Arkeos	X80	80 kg.h <sup>-1</sup> NH <sup>3</sup> /NH <sup>4+</sup>		Thizay	3
Obiwan					Parfois 4
Chevignon					
Rubisko					

Figure 5 : Tableau explicatif des modalités

Tableau récapitulatif des types de données et des différents niveaux

La Figure 5 représente les différentes variables d'intérêt présent dans le tableau source. Il résume la répartition des données et les niveaux des facteurs.

Trois objectifs ont donc été réalisés :

- 1- Explication du choix des fonctions testés
  - a. Gompertz
  - b. Logistique 'classique' (Verhulst (1838))
  - c. Logistique 'forte'
  - d. Logistique Gompertz
  
- 2- Méthodologie de régression
  
- 3- Obtention de 3 variables d'intérêts avec pour chaque un calcul empirique de l'intervalle de confiance pour une combinaison 'conditions de cultures X cultivars expérimentés'.
  - a. Obtention de la perte de NDVI entre l'asymptote haute et basse => Utilisation des paramètres asymptote haute et basse de l'équation et calcul d'un IC (expliqué après)
  - b. Obtention du paramètre de mi pente => Utilisation du paramètre de mi pente et calcul d'un IC (expliqué après)
  - c. Obtention du paramètre de pente max => Utilisation de la pente max sur la moyenne des 3 ou 4 courbes de chaque cinétique en chaque points (10000 de -30 à 60 jours après Z55) et calcul d'un IC (expliqué après)

## 2 – 1 : Choix des fonctions

Pour choisir le meilleur modèle on a dans un premier temps utilisés les équations disponibles dans la littérature scientifique pour les modèles de croissance (Paine et al. (2012)) ou sont mis en perspective deux équations intéressantes : la Logistique (dont la version symétrique à 3 paramètres a déjà utilisée dans la sénescence du blé (Crain, Reynolds, et Poland (2017))) et la fonction de Gompertz (déjà utilisée également (Kouadio (2012))).

J'ai ensuite ajouté deux fonctions pour élargir les champs de recherche habituellement utilisée pour décrire la sénescence. J'ai ainsi créé une fonction appelée 'Logistique forte' cette fonction permet de décrire une différence entre la force de l'accélération faible du début de la pente et la décélération forte de la fin de la pente plus importante encore que le déséquilibre décrit par Gompertz.

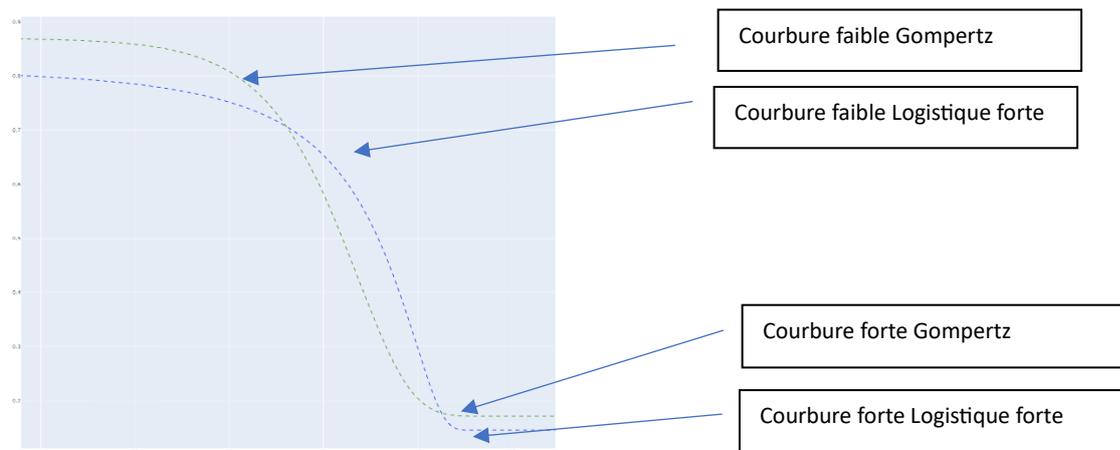


Figure 6 : Illustration de la courbe 'Logistique forte' par rapport à la courbe de 'Gompertz'

Courbe de Gompertz (vert) et 'Logistique forte' (bleu) dans leurs phases descendantes. Explications pour chaque partie de la courbe à droite.

La Figure 6 montre en effet l'apport exploratoire que permet l'essai d'une modélisation par la fonction sigmoïdale 'Logistique forte' créée. Cela permet de décrire une sigmoïde dont le rapport de la force de courbure du bas de pente sur le haut de la pente est supérieur à celui observé sur une courbe générée par la fonction de Gompertz. Ce rapport est toujours positif pour Gompertz et constant, il fixe en effet la hauteur du point d'inflexion de la sigmoïde entre l'asymptote min et max, non à 50 % comme le ferait une fonction sigmoïde Logistique classique, mais à 37.5 % au-dessus de l'asymptote basse.

Enfin, une quatrième fonction utilisée est une fonction qui combine la fonction Logistique forte et la fonction de Gompertz appelée 'Logistique-Gompertz'.

Pour toutes les équations qui suivent ; les paramètres utilisés sont les suivants :

- a = Asymptote haute
- b = force de la pente
- c = force de la pente de début par rapport à celle de fin
- d = position de la pente sur l'axe des abscisses
- e = Asymptote basse

*Gompertz*

$$f(x) = e + (a - e) \times e^{-1 \times e^{-b(x-d)}}$$

*Logistique*

$$f(x) = e + \frac{a - e}{(1 + c \times (e^{b(x-d)}))^{1/c}}$$

*Logistique 'forte'*

$$f(x) = e + \frac{(a - e) \times e^1}{(1 + c \times (e^{b \times (1 + c \times (e^{b(x-d)})^{1/c}}))^{1/c}}$$

*'Logistique-Gompertz'*

$$f(x) = \frac{e + (a - e) \times e^{-1 \times e^{-b(x-d)}}}{2} + \frac{e + \frac{(a - e) \times e^1}{(1 + c \times (e^{b \times (1 + c \times (e^{b(x-d)})^{1/c}}))^{1/c}}}{2}}$$

## 2- 2 : Méthodologie de régressions

Les régressions sont faites sur chaque série temporelle de mesure de NDVI réalisée sur une micro-parcelle (ou bloc). On compte entre 3 et 4 cinétiques uniques pour une même combinaison de conditions de cultures (environnement et modalité) avec la variété utilisée, c'est-à-dire une répétition. Les points bruts ajustés ont pour ordonnée la valeur du NDVI et pour abscisse la durée en jour entre cette mesure et la date du stade phénologique Z55 remarquée pour la variété cultivée (dans une micro-parcelle donnée). Z55 est le stade phénologique correspondant à la moitié de l'inflorescence émergée chez les céréales sur l'échelle de Zadock (Zadoks, Chang, et Konzak (1974)). Les cinétiques ont été recentré autour de ce stade phénologique car c'est un standard largement utilisé notamment pour la comparaison des mesures temporelles chez les céréales.

La fonction `curve_fit` du package `scipy.optimize` a été utilisé pour toutes ces régressions.

Uniquement les répétitions avec 6 points ou plus ont été utilisés pour l'ajustement.

Pour ajuster les paramètres de ces fonctions, les paramètres initiaux suivant ont été utilisés :

- Gompertz :  $a = 0.87$ ,  $e = 0.145$ ,  $c = 0.165$ ,  $d=43$
- Logistique :  $a = 0.7$ ,  $e = 0.12$ ,  $b = 0.172$ ,  $c=0.001$ ,  $d=37$
- Logistique forte :  $a = 0.722$ ,  $e = 0.14$ ,  $b=0.14$ ,  $c=1.05$ ,  $d=25.7$
- Logistique-Gompertz :  $a = 0.84$ ,  $e = 0.1425$ ,  $b=0.152$ ,  $c=1.05$ ,  $d=31.2$

Les limites suivantes ont été utilisées pour la fonction 'Logistique forte' et la fonction 'Logistique-Gompertz' :

- $0.4 < a < 1.3$
- $0.05 < e < 0.25$
- $0.02 < b < 1.1$
- $0.0001 < c < 1.1$
- $5 < d < 40$

Pour les fonctions 'Logistique forte' et 'Logistique-Gompertz', on a utilisé la méthode 'dogbox' de `curve_fit` pour performer l'ajustement aux données brutes.

Une fois les régressions réalisées, les paramètres de l'équation et la matrice de variances des paramètres sont récupérées pour chaque cinétique ajustée.

On calcule ensuite différents indicateurs qui permettront de connaître la performance de ces régressions pour chaque cinétique. Ces indicateurs ont été calculés comme suit :

- Calcul du coefficient de détermination ( $R^2$ ) : 
$$R^2 = 1 - \frac{\sum_{i=1}^{nbptscin} (NDVIBRUT_i - NDVIPREDIS_i)^2}{\sum_{i=1}^{nbptscin} (NDVIBRUT_i - NDVIMEAN)^2}$$
- Calcul du maximum de vraisemblance (LL) :
- $$LL = -0.5 \times (nbptscin \times \log(2 \times \pi)) + nbptscin \times \log\left(\frac{R^2}{nbptscin}\right) + nbptscin$$
- Calcul du critère d'information d'Akaike (AIC) :  $2 \times k - 2 \times LL$
- Calcul du critère d'information bayésien (BIC) :  $k \times \log(nbptscin) - 2 \times LL$

Nous choisissons ensuite selon ces paramètres la meilleure fonction permettant d'ajuster les données.

Après une étude visuelle des courbes obtenues par ces régressions sélectionnées, 3 paramètres semblent réellement indépendants et intéressants à explorer. Il s'agit :

- 1- Le drop de NDVI
- 2- La position de la sigmoïde sur l'axe des abscisses
- 3- La pente maximale de la courbe

- **3 méthodes de récupérations des données ont été utilisées**

## 2 – 3 : Méthode 1 : récupération des données depuis les régressions

On récupère les trois variables précédemment mentionnées ainsi :

- 1- Le drop de NDVI issu de la soustraction du facteur d'asymptote haute par l'asymptote basse
- 2- La position de la sigmoïde sur l'axe des abscisses, par le paramètre dédié dans la régression
- 3- La pente maximale de la courbe, par la valeur minimale de sa dérivée première

## 2 – 4 : Méthode 2 : récupération des données depuis les régressions sur un même EnvXVar

Voir annexes

## 2 – 5 : Méthode 3 : combinaison des deux méthodes

Les données issues des méthodes 1 et 2 ont été récupéré et assemblé dans un même tableau.

## 2 – 6 : Nettoyage des données et création des nouvelles variables

On a appliqué à l'ensemble des valeurs issues des 3 méthodes, un nettoyage des valeurs aberrantes. Dans un premier temps les valeurs de Drop\_NDVI ont été borné de 0,2 à 0,82. Il est à noter que pour la méthode 3, les valeurs de Drop\_NDVI n'ont pas été capé pour la borne inférieure à 0,2. Les valeurs de Stay-Green ont été borné à 55 pour la valeur maximale.

Avec les données suivantes :

- D1 correspond au décile inférieur de la variable
- D9 au décile supérieur
- Pm à la Pente\_MAX
- Pa à la pente ajustée et Sg au Stay-Green

Les nouvelles variables créées répondent aux équations suivantes :

- Pente ajustée =  $\frac{PenteMAX}{DropNDVI}$
- Score\_Pente\_MAX =  $\max(\min(f(Pm), 140), 0) \mid f(Pm) = 20 + \left(\frac{Pm - Pm_{D9}}{Pm_{D9} - Pm_{D1}}\right) \times 100$
- Score\_Pente\_ajustée =  $\max(\min(f(Pa), 140), 0) \mid f(Pa) = 20 + \left(\frac{Pa - Pa_{D9}}{Pa_{D9} - Pa_{D1}}\right) \times 100$
- Score\_Stay-Green =  $\max(\min(f(Sg), 140), 0) \mid f(Sg) = 20 + \left(\frac{Sg - Sg_{D9}}{Sg_{D9} - Sg_{D1}}\right) \times 100$
- Score de performance =  $\frac{ScorePenteAjustée}{ScoreStayGreen}$

## 2 – 7 : Méthodes d'analyses statistiques

Une analyse statistique a été menée pour chaque environnement afin de déterminer s'il existait des différences significatives entre les variétés. Cette analyse s'est déroulée en deux étapes :

1. Test initial : Le test de Kruskal-Wallis a été appliqué pour évaluer globalement les différences entre les variétés. Ce test non paramétrique est particulièrement adapté pour comparer plusieurs groupes indépendants.
2. Analyse post-hoc : Lorsque le test de Kruskal-Wallis révélait des différences significatives ( $p\text{-value} < 0,05$ ), une analyse post-hoc a été effectuée. Cette étape supplémentaire permet d'identifier précisément quelles variétés diffèrent les unes des autres.

Les positions de ces ensembles, pour une méthode donnée et un environnement donné ont été conservé pour établir les différentes positions dans le classement dans des tableaux. Ces tableaux ont dû être adapté pour la correspondance des tests post-hoc entre environnement. Différentes méthodes de répartition des clusters vides entre les environnements pour les tableaux de classement inter-environnements ont été testé lorsque le nombre d'ensembles récupérés entre chaque environnement n'était pas toujours le même :

### *Simplification méthodologique 1*

- 1- Les ensembles seuls ont été placé au milieu ou sur la place juste à droite du milieu du tableau lorsque le milieu n'existait pas (nombre impair de cluster maximum du tableau)
- 2- Les ensembles en nombre pair ont été divisé en deux et placé aux extrémités
- 3- Les ensembles impairs en nombre supérieur à 1 ont été séparé par un de plus à l'avant des clusters

### *Simplification méthodologique 2*

- 1- Les ensembles seuls ont été placé au milieu ou sur la place juste à gauche du milieu du tableau lorsque le milieu n'existait pas (nombre impair de cluster maximum du tableau)
- 2- Les binômes ont été placé au position 2 et 4 pour les tableaux de 5 ensembles
- 3- Les trios ont été placé au position 1, 3 et 5
- 4- 4 ensembles ont été placé aux extrémités

L'apparition des variétés dans les ensembles a été utilisé pour créer des graphiques de densité de classement.

Les annovas, les tests post-hoc et les tests de Kruskal-Wallis ont été performé avec le package statmodels de python

Le test de Shapiro a été réalisé avec le package scipy

Les clusterisation ont été réalisé avec les méthodes DBSCAN et RandomForestClassifier de sklearn

### 3 : Résultats

#### 3 – 1 : Modélisation

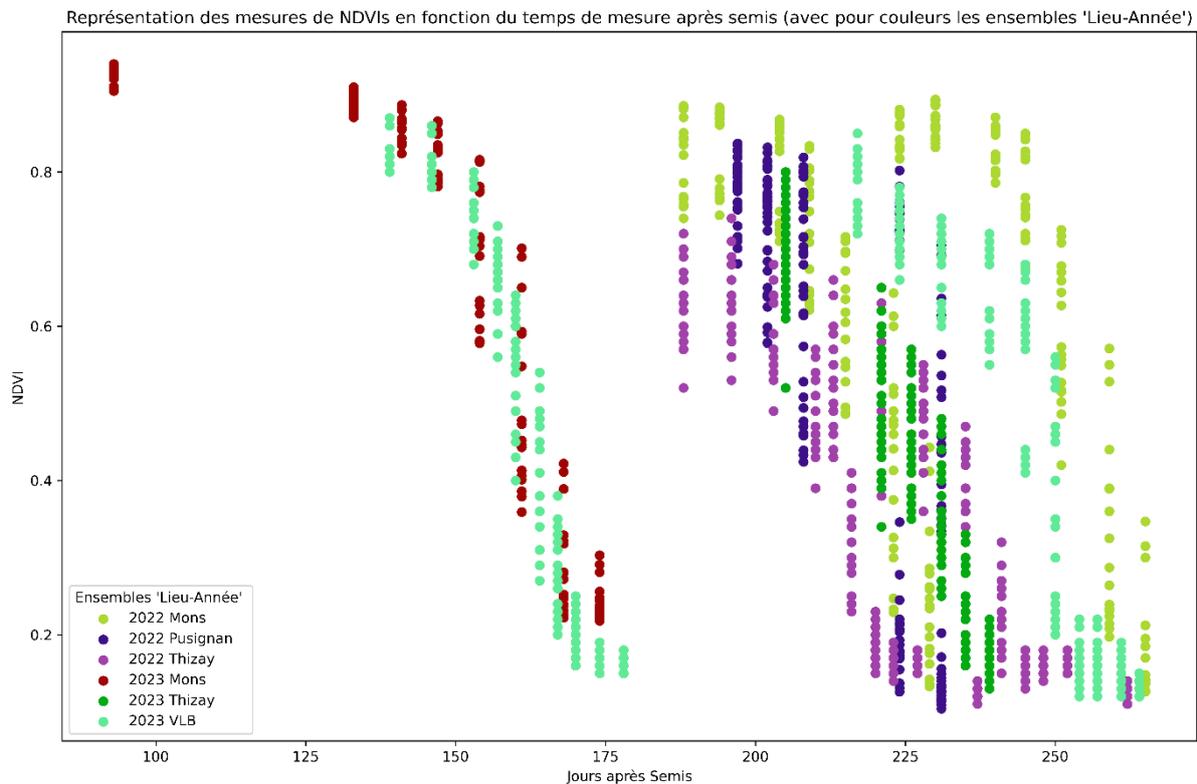


Figure 7 : Nuages de points représentant l'ensemble des mesures de NDVI depuis la date de semis

L'axe des ordonnées représente les valeurs de NDVI et l'axe des abscisses la position de cette mesure en nombre de jours après le semis. Les couleurs représentent un ensemble année et lieu.

L'analyse des données brutes, classées par couleur selon l'année et le lieu, révèle une variabilité significative dans la durée de vie des différentes variétés de plantes. Dans les environnements de VLB et Mons, deux périodes de semis ont été testées : précoce et tardive. Le semis tardif a résulté en des cycles de vie plus courts pour les plantes. Ce phénomène s'explique par le fait que la sénescence est influencée non seulement par la durée de vie intrinsèque de la plante, mais aussi par des facteurs saisonniers tels que la photopériode et la température.

Cette observation du décalage dans les cycles de vie souligne l'importance cruciale de recalibrer la comparaison des mesures en utilisant un stade phénologique spécifique comme point de référence. Dans ce cas, le stade de mi-épiaison, également connu sous le code Z55, est proposé comme repère pour standardiser les comparaisons entre les différentes conditions de croissance et variétés.

L'affichage des points bruts sur cette première figure ne permet pas de différencier, par simple lecture graphique, des différences particulières sur des paramètres de pente ou de différence de NDVI entre l'asymptote haute et basse. L'analyse de points bruts est difficile en l'état et l'étude de la cinétique par régression sera plus à même de récupérer des points d'intérêts utiles pour la caractérisation des variétés.

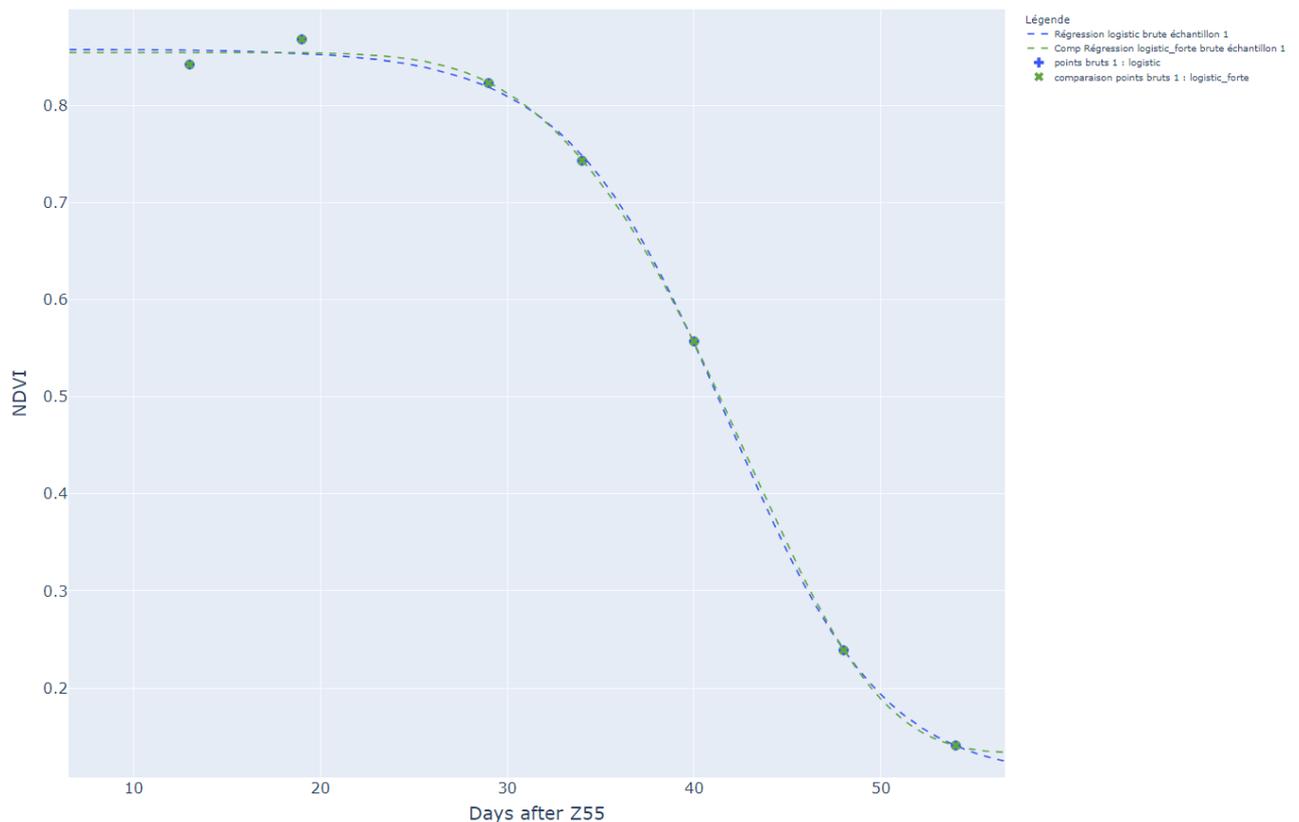


Figure 8 : Nuages de points du NDVI en fonction du nombre de jours après Z55 de la mesure et courbes de régressions comparant les fonctions Logistique et Logistique forte sur une répétition (cinétique unique)

L'axe des abscisses représente le nombre de jours après le stade phénologique Z55 constaté pour le cultivar cultivé sur la micro-parcelle considérée de la mesure effectuée. Sur l'axe des ordonnées est représenté la valeur de NDVI de cette mesure effectuée. Le réplica utilisé : Bloc de répétition (non standardisé (numéro 1)), Variété 'Advisor', lieu 'Mons', année '2024', modalité de culture 'semi précoce ou D1'.

Les analyses de la figure 8 nous montre que l'évolution des cinétiques est très proche des points bruts pour toutes les fonctions. La visualisation d'une sigmoïde de sénescence est largement visible. On remarque une courbure plus faible au début de la pente et une inflexion plus forte à la fin de la pente.

Les valeurs de NDVI observées suivent une évolution typique, allant d'un maximum de 0,7 à 0,9 (caractéristique des plantes en pleine croissance) jusqu'à 0,1 à 0,2 (indiquant la sénescence). La période principale de cette transition, représentée par une courbe sigmoïde, s'étend généralement sur environ 30 jours après le stade Z55 (mi-épiaison).

La figure 8 compare deux modèles mathématiques couramment utilisés dans la littérature pour décrire cette évolution : la fonction logistique et la fonction de Gompertz. Bien que les deux modèles présentent des comportements très similaires, la fonction logistique semble avoir légèrement plus de difficulté à s'ajuster précisément au point observé à 29 jours comparée à la fonction de Gompertz. Cependant, ces différences sont minimales et doivent être interprétées avec prudence.

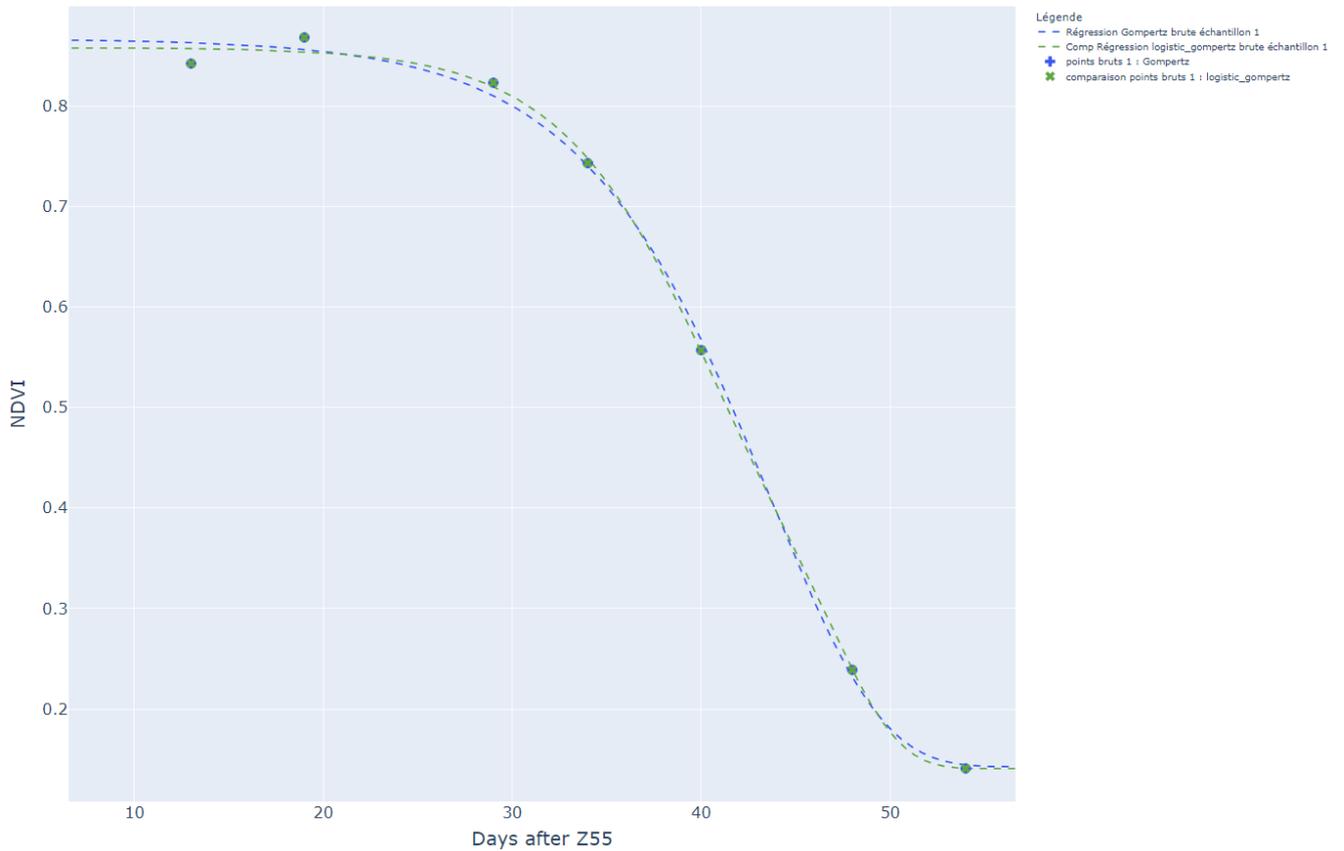


Figure 9 : Nuages de points du NDVI en fonction du nombre de jours après Z55 de la mesure et courbes de régressions comparant les fonctions Gompertz et Logistique-Gompertz sur une répétition (cinétique unique)

Même graphique que le graphique précédent. Seules les fonctions utilisées pour réalisées les régressions permettant de représenter les courbes changent. Ici sont visibles : Logistique-Gompertz en vert et Gompertz en bleu, les points représentent les données brutes utilisées

Les valeurs de NDVI évoluent selon l'axe du temps en forme de sigmoïde mais il est intéressant de comparer les sigmoïdes que l'on peut générer. Sur les exemples donnés dans les Figure 8 et Figure 9 on remarque que les différences majeures dans les comportements des courbes se situent sur les inflexions de haut et de bas de pente. L'inflexion du haut de la pente semble être la plus atteinte, de toutes les parties de la courbe, par les différences entre les modèles testés.

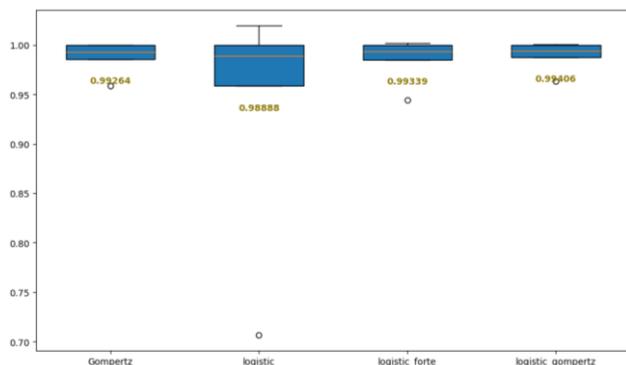


Figure 10: Diagramme en 'boite à moustache' du coefficient de détermination ( $R^2$ ) sur l'ensemble des cinétiques en fonction de l'équation utilisée dans la régression

Dispersion et moyenne des coefficients de détermination calculé pour l'ensemble des réplicas. Valeurs moyenne par fonctions utilisées :

Gompertz : 0.99264 ; Logistique : 0.98888

Logistic\_forte : 0.99339 ; logistic\_Gompertz : 0.99406

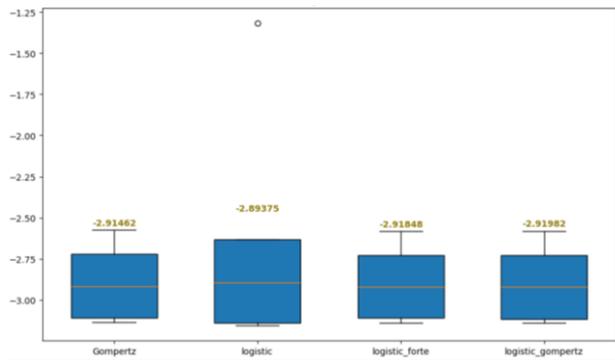


Figure 11: Diagramme en 'boite à moustache' du maximum de vraisemblance (LL) sur l'ensemble des cinétiques en fonction de l'équation utilisée dans la régression

Dispersion et moyenne des maximums de vraisemblance calculé pour l'ensemble des répliques. Valeurs moyenne par fonctions utilisées :

Gompertz : -2.91462 ; Logistique : -2.89375

Logistic\_forte : -2.91848 ; logistic\_Gompertz : -2.91982

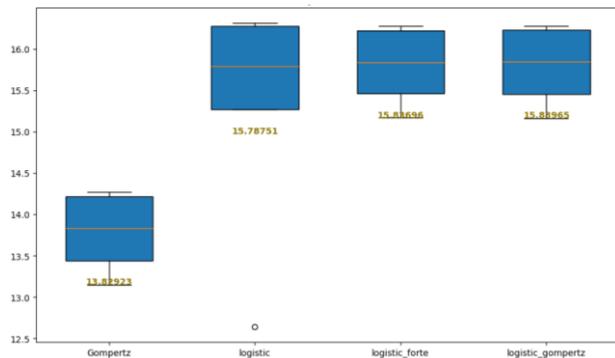


Figure 13: Diagramme en 'boite à moustache' du critère d'information d'Akaike (AIC) sur l'ensemble des cinétiques en fonction de l'équation utilisée dans la régression

Dispersion et moyenne du critère d'information d'Akaike calculé pour l'ensemble des répliques. Valeurs moyenne par fonctions utilisées :

Gompertz : 13.82923 ; Logistique : 15.78751

Logistic\_forte : 15.83696 ; logistic\_Gompertz : 15.88965

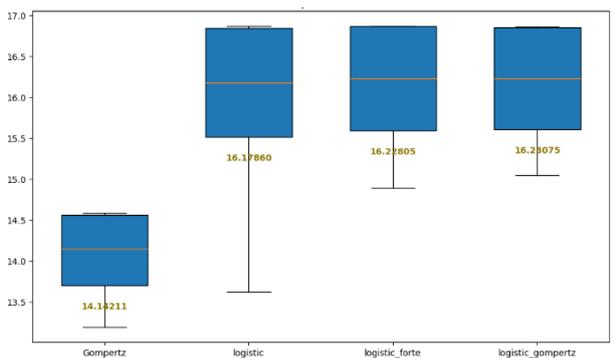


Figure 12: Diagramme en 'boite à moustache' du critère d'information bayésien (BIC) sur l'ensemble des cinétiques en fonction de l'équation utilisée dans la régression

Dispersion et moyenne du critère d'information bayésien calculé pour l'ensemble des répliques. Valeurs moyenne par fonctions utilisées :

Gompertz : 14.14211 ; Logistique : 16.17860

Logistic\_forte : 16.22805 ; logistic\_Gompertz : 16.23075

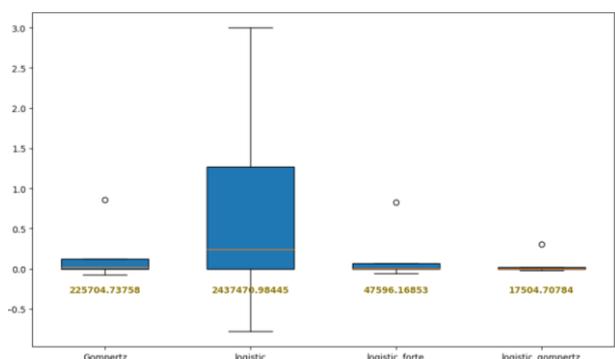


Figure 14: Diagramme en 'boite à moustache' de l'erreur global (Voir mat et med) sur l'ensemble des cinétiques en fonction de l'équation utilisée dans la régression

Dispersion et moyenne des maximums de vraisemblance calculé pour l'ensemble des répliques. Axe des ordonnées en logarithme de base 10<sup>7</sup>. Valeurs moyenne par fonctions utilisées :

Gompertz : 2.26\*10<sup>5</sup> ; Logistique : 2.44\*10<sup>6</sup>

Logistic\_forte : 4.76\*10<sup>4</sup> ; logistic\_Gompertz : 1.75\*10<sup>4</sup>

Pour choisir ces modèles sur des données numériques, une simple étude graphique étant largement insuffisante, nous utilisons donc des comparateurs de performances. Nous avons ici représenté graphiquement la dispersions de ces valeurs de comparaisons sur l'ensemble des 174 cinétiques utilisées. Le coefficient de détermination est très bon pour les 4 fonctions testés. Les sigmoïdes sont

en effet représentées par relativement peu de point ce qui permet un ajustement plus facile car il y'a peu de degré de liberté.

La fonction additionnant 'la Logistique forte' à la fonction de Gompertz, la fonction 'Logistique-Gompertz', a le meilleur coefficient de détermination. Ensuite, la fonction Logistique forte puis la fonction de Gompertz complète par ordre décroissant les fonctions décrites ici. Enfin, la fonction Logistique à 4 paramètres avec moins de 0,99 de moyenne de coefficient de détermination, termine le classement.

Après avoir présenté les forces et faiblesses relatives des modèles basés sur le  $R^2$ , il est pertinent d'examiner une autre approche d'évaluation de la qualité d'ajustement : le maximum de vraisemblance. Contrairement au  $R^2$  qui se concentre sur la variance expliquée, le maximum de vraisemblance offre une perspective différente en estimant les paramètres qui maximisent la probabilité d'observer les données.

Ici le classement sur les valeurs absolus reste inchangé mais le critère pour le choix du modèle est inversé. Ces deux paramètres seuls donnent donc des données contradictoires qui pourront être discutées. Ils ne permettent pas de sélectionner le meilleur modèle qui sera utilisé pour la suite des analyses. Nous continuons donc avec l'AIC et le BIC. Les critères d'information d'Aïkike et le critère d'information bayésien indiquent tous deux, que le modèle décrit par la fonction de Gompertz est le meilleur pour ajuster les 174 cinétiques.

Enfin la *Figure 14*, illustre la dispersion de la somme des écarts-types des paramètres estimés pour différents modèles. Une valeur élevée de cette somme indique une plus grande incertitude globale sur les paramètres, ce qui peut se traduire par des ajustements de qualité similaire obtenus avec des combinaisons de paramètres différentes.

Dans un contexte où nous disposons d'un nombre limité de données pour établir les cinétiques, comparativement au nombre de paramètres utilisés dans les modèles, et où tous les modèles testés présentent des coefficients de détermination très élevés, il devient crucial d'utiliser des critères supplémentaires pour sélectionner le modèle le plus approprié.

L'analyse des erreurs globales des paramètres estimés révèle des différences significatives entre les modèles :

1. Les fonctions 'Logistique forte' et 'Logistique-Gompertz' se distinguent par des erreurs globales nettement inférieures à celles des autres modèles.
2. Le modèle de Gompertz présente des erreurs globales modérées.
3. La fonction Logistique simple se démarque par des erreurs globales non seulement très élevées, mais aussi fortement dispersées.

Ces observations sur les erreurs globales fournissent un critère de discrimination important entre les modèles, permettant une sélection plus éclairée du modèle le plus robuste et le mieux adapté aux données disponibles, au-delà de la seule considération des coefficients de détermination.

Modalité	Année	Lieu	Variété	Bloc	Staygreen	Drop_NDM	Pente_MAX	Pente_ajustée	Score_Pente_MAX	Score_Pente_ajustée	Score_Staygreen	Score
X80	2023	Thizay	ADVISOR	1	28.5370	0.7662	-0.0199	-0.0260	6.0841	4.2413	14.4536	9.3475
X80	2023	Thizay	ADVISOR	2	29.5706	0.6832	-0.0204	-0.0299	7.2445	8.9633	19.6436	14.3034
Normal	2023	Thizay	ADVISOR	1	30.9799	0.7956	-0.0240	-0.0302	15.8088	9.2815	26.7196	18.0006
X80	2023	Thizay	RUBISKO	1	32.0353	0.5707	-0.0172	-0.0302	0.0000	9.3593	32.0188	20.6890
X80	2023	Thizay	ARKEOS	1	31.7392	0.5782	-0.0193	-0.0333	4.4911	13.1678	30.5318	21.8498
X80	2023	Thizay	OBIWAN	1	33.8945	0.6564	-0.0182	-0.0277	1.8906	6.3064	41.3536	23.8300
Normal	2023	Thizay	ADVISOR	3	30.5849	0.6627	-0.0283	-0.0427	26.1966	24.6361	24.7361	24.6861
X80	2023	Thizay	ARKEOS	2	31.5544	0.5230	-0.0213	-0.0406	9.2493	22.0889	29.6041	25.8465

Figure 15 : Tableau des valeurs récupérées par la méthode numéro 1 et nouvelles variables calculées

Premières lignes du tableau des valeurs récupérées sur les régressions des cinétiques uniques par les méthodes décrites page. 18, les colonnes à gauche de 'pente\_MAX', à partir de la 9<sup>ème</sup> colonne sont calculés par les méthodes décrites page. 18. Le tableau a été classé avec la dernière colonne, le score, sensé indiquer la performance générale de la variété sur un Stay-Green long et une sénescence rapide.

Les données récupérées sur les cinétiques sont compilées dans des tableaux. Figure 15, nous présentons les premières lignes classé par ordre décroissant de score de performance. La méthode 1 a donc produit 174 individus uniques réparties en 54 EnvXVars. Le classement par le score de performance permet de voir sur la variable construite ayant fait intervenir le plus d'opérations intermédiaires, les éventuelles valeurs aberrantes.

Modalité	Année	Lieu	Variété	Rep	Pente_MAX	Staygreen	Drop_NDM	Pente_ajustée	Score_Pente_MAX	Score_Pente_ajustée	Score_Staygreen	Score
X80	2023	Thizay	ADVISOR	3	-0.015	28.357	0.581	-0.025	0.000	1.163	14.453	7.808
X80	2023	Thizay	ADVISOR	1	-0.018	28.989	0.726	-0.025	7.332	1.163	18.472	9.818
X80	2023	Thizay	ADVISOR	2	-0.022	30.078	0.820	-0.027	16.159	3.224	25.394	14.309
X80	2023	Thizay	ARKEOS	3	-0.014	29.730	0.424	-0.034	0.000	12.078	23.181	17.630
Normal	2022	Thizay	ARKEOS	3	-0.025	26.516	0.493	-0.050	22.666	33.675	2.748	18.211
Normal	2023	Thizay	ADVISOR	3	-0.021	31.087	0.664	-0.032	14.629	10.192	31.812	21.002

Figure 16 : Tableau des valeurs récupérées par la méthode numéro 2 et nouvelles variables calculées

Premières lignes du tableau des valeurs récupérées par la méthode sur un EnvXVar avec une moyenne et deux seuils ayant permis d'établir un écart-type voire méthode page.55, les méthodes utilisées sont les mêmes qu'expliqué sous la Figure 15 pour calculer les variables à partir de la colonne 9. Le tableau a été également classé par score décroissant.

Le tableau présenté ensuite, Figure 16, est-celui issu de la méthode 2. La même méthodologie appliquée nous permet de vérifier la présence d'outliers. Cela illustre également les données disponibles sur des individus potentiellement particuliers et permet d'avoir une vision des valeurs utilisées dans la suite des analyses statistiques.

modalité	Année	lieu	Variété	rep	Pente_MAX	Staygreen	Drop_NDM	Pente_ajustée	Score_Pente_MAX	Score_Pente_ajustée	Score_Staygreen	Score
D1	2022	Thizay	ADVISOR	1	-0.0289352	24.8584237	-0.50653379	0.05712393	31.191421	0	0	0
Normal	2022	Thizay	ARKEOS	3	-0.0234396	24.6220914	-0.46742028	0.05014673	18.0257635	0	0	0
Normal	2022	Thizay	ARKEOS	2	-0.04142697	26.7140962	-0.44342489	0.093425	61.1175938	0	5.37288031	2.68644015
D1	2022	Thizay	CHEVIGNON	2	-0.03618345	26.7763375	-0.40384371	0.08959766	48.5558477	0	5.70586386	2.85293193
X80	2023	Thizay	ADVISOR	6	-0.01475903	28.356982	0.58077141	-0.02541281	0	3.17889924	14.1621186	8.67050893
X80	2023	Thizay	ADVISOR	1	-0.01993993	28.5369598	0.7661714	-0.02602542	9.64170866	3.96413224	15.1249778	9.54455501
X80	2023	Thizay	ADVISOR	4	-0.01844879	28.9891031	0.72596426	-0.02541281	6.06943417	3.17889924	17.5438892	10.3613942
X80	2023	Thizay	ADVISOR	5	-0.02213855	30.0777096	0.82	-0.02699823	14.9088842	5.21108298	23.3678012	14.2894421

Figure 17 : Tableau des valeurs récupérées par la méthode numéro 3 et nouvelles variables calculées

Premières lignes du tableau combinés. Les valeurs des 3 variables d'intérêt ont été concaténé dans un premier temps et les déciles de linéarisation des scores ont été calculé dans un second temps offrant des variables calculées de score légèrement différente que celle des réplcas correspondant dans les méthodes précédentes. Le tableau a été classé de la même manière que précédemment.

Enfin, ce tableau présente la méthode 3, ici l'absence de sélection de bornage sur les valeurs de drop\_NDVI a laissé apparaître des drops\_NDVI négatifs. Ce défaut pourra montrer une résilience de cette méthode aux outliers. Cette résilience pourra être corroboré par comparaison avec les 2 autres méthodes.

Une fois les données récupérées des régressions présentées, nous pouvons analyser les facteurs influençant le plus ces données (récupérées pour les 3 méthodes).

### 3 – 2 : Analyse univariée

A	Variable	modalité Variance	modalité p-value	lieu Variance	lieu p-value	Année Variance	Année p-value	Variété Variance	Variété p-value
	Pente_MAX	0.99	4.3E-01	0.77	5.2E-01	0.07	7.3E-01	17.94	3.0E-06
	Stay-green	4.26	2.4E-02	21.90	6.6E-10	0.56	3.2E-01	9.88	3.4E-03
	Drop_NDVI	14.89	1.0E-06	18.10	3.9E-08	11.68	3.9E-06	3.20	3.6E-01
	Pente_ajustée	4.41	2.1E-02	1.70	2.3E-01	1.65	9.1E-02	7.18	2.7E-02
	score	5.77	6.2E-03	4.92	1.3E-02	0.04	8.0E-01	22.67	2.8E-08
	score_Pente_ajustée	14.46	1.6E-06	7.29	1.5E-03	1.02	1.8E-01	23.26	1.5E-08
	score_Stay-green	3.80	3.6E-02	23.98	6.6E-11	0.47	3.7E-01	9.66	4.1E-03

B	Variable	modalité Variance	modalité p-value	lieu Variance	lieu p-value	Année Variance	Année p-value	Variété Variance	Variété p-value
	Pente_MAX	3.71	4.9E-02	3.78	4.67E-02	0.11	6.7E-01	23.82	4.0E-08
	Stay-green	5.63	1.0E-02	54.55	5.96E-28	0.06	7.6E-01	9.73	6.5E-03
	Drop_NDVI	7.53	2.0E-03	29.57	7.90E-13	5.51	2.6E-03	3.08	4.3E-01
	Pente_ajustée	7.88	1.5E-03	3.31	6.86E-02	1.28	1.5E-01	19.80	1.7E-06
	score	19.58	3.0E-08	21.89	2.94E-09	0.45	4.0E-01	13.96	2.5E-04
	score_Pente_ajustée	11.16	8.2E-05	7.38	2.24E-03	0.45	4.0E-01	24.68	1.7E-08
	score_Stay-green	5.45	1.2E-02	54.68	4.76E-28	0.04	8.1E-01	9.63	7.0E-03

C	Variable	modalité Variance	modalité p-value	lieu Variance	lieu p-value	Année Variance	Année p-value	Variété Variance	Variété p-value
	Pente_MAX	1.77	5.1E-02	1.64	6.33E-02	0.09	5.8E-01	19.25	6.8E-14
	Stay-green	1.38	1.0E-01	31.65	3.04E-28	0.14	5.0E-01	7.91	5.0E-05
	Drop_NDVI	6.93	6.5E-06	17.34	1.69E-14	7.91	1.6E-07	2.27	1.8E-01
	Pente_ajustée	4.14	8.8E-04	1.28	1.17E-01	0.40	2.5E-01	9.19	6.1E-06
	score	9.88	3.0E-08	11.94	6.39E-10	0.00	9.1E-01	18.59	2.4E-13
	score_Pente_ajustée	12.08	4.9E-10	7.54	2.14E-06	0.13	5.1E-01	25.00	5.0E-19
	score_Stay-green	1.57	7.2E-02	35.96	5.92E-33	0.07	6.2E-01	8.17	3.2E-05

Figure 18 : Tableaux des résultats des anovas à une variable dépendante pour les 3 méthodes et un facteur pour toutes les variables (récupérées et calculées) (variables dépendantes) et toutes les catégories (facteurs). Successivement % de Variance expliquée et p-value associée.

La colonne de gauche montre les variables dépendantes utilisées pour l'annova et les colonnes les facteurs avec successivement le % de variance expliquée et la p-value associée. En rouge sont représentés les p-value inférieure à 0,05 en vert clair, les p-value inférieure à 1 millièème. Les variances correspondant à ces p-value sont coloriés du rouge clair 1% au vert foncé 30 %. A méthode 1 B méthode 2 C méthode 3

La figure 18 montre un grand nombre de p-value largement significative (inférieur à 0,001) pour les modalités et les variables testées séparément. Parmi ces modalités, la variété semble expliquer le plus les variations dans ces annovas. Cependant, le découpage de l'environnement en modalité, année et lieu avec peu de degrés de libertés peut faire diminuer la visibilité des impacts environnementaux. On remarque que le score de pente ajustée semble être la variable qui sépare le mieux les variétés tandis ce que le Stay-Green sépare le mieux les lieux. Le score de pente ajustée et le score globale semblent être les plus variables en fonction de la modalité. Les annovas à un facteur n'expliquent souvent pas

des modèles complexes comme ceux étudiés ici et il est important de considérer les interactions entre facteurs. De plus, pour savoir si les facteurs expliquent le modèle, il faut tester la normalité des résidus.

C'est ce qui sera réalisé dans l'étape suivante et qui permettra ainsi, une analyse plus complète.

Variable	Var-Variance	Var-PValue	Env-Variance	Env-Pvalue	Normality-PValue
Pente_MAX	19.25	7.0E-15	7.63	8.4E-05	1.8E-23
Duree_Z55_A_Pente_Max	7.91	1.7E-12	56.07	3.8E-61	4.5E-18
Drop_NDVI	2.27	9.5E-03	50.56	1.2E-46	1.4E-25
Pente_ajustée	9.19	2.0E-06	8.79	5.9E-05	2.6E-29
score	18.59	2.2E-20	31.13	8.5E-30	7.9E-11
score_Pente_ajustée	25.00	2.0E-25	22.33	3.9E-21	1.9E-06
score_Duree_Z55_A_Pente_Max	8.17	8.6E-15	60.59	8.3E-71	2.3E-16

Figure 19 : Anova sur les niveaux groupés des facteurs d'année, de lieu et de modalité de culture, en Environnement (Env) par rapport à la Variété (Var)

Les effets de 7 variables dépendantes à gauche sur 2 variables catégorielle Variétés et Environnement avec dans l'ordre les colonnes indiquant successivement le % de variance expliquée et la p-value du test. La dernière colonne indique la p-value d'un test de Shapiro sur les résidus.

Il est important de vérifier l'impact d'un environnement unique par rapport à l'effet de la variété c'est-à-dire une même combinaison de lieuXannéeXmodalité-de-culture. Donc, nous avons réalisé (Figure 19), des annovas pour les 7 variables récupérées et générées à 2 facteurs :

- La combinaison des facteurs environnementaux, contenu dans une interaction modalité-année-lieu
- La variété

On remarque que les paramètres de pente sont les seuls à avoir une variance plus expliquée par la variété que l'environnement. L'analyse des résidus nous montre que cette combinaison des paramètres est suffisante pour expliquer notre modèle.

Les annovas réalisées nous renseignent donc plus en profondeur sur l'effet réel de la variété sur nos variables. Toutefois, Il est important de savoir, si les effets des variétés sont transmissibles ou stables entre les environnements et que le classement, croissant ou décroissant, sur une variable donnée de ces variétés, ne changera pas.

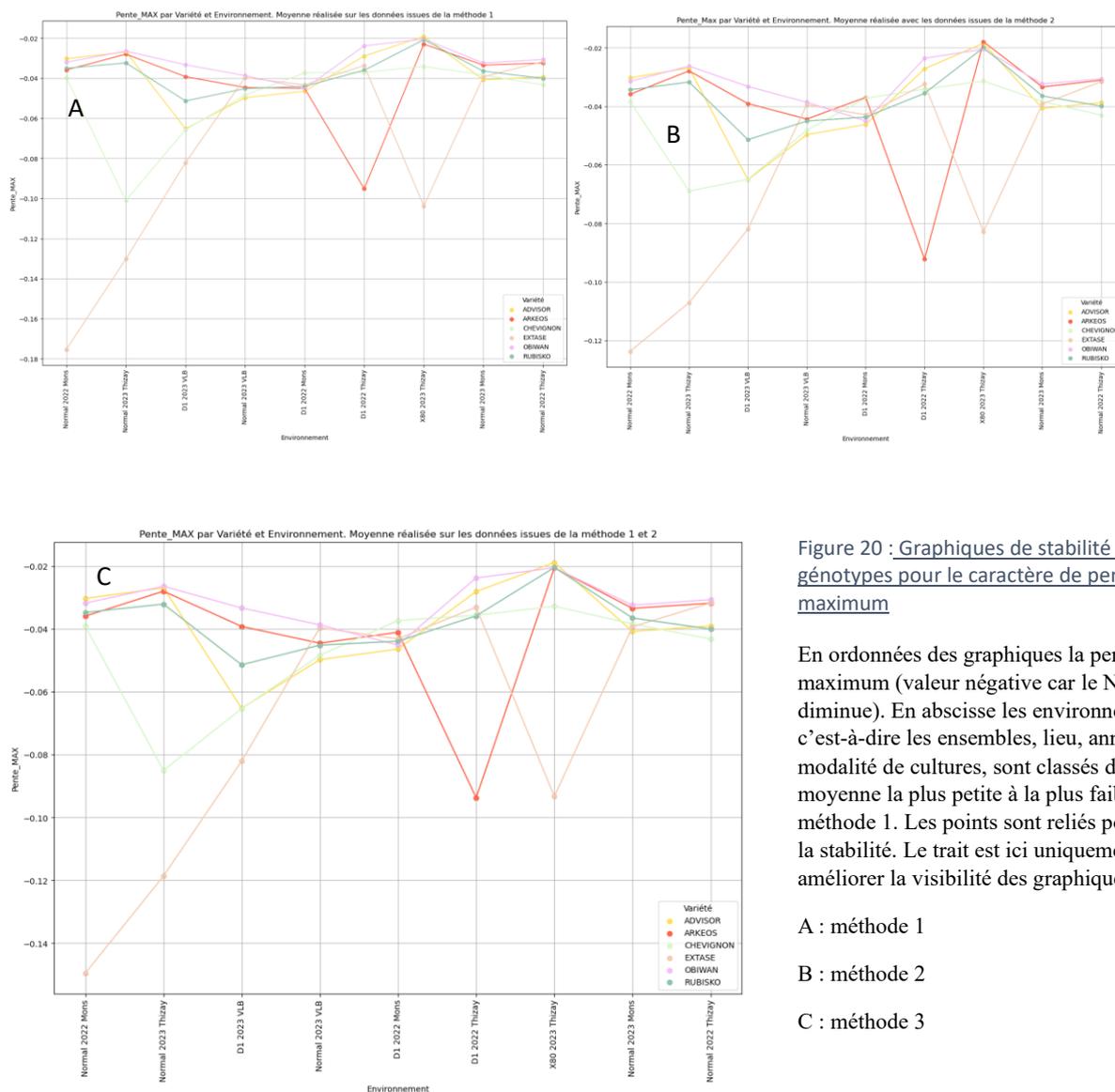


Figure 20 : Graphiques de stabilité des génotypes pour le caractère de pente maximum

En ordonnées des graphiques la pente maximum (valeur négative car le NDVI diminue). En abscisse les environnements c'est-à-dire les ensembles, lieu, année, modalité de cultures, sont classés de la valeur moyenne la plus petite à la plus faible avec la méthode 1. Les points sont reliés pour noter la stabilité. Le trait est ici uniquement pour améliorer la visibilité des graphiques.

- A : méthode 1
- B : méthode 2
- C : méthode 3

Nous observons donc, ici, *Figure 20*, la stabilité des génotypes entre les environnements, sur la variable de pente maximum, (ou la force de sénescence). Nous voyons une grande variabilité avec la présence de variétés particulières, acceptant une forte variabilité entre environnements. C'est le cas d'Extase. On note également des valeurs très importantes, pouvant faire penser à des valeurs aberrantes. Nous voyons que ces comportements extrêmes sont légèrement tempérés sur la méthode 2 de comparaison des données (attention les échelles des deux graphiques ne sont pas identiques).

Même si de nombreux croisements ont lieu, la variété Obiwan semble avoir en moyenne une valeur de pente maximale plus élevée que les autres, à l'exception de l'environnement 'D1-2022-Mons' où le classement semble pratiquement s'inverser. Les effets environnementaux sont forts et les 3 méthodes semblent indiquer des comportements globalement similaires. Nous remarquons que Chevignon et Rubisko ont globalement des valeurs de pentes maximales plus faibles. Notons que la pente est négative car il s'agit d'une décroissance, plus la pente est faible plus la force de sénescence est importante. Ainsi, ici, Obiwan à la force de sénescence la plus faible tandis que des forces extrêmes de sénescence seraient plutôt liées à la variété Extase.

La valeur de pente maximale indique le NDVI perdu en une journée sur les courbes modélisées. Le temps que les plantes mettent pour sécher n'est pas directement influencée par cette valeur, car cette valeur est dépendante du NDVI perçu en début et en fin de sigmoïde. C'est pourquoi nous utiliserons la différence entre l'asymptote haute et basse pour créer la pente ajustée.

Nous observerons dans un premier cette variable utilisée dans le calcul d'ajustement, le Drop\_NDVI et la deuxième variable d'intérêt, le Stay-Green<sup>1</sup>.

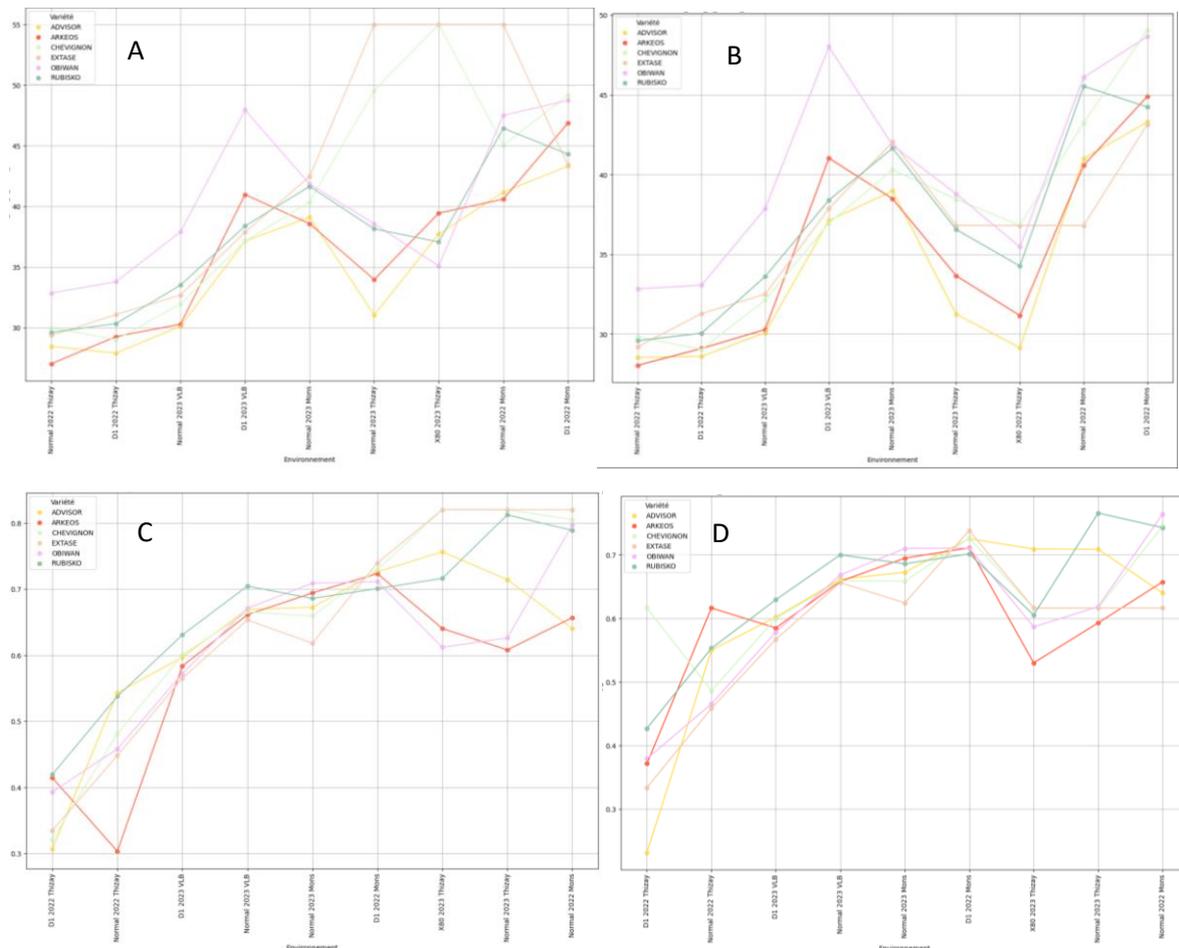


Figure 21 : Graphiques de stabilité des génotypes pour le Drop\_NDVI et le Stay-Green

En ordonnée les variables représentés (A-B) Drop\_NDVI, (C-D) Stay-Green. A et C représentent les données issues de la méthode 1, C et D les données issues de la méthode 2. Le jaune correspond à la variété Advisor, le rouge à Arkeos, le vert clair à Chevignon, le marron à extase, le violet à obiwan, le vert foncé à Rubisko

Les graphiques présents dans la figure 20 montrent la stabilité des génotypes pour le Drop\_NDVI et le Stay-Green entre les environnements, pour les deux premières méthodes. Pour la durée en jours de Z55 à la senescence, Obiwan semble clairement se démarquer pour les 5 premiers environnements tandis ce que pour le seul environnement avec une modalité azoté 'X80-2023-Thizay' la sénescence semble accélérée par rapport aux autres variétés. On remarque des valeurs extrêmes sur la méthode 1, pour extase et Chevignon, qui semblent avoir été totalement effacées par la méthode 2.

<sup>1</sup> la durée en jour du stade phénologique Z55, jusqu'au point de pente maximale de la modélisation

Le drop NDVI semble être fortement dépendant de l’environnement avec des valeurs très faibles en 2022 à Thizay et en 2023 à VLB, toutes modalités de cultures confondues. Si les annovas sur les 3 méthodes ont indiqué de p-value supérieur à 0,05 il semble que Rubisko est un Drop\_NDVI plus important que les autres variétés sur de nombreux environnement. Une analyse supplémentaire, au vu de la participation du Drop\_NDVI dans l’ajustement des pentes a été faite sur cette variable en annexe.

Ces graphiques de stabilité des génotypes, ne permettent pas de classer efficacement les génotypes entre les environnements. Les analyses qui vont suivre tentent de simuler les différences significatives entre chaque variétés inter-environnement. On cherche à comprendre l’organisation des génotypes entre eux pour une même variable entre les environnements.

### 3 – 3 : Classement inter-environnement sur les variables d’intérêt

Environnement	faible pente					forte pente
	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	Ensemble_6
D1 2022 Mons	ARKEOS CHEVIGNON	ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS EXTASE OBIWAN RUBISKO	ADVISOR EXTASE OBIWAN RUBISKO
D1 2022 Thizay	ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO					ARKEOS
D1 2023 VLB	ARKEOS OBIWAN	RUBISKO			ADVISOR CHEVIGNON	EXTASE
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE
Normal 2022 Thizay	ARKEOS EXTASE OBIWAN	ADVISOR ARKEOS EXTASE OBIWAN			ADVISOR ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON RUBISKO
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	EXTASE	ADVISOR
Normal 2023 VLB	EXTASE OBIWAN	ARKEOS CHEVIGNON RUBISKO			ADVISOR ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON RUBISKO
Normal 2023 Thizay	ADVISOR ARKEOS OBIWAN RUBISKO					CHEVIGNON EXTASE
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE

Figure 22 : Tableau de comparaison entre environnements des ensembles de variétés réalisant des différences significativement différentes classés du plus grand au plus petit pour la valeur de pente chevignon vs Arkeos:

Ce tableau représente les ensembles de variété admettant des différences significatives entre les environnements (lignes). Les colonnes correspondent au nombre maximum d’ensembles formant des différences significatives sur la variable de pente. Les ensembles sont classés de la valeur moyenne la plus grande à la plus petite. La méthode de récupération des données des régressions utilisées pour cette analyse est la méthode 3. Le rouge correspond à la variété Arkeos lorsqu’elle n’apparaît pas dans cet environnement dans un même ensemble que Chevignon. Chevignon apparaît en vert clair lorsque cette variété n’apparaît pas dans un même ensemble qu’Arkeos pour cet environnement. Les cases grises correspondent à l’apparition conjointe d’Arkeos et de Chevignon et marquent une absence de différenciation significative. Les cases noires, les ensembles n’existant pas. Ces cases noires apparaissent entre les ensembles selon la méthode de répartition des ensembles décrit dans le matériel et méthodes. La répartition des cases noires se fait pour les environnements ayant obtenu moins d’ensembles que le maximum d’ensembles générés sur tous les environnements. Ces ensembles sont issus de post-hoc unique par environnement.

Nous avons donc réalisé un tableau de comparaison des génotypes qui classe les ensembles issus d'un test post-hoc montrant les groupes de variétés significativement différent d'au moins une autre variété. L'ensemble de ces groupes est ensuite réparti par moyenne de la variable étudiée. Ils sont donc rangés de la colonne la plus à gauche à la colonne la plus à droite par ordre décroissant.

La gestion des différences de nombre d'ensemble générés entre environnement est expliquée ici : Simplification méthodologique 1, page. 19 et Simplification méthodologique 2, page. 19. Les lignes représentent les environnements. Comme il ne s'agit pas de différences significatives strictes entre chaque groupe, mais d'au moins une différence significative d'une variété avec les autres, il est difficile de parler de classement à proprement parler. La puissance statistique sur un même environnement pour classer 6 variétés n'était, en effet, pas suffisante. Toutefois, il est possible d'établir des tendances visuelles sur ces tableaux qui seront ensuite retranscrit graphiquement pour essayer de supposer le rang d'un génotype sur une variable particulière et si une tendance, au moins se dessine.

Pour pallier ce manque de puissance statistique et la complexité d'interprétation de ces tableaux, on réalise des oppositions entre deux variétés. Lorsque que les génotypes n'apparaissent pas dans une même cellule du tableau pour un environnement donné, les deux couleurs de ces deux variétés ont été mis en avant pour signifier la différence significative et la position dans le classement des ensembles. D'autres oppositions entre variétés sur différentes variables ont été réalisé en annexes ici : D'abord, la Simplification méthodologique 7 a réindexé de manière artificielle les simulations de pente maximum au-dessus de la moyenne. Cela était dû à :

- 1- La Simplification méthodologique 3 qui a généré une mauvaise translation des courbes pour minimiser les effets compensatoires de la variabilité de la pente entre les réplicas.
- 2- La Simplification méthodologique 5 de lissage des bornes qui a fait passer la borne supérieure en-dessous de la moyenne créée

Les perspectives pour résoudre ces problèmes seraient :

- 1- D'utiliser un algorithme plus aboutit pour la création des centres de gravité en prenant en compte le point d'annulation de la dérivée première pour éviter un biais lié au choix de la plage d'abscisse de création des images des courbes.
- 2- De mieux comprendre les comportement univariés sur l'établissement d'un intervalle de confiance entre la moyenne et le réplica : Simplification méthodologique 4
- 3- D'améliorer le lissage des courbes (Simplification méthodologique 5) en prenant notamment en compte (dans l'algorithme) la moyenne en tous points de la génération des bornes pour contrôler l'écart avec la moyenne.

Par ailleurs, deux éléments pouvaient être réalisé pour améliorer la création des simulations une fois les bornes translattées générées en choisissant plus intelligemment les points utilisés pour les régressions destinés à simuler les pentes seuils max et min :

- 4- Améliorer par le choix des coupures destinées à faire passer les simulations de pente minimale et maximale entre les bornes : Voir Simplification méthodologique 6

- 5- En créant aléatoirement des points au-dessus et au-dessous des bornes sur des faibles plages d'intervalles. Puis :
  - a. Réaliser plusieurs régressions (pour chaque génération aléatoire de points) pour les deux simulations de pente maximale et minimale
  - b. Prendre la moyenne de la valeur maximale des dérivées (de chaque équation de simulation créée) pour établir le seuil de pente maximale et minimale

D'autres améliorations et d'autres méthodologies pourraient être envisagées.

7 – 5 : Classement des géotypes par cluster, page. 70.

C'est dans ce cadre-ci, que pour la variable de pente, une des principales caractéristiques d'intérêt pour la sélection variétale, nous avons réalisé une opposition entre la variété Chevignon et Arkeos, Figure 22. Des différences significatives ont été observé dans 4 environnements. Deux à Thizay : un en 2022, en semi précoce et un en 2023 en semis 'normal', un à Mons en 2023 et un à VLB en semis précoce.

Sur 4 environnements, 3 classent Chevignon avec une pente plus forte qu'Arkeos à l'exception de l'environnement D1 2023 VLB. On remarque sur la Figure 20 que c'est un environnement où Arkeos semble admettre une moyenne de valeur de pente max, aberrante et très faible. On peut le voir dans le tableau où Arkeos est isolé, tandis que les autres variétés sont toutes groupées dans le même ensemble (y compris dans les 2 autres méthodes voir Figure 47).

Après avoir analysé la pente, nous analyserons la pente ajustée, pour voir l'incidence de l'intervention du Drop\_NDVI dans la nouvelle variable calculée.

Environnement	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4	Ensemble 5
D1 2022 Mons			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2023 VLB	ARKEOS OBIWAN ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ARKEOS OBIWAN RUBISKO	ARKEOS RUBISKO	ADVISOR CHEVIGNON	EXTASE
Normal 2022 Mons					EXTASE
Normal 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
Normal 2023 Mons	ARKEOS OBIWAN	RUBISKO		ADVISOR CHEVIGNON	EXTASE
Normal 2023 VLB	EXTASE OBIWAN RUBISKO	ARKEOS EXTASE OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS CHEVIGNON
Normal 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO				EXTASE

Environnement	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4	Ensemble 5	Ensemble 6
D1 2022 Mons	ARKEOS CHEVIGNON	EXTASE				ADVISOR OBIWAN
D1 2022 Thizay	CHEVIGNON OBIWAN RUBISKO	CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR EXTASE RUBISKO		ADVISOR EXTASE ADVISOR CHEVIGNON RUBISKO	ARKEOS
D1 2023 VLB	ARKEOS OBIWAN RUBISKO ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO				EXTASE
Normal 2022 Mons						EXTASE
Normal 2022 Thizay	ARKEOS OBIWAN	ADVISOR ARKEOS EXTASE OBIWAN RUBISKO			ADVISOR EXTASE OBIWAN RUBISKO	CHEVIGNON
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	ADVISOR	EXTASE
Normal 2023 VLB	ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO				ADVISOR ARKEOS CHEVIGNON EXTASE
Normal 2023 Thizay	ADVISOR ARKEOS OBIWAN RUBISKO	CHEVIGNON				EXTASE
X80 2023 Thizay	ADVISOR	ARKEOS OBIWAN RUBISKO			CHEVIGNON	EXTASE
Environnement	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4	Ensemble 5	Ensemble 6
D1 2022 Mons	ARKEOS CHEVIGNON	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO				ADVISOR ARKEOS EXTASE OBIWAN RUBISKO
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
D1 2023 VLB	ARKEOS OBIWAN	RUBISKO			ADVISOR CHEVIGNON	EXTASE
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE
Normal 2022 Thizay	ARKEOS					ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	ADVISOR	EXTASE
Normal 2023 VLB	EXTASE OBIWAN	EXTASE OBIWAN RUBISKO	ARKEOS EXTASE RUBISKO	ARKEOS CHEVIGNON RUBISKO	ADVISOR ARKEOS CHEVIGNON	ADVISOR CHEVIGNON
Normal 2023 Thizay	ADVISOR ARKEOS OBIWAN RUBISKO	CHEVIGNON				EXTASE
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE

Figure 23 : Tableaux de comparaison sur la pente ajustée des ensembles de variétés ayant admis des différences significatives lors d'un post-hoc test.

Ces tableaux mettent en lumière les différences significatives entre Chevignon et Arkeos comme dans la figure 20. A : réalisé sur les données issues de la méthode de récupération des données 1, B : 2, C : 3

Nous continuons donc l'analyse avec la pente ajustée et nous voyons *Figure 23* que 4 environnements demeurent différenciant entre les variétés mais qu'un environnement a changé dans les 4. En effet, si nous nous concentrons sur la méthode 3 *Figure 22.C*, 'D1 2023 VLB' ne semble plus être discriminant tandis que Normal 2022 Thizay est à présent différenciant et indique une même orientation, à savoir, que Chevignon a une pente plus négative qu'Arkeos dans 4 environnements. Il semblerait donc que dans le classement des géotypes nous ayons eu une amélioration par rapport à la pente brute, grâce à l'ajustement par le Drop\_NDVI.

Par ailleurs, nous remarquons que la méthode 3 apporte une clarification du classement par rapport aux deux méthodes précédentes. En effet, la méthode 1 classe sur seulement 2 environnements, ou classe, méthode 2, sur 5 environnements mais dont 1 environnement inverse le sens de classement.

En suivant ce match entre Chevignon et Arkeos et le contrôle des modifications apportées aux variables, voyons à présent ce que la création d'un score de pente ajustée génère sur le classement.

Environnement	Ensemble 1	Ensemble 2	Ensemble 3	Ensemble 4	Ensemble 5	Ensemble 6
D1 2022 Mons	ADVISOR ARKEOS EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO				ARKEOS CHEVIGNON
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
D1 2023 VLB	EXTASE	ADVISOR CHEVIGNON	RUBISKO		ARKEOS	OBIWAN
Normal 2022 Mons	EXTASE	ARKEOS CHEVIGNON	ADVISOR ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON RUBISKO	ADVISOR CHEVIGNON OBIWAN RUBISKO	OBIWAN RUBISKO
Normal 2022 Thizay	CHEVIGNON	ADVISOR EXTASE OBIWAN RUBISKO				ARKEOS
Normal 2023 Mons	EXTASE	ADVISOR	CHEVIGNON	RUBISKO	ARKEOS	OBIWAN
Normal 2023 VLB	ADVISOR CHEVIGNON	ADVISOR ARKEOS CHEVIGNON	ARKEOS CHEVIGNON RUBISKO	ARKEOS EXTASE RUBISKO	EXTASE OBIWAN RUBISKO	EXTASE OBIWAN
Normal 2023 Thizay	CHEVIGNON EXTASE					ADVISOR ARKEOS OBIWAN RUBISKO
X80 2023 Thizay	EXTASE	ARKEOS CHEVIGNON OBIWAN RUBISKO			ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS OBIWAN RUBISKO

Figure 24 : Tableau de comparaison entre environnements des ensembles admettant des différences significatives pour la valeur de score pente ajustée sur les données issues de la méthode de récupération des informations de la régression numéro 3

Ce tableau met en lumière les différences entre Chevignon et Arkeos pour le score de pente ajustée. La structure est rappelée dans la figure 20.

Le score est sensé décrire une performance agricole, nous avons donc pris l'inverse de la pente ajustée (négative) pour que des valeurs hautes représentent une bonne force de pente et donc de sénescence. Dans la Figure 24, nous pouvons voir que l'environnement D1 2022 Thizay n'est plus différenciant (par rapport à la figure précédente), tandis ce qu'Arkeos classé soit du côté opposé, soit confondu est à présent différencié du même côté que les 3 autres environnements. Notons que D1 2022 Thizay dans les 4 environnements est remplacé par Normal 2023 Mons (ou les 6 variétés sont séparées significativement pour le score de pente ajustée).

La force de la pente n'est pas la seule donnée d'intérêt pour la fin de vie d'une plante. La durée de vie avant le moment où se produit cette pente dans le temps semble également importante pour conserver un bon rendement. Voilà pourquoi nous établirons un score global qui illustre le comportement d'une micro parcelle dans sa capacité à rester verte longtemps tout en desséchant rapidement (ou Stay-Green). Ce score s'appellera le score de performance.

Pour établir ce score, nous faisons la moyenne du score de pente ajustée et du score de Stay-Green. Nous allons donc voir, comment nos deux variétés illustrées ici, Chevignon et Arkeos se sont comportés pour ce score de Stay-Green.

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5
D1 2022 Mons			CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2022 Thizay	EXTASE OBIWAN RUBISKO	CHEVIGNON EXTASE OBIWAN RUBISKO		CHEVIGNON EXTASE RUBISKO	
D1 2023 VLB	OBIWAN	ARKEOS		EXTASE RUBISKO	ADVISOR CHEVIGNON
Normal 2022 Mons	EXTASE	CHEVIGNON OBIWAN RUBISKO		CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON
Normal 2022 Thizay	OBIWAN				CHEVIGNON EXTASE RUBISKO
Normal 2023 Mons	EXTASE OBIWAN			CHEVIGNON	ADVISOR ARKEOS
Normal 2023 VLB	OBIWAN	EXTASE RUBISKO	CHEVIGNON EXTASE RUBISKO	CHEVIGNON EXTASE	ADVISOR ARKEOS
Normal 2023 Thizay	CHEVIGNON EXTASE	CHEVIGNON EXTASE OBIWAN		ADVISOR ARKEOS OBIWAN RUBISKO	CHEVIGNON OBIWAN RUBISKO
X80 2023 Thizay			CHEVIGNON EXTASE OBIWAN RUBISKO		

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4
D1 2022 Mons	ARKEOS CHEVIGNON OBIWAN		CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS EXTASE RUBISKO
D1 2022 Thizay	EXTASE OBIWAN RUBISKO	CHEVIGNON EXTASE OBIWAN RUBISKO		CHEVIGNON EXTASE RUBISKO
D1 2023 VLB	OBIWAN	ARKEOS	EXTASE RUBISKO	ADVISOR CHEVIGNON
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE
Normal 2022 Thizay	OBIWAN			CHEVIGNON EXTASE RUBISKO
Normal 2023 Mons	RUBISKO	CHEVIGNON		ADVISOR ARKEOS
Normal 2023 VLB	OBIWAN	RUBISKO	CHEVIGNON EXTASE	ADVISOR ARKEOS
Normal 2023 Thizay		CHEVIGNON EXTASE OBIWAN RUBISKO		
X80 2023 Thizay		ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5
D1 2022 Mons	CHEVIGNON OBIWAN				ADVISOR ARKEOS EXTASE RUBISKO
D1 2022 Thizay	OBIWAN	ARKEOS EXTASE RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS CHEVIGNON RUBISKO	
D1 2023 VLB	OBIWAN	ARKEOS	RUBISKO	EXTASE	ADVISOR CHEVIGNON
Normal 2022 Mons			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN		
Normal 2022 Thizay	OBIWAN	ADVISOR CHEVIGNON EXTASE RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE	ADVISOR ARKEOS
Normal 2023 Mons	EXTASE OBIWAN	EXTASE OBIWAN RUBISKO	OBIWAN RUBISKO	CHEVIGNON	ADVISOR ARKEOS
Normal 2023 VLB	OBIWAN	RUBISKO		CHEVIGNON EXTASE	ADVISOR ARKEOS
Normal 2023 Thizay	CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS OBIWAN RUBISKO
X80 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN		

Figure 25 : Tableau de comparaison des génotypes entre environnements pour le score de Stay-Green

Ces tableaux, dont les modalités de construction et les méthodologies utilisées ont été décrites dans les figures précédentes du même style, illustrent les différenciations possibles des génotypes entre les scores (de durée en jours du stade phénologique Z55 au temps d'obtention de la pente maximale de la sigmoïde) ou Stay-Green. Le score est établi selon un principe de linéarisation entre les déciles, pour une comparaison avec d'autres variables, et de suppression des valeurs extrêmes.

La figure 24, illustre la stabilité du classement des génotypes entre les environnements pour la valeur de score de Stay-Green. On remarque que la méthode 1 et 2 permettent de différencier par la variable du Stay-Green significativement seulement pour 3 environnements, là où la méthode combinatoire permet de différencier sur 4 environnements. Le Stay-Green semble meilleur pour la variété Chevignon que pour la variété Arkeos. On a donc la même tendance que sur le score de force de sénescence.

Pour comprendre le comportement de ces deux scores l'un avec les autres, représentons la dispersion des valeurs.

### 3 – 4 : Analyse bivariée des comportements des variables d'intérêts ajustées

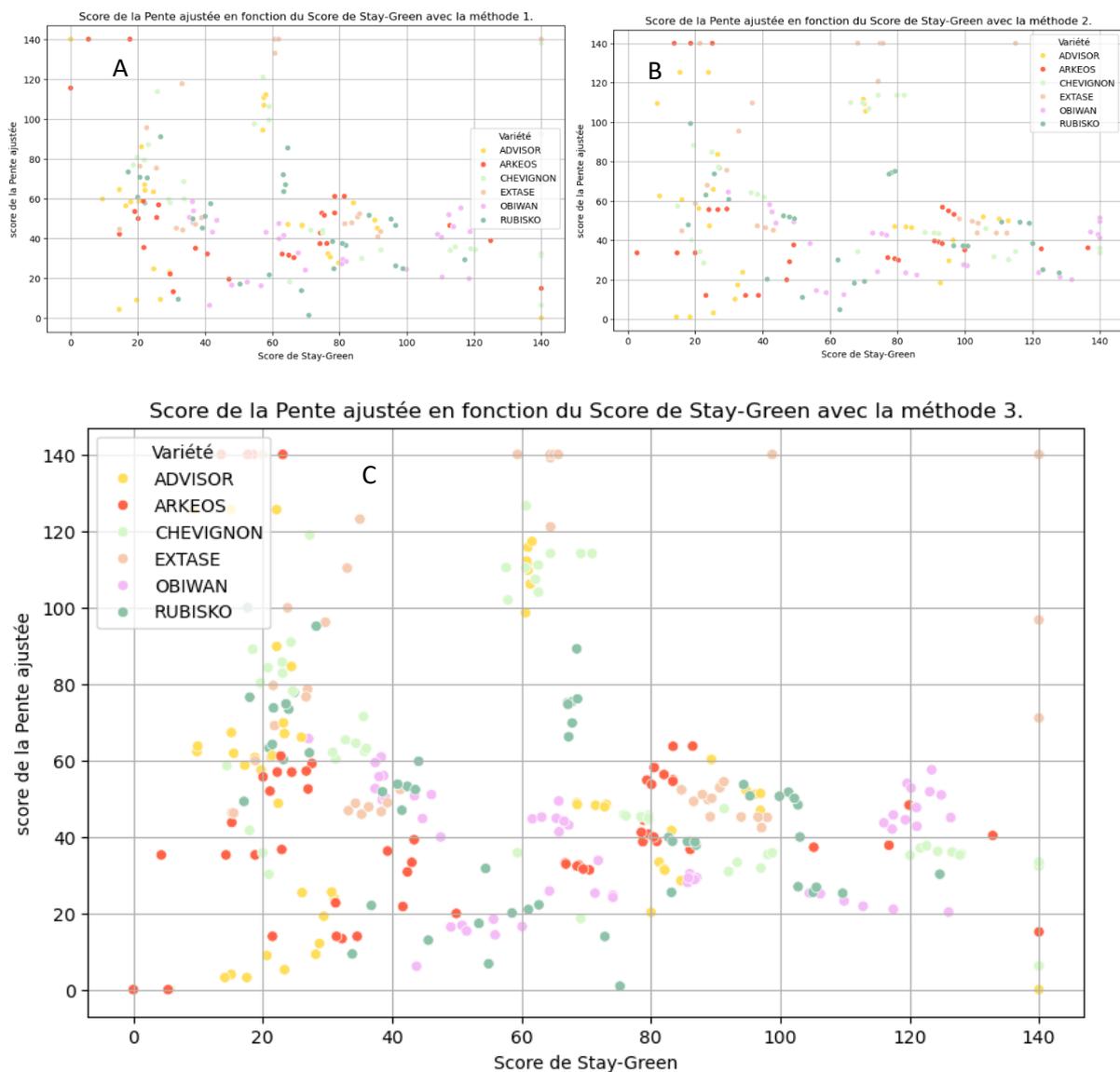


Figure 26 : Représentation graphique des variétés selon le score de pente ajustée et de Stay-Green

Graphiques représentant chaque individu en fonction de son score de Stay-Green en abscisse et de pente ajustée en ordonnée. Les couleurs représentent les variétés, indication des couleurs dans la légende. Le graphique A : représente la méthode 1, B : 2, C : 3.

La dispersion des valeurs du score de Pente en fonction du Stay-Green montre un comportement intéressant. En effet, les valeurs d'intérêts pour un bon génotype en agriculture : très forte pente et très longue durée de Stay-Green, semblent absentes du graph. Nous avons en effet, mis-à-part quelques extremum, un carré vide pour les valeurs fortes des deux paramètres au-dessus d'un score de 80 pour le Stay-Green et de 60 pour le score de pente ajustée.

On remarque une dispersion dans le reste du tableau qui semble former des groupes. Cela peut-être une indication soit du comportement d'un génotype particulier, soit de comportements très différents entre les environnements. Pour mieux comprendre dans un premier temps la dispersion des données en elle-même nous réalisons sur ces deux variables, un clustering de proximité géographique sans à priori sur le nombre de groupe, Figure 27.

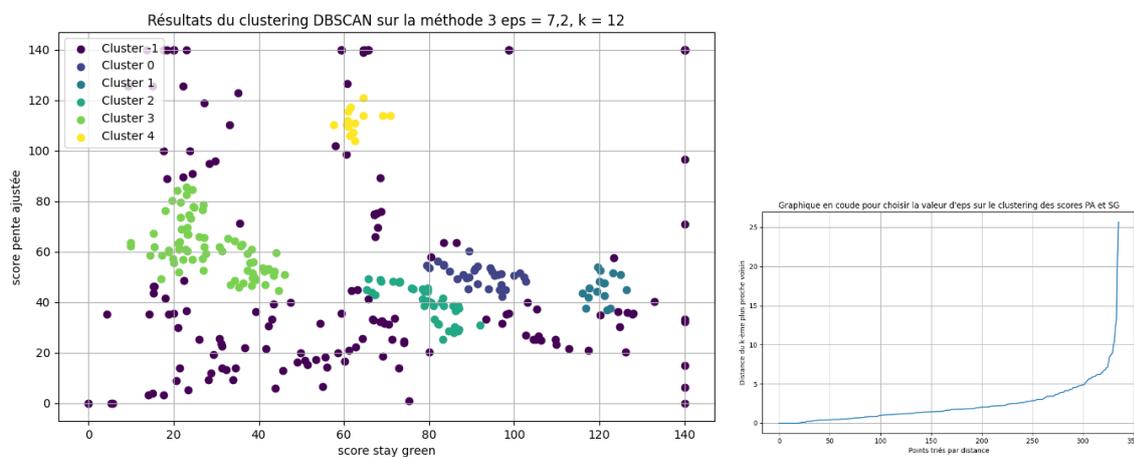


Figure 27 : Représentation des clusters, sans apriori sur leur nombre, découverts par machine learning (algorithme dbscan) sur l'ensemble des variétés

Deux graphiques concernant la clusterisation DBSCAN A : Représentation de chaque individus (données issues des régressions par la méthode 3) selon deux paramètres le score de Stay-Green et en abscisse et le score de pente ajustée en ordonnée (force de la pente de senescence ajustée par le Drop\_NDVI) En violet les individus qui n'ont pas pu être intégré à un cluster dbscan, et dans les autres couleurs, les 5 cluster de 0 à 5, découvert par DBSCAN avec k, nombre minimum d'individus souhaité par cluster 12, et 7,2 la distance moyenne de séparation des cluster. B Graphique en coude représentant en abscisse les points triés par distance moyenne des 12 voisins les plus proches, en ordonnée, la distance moyenne du point considéré avec ses 12 voisins les plus proches. Permettant de décider de la valeur de eps.

Sur cette figure nous observons la formation de 5 clusters. Le choix de 12 individus par groupe minimum s'explique par la volonté de ne pas générer de cluster ne représentant qu'une seule variété dans une seule modalité EnvXVar. Le choix de la valeur d'eps se justifie elle par le graphique en coude ou à 7,2 de distance (ordonnée), une montée plus importante se remarque Figure 27. Ces 5 clusters ont des tailles différentes et se répartissent sur une grande partie du graphique.

Il est intéressant de comprendre ce qui a pu générer ces concentrations de points avec un algorithme de machine learning de classification, le Random Forest afin de découvrir le facteur qui explique la concentration.

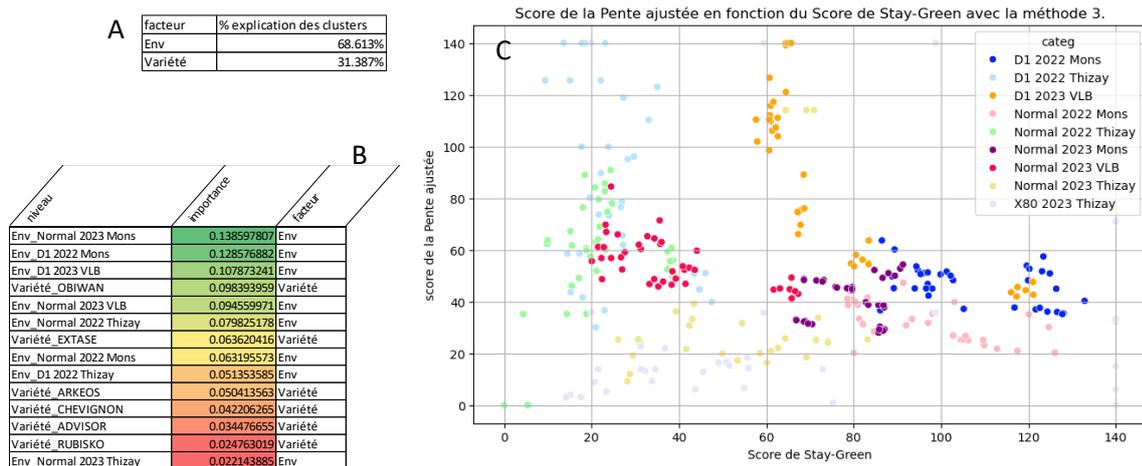


Figure 28 : Tableaux et nuage de points sur le facteur principale trouvé par Random Forest le plus à même de classifier les clusters précédemment créés

Ensembles de résultats après avoir performé une clusterisation par Random Forest. A : tableau récapitulatif des % d'explication par facteurs ; B : tableau récapitulatif par niveau de facteurs ; C représentation du facteur principal trouvé (avec niveau du facteur en différentes couleur). La représentation des individus est faite dans l'espace de répartition des points sur le score de Stay-Green et le score de force de pente ajustée. Cet espace est celui précédemment utilisé pour réaliser les clusters spatiaux.

L'observation des résultats de Machine learning pour comprendre la classification en cluster montre que l'environnement génère cette clusterisation majoritairement (Figure 28. A). La figure 28. B montre que ce sont les niveaux de facteurs 'Normal 2023 Mons', 'D1 2022 Mons', 'D1 2023 VLB', 'Normal 2023 VLB' qui explique le plus les clusters + de 9 % d'explication. La variété Obiwan est la seule variété avec plus de 9 % d'explications des clusters.

Sur la figure 28. C on voit que D1 2022 Mons semble particulièrement lié au cluster 1 et 0, tandis que l'environnement Normal 2023 Mons semble plus lié au cluster 2. D1 2023 VLB semble bien expliqué le cluster excentré numéro 4, tandis que le grand cluster 3 avec les mauvais scores de Stay-Green et relativement mauvais score de pente s'explique principalement par Normal 2023 VLB.

Cette étude étant à présent réalisé de l'organisation entre-elles de ces deux variables. Il semble intéressant de s'intéresser au score moyen de ces variables. Un graphique de stabilité Figure 29 permettra de comprendre la tendance générale de cette variable pour les facteurs de variétés et d'environnement.

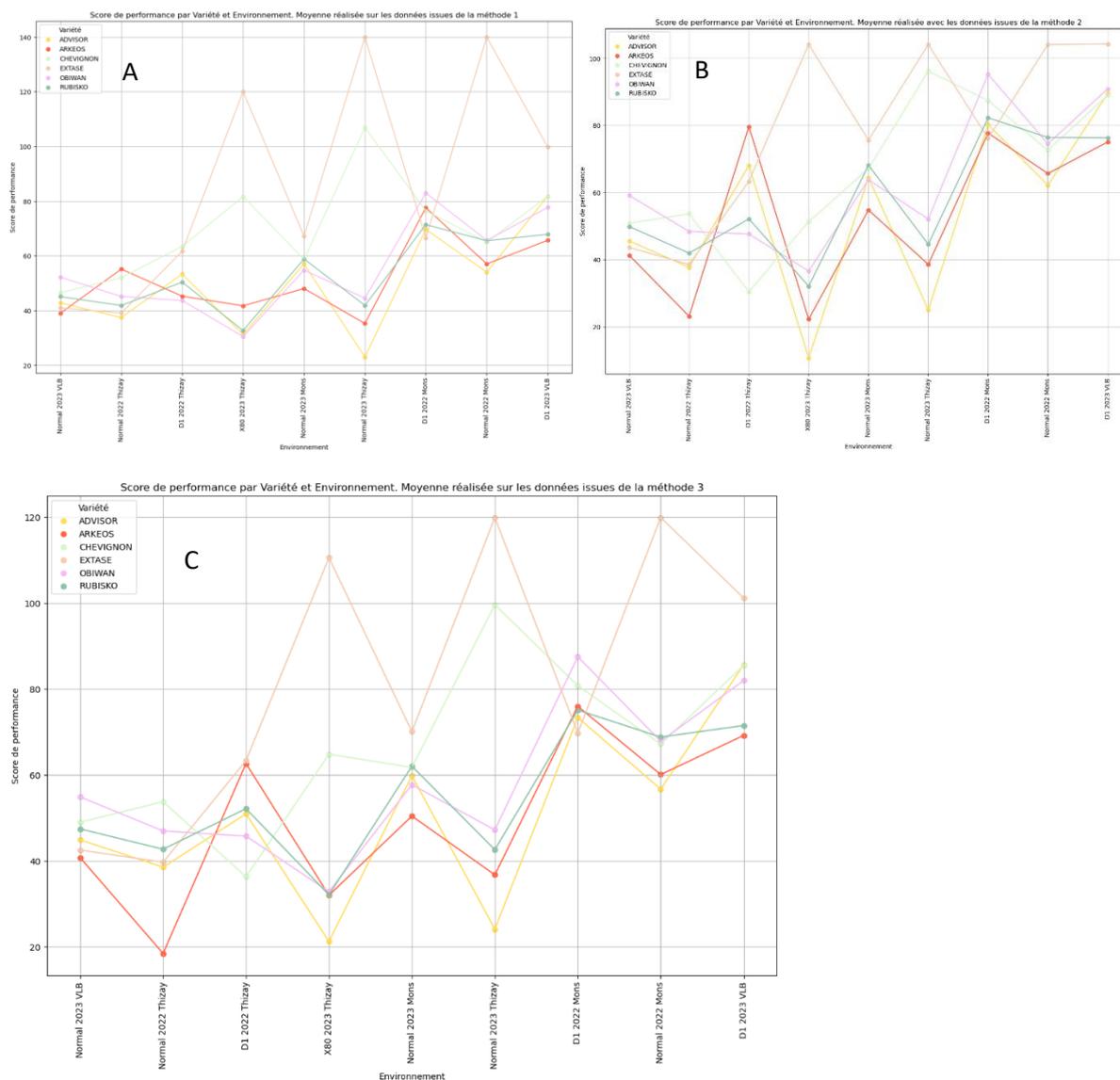


Figure 29 : Graphiques de stabilité des génotypes sur la valeur du score de performance combinant le score de Stay-Green et de force de pente

Ces graphiques de stabilités des génotypes montrent pour chaque environnement, en abscisse, la stabilité du classement des variétés (les couleurs) entre elles pour la valeur du score de performance. Les lignes sont ici pour assurer l'absence de croisements ou leurs faibles nombres ils ne sont pas normalement présent sur un graphique avec des variables discrètes catégories non ordinales (l'orientation des environnements en abscisses est basé sur la moyenne de la valeur pour la méthode 1)

La figure 29 montre que le comportement du score de performance semble légèrement moins stabilisé entre les méthodes que pour les variables précédemment étudiées. Toutefois les tendances générales de l'organisation des variétés entre-elles semblent sensiblement identiques. Extase est plus haut sur la majorité des environnements avec une grande variabilité entre environnements de son score de performance global. La variété Chevignon semble également être au-dessus de la plupart des variétés sans admettre des valeurs extrêmes. On remarque un changement dans le classement de cette variété pour l'environnement D1 2022 Thizay méthode 1 et 2. Les autres variétés semblent difficilement se différencier les unes des autres.

### 3 – 5 : Classement des variétés sur un score de performance global

La Figure 49, page. 79 classe les génotypes avec les tableaux de comparaison comme précédemment réalisé mais s'intéresse cette fois-ci à l'occurrence unique des variétés, plutôt qu'à l'opposition de deux variétés entre elles. En effet, la puissance statistique ne permettait pas une telle opposition. La variété Arkeos, Chevignon et Extase ont été ici étudié. La compilation des occurrences brutes a été fait *Figure 30* et permet de voir pour les 3 méthodes la tendance générale dans le classement des variétés sur la variable du score de performance.

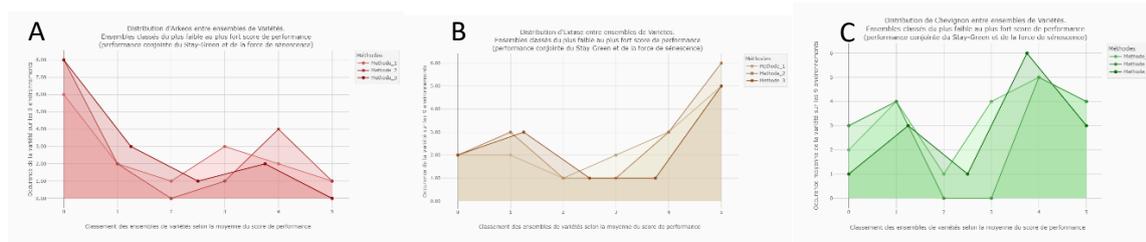


Figure 30 : Graphiques de densité d'apparitions dans les tableaux de comparaison des variétés (comptes bruts)

Ces graphiques représentent la densité d'apparition brute. En abscisse, la position dans les tableaux de comparaison entre les colonnes, et en ordonnée le nombre d'occurrence de la variété pour les 9 environnements. Les plots des 3 méthodes ont été réalisé avec et 3 variétés graphique (A, B, C) des mentions de la variété entre les environnements dans les ensembles classés dans les tableaux de comparaison précédents.

Les figures 30 et 31 présentent une analyse comparative des performances de trois variétés (A, B et C) sur 9 environnements différents.

La figure 30 illustre la fréquence d'apparition de chaque variété dans les différentes positions du classement. L'axe horizontal représente le nombre total d'occurrences. Cette visualisation révèle des tendances distinctes :

La variété Arkeos se retrouve fréquemment dans les dernières positions du classement.

Extase, en revanche, apparaît plus souvent dans la première moitié du classement.

Chevignon montre une légère tendance à se positionner dans la partie supérieure du classement, bien que cette tendance soit moins prononcée que pour Extase.

On observe des espaces vides au centre des graphiques, résultant d'un nombre réduit de clusters dans les positions médianes des tableaux de comparaison.

Pour corriger ce biais visuel, la *Figure 31* présente une version standardisée des mêmes données. Cette standardisation est basée sur le nombre moyen d'occurrences des variétés entre les différentes positions du classement. Cette approche permet une représentation plus équilibrée de la distribution des variétés à travers l'ensemble du classement.

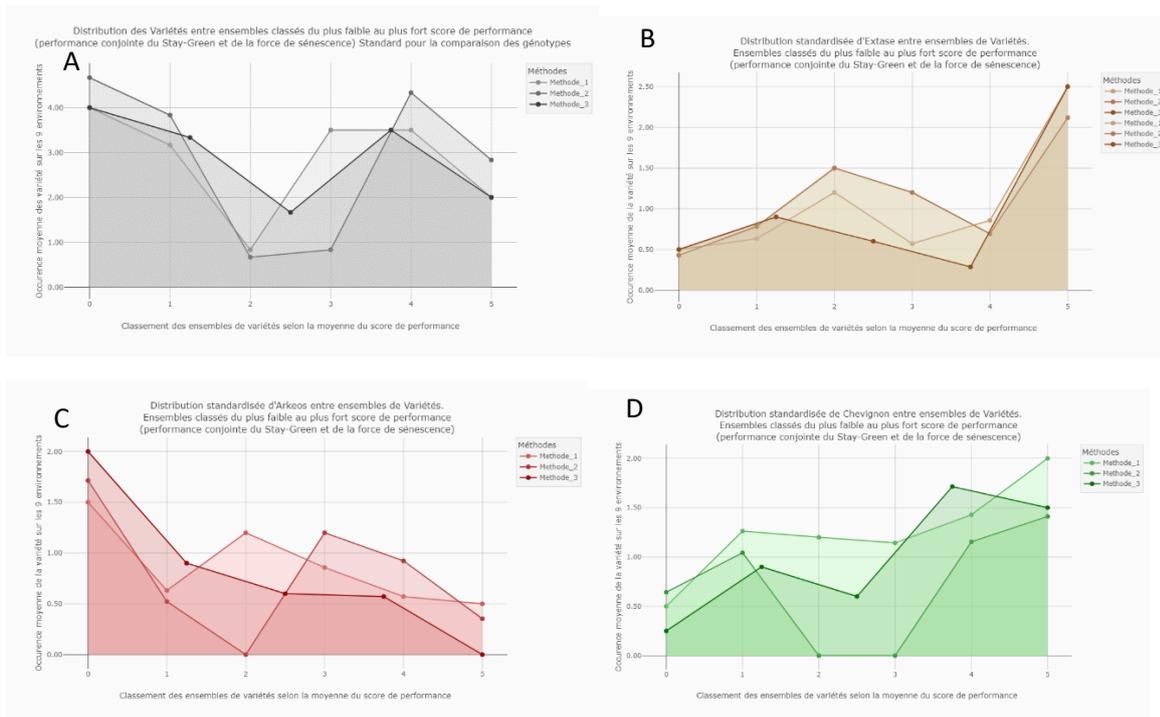


Figure 31 : Graphiques de densité d'apparitions dans les tableaux de comparaison des variétés et courbes de standardisation (résultats standardisés)

Courbe de densité représentant 3 variétés et un 1 standard. Les ordonnées représentent la densité d'apparition standardisée par le graph A. L'abscisse représente la position dans le classement des tableaux de comparaison sur la variable du score de performance. Un score élevé est à droite tandis qu'un score faible est à gauche. Du plus clair au plus foncé la méthode 1, 2 et 3.

La figure 31 présente des graphiques d'occurrences standardisées, où l'occurrence totale de chaque variété est normalisée par le nombre total d'occurrences de toutes les variétés. La figure 31A illustre la moyenne d'apparition des variétés dans les différentes positions du classement. Cette standardisation a atténué, sans toutefois l'éliminer complètement, le phénomène de zones vides entre les extrémités du classement. L'utilisation de la méthode 3 semble particulièrement efficace pour lisser les courbes et mettre en évidence des tendances générales :

Pour Arkeos, la tendance devient plus claire et plus facilement interprétable.

Extase montre également une amplification de sa tendance générale avec cette méthode.

Chevignon présente une tendance, bien que moins prononcée que pour les deux autres variétés.

Face à ces résultats, nous avons opté pour la la Simplification méthodologique 2, jugée plus parcimonieuse, plutôt que la Simplification méthodologique 1 (page. 19).

La Figure 32 illustre spécifiquement l'application de la méthode 3 à l'organisation du tableau de comparaison et à l'occurrence de Chevignon. Cette représentation offre une visualisation claire de la performance de Chevignon à travers les différents environnements étudiés.

Cette approche analytique permet une compréhension plus nuancée et précise des performances relatives des trois variétés de blé dans les divers environnements testés.

Categorie	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5
D1 2022 Mons	CHEVIGNON OBIWAN	ARKEOS CHEVIGNON OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS EXTASE RUBISKO
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2023 VLB	EXTASE		ADVISOR CHEVIGNON OBIWAN		ARKEOS RUBISKO
Normal 2022 Mons		EXTASE		ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	
Normal 2022 Thizay	CHEVIGNON OBIWAN RUBISKO	ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO		ADVISOR EXTASE OBIWAN RUBISKO	ARKEOS
Normal 2023 Mons	EXTASE	CHEVIGNON RUBISKO	ADVISOR	OBIWAN	ARKEOS
Normal 2023 VLB	OBIWAN	ADVISOR CHEVIGNON RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS EXTASE
Normal 2023 Thizay		CHEVIGNON EXTASE		ADVISOR ARKEOS OBIWAN RUBISKO	
X80 2023 Thizay	EXTASE	ARKEOS CHEVIGNON OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS OBIWAN RUBISKO

Figure 32 : Tableau de comparaison des ensembles de variétés significativement différents sur le score de performance

Tableau de comparaison du score de performance arrangée selon la (Simplification méthodologique 2, page 19) jugée plus homogène. Les ensembles sont classés du plus grand au plus petit et en ligne sont représentés les environnements. Les cases noires représentent les ensembles absents car moins d'ensembles significativement différents que les autres découverts sur cet environnement lors du test post-hoc. L'assemblage de ces groupes de variétés dans le classement discret plus grand du tableau complet (les 5 colonnes) a été fait en suivant la méthode de répartition des ensembles de la Simplification méthodologique 2. En vert, les ensembles où sont apparus la variété Chevignon.

Sur ce tableau, *Figure 32*, de comparaison nous voyons que sur 5 positions de classement, Chevignon apparaît préférentiellement à la deuxième position, 'cluster 2'. Toutefois, une apparition importante se fait dans le cluster 4. Il est important de standardiser, comme nous l'avons vu précédemment les occurrences par l'occurrence moyenne d'apparition des variétés par positions de classement. Une fois cette standardisation réalisée nous obtenons la *Figure 33*, qui montre la densité d'apparition de Chevignon, pour une nouvelle méthode de construction du tableau de comparaison.

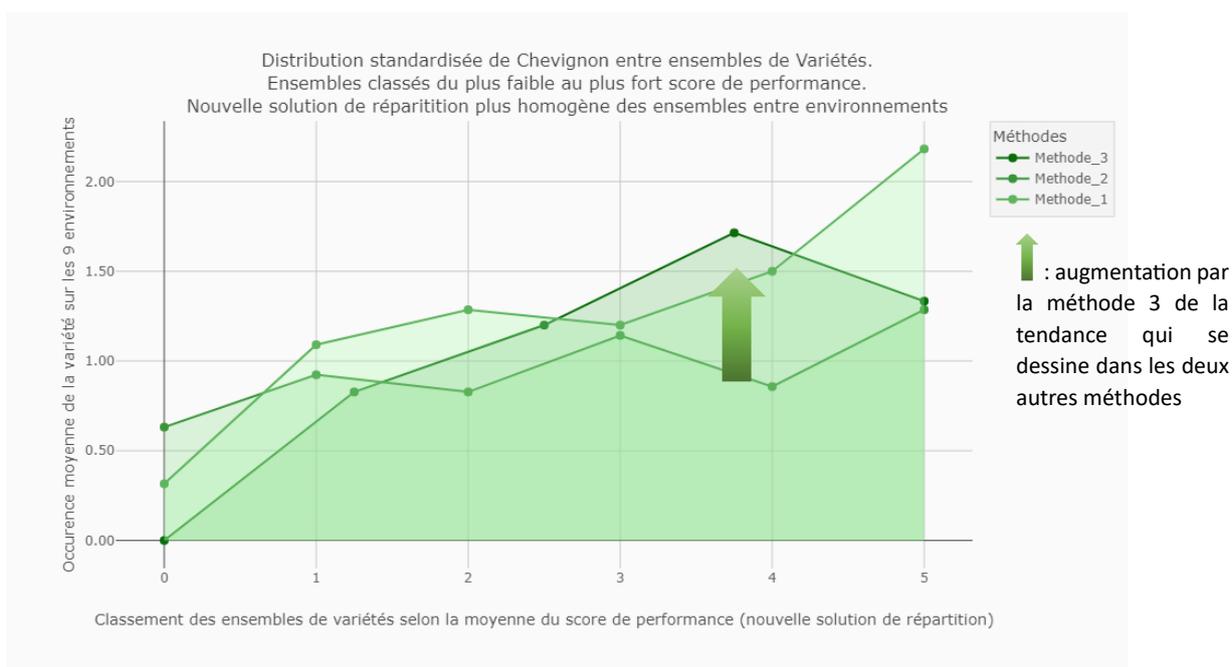


Figure 33 : Courbes de densité de l'occurrence d'apparition de la variété chevignon sur le tableau de comparaison assemblé par la simplification méthodologique 2, du score de performance

Ce graphique présente en abscisses les positions du plus petit au plus fort score de performance, lors d'un classement des ensembles significativement différent sur cette variable pour un environnement donné. L'ensemble des occurrences de Chevignon sur les 9 environnements ont été compilé et standardisé avec le nombre d'apparition moyen des variétés pour un rang du classement donné. C'est ce qui est représenté en ordonnée. Les 3 méthodes de récupération des données issues des cinétiques sur lesquels ont été généré les tableaux de comparaison sont représentées. Du vert clair au vert foncé, la méthode 1 à 3.

Sur la Figure 33 on voit que Chevignon s'étage sur toutes les positions du classement. La méthode 2 en particulier montre un étagement quasi constant avec une tendance à la hausse légère tandis ce que la méthode 1 présente une position de Chevignon sur toutes les positions du classement et en particulier sur la première position. La combinaison des méthodes, méthode 3, montre, encore une fois une accentuation de la tendance vers la seconde position du classement sur 5. Toutefois le pic étant très peu net, cela demande d'être particulièrement vigilant pour se prononcer sur la possibilité réelle de mesurer sur mesurer sur un nouvel environnement, ou de confirmer que Chevignon est bien environ seconde ou troisième dans le classement des 6 variétés étudiées pour son score de performance.

Le choix de Chevignon plutôt qu'Extase pour illustrer une variété performante en termes de sénescence et de Stay-Green se justifie par des raisons méthodologiques. Lors de l'analyse des données issues des régressions, Extase a généré un nombre important de valeurs extrêmes, notamment des pentes très fortes ou des durées de Stay-Green exceptionnellement longues. Ces valeurs atypiques ont potentiellement influencé le calcul du score de performance.

Chevignon, en revanche, a présenté des données plus cohérentes et représentatives, ce qui en fait un meilleur exemple pour illustrer les caractéristiques de sénescence et de Stay-Green sans les biais potentiels observés avec Extase.

Cette situation souligne l'importance d'avoir comparé différentes méthodes d'analyse et met en lumière certaines limites de l'étude. Ces aspects, ainsi que l'impact des variables extrêmes sur la génération du score, seront abordés plus en détail dans la section discussion qui suit.

## 4 : Discussion

### 4 – 1 : Modélisation et récupération des données issues des régressions

#### 4 – 1 – A : Standardisation

Les différences de temps de croissance Figure 7, page 20 entre les environnements et les modalités de culture ont corroboré la nécessité standardiser les données par un standard plus performant que le temps de croissance totale depuis la date de semis. L'utilisation de la standardisation du calcul du Stay-Green par le stade phénologique du Z55 a été choisie. Toutefois, les différences entre environnements du Stay-Green n'influencent pas, que la date de stade phénologique Z55, mais également le temps entre cette date et le début de la sénescence. C'est pourquoi, utiliser des jours plutôt que des degrés jours est une limite certaine dans le calcul de la performance de Stay-Green. On sait en effet que le blé répond à des périodes de croissances liées à des degrés jours adaptés au modèle de croissance de cette espèce (Bauer, Frank, et Black (1984)). Une perspective d'approche serait donc d'essayer différents stades phénologiques pour la standardisation et d'essayer de modéliser le NDVI en fonction des degrés jours plutôt que des jours.

#### 4 – 1 – B : Choix des modèles

Les différentes représentations des courbes de régressions Figure 8 page. 21 et Figure 9 page. 22 montrent que les 4 modèles utilisés fonctionnent en réalité tous bien. Les résultats complémentaires

(Figure 10, Figure 11, Figure 12, Figure 13, Figure 14) montrent en effet que même avec des critères d'informations chiffrées la distinction entre les modèles est restée complexe.

Notre analyse précédente a révélé une contradiction intéressante entre les critères d'évaluation des modèles, en particulier entre le coefficient de détermination et le maximum de vraisemblance. Cette observation souligne la complexité de la sélection du modèle optimal, particulièrement avec un ensemble de données hétérogènes comme nos 174 cinétiques. Les différences entre les critères peuvent s'expliquer par leurs sensibilités variables aux nombres de points dans chaque cinétique. Cette situation met en évidence l'importance d'une approche d'évaluation multidimensionnelle, prenant en compte non seulement les performances moyennes, mais aussi la robustesse des modèles face à la variabilité des données expérimentales. Le maximum de vraisemblance, contrairement au coefficient de détermination prend en compte le nombre de points de la cinétique.

Comme le coefficient de détermination ne prend pas en compte le nombre de points bruts et que la fonction 'Gompertz' a le plus mauvais score sur ce critère par rapport aux autres fonctions, la fonction de 'Gompertz' ajuste plus mal les données que les trois autres fonctions d'un point de vue de la moyenne des ajustements sur toutes les cinétiques testées. En revanche, si nous prenons en compte le nombre de points bruts utilisés pour les régressions, le maximum de vraisemblance nous indique que quand le nombre de points augmente, la fonction de Gompertz semble finalement plus performante que les autres modèles pour ajuster les points, ce qui est plus intéressant à prendre en compte dans le choix d'un modèle, car l'on sait qu'un modèle peut ajuster 'par hasard' des points. C'est une indication prise en compte dans les 3 autres critères testés, qui évaluent la parcimonie, notamment avec la pris en compte du nombre de paramètres du modèle évalué qui doit être le plus faible possible.

#### 4 – 1 – C : Augmentation de la puissance statistique

L'importance d'un nombre de mesures temporelles suffisante pour la création des modèles de croissance des plantes est confirmée dans la littérature. Ainsi, pour améliorer la précision des mesures et discriminer le plus finement possible les génotypes, il serait intéressant d'utiliser préférentiellement un grand nombre de données temporelles, plutôt que la multiplication des réplicas (Paine et al. (2012)).

En effet, le travail réalisé ici, s'est employé à augmenter le nombre de points individuels disponible pour un même environnement et une même variété. Les répétitions de 3 à 4 cinétiques ne reflètent en réalité par la puissance statistique disponible dans les régressions. L'obtention par une méthode indépendante de celle se basant sur la récupération unique depuis les cinétiques est une méthode intéressante d'augmentation du nombre de réplicas. Nous avons 3 réplicas multipliés par 6 points (pour l'ensemble lieu-année Thizay 2022) ou 4 réplicas multipliés par 10 points pour l'environnement (pour l'ensemble lieu-année VLB 2023). Ce qui fait que le nombre de mesures uniques pour un même environnement et une même variété n'étaient pas de 3 à 4 mais de 18 à 40.

L'augmentation du nombre de réplicas pourrait être perçue comme artificielle. Néanmoins, les trois méthodes ont élaboré de façon indépendante des scores basés sur les déciles supérieurs et inférieurs de l'ensemble des résultats de la méthode Figure 17. Cette variété dans l'obtention des données semble corroborer l'idée que la méthode 3 ne se contente pas d'amplifier artificiellement les résultats, mais qu'elle renforce probablement la tendance statistique principale observée dans les deux

premières méthodes. En produisant de nouvelles données de manière relativement autonome, cette approche pourrait permettre une compréhension plus nuancée des phénomènes étudiés.

## 4 – 2 : Annovas

### 4 – 2 – A : Approches globales

Nous avons donc, pour estimer la pertinence des méthodes effectuées, réalisé la majorité des analyses ultérieures sur les 3 méthodes d'acquisition des données. Dans un premier temps les annovas à un facteur nous ont montré que les facteurs pris séparément permettaient d'obtenir de nombreuses p-values faibles. Dans un second temps, nous avons remarqué de nombreuses interactions entre les facteurs dû à un faible degré de liberté (remarquable notamment avec le facteur 'année'). En effet, la présence d'uniquement une année pour deux lieux rend l'interaction entre les facteurs trop forte. C'est pourquoi dans la suite des analyses l'accent a été mis sur l'interaction des 3, lieu-année-modalité, appelé 'environnement', comparativement à la variété.

### 4 – 2 – B : Approche par environnement

Les annovas réalisées avec ces combinaisons nous montrent en effet, *Figure 19*, une moins grande capacité de la 'Variété' à expliquer la variance des variables dépendantes. L'environnement semble avoir une très grande incidence sur le Stay-Green notamment. Ce qui est logique, puisque l'utilisation du temps au lieu des degrés jours est un biais (Bauer, Frank, et Black (1984)), déjà discuté plus tôt. On remarque que les paramètres de pentes sont les seuls à mieux discriminer les génotypes que les environnements. Nous avons donc des paramètres intéressants qui décrivent la force de sénescence. La question est de savoir si le classement des variétés ne diffère pas entre les environnements

## 4 – 3 : Classement des génotypes

### 4 – 3 – A : Apport de la méthode 2 dans ce classement

L'analyse de la transmission du classement des variétés, *Figure 21*, réalisée sur la pente entre les environnements révèle une relative stabilité, ponctuée de comportements extrêmes. La méthode 2 semble avoir atténuée ces comportements extrêmes. Cette atténuation peut s'expliquer par l'approche moins sensible aux variations inter-réplicas apparue dans la méthode 1. En effet la méthode 1 utilise la valeur maximale de la dérivée première des équations ajustées. Ces valeurs sont fortement variables entre les réplicas.

Cette forte variabilité peut s'expliquer par le fait que la variation de la pente maximale dépend principalement d'une ou deux mesures prises à l'approche de la courbe. Ainsi, une légère fluctuation du NDVI pourrait significativement impacter la valeur de pente maximale d'une cinétique unique.

En revanche, la méthode 2 utilise

- La combinaison des images des cinétiques retracées
- Une dérivée partielle sur une dizaine de points (Équation 4, page. 56 ),

- Des seuils (avec un risque alpha calculé à environ 95% (bas et haut))

Ces points sont autant d'obtentions indépendantes de la méthode 1, se basant sur les paramètres des équations issues des régressions, justifiant l'augmentation du nombre de réplicas.

Par ailleurs, la méthode 2 semble offrir une représentation plus fidèle de la variation réelle de la pente au sein d'un même EnvXVar. Cette approche paraît moins sensible aux variations mineures et pourrait donc fournir une évaluation plus robuste des tendances générales.

Pour comprendre la systématique inter-environnement ordonnant réellement les variétés entre-elles, une technique de classification a été utilisée Simplification méthodologique 1. Cette technique permet de représenter graphiquement des tests de groupement de variétés significativement différents d'au moins une autre variété sur une même variable (test post-hoc), et sur un même environnement. Il peut être un bon indicateur pour une entreprise semencière de la faisabilité et de la répétabilité des mesures générées pour l'élaboration d'un nouveau trait.

#### 4 – 3 – B : Apport sur la variable de pente

Pour la variable de pente, ce tableau de comparaison fait apparaître un étagement entre les variétés relativement semblable. La variation principale entre les environnements n'est pas tant dans l'ordre du classement des variétés que dans leur capacité de séparation des variétés entre-elles de manière significative. Si l'on se penche sur les 4 environnements qui ont discriminé significativement Arkeos et Chevignon on remarque que ceci est dû notamment au nombre de points utilisés dans les cinétiques de ces lieux : 10 points de cinétique à VLB en 2023, 9 à Thizay en 2022 et 8 à Mons en 2023.

La Figure 23 qui étudie la pente ajustée et compare les méthodes entre-elles montre qu'une meilleure performance dans la séparation des génotypes, de la méthode 2 par rapport à la méthode 1, alors même que la méthode 2 admet un maximum de 3 réplicas par EnvXVar par rapport à la méthode 1 qui admet plus de réplicas (de 3 à 4 réplicas). Toutefois, ce fait doit être tempéré par la Simplification méthodologique 7, en effet, cette simplification méthodologique dû à la Simplification méthodologique 3 et à la Simplification méthodologique 5 ont créé des réplicas supérieurs à la moyenne très artificiellement proche de cette moyenne. Ce biais a engendré une sous-estimation de la variance au sein d'un EnvXVar, (en particulier vers la valeur supérieur).

Comment expliquer ce phénomène ?

Voir :

#### 4 – 3 – C : Analyse bivariée des scores

Si nous continuons l'analyse, figure 24, nous voyons que le score de Stay-Green est différenciant également pour les 4 environnements déjà vu précédemment <sup>2</sup>. Arkeos et Chevignon inversent leur classement pour l'environnement D1-2023-VLB par rapport aux autres environnements, cela pourrait en effet s'expliquer par un comportement inhabituel de cet environnement. Il est possible que des

---

<sup>2</sup> D1 2023 VLB, Normal 2023 VLB (10 points) Normal 2023 Thizay (9 points), D1 2022 Thizay (8 points),

interactions génotypes-environnement soient particulièrement fortes. Dans ces interactions, il serait intéressant de savoir si ces environnements créant ces différences par rapport à la 'pluspart' des autres environnements généreraient des comportements inhabituels sur un grand nombre de variétés ou si au contraire ces changements sont spécifiques à certaines variétés avec cet environnement. 3 raisons peuvent en effet expliquer une interaction génotypeXenvironnement :

- 1- La variable mesurée est peu héritable et bénéficie d'une forte variation dû à l'environnement et d'une interaction des génétiques testées faibles ou complexes
- 2- L'environnement est particulier et tend à inverser les génotypes par rapport aux autres environnements
- 3- Un génotype est particulier et tend à être plus influencé que les autres par l'environnement

Enfin, bien-sûr, il peut s'agir d'une combinaison des 3. S'il est complexe avec un nombre de données relativement restreint de savoir ce qui est le plus en jeu pour expliquer les modifications dans le classement des variétés, il peut être intéressant de prendre en compte plusieurs variables pour comprendre les dynamiques qui font interagir les variables mesurées, l'environnement de culture et la variété.

Utiliser, par exemple le comportement conjoint du score de Stay-Green et du score de pente ajustée pourrait nous permettre dans une analyse bivariée de mieux comprendre les biais, les artefacts expérimentaux ou les tendances réelles créés par les environnements étudiés et les variétés étudiées.

Afin de comprendre les comportements inhabituels de certains environnements nous voyons *Figure 25* l'étude comparative des scores et nous observons une dispersion des points qui n'est pas homogène mais clusterisé. L'étude de ces clusters *Figure 26* et la compréhension de ces clusters vis-à-vis des variétés et des environnements *Figure 27* montrent en effet que D1-2023-VLB (qui a produit un inversement du comportement d'Arkeos et de Chevignon sur le score de Stay-Green) est également un environnement qui semble avoir généré des mesures de haut score de Stay-Green et de haut score de force de pente, conjointement. L'étude météorologique en 2023 à VLB et de l'impact du semis précoce (D1) pour comprendre plus en profondeur ce cluster-environnement et sa position dans le graphique des scores, pourraient être intéressants.

Plus pragmatiquement, cette forte clusterisation des environnements sur les variables de scores indique qu'il serait préférable de moyenniser les effets de ces clusters. Cela pourrait, par exemple être fait, en translatant tous les individus. Cette translation pourrait s'effectuer par les vecteurs entre le barycentre, des centres de gravités des environnements et le centre de gravité de l'environnement considéré.

#### 4 – 3 – D : Classement sur le score de performance

La *Figure 26* montre des comportements extrêmes d'Extase avec des scores manifestement très élevés à la fois pour le Stay-Green et la force sénescence. Ces valeurs probablement aberrantes se retrouvent naturellement *Figure 29* avec la présence d'Extase dans des scores deux fois plus élevés que les autres variétés qui peinent à s'étager autour de plus de 40 points de score sur 140.

Ces valeurs extrêmes font que l'intérêt s'est plutôt porté sur Chevignon bien qu'Extase a été également analysé pour le classement. Chevignon semble *Figure 33* être entre la 3<sup>ème</sup> et la 4<sup>ème</sup> place. Toutefois, Extase a probablement généré des valeurs extrêmes pour des raisons techniques particulières (dont les raisons doivent être approfondies), ce qui pourrait faire passer Chevignon à la première place de ce classement.

Par ailleurs, la *Figure 33*, ainsi que les figures (*Figure 30*, *Figure 31*) nous montrent que la méthode 2 et la combinaison par la méthode 3 semblent réellement pertinentes car elles confirment ou augmentent les tendances de la méthode 1.

#### 4 – 4 : Note sur les génotypes et explications des comportements

##### 4 – 4 – A : Anti-corrélation entre le Stay-Green et la force de sénescence

L'amélioration conjointe du Stay-Green et de la sénescence devra donc prendre en compte l'environnement et se confrontera probablement à des anti-corrélations génétiques comme l'a suggéré la figure de plot des valeurs de score de Stay-Green et de force de sénescence et comme tendent à le prouver les études sur les QTL de la sénescence chez le maïs (Zhang et al. (2019)). Il peut être intéressant d'étudier les caractéristiques d'Extase expliquant la formation de ces valeurs extrêmes et de répéter l'expérience si ces valeurs s'avèrent concordantes.

Par ailleurs, des analyses complémentaires et plus en profondeur doivent être réalisés notamment sur le Stay-Green qui admet différents 'modes' influençant le rôle dans la photosynthèse et le rendement. On peut noter en particulier le Stay-Green 'cosmétique' ou la plante accumule de la chlorophylle sans pour autant réaliser de la photosynthèse utile au rendement (Thomas et Howarth (2000); Thomas et Ougham (2014)).

##### 4 – 4 – B : Qualités des variétés, Stay-Green et sénescence

Il est important de garder à l'esprit que la force de sénescence et le Stay-Green, bien que capable en partie, de prédire le rendement ne sont pas les seules qualités recherchées chez le blé. Si l'on compare les qualités de Chevignon et d'Arkeos, on remarque qu'Arkeos a une qualité biscuitière qui demande de répondre à un cahier des charges très important et que de nombreuses anti-corrélations génétiques probables peuvent exister entre ces qualités et l'amélioration conjointe du Stay-Green et de la force de la pente.

#### 4 – 5 : Perspectives supplémentaires

Différentes améliorations sont envisageables pour mieux comprendre et mieux caractériser le comportement des plantes par une cinétique sur un indice de télémétrie.

En amont de la prise de résultats :

- 1- Augmenter le nombre de points cinétiques, en particulier sur le début de l'inflexion et dans la pente de la sigmoïde
- 2- Augmenter le nombre d'environnements testés pour augmenter les DDL

- 3- Utiliser différents indices (en plus du NDVI) et notamment le PSRI

Une fois les résultats obtenus

- 4- Réaliser le travail de régression et de récupération des résultats sur d'autres équations
  - a. Des équations à 5 paramètres comme la Logistique forte
  - b. Des équations capables d'intégrer une fonction affine avant la sigmoïde de sénescence, pour caractériser plus finement le Stay-Green (une autre option pourrait être de trouver une méthode retirant les points trop en amont de la sigmoïde et perturbant la régression)
- 5- Améliorer les simplifications méthodologiques sur l'obtention des répliques par la méthode 1
- 6- Améliorer les simplifications méthodologiques pour le ranking inter-environnement
- 7- Grâce à une puissance statistique augmentée, tenter de comprendre les comportements spécifiques de l'interaction génotypes environnements

## 5 : Conclusion

Les objectifs de cette étude étaient de comprendre comment modéliser des cinétiques de sénescence de blé tendre d'hiver et de comprendre quels caractères des plantes sont décelables entre les variétés sur ces cinétiques. Nous avons vu que pour le premier objectif, la réponse pouvait être complexe à apporter même si des résultats satisfaisants pour les 4 fonctions testées ont été obtenus. D'un point de vue des grandeurs utilisées pour les modélisations, il serait préférable, dans une prochaine étude, de modéliser avec des degrés jours. Pour le second objectif, il semblerait que s'il est possible de distinguer les variétés par le stay-green et la force de pente, d'importantes questions demeurent sur le rôle de l'environnement et de ses différentes modalités. La puissance statistique et la couverture modale devraient être plus importantes dans les expériences à suivre.

Pour l'augmentation de la puissance statistique, la méthode 3 s'est montrée satisfaisante mais des confirmations doivent être apportées pour s'assurer que nous ne sommes pas dans le cas d'une augmentation artificielle des répliques. En effet, nous avons vu que de nombreuses améliorations pouvaient être apportées à la méthode 2. Le comportement conjoint du score de Stay-Green et de la sénescence montre l'importance de l'incidence de l'environnement sur la dynamique de ces scores. Par ailleurs, il sera important de tester un grand nombre d'environnements pour mieux comprendre les nombreuses interactions génétiques-environnement pouvant exister.

Nous avons déjà pu remarquer que certaines variétés avaient des comportements particuliers comme Obiwan avec un long Stay-Green ou Cheviron avec un Stay-Green et une force de sénescence moyenne mais une combinaison des deux relativement meilleure que la variété Arkeos. Cette étude a donc offert quelques premiers résultats et tendances intéressantes sur les possibilités de modélisation, de caractérisation et de différenciation des génotypes. De nombreuses perspectives d'améliorations pour de futures études ont été envisagées et devraient permettre aux semenciers et à Limagrain d'améliorer leurs méthodologies et leurs outils d'aide à la décision pour les personnes responsables des croisements.

## 6 : Bibliographie

- Ach, Yves-Alain. 2006. *Echelle BCHH du blé*. NOUVELLE GENERATION DE PRESSE AGRICOLE. <https://www.terre-net.fr/2006/article/30491/maladie-ble-oidium>.
- Alt, Douglas S., Pierce A. Paul, Alexander J. Lindsey, et Laura E. Lindsey. 2019. « Early Wheat Harvest Influenced Grain Quality and Profit but Not Yield ». *Crop, Forage & Turfgrass Management* 5 (1): 190001. <https://doi.org/10.2134/cftm2019.01.0001>.
- Anderegg, Jonas, Kang Yu, Helge Aasen, Achim Walter, Frank Liebisch, et Andreas Hund. 2020. « Spectral Vegetation Indices to Track Senescence Dynamics in Diverse Wheat Germplasm ». *Frontiers in Plant Science* 10 (janvier):1749. <https://doi.org/10.3389/fpls.2019.01749>.
- Baillet, Nadège. 2020. « Distribution de la masse individuelle du grain de blé tendre: analyse de la variabilité de la masse en fonction de la position au sein de l'épi ».
- Bauer, Armand, A. B. Frank, et A. L. Black. 1984. « Estimation of Spring Wheat Leaf Growth Rates and Anthesis from Air Temperature ». *Agronomy Journal* 76 (5): 829-35. <https://doi.org/10.2134/agronj1984.00021962007600050027x>.
- Calderini, Daniel F., et Gustavo A. Slafer. 1998. « Changes in Yield and Yield Stability in Wheat during the 20th Century ». *Field Crops Research* 57 (3): 335-47. [https://doi.org/10.1016/S0378-4290\(98\)00080-X](https://doi.org/10.1016/S0378-4290(98)00080-X).
- Chang, Sungyul, Unseok Lee, Min Jeong Hong, Yeong Deuk Jo, et Jin-Baek Kim. 2021. « Time-Series Growth Prediction Model Based on U-Net and Machine Learning in Arabidopsis ». *Frontiers in Plant Science* 12 (novembre):721512. <https://doi.org/10.3389/fpls.2021.721512>.
- Couturier, Christian. 2014. « Blé dur d'hiver Triticum aestivum ». <https://www.herbea.org/fr/culture/11954/BI%C3%A9-dur-d%27hiver>.
- Crain, Jared, Matthew Reynolds, et Jesse Poland. 2017. « Utilizing High-Throughput Phenotypic Data for Improved Phenotypic Selection of Stress-Adaptive Traits in Wheat ». *Crop Science* 57 (2): 648-59. <https://doi.org/10.2135/cropsci2016.02.0135>.
- De Meyer, Sam, Daniel Felipe Cruz, Tom De Swaef, Peter Lootens, Jolien De Block, Kevin Bird, Heike Sprenger, et al. 2023. « Predicting Yield of Individual Field-Grown Rapeseed Plants from Rosette-Stage Leaf Gene Expression ». Édité par Kiran Raosaheb Patil. *PLOS Computational Biology* 19 (5): e1011161. <https://doi.org/10.1371/journal.pcbi.1011161>.
- Diaz-Mendoza, Mercedes, Blanca Velasco-Arroyo, M. Estrella Santamaria, Pablo González-Melendi, Manuel Martinez, et Isabel Diaz. 2016. « Plant Senescence and Proteolysis: Two Processes with One Destiny ». *Genetics and Molecular Biology* 39 (3): 329-38. <https://doi.org/10.1590/1678-4685-GMB-2016-0015>.
- Distelfeld, A., R. Avni, et A. M. Fischer. 2014. « Senescence, Nutrient Remobilization, and Yield in Wheat and Barley ». *Journal of Experimental Botany* 65 (14): 3783-98. <https://doi.org/10.1093/jxb/ert477>.
- Dolan, K.D., L. Yang, et C.P. Trampel. 2007. « Nonlinear Regression Technique to Estimate Kinetic Parameters and Confidence Intervals in Unsteady-State Conduction-Heated Foods ». *Journal of Food Engineering* 80 (2): 581-93. <https://doi.org/10.1016/j.jfoodeng.2006.06.023>.
- Farooq, Muhammad, Helen Bramley, Jairo A. Palta, et Kadambot H.M. Siddique. 2011. « Heat Stress in Wheat during Reproductive and Grain-Filling Phases ». *Critical Reviews in Plant Sciences* 30 (6): 491-507. <https://doi.org/10.1080/07352689.2011.615687>.
- Gompertz, B. 1825. « On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. », *Philosophical Transactions of the Royal Society of London*, 115:513-83.
- Guo, Zifeng, Dijun Chen, et Thorsten Schnurbusch. 2018. « Plant and Floret Growth at Distinct Developmental Stages During the Stem Elongation Phase in Wheat ». *Frontiers in Plant Science* 9 (mars):330. <https://doi.org/10.3389/fpls.2018.00330>.

- Hara, Nathan C., et Eric B. Ford. 2023. « Statistical Methods for Exoplanet Detection with Radial Velocities ». *Annual Review of Statistics and Its Application* 10 (1): 623-49. <https://doi.org/10.1146/annurev-statistics-033021-012225>.
- Kirby, E.J.M., et M Appleyard. 1984. *Cereal Development Guide*. NAC Cereal Unit. Stoneleigh, Kenilworth, UK. [https://books.google.fr/books/about/Cereal\\_Development\\_Guide.html?id=hJYiuQAACAAJ&redir\\_esc=y](https://books.google.fr/books/about/Cereal_Development_Guide.html?id=hJYiuQAACAAJ&redir_esc=y).
- Kouadio, Amani Louis. 2012. « Prévisions des rendements du blé d'hiver à échelle régionale par modélisation de la courbe de chute de l'indice foliaire ». Thèse présentée en vue de l'obtention du grade de Docteur en Sciences, Université de Liège Faculté des Sciences Département des Sciences et Gestion de l'Environnement: Université de Liège. <https://orbi.uliege.be/handle/2268/115714>.
- Lindsey, Laura E., Pierce Paul, et Edwin Lentz. 2017. « Wheat Growth Stages and Associated Management ». Agriculture and Natural Resources. The Ohio State University College of Food, Agricultural, And Environmental, Sciences. <https://ohioline.osu.edu/factsheet/agf-126>.
- Omrane, Selim, Colette C Audeon, Amandine Ignace, Clémentine Duplaix, Hind Sghyer, Lamia Aouini, Gert Kema, Sabine Fillinger, et Anne-Sophie Walker. 2015. « La résistance de type MDR chez l'agent de la septoriose du blé: dernières avancées sur un mode d'action original ». Conférence présentée à AFPP – ONZIÈME CONFÉRENCE INTERNATIONALE SUR LES MALADIES DES PLANTES, Tours, décembre 9. <https://hal.inrae.fr/hal-02740700/document>.
- Paine, C. E. Timothy, Toby R. Marthews, Deborah R. Vogt, Drew Purves, Mark Rees, Andy Hector, et Lindsay A. Turnbull. 2012. « How to Fit Nonlinear Plant Growth Models and Calculate Growth Rates: An Update for Ecologists ». *Methods in Ecology and Evolution* 3 (2): 245-56. <https://doi.org/10.1111/j.2041-210X.2011.00155.x>.
- Pomerantsev, Alexey L. 1999. « Confidence Intervals for Nonlinear Regression Extrapolation ». *Chemometrics and Intelligent Laboratory Systems* 49 (1): 41-48. [https://doi.org/10.1016/S0169-7439\(99\)00026-X](https://doi.org/10.1016/S0169-7439(99)00026-X).
- Roby, Charlotte. 2023. « Le phénotypage : processus clé en amélioration des plantes ». *Vegenov*, 20 novembre 2023. <https://blog.vegenov.com/2023/11/le-phenotypage-processus-cle-en-amelioration-des-plantes/>.
- Rouse, W, et R H Haas. 1974. « MONITORING VEGETATION SYSTEMS IN THE GREAT PLAINS WITH ERTS ».
- Sanger, Jon E. 1971. « Quantitative Investigations of Leaf Pigments From Their Inception in Buds Through Autumn Coloration to Decomposition in Falling Leaves ». *Ecology* 52 (6): 1075-89. <https://doi.org/10.2307/1933816>.
- Saphir, Nicholas. 2020. « Analyst Insight: Sell or Store Your Grain during Harvest Time? » <https://ahdb.org.uk/news/analyst-insight-sell-or-store-your-grain-during-harvest-time>.
- Schauberger, Bernhard, Tamara Ben-Ari, David Makowski, Tomomichi Kato, Hiromi Kato, et Philippe Ciais. 2018. « Yield Trends, Variability and Stagnation Analysis of Major Crops in France over More than a Century ». *Scientific Reports* 8 (1): 16865. <https://doi.org/10.1038/s41598-018-35351-1>.
- Shewry, Peter R., Rowan A.C. Mitchell, Paola Tosi, Yongfang Wan, Claudia Underwood, Alison Lovegrove, Jackie Freeman, Geraldine A. Toole, E.N. Clare Mills, et Jane L. Ward. 2012. « An Integrated Study of Grain Development of Wheat (Cv. Hereward) ». *Journal of Cereal Science* 56 (1): 21-30. <https://doi.org/10.1016/j.jcs.2011.11.007>.
- Thomas, Howard, et Catherine J. Howarth. 2000. « Five Ways to Stay Green ». *Journal of Experimental Botany* 51 (suppl\_1): 329-37. [https://doi.org/10.1093/jexbot/51.suppl\\_1.329](https://doi.org/10.1093/jexbot/51.suppl_1.329).
- Thomas, Howard, et Helen Ougham. 2014. « The Stay-Green Trait ». *Journal of Experimental Botany* 65 (14): 3889-3900. <https://doi.org/10.1093/jxb/eru037>.
- Verhulst, Pierre François. 1838. « Notice sur la loi que la population suit dans son accroissement. Correspondance mathématique et physique » 10:113-21.

- Zadoks, J. C., T. T. Chang, et C. F. Konzak. 1974. « A Decimal Code for the Growth Stages of Cereals ». *Weed Research* 14 (juin):415-21. <https://doi.org/10.1111/j.1744-7348.1991.tb04895.x>.
- Zeng, Xueqin, Walter Durka, Erik Welk, et Markus Fischer. 2017. « Heritability of Early Growth Traits and Their Plasticity in 14 Woody Species of Chinese Subtropical Forest ». *Journal of Plant Ecology* 10 (1): 222-31. <https://doi.org/10.1093/jpe/rtw086>.
- Zhang, Jun, Kevin A. Fengler, John L. Van Hemert, Rajeev Gupta, Nick Mongar, Jindong Sun, William B. Allen, et al. 2019. « Identification and Characterization of a Novel Stay-green QTL That Increases Yield in Maize ». *Plant Biotechnology Journal* 17 (12): 2272-85. <https://doi.org/10.1111/pbi.13139>.

## 7 : Annexes

### 7 – 1 : Autres équations testées

Deux autres fonctions ont également été testé et tentent de décrire certaines situations ou le Stay-Green qui précède la décélération en forme sigmoïdale ne semble pas être constant mais décroissant, c'est-à-dire suivant une courbe affine.

Ajouter les 2 paramètres aux équations de sigmoïdes pour décrire les pentes affines des Stay-Green a dû être fait avec des équations de 7 paramètres. Cela élimina d'office toutes les cinétiques inférieures à 7 points et génère lors de l'ajustement avec `curve_fit` pour les cinétiques avec 7 points ou plus des matrices de variances des paramètres avec uniquement des valeurs infinies. Le modèle n'était donc pas assez parcimonieux.

C'est pourquoi il a été préféré de mixer les paramètres et d'ajuster le rapport entre ces paramètres pour que soit commun les paramètres de la pente du Stay-Green avec certains paramètres de la sigmoïde de sénescence.

Le terme ' $x^2/d^2$ ' a été utilisé pour supprimer progressivement l'influence du Stay-Green décroissant sur le modèle et augmenter progressivement l'importance de l'influence de l'équation de la courbe de Gompertz. Afin de diminuer le nombre de paramètres utilisés, l'asymptote minimale a été fixé à la moyenne de NDVI des derniers points de cinétiques dans le jeu de données brutes filtré pour l'ajustement ; ici ce nombre est 0.17204.

Il est à noter qu'ici la partie fonction de Gompertz utilisée comprends un paramètre supplémentaire dans l'équation (le facteur b devant la seconde exponentielle) qui est normalement de -1 dans l'équation de Gompertz décrivant une décroissance classiquement. Il n'y a pas de raison particulière à cette utilisation et les influences de la mobilité de ce paramètre n'ont pas été mesuré ici sur la sigmoïde proprement dite.

En revanche ce paramètre a été utilisée dans le croisement empirique des paramètres de sigmoïde de la sénescence et ceux de la droite affine décrivant le Stay-Green.

On peut donc ajouter dans la définition des paramètres précédemment donnée :

- b = influence inconnue sur la sigmoïde pour la fonction de Gompertz, et force de la pente pour la fonction Logistique et force de la pente du Stay-Green dans les deux cas

- $c$  = force de la pente de la sigmoïde pour la fonction de Gompertz, déséquilibre entre l'inflexion de début et de fin de la courbe sigmoïde pour la fonction Logistique et départ de la pente du Stay-Green dans les deux cas

Les meilleurs rapports obtenus empiriquement sont écrits dans les équations qui suivent :

*Gompertz avec Stay-Green*

$$\frac{0.00484 \times bx + 0.295 \times (c - x \times c/60)}{1 + x^2/d^2} + 0.17204 + e \times e^{b \times e^{c(x-d)}}$$

*Logistique avec Stay-Green*

$$\frac{0.00495 \times bx + 0.0385 \times (c - x \times 6000/c)}{1 + (x^2/(d^2 + 0.01 \times (1/c)))} + 0.17204 + \frac{a-0.17204}{1 + (e^{b(x-0.001 \times (1/c))})^{1/c}}$$

Toutefois, les modélisations, même en croisant les paramètres pour limiter leurs nombres, ne s'ajustaient pas bien. Le travail suivant, d'ajustement de paramètres généraux, n'a pas, n'ont plus, été suffisamment convainquant. Dans le cas d'un travail avec moins de 10 points cinétiques, les équations présentées dans cette partie annexe n'étaient donc pas utilisables. Elles n'ont donc pas servi pour la suite des travaux.

## 7 – 2 : Formules des dérivées de Gompertz

Toutes les dérivées découleront de cette équation qui admet un paramètre supplémentaire, le paramètre b.

$$f(x) = e + (a - e) \times e^{-b \times e^{-c(x-d)}}$$

Dérivée d'ordre 1 :

$$f'(x) = -a \times b \times c \times e^{-b \times e^{c \times (x-d)}} \times e^{c \times (x-d)}$$

Dérivée d'ordre 2 :

$$f''(x) = -a \times b \times c \times \left[ e^{-b \times e^{c \times (x-d)}} \times e^{c \times (x-d)} + c \times e^{-b \times e^{c \times (x-d)} \times e^{c \times (x-d)}} \right]$$

Dérivée d'ordre 3 :

$$f'''(x) = -a \times b \times c \times (A + B + C)$$

Avec :

$$A = -3 \times b \times c^2 \times e^{-b \times e^{c \times (x-d)}} \times e^{2 \times c \times (x-d)}$$

$$B = b^2 \times c^2 \times e^{-b \times e^{c \times (x-d)}} \times e^{3 \times c \times (x-d)}$$

$$C = c^2 \times e^{-b \times e^{c \times (x-d)}} \times e^{c \times (x-d)}$$

Dérivée d'ordre 4 :

$$\begin{aligned} f''''(x) &= -a \times b \times c^4 \times e^{-b \times e^{c \times (x-d)}} \\ &\times \left[ -3 \times b \times e^{2 \times c \times (x-d)} \times (2 - b \times e^{c \times (x-d)}) \right. \\ &\left. + b^2 \times e^{3 \times c \times (x-d)} \times (3 - b \times e^{c \times (x-d)}) + e^{c \times (x-d)} \times (1 - b \times e^{c \times (x-d)}) \right] \end{aligned}$$

Dérivée d'ordre 5 :

$$f''''''(x) = Ud \times (V + W + Y) + U \times (Vd + Wd + Yd)$$

Avec :

$$U = g \times h$$

$$Ud = gd \times h + g \times hd$$

$$V = i \times j$$

$$Vd = id \times j + i \times jd$$

$$W = k \times l$$

$$Wd = kd \times l + k \times ld$$

$$Y = m \times n$$

$$Yd = md \times n + m \times nd$$

Avec :

$$g = -a \times b \times c^4 \text{ et } gd = 0$$

$$h = e^{-b \times e^{c \times (x-d)}} \text{ et } hd = -b \times c \times e^{-b \times e^{c \times (x-d)}} \times e^{c \times (x-d)}$$

$$i = -3 \times b \times e^{2c \times (x-d)} \text{ et } id = -6 \times b \times c \times e^{2c \times (x-d)}$$

$$j = 2 - b \times e^{c \times (x-d)} \text{ et } jd = -b \times c \times e^{c \times (x-d)}$$

$$k = b^2 \times e^{3c \times (x-d)} \text{ et } kd = 3 \times b^2 \times c \times e^{3c \times (x-d)}$$

$$l = 3 - b \times e^{c \times (x-d)} \text{ et } ld = -b \times c \times e^{c \times (x-d)}$$

$$m = e^{c \times (x-d)} \text{ et } md = c \times e^{c \times (x-d)}$$

$$n = 1 - b \times e^{c \times (x-d)} \text{ et } nd = -b \times c \times e^{c \times (x-d)}$$

### 7 – 3 : Méthode numéro 2 de récupération des données

On suppose déterminer une loi de probabilité de type normale avec un seuil haut et bas pour chaque EnvXVar et pour les 3 variables que l'on comparera par la suite en générant trois échantillons :

- Un pour la moyenne
- Deux pour représenter l'écart-type

La moyenne de ces lois sera déterminée par un système de pondération des répliques entre eux en fonction de la fiabilité de la régression, on crée ensuite un système unique de calcul pour chaque variable. Ce système sera unique pour le calcul de la moyenne et de des valeurs de seuil supérieur et inférieur.

Pour réaliser cette méthode nous calculons donc des paramètres qui nous seront utiles pour la pondération des résultats.

Nous ajoutons, aux 4 paramètres déjà obtenus entre les résultats prédits et les résultats bruts des régressions (le coefficient de détermination ( $R^2$ )) et le nombre de paramètres des fonctions choisies et les nombres de points sur lesquels elles ont été ajustées (LL, BIC et AIC) deux critères supplémentaires. Ces critères de discrimination des régressions uniques sont obtenus à partir de la matrice de variances entre les paramètres issu des résultats obtenu par l'algorithme de régression de la fonction 'curve\_fit' :

$$\text{Calcul de la variance des paramètres (Globalerror)} : \text{Globalerror} = \sum \sqrt{\text{diag} \left[ \begin{matrix} var_a & \cdots & var_{da} \\ \vdots & \ddots & \vdots \\ var_{ad} & \cdots & var_d \end{matrix} \right]}$$

Calcul d'un indice de comparaison de cette variance des paramètres (Globalerrorcomp) :

$$\text{Globalerrorcomp} = (\log \text{Globalerror})$$

Calcul d'un comparateur sur la qualité de chaque régression prenant en compte tous les paramètres (Les 3 premiers sont ceux ayant été utilisé pour le calcul du BIC)

- Nombre de points utilisés dans la cinétique
- Nombre de paramètre de l'équation de régression
- Le coefficient de détermination
- La matrice de variance des paramètres récupérée lors de l'ajustement itératif réalisé par l'algorithme (finalcomp) :

Équation 1

$$\text{finalcomp} = \text{globalerrorcomp} \times \text{BIC}$$

Ensuite, nous réalisons un calcul d'une fiabilité pour chaque cinétique permettant de représenter un indice de confiance de cette cinétique. Cet indice de confiance sera ensuite comparable entre toute les autres cinétiques (*fiab*). Nous réalisons dans un premier temps un calcul intermédiaire (*fiabintermediate*). Voici le calcul :

Avec : *min* = valeur minimale de *finalcomp*

$$\text{fiabintermediate} = \frac{1}{\log(-\text{min} + 2 + \text{finalcomp})}$$

Le comparateur *fiab* est ensuite ainsi calculé. Ce comparateur permettra d'évaluer la fiabilité d'une cinétique au sein d'une combinaison 'Condition de culture X Variété' :

$$\text{fiab} = \frac{\text{fiabintermediate}}{\sum_{i=1}^{\text{Comb}(CDCltXVar)} \text{fiabintermediate}_i}$$

Une fois ces indicateurs calculés pour chaque cinétique, on procède à la création des courbes. On génère pour cela 10.000 points de la valeur minimale de la durée entre le stade Z55 de la plante phénotypée et la première mesure de NDVI réalisée parmi toute les cinétiques (ici -30 jours) et la valeur maximale de cette durée avec la dernière mesure de NDVI réalisée (ici 60 jours).

On cherche à présent à obtenir des paramètres sur ces cinétiques. Dans cette méthode, le choix a été fait de se concentrer sur le calcul d'un seul paramètre par combinaison plutôt que par cinétique. En effet, en première lecture graphique des courbes on a pu observer une grande similitude entre les mêmes combinaisons (EnvXVar) des courbes obtenues.

Dans le cadre d'une approche exploratoire, il paraît intéressant d'illustrer cette grande répétabilité dans le cadre d'une même combinaison. En effet, les régressions offrent la possibilité de s'intéresser à un phénomène en deux dimensions. C'est-à-dire que le nombre de réplica disponibles pour la mesure de la confiance en la donnée obtenue peut potentiellement être dépassé par l'obtention d'un grand nombre de points par calculs des valeurs prédites, avant et après un moment d'intérêt.

La régression offre la possibilité d'augmenter, parfois bien au-delà de l'erreur technique d'un appareil, la mesure d'une donnée quand l'utilisateur dispose d'un très grand nombre de données temporelles (Hara et Ford (2023)).

Grâce à un modèle prédictif cohérent, on peut en effet dépasser certaines limites techniques de mesure de l'erreur et cet effet sera ici en parti utilisé. Il a donc été décidé d'illustrer cette variabilité faible sur chaque combinaison par le calcul empirique d'un IC à environ 95%.

Il a été décidé d'utiliser une méthode empirique car il est compliqué d'implémenter un intervalle de confiance dans les régressions non linéaires (Paine et al. (2012)), (Dolan, Yang, et Trampel (2007)), (Pomerantsev (1999)). Chacun des calculs empiriques de ces intervalles de confiance seront spécifiques pour chaque paramètre créé. Ils seront expliqués plus loin dans ce matériel et méthode.

Dans un premier temps voici comment sont récupérés les valeurs moyennes pour chacun de ces paramètres :

7 – 3 – A : Récupération des valeurs moyennes

*7 – 3 – A – a : Récupération du Drop NDVI pondérée*

Pour le calcul du drop de NDVI, on pondère en utilisant comme poids la fiabilité (l'indice qui est comparateur des réplicas entre eux)

A partir des deux paramètres d'asymptote haute et basse on a calculé un drop :

$$\text{Drop.NDVI}_n = \text{Parm.asymp.haute}_n - \text{Parm.asymp.basse}_n$$

On a ensuite calculé pour chaque cinétique un poids à partir de la fiabilité en renforçant la prise en compte de la variance des paramètres d'asymptotes (obtenu par la matrice de covariance de l'algorithme de régression `curve_fit`).

L'équation de calcul des poids est la suivante :

Avec :

- Mn la matrice de covariance des paramètres d'une régression d'une cinétique donnée n
- *ass.basse* la position dans la matrice du paramètre d'asymptote basse et *ass.haute* la position du paramètre d'asymptote haut.

Équation 2

$$Drop.NDVI.weight_n = \frac{fiab_n}{Mn_{ass.basse,ass.basse} + Mn_{ass.haute,ass.haute} + Mn_{ass.basse,ass.haute}}$$

$$ValeurDropNDVI_{EnvXVar} = \frac{\sum_{i=1}^{nbreplacatEnvXVar} Drop.NDVI.weight_i \times Drop.NDVI_i}{\sum_{i=1}^{nbreplacatEnvXVar} Drop.NDVI.weight_i}$$

7-3-A-b : Récupération de la position de la sigmoïde sur l'axe des abscisses pondérée (Stay-Green)

Nous calculons à présent le nombre de jours après Z55 pour une combinaison de cinétique appartenant à même environnement et variété (EnvXVar). Ici nous nous intéressons au calcul de la valeur moyenne. Pour augmenter la confiance dans cette valeur nous utilisons comme précédemment un poids à partir de la matrice de covariance des paramètres issues de l'algorithme de régression.

Pour calculer ce poids on a donc ici utilisé la variabilité du paramètre de décalage de la courbe sigmoïde qui est directement dépendante de notre valeur d'intérêt.

Équation 3

Voici donc l'équation suivante :

Avec

- *pos.sigmo* : la position dans la matrice du paramètre de position de sigmoïde sur l'axe des x

$$Pos.sigmo.weight_n = \frac{fiab_n}{Mn_{pos.sigmo,pos.sigmo}}$$

$$ValeurPossigmo_{EnvXVar} = \frac{\sum_{i=1}^{nbreplacatEnvXVar} Pos.sigmo.weight_i \times Param.pos.sigmo_i}{\sum_{i=1}^{nbreplacatEnvXVar} Pos.sigmo.weight_i}$$

7-3-A-c : La valeur de la pente maximale pondérée

Cette valeur a été calculé sur la dérivée partielle de la moyenne des courbes translattées. En effet, les courbes translattées seront utilisées pour le calcul de l'intervalle de confiance de la pente maximum. Elles sont donc utilisées pour le calcul de la moyenne de la pente maximum. Cela n'a pratiquement

aucune incidence puisque la courbe moyenne de la combinaison des courbes brutes et la courbe moyenne de la combinaison des courbes translattées sont quasiment identiques<sup>3</sup>.

Translations des cinétiques d'une même combinaison autour de leur centre de gravité commun

- 1- Calcul du centre de gravité de la courbe de chaque régression issue d'une série temporelle unique

*Simplification méthodologique 3*

On réalise le calcul du centre de gravité de la courbe sur l'ensemble des points de la courbe générée de -30 jours à 60 jours, ce qui peut causer des problèmes de normalisation entre les courbes.

```
# moments statiques Mx et My (contribution de chaque point à la position du centre de gravité)
Mx = 0
My = 0
L = 0
for i in range(len(x_coords)-1):
    dx = x_coords[i+1] - x_coords[i]
    #constante ici
    dy = y_coords[i+1] - y_coords[i]
    #différence de NDVI entre deux points consécutifs
    ds = (dx**2 + dy**2)**0.5
    #distance entre ces deux points
    Mx += x_coords[i] * ds
    #moment cinétique en abscisse
    My += y_coords[i] * ds
    #moment cinétique en ordonnée
    L += ds

# Coordonnées du centroïde
xG = Mx / L
yG = My / L
```

Figure 34 : algorithme utilisé pour le calcul des centres de gravité

- 2- Calcul du centre de gravité commun

On a pris pour chaque EnvXVar le centre de gravité de toutes les courbes uniques pour chaque réplica correspondant. On a ensuite appliqué ce calcul

Avec :

- xGc = coordonnée x du centre de gravité commun des courbes
- yGc = coordonnée y du centre de gravité commun des courbes
- xGi = coordonnée x du centre de gravité de la courbe i
- yGi = coordonnée y du centre de gravité de la courbe i

$$xGc, yGc = \frac{\sum_{i=1}^n xGi}{n}, \frac{\sum_{i=1}^n yGi}{n}$$

On a ensuite calculé le vecteur  $\vec{V}$  entre chaque centre de gravité de chaque courbe réplica unique vers son centre de gravité général EnvXVar.

Avec :

<sup>3</sup> L'utilisation de la translation se justifie par la volonté de minimiser les effets micro-parcelles de décalage de la courbe sur les deux axes ce qui crée des biais sur le calcul de la variabilité de la pente. Ces effets de décalage sont ainsi minimisés par les translations.

- $x_{DGi}$  : distance du centre de gravité de la courbe considérée vers le centre de gravité des courbes EnvXVar selon l'axe x
- $y_{DGi}$  : distance du centre de gravité de la courbe considérée vers le centre de gravité des courbes EnvXVar selon l'axe y

$$\vec{V}_{x_{DGi},y_{DGi}} = x_{Gc} - x_{Gi}, y_{Gc} - y_{Gi}$$

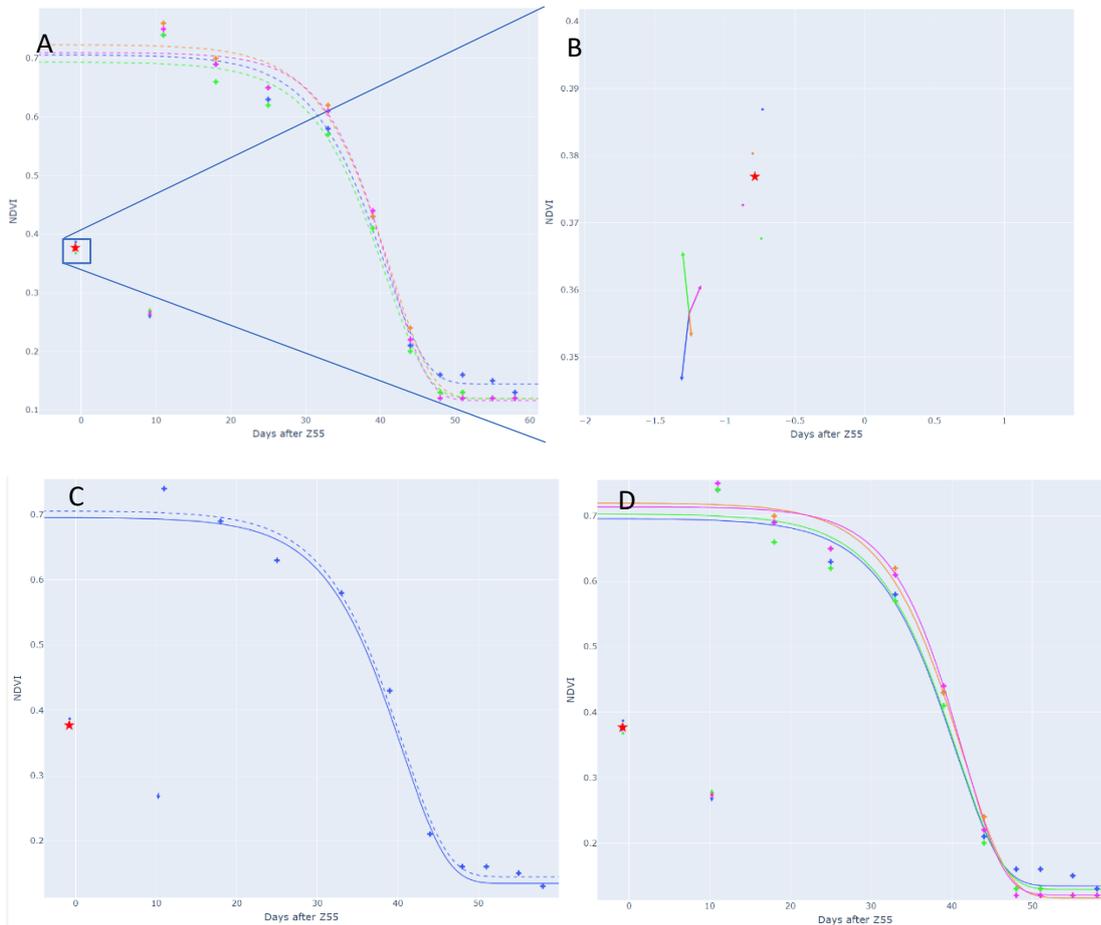


Figure 35 : Etapes de construction pour la translation des courbes vers leur centre de gravité commune EnvXVar

- A : Représentation des 4 réplicas pour une condition EnvXVar donnée (ici VLB 2023 modalité D1 Variété Arkeos)  
 + : données brutes représentée dans l'axe des x par le nombre de jours après la date de Z55 du génotype correspondant de la mesure effectuée et sur l'axe des y le NDVI mesurée. Les 4 couleurs représentent les 4 réplicas.  
 -- : Régressions de Gompertz réalisées et représentations des mesures de NDVI prédites en fonction du temps par rapport aux valeurs ajustées. La couleur des pointillées correspondent aux réplicas.  
 • : centres de gravités de chaque courbe avec pour couleur le réplica représenté  
 \* : centre de gravité des centres de gravités  
 → : vecteurs appliqués à chaque courbe pour la recentrer vers son centre de gravité commun.
- B : Zoom sur les centres de gravités avec les 4 vecteurs de translation (le réplica orange est invisible car son centre de gravité est confondu avec le centre de gravité commun).
- Les vecteurs montrent le sens de translation pour chaque réplica afin d'être recentré sur la moyenne des réplicas.
- C : Représentation de la translation effectuée entre la courbe brute du réplica bleu et sa courbe tradatée  
 -- : Courbe de la régression de Gompertz : valeurs brutes  
 - : Courbe de la régression de Gompertz : valeurs tradatées par leur vecteur de recentrage

On a ensuite tradaté les courbes par les vecteurs calculés pour chaque réplica (voir les étapes dans Figure 35, ci-dessus).

On a ensuite calculé une moyenne de ces courbes translatées en prenant en compte l'indice de fiabilité de chaque courbe (*fiab*). Le NDVI prédit pour une même courbe a donc été calculé de la manière suivante pour les 10.000 points de temps répartis entre -30 jours et 60 jours après le stade phénologique Z55, pour un même EnvXVar.

En partant de chaque NDVI, prédits et translatés comme noté auparavant :

On a la formule pour les 10000 points  $i$  prédits :

$$\text{Moyenne. EnvXVar. Trsl}_i = \sum_{n=1}^{n \rightarrow \text{nb.replicats.EnvXVar}} \text{NDVI}_n \cdot \text{Perdis. Trsl}_i \times \text{fiab}_n$$

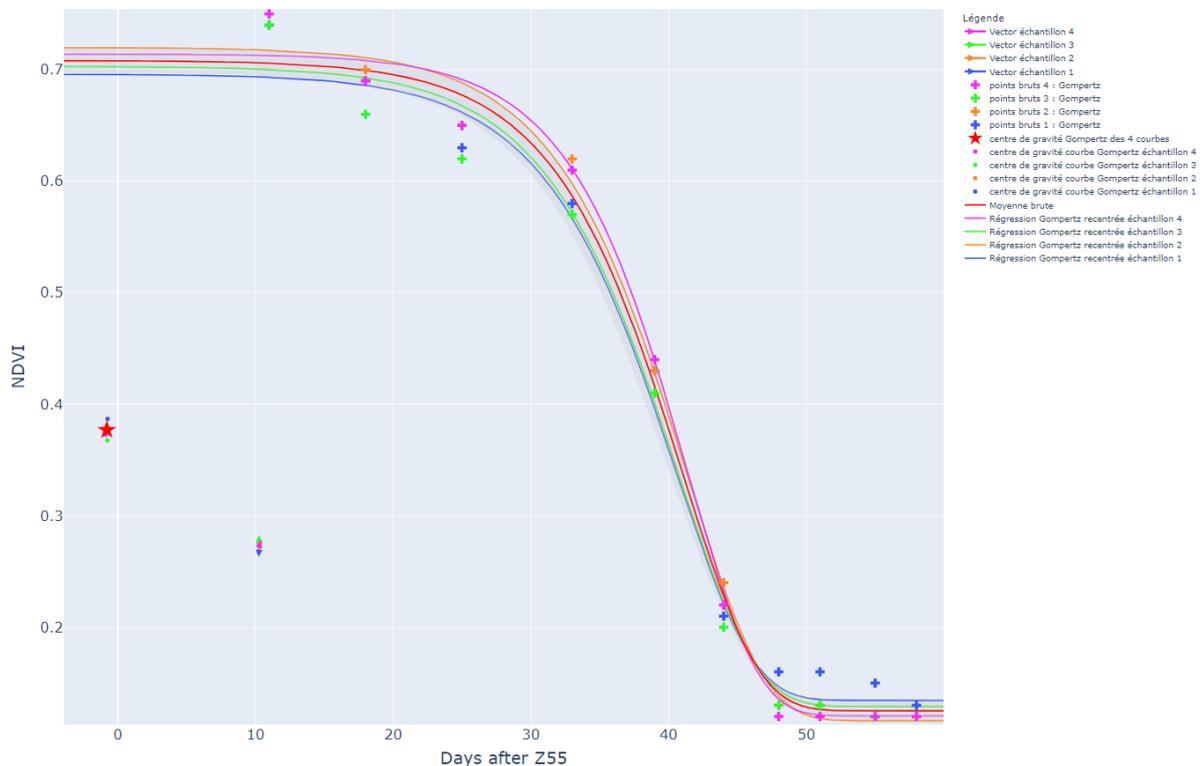


Figure 36 : Illustration de la moyenne translatée d'un même EnvXVar

En abscisse le nombre de jours après Z55 et en ordonnée le NDVI prédits. Moyenne (pondérée par la fiabilité *fiab* correspondant à la fiabilité du réplica au sein du EnvXVar représenté) des courbes translatées en rouge. En bleu, orange et vert et orange des 4 réplicas translatés.

On a ensuite récupéré la pente max de cette courbe translatée moyenne avec la formule suivante, en faisant varier  $i$  (un itérateur représentant les 10.000 points dans le temps générés entre -30 jours et 60 jours après le Z55 des variétés), de 500 à 9600<sup>4</sup>, on a donc :

Équation 4

$$f(i) = \frac{\text{Moyenne.EnvXVar.Trsl}_i - \text{Moyenne.EnvXVar.Trsl}_{i+10}}{\text{JoursAprèsZ55Prediction}_i - \text{JoursAprèsZ55Prediction}_{i+10}}$$

$$\text{ValeurPenteMax}_{\text{EnvXVar}} = \max_i f(i)$$

### 7 – 3 – B : Méthode de récupération des bornes

Afin de récupérer des bornes pour nos paramètres, nous établissons dans un premier temps des bornes autour des images des fonctions ajustées pour chaque réplica. Nous utiliserons les bornes créées autour des images des fonctions translattées pour obtenir des intervalles de confiance des pentes. Nous utiliserons, en revanche, les bornes créées autour des images brutes pour générer les intervalles de confiance autour des deux autres variables : le drop NDVI et la position de la sigmoïde dans le temps ou ‘Stay-Green’.

Ces bornes seront calculées selon l’axe du NDVI en prenant en compte :

- La moyenne pondérée pour cet axe le nombre de valeurs
- Les valeurs prédites de NDVI positionnés en-dessous ou au-dessus de la moyenne

Admettant, en effet, que nous ayons à la fois :

- a. Entre 18 et 40 points bruts pour un même ensemble EnvXVar
- b. Réduit la possibilité théorique de variabilité de nos valeurs en administrant la bonne fonction de régression sensée décrire réellement le phénomène observé pour l’ajustement de nos données

Nous cherchons à approximer l’intervalle de confiance autour de la prédiction du NDVI. Plus précisément nous nous intéressons à l’intervalle de confiance de l’image générée par la fonction en chaque point. Nous prendrons ici l’écart entre les images des différentes courbes d’ajustement pour chaque réplica, comme des indicateurs de l’écart-type définissant d’une loi normale à même de définir le risque d’approximation.

On calculera donc en chaque valeur de NDVI des 10.000 points générés sur l’axe des abscisses, la valeur de borne correspondante par le nombre de réplica au-dessus ou au-dessous de la moyenne en ce point-précis. Comme nous approximons par la loi normale l’intervalle de confiance, les valeurs suivantes ont alors été utilisé : Voir Simplification méthodologique 4.

---

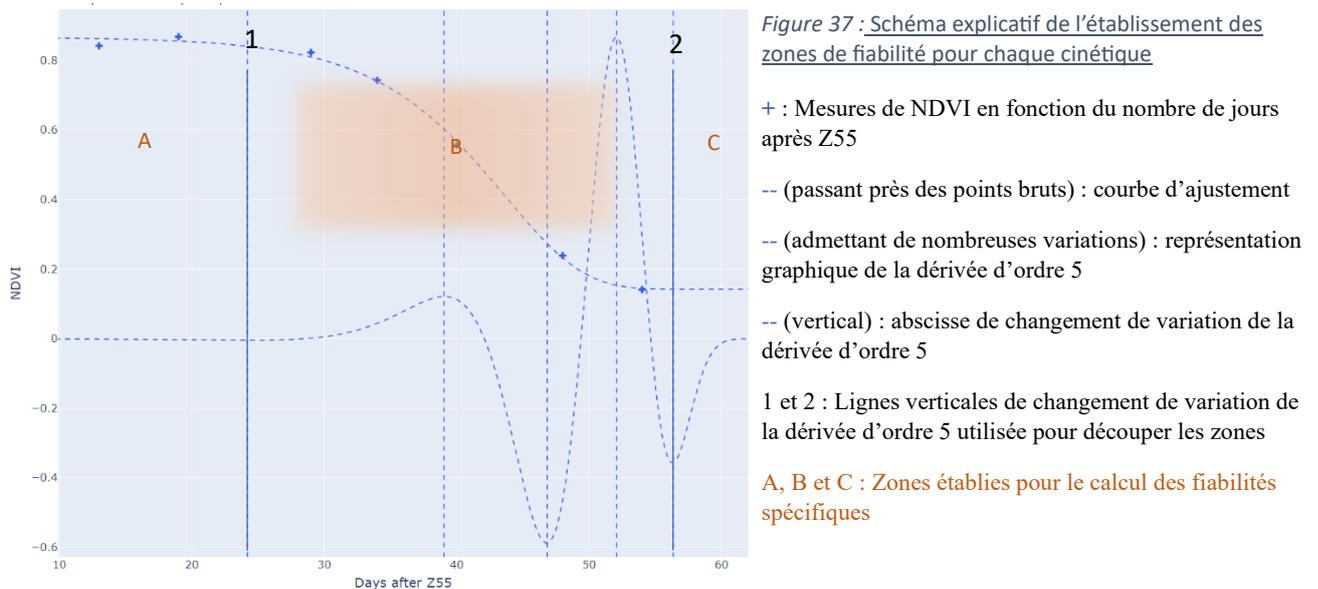
<sup>4</sup> On ne prend pas  $i$  entre 0 et 10.000 car les fonctions translattées peuvent admettre des valeurs inexistantes pour des valeurs de temps, proches des extrêmes (des bornes de calculs fixées pour le calcul des valeurs prédites). En effet, les courbes translattées l’ont été dans les deux dimensions. Voir *Figure 34* ‘calcul du centre de gravité’ ayant permis d’obtenir des décalages entre les courbes dans l’axe du NDVI mais également dans l’axe du temps.

La moyenne des bornes brutes sera calculée comme précédemment pour la moyenne des courbes translataées.

$$Moyenne.EnvXVar_i = \sum_{n=1}^{n \rightarrow nb.replicats.EnvXVar} NDVI_n.Predis_i \times fiab_n$$

Les bornes sont ensuite déterminées avec des zones découpées par les dérivées de la fonction de Gompertz<sup>5</sup>.

Chaque cinétique se verra assignée une fiabilité pour 3 zones, qui seront découpées à l'aide du premier et du dernier points de changements de directions de la dérivée cinquième.



Les zones 1 et 3 seront utilisées pour délimiter les zones de courbes liées aux asymptotes et la zone 2 sera utilisée comme zone liée à la pente. Ainsi, nous utiliserons les résultats de la matrice de covariance lors de l'utilisation de l'algorithme `curve_fit` décrivant la fiabilité de l'obtention des paramètres :

- Le paramètre d'asymptote haute pour la zone 1,
- Le paramètre de pente sera utilisé pour la zone 2
- Le paramètre d'asymptote basse pour la zone 3.

Plus ce paramètre est variable, moins il a été contraint lors de l'utilisation de l'algorithme et ainsi moins sa fiabilité est grande. Les valeurs de variances issues des matrices seront normalisées avant d'être utilisés dans les tenseurs spécifiques des zones.

### 7-3-B-a : Création des tenseurs

Nous créerons ainsi des tenseurs spécifiques pour chaque cinétique et chaque zone afin de calculer une borne. Pour ce faire, nous établissons déjà la partie spécifique du tenseur à la borne, à partir du paramètre de variation issue de la matrice de covariance obtenue lors de la régression.

<sup>5</sup> Voir annexes pour la formule des dérivées de la fonction de Gompertz

Soit D1 le premier décile et D9 le dernier décile on assigne pour valeur cible

- $target_{D1} = 0,7$
- $target_{D9} = 1,1$

On obtient ainsi la transformation linéaire suivante, pour une cinétique donnée i.

$$speczone = a * variab.param * b$$

Avec :

$$a = \frac{target_{D9} - target_{D1}}{D9 - D1}$$

$$b = target_{D1} - a \times D1$$

Le tenseur pour chaque zone prendra également compte la fiabilité globale de la fonction précédemment calculée (p16 'finalcomp') ces paramètres s'étagent de 13 à 32 pour les 174 régressions. Plus l'indice est faible, plus la confiance dans la régression est grande.

Nous choisissons ici 25 comme facteur limite entre une régression plus informative qu'un point brute et moins informative.

Ainsi, ces indices seront divisés par 25 puis multipliés pour chaque zone au terme décrivant la fiabilité spécifique de la zone calculée précédemment. Nous avons ainsi pour i une cinétique et zn une zone donné le tenseur suivant

$$tenseur_{i_{zn}} = (finalcomp_i / 25) \times speczone_{i_{zn}}$$

A présent, en chaque points choisis seront établis une borne supérieure et inférieure en fonction du nombre de réplica en-dessous et au-dessus de la moyenne. Les tenseurs, spécifiques du nombre de réplicas, sont les suivants. Il faut savoir qu'il ne peut y'avoir plus de 3 réplicas en dessous ou au-dessus de la moyenne en même temps. On formule les hypothèses suivantes pour le calcul des IC à partir d'un nombre d'échantillon de part et d'autre de la moyenne.

#### Simplification méthodologique 4

Nb Ech côté de la moyenne	Facteur de multiplication de l'écart-type univarié pour un IC de 95 % d'un côté de la moyenne	IC 90%
1	1.96	1.645
2	1.39	1.16
3	1.13	0.95

A chaque point, de chaque côté de la moyenne de l'intervalle considéré (en dessous ou au-dessus), on calcule ainsi deux intervalles avec comme entrées :

- L'écart-type univarié (la distance à la moyenne de ces échantillons)
- La moyenne des tenseurs affectée aux courbes-réplicas

- Le nombre d'échantillons impliqués

On applique ainsi la formule pour calculer la borne d'un côté, dessus ou dessous, d, de la moyenne, pour un x donnée. On récupère :

Les courbes d'ajustement de ce côté de la moyenne dans une liste : 'Ca.d'

- $y$  : l'image calculé par la fonction de Gompertz ajusté pour le réplica 'Car', en un x donné
- $\bar{y}$  : La moyenne des y des courbes d'ajustements considérées
- $\overline{tenseurs_{znx}}$  : La moyenne des tenseurs de la zone de la courbe pour un x correspondant des courbes d'ajustements considérées
- FMIC95 : le facteur de multiplication selon le nombre d'échantillons

$$y. \text{Borne. } d_x = y. \text{Moyenne. EnvXVar}_x \pm [|\bar{y}Ca. d_x - y. \text{Moyenne. EnvXVar}_x| \times \text{FMIC95}_{\dim(Ca.d_x)} \times \overline{tenseurs_{znx}}Ca. d]$$

*Simplification méthodologique 5*

Ces deux bornes brutes sont ensuite lissées avec une fonction de moyenne mobile exponentielle, utilisant les points amonts et en aval de chaque point lissé. On utilise pour paramètre alpha = 0.008.

On réalise cela sur les courbes ajustées brutes et translitées. On obtient ainsi deux bornes.

7 – 3 – C : Méthode de récupération des valeurs de seuils pour chaque variable

*7 – 3 – C – a : Récupération des valeurs de bornes pour la variable Drop NDVI*

Les bornes inférieures et supérieures autour de la valeur moyenne sont supposées équivalentes. On calcule d'abord une variance pondérée pour chaque groupe en utilisant les poids  $w_i$  utilisés précédemment (Équation 2, page. 58).

Avec

La moyenne pondérée calculée précédemment :  $\bar{x}_p$

*Équation 5*

$$\sigma_p^2 = \frac{\sum_i w_i \times (x_i - \bar{x}_p)^2}{\frac{n-1}{n} \times \sum_i w_i}$$

Une fois la variance calculée, on calcule l'écart-type qui a pour formule

*Équation 6*

$$\sigma_p = \sqrt{\sigma_p^2}$$

Deux paramètres de différences entre les intervalles de confiance précédemment calculés ont été utilisés. Celui entre la borne basse de l'intervalle de confiance et la borne haute au point 0, puis la même chose au point correspondant au 64 jours de la simulation.

Avec :

- On note le premier paramètre : `incertitude.assympt.haute`
- On note le second paramètre : `incertitude.assympt.basse`

On a ainsi pour le seuil bas du drop NDVI :

$$\text{SeuilBasDropNDVI}_{EnvXVar} = \text{ValeurDropNDVI}_{EnvXVar} - (\text{incertitude.assympt.haute} + \text{incertitude.assympt.basse} \times 2 + \frac{\sigma_p}{3})$$

$$\text{SeuilHautDropNDVI}_{EnvXVar} = \text{ValeurDropNDVI}_{EnvXVar} - (\text{incertitude.assympt.haute} \times 2 + \text{incertitude.assympt.basse} + \frac{\sigma_p}{3})$$

### 7-3-C-b : Récupération des bornes Stay-Green

On calcule comme précédemment une moyenne et un écart-type à partir des valeurs des paramètres issues des régressions. Les poids utilisés  $w_i$  sont ceux utilisés précédemment pour le calcul de la valeur moyenne du Stay-Green (Équation 3, page. 58). On applique ensuite ces deux équations (Équation 5 et Équation 6, page. 65 et page. 65) pour obtenir l'écart-type pondéré.

Un travail aura ensuite lieu sur les bornes graphiques générées précédemment pour récupérer les valeurs de seuils haut et bas de la variable de Stay-Green :

Nous récupérons les valeurs de l'image de la moyenne brute tracée pour respectivement la valeur de Stay-Green (moyenne) précédemment calculée moins une fois l'écart type pondéré divisé par 4 et plus une fois l'écart-type pondéré divisé par 4. Ces valeurs d'images sont des valeurs de NDVI.

Nous récupérons respectivement pour la première et la seconde image de la courbe moyenne brute, les images de la borne supérieure et inférieure. Nous obtenons ces images en cherchant la valeur d'image de la fonction la plus proche dans les 10.000 points calculés pour chacune de ces bornes.

Pour ces deux images ainsi trouvées, nous récupérons leur valeur d'abscisse qui seront les valeurs de seuil inférieurs et de seuil supérieurs pour la variable du Stay-Green ou de 'la durée en jours du Z55 à la mi pente'.

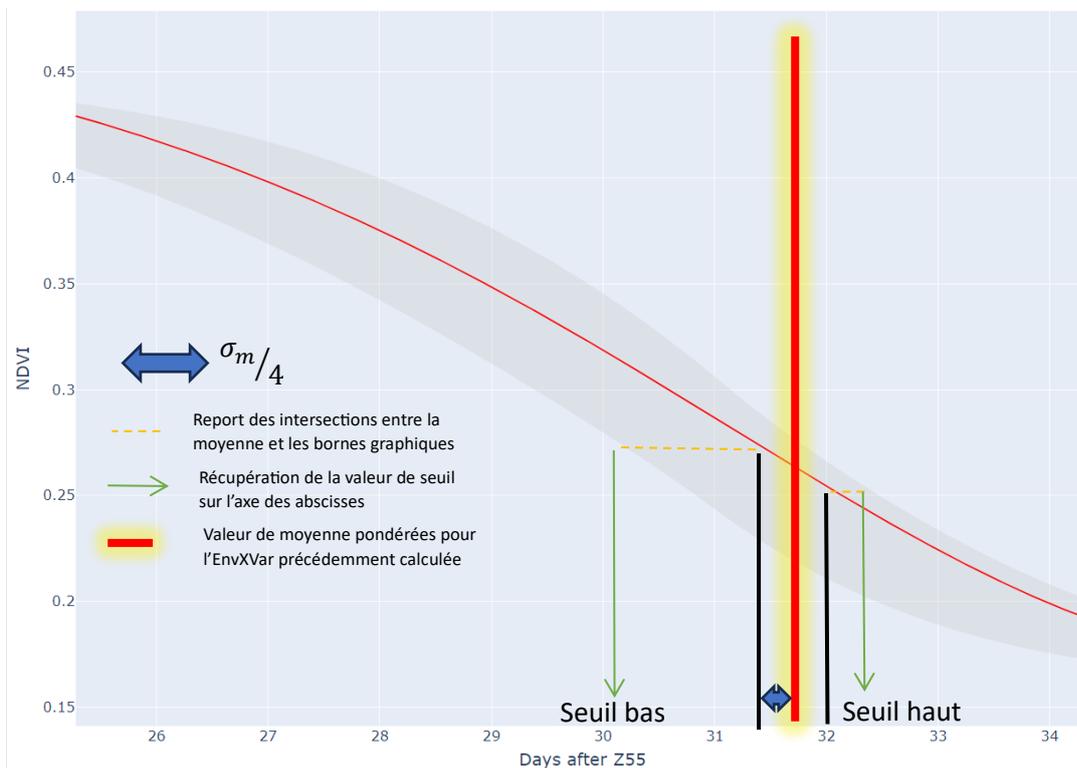


Figure 38 : Illustration de la méthode de récupération des seuils pour la variable de Stay-Green à partir de la 'valeur moyenne'

Illustration de la méthode de récupération des seuils pour la variable de Stay-Green à partir de la 'valeur moyenne' calculée précédemment. Illustration d'une cinétique de Gompertz d'un EnvXVar avec les bornes brutes. La ligne rouge correspond à la valeur moyenne de Stay-Green, la flèche bleue représente un écart de  $1/4$  de l'écart-type pondéré. La barre noire la valeur récupérée sur la moyenne de NDVI pour récupérer l'image correspondante la plus proche sur les bornes (ligne jaune pointillées). Ensuite la valeur de  $x$  de ces deux valeurs (flèches vertes) sont récupérées pour établir le seuil.  $1/4$  d'écart-type vers l'avant récupéré sur la borne supérieure correspond au seuil-haut,  $1/4$  d'écart-type vers l'arrière sur la borne inférieure correspond au seuil bas.

### 7 – 3 – C – d : Récupération des bornes de la force de la pente

Pour obtenir les bornes des pentes une régression a été faite entre la borne supérieure sur la partie asymptote haute de la courbe vers la borne inférieure de la partie asymptote basse de la courbe. Cette régression est sensée établir le seuil supérieur de la pente max. La logique inverse, la borne inférieure de la partie asymptote haute de la courbe et la borne supérieure de la partie asymptote basse de la courbe ont été utilisées dans une régression afin de connaître le seuil inférieur de la pente max.

#### Simplification méthodologique 6

Pour différencier les parties asymptotes hautes et asymptotes basses mais conserver une plus grande quantité de points déterminant les débuts des inflexions supérieures et inférieures, il a été décidé d'utiliser les points d'abscisses de changement de direction de la dérivée 3<sup>ème</sup>.

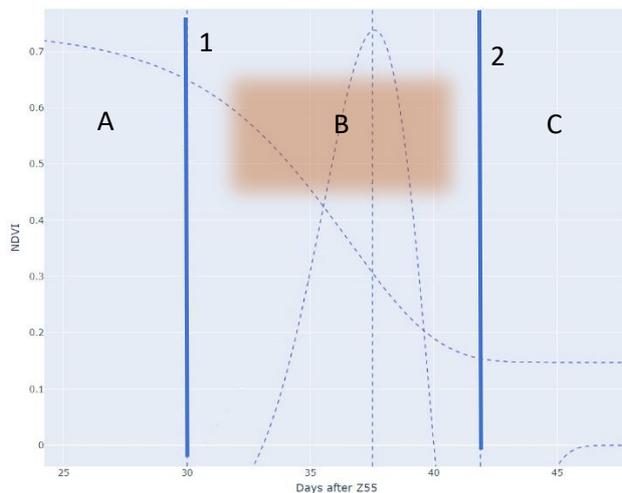


Figure 39 : Schéma explicatif de l'établissement des zones de fiabilité pour chaque cinétique

-- (passant près des points bruts) : courbe d'ajustement  
 -- (admettant de nombreuses variations) : représentation graphique de la dérivée d'ordre 3

-- (vertical) : abscisse de changement de variation de la dérivée d'ordre 3

1 et 2 : Lignes verticales de changement de variation de la dérivée d'ordre 3 utilisée pour découper les zones

A, B et C : Zones établies pour le calcul des fiabilités spécifiques

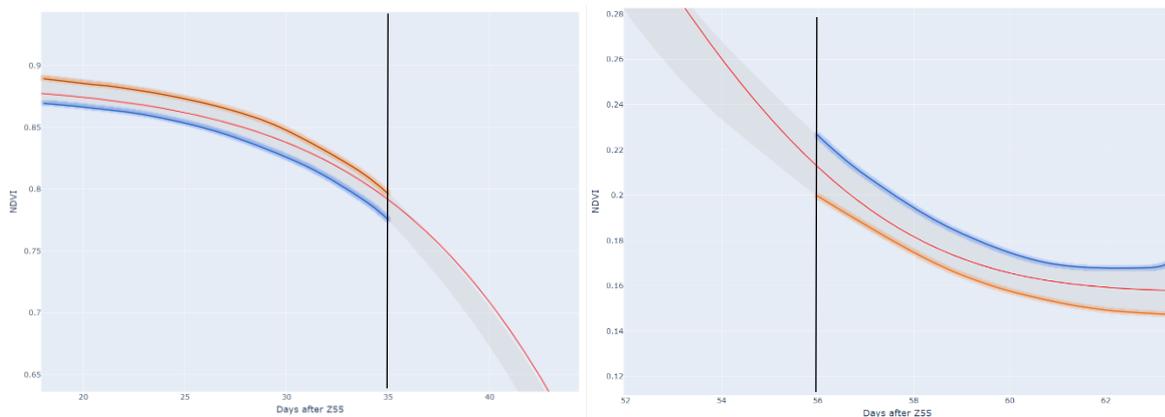


Figure 40 : illustration de la méthode de régression pour établir les seuils (supérieur et inférieur)

L'axe des ordonnées représente le NDVI, l'axe des abscisses le temps de la simulation. A : zoom sur la partie supérieure de l'inflexion de courbe, B : zoom sur la partie inférieure de l'inflexion de la courbe. 1 : lignes verticales de changement de direction de la dérivée 3<sup>ème</sup> servant de coupures pour l'utilisation des points pour réaliser les régressions.

Pour réaliser la régression sensée établir la simulation de pente maximale on utilise les points sur les bornes translattées et les parties dessinées en orange. Pour réaliser la régression sensée établir la simulation de pente minimale, on utilise les parties dessinées en bleu.

Les régressions simulant le seuil maximal de la variable de pente max et le seuil minimale de cette même variable une fois réalisées sont dérivées.

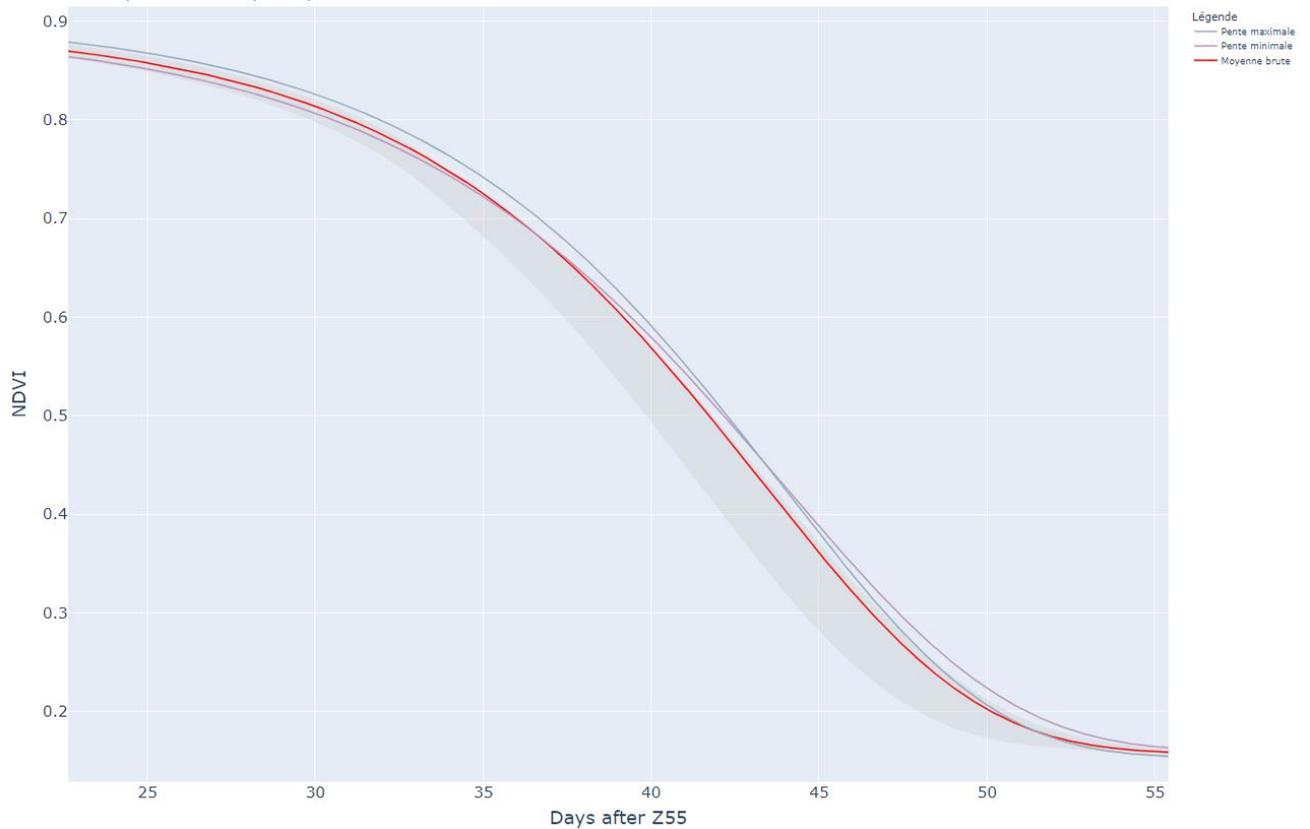


Figure 41 : *Illustration des simulations de pentes minimales et maximales*

La simulation de la pente maximale en bleu légèrement foncé et la simulation de la pente minimale en bordeaux sont ici représentées. En rouge la moyenne des images des courbes cinétiques translétées et en gris l'espace entre les bornes inférieures et supérieures calculées précédemment pour l'ensemble des réplicas translétés.

Les valeurs des dérivées maximums de ces deux simulations sont utilisées.

#### *Simplification méthodologique 7*

Malheureusement, ces valeurs ne peuvent être utilisées directement. En effet, les simplifications méthodologie (Simplification méthodologique 3, page. 59, la Simplification méthodologique 5, page. 65 et la Simplification méthodologique 6, page. 67) ont créé des biais dans l'obtention des bornes et dans l'obtention des simulations des pentes minimum et maximum qui ont utilisé ces bornes.

Le problème principal est que les valeurs de seuil haut de pente max de l'EnvXVar sont inférieures aux valeurs de pente moyenne précédemment obtenues. Pour pallier ce problème, nous avons décidé de réindexer la moyenne des seuils sur la valeur moyenne. Nous avons donc multiplié chaque seuil par le rapport de la moyenne des seuils avec la moyenne obtenue. L'équation peut s'écrire ainsi avec :  $ValeurPenteMax_{EnvXVar}$  pour la valeur moyenne de la courbe et  $SeuilHautPenteMax_{EnvXVar}$  et  $SeuilBasPenteMax_{EnvXVar}$  pour la nouvelle valeur de seuil calculé  $N$  et  $A$  l'ancienne (faisant référence au seuil bas ou haut selon le calcul effectué) :

$$N = A \times \frac{2 \times ValeurPenteMax_{EnvXVar}}{SeuilHautPenteMax_{EnvXVar} + SeuilBasPenteMax_{EnvXVar}}$$

On sélectionne ensuite toutes les valeurs de pentes seuil haut et seuil bas qui ne sont pas convenablement placés. C'est-à-dire que nous avons sélectionné toutes les EnvXVar dont les valeurs de pente seuil haut et pentes seuil bas ne sont pas respectivement placés au-dessus et en-dessous de la valeur moyenne.

Nous avons ensuite pris la décision que si la différence de seuil haut avec la moyenne excédée 0,01, la valeur n'était pas générée. On a ensuite recentré ces valeurs par la différence entre les seuils.

Pour les autres, on a soustrait la différence entre le seuil haut et la moyenne pour garder l'erreur informative d'un décalage des valeurs vers des valeurs faibles.

Enfin, comme on suppose les bornes représenter un intervalle de confiance à 95 % qu'un comportement de la loi réelle soit effectivement autour de la loi théorique, nous avons calculé l'encadrement théorique admis par ces nouvelles simulations de pentes.

La probabilité, (pour qu'une loi réelle ait en réalité pour paramètre d'asymptote haute la borne supérieure à 95 % de la loi théorique et la borne inférieure pour son asymptote basse de la loi théorique), est de  $0,05 \times 0,05$  ; soit : 0,0025. C'est-à-dire, qu'en supposant que la stochastique de la mesure et des résultats obtenus par les calculs précédemment réalisés soit normaux, la probabilité que le paramètre de pente maximum réelle soit réellement compris entre le seuil haut et bas ainsi simulé représente 99.75 % des probabilités.

On divise donc en suivant le z-score obtenu par la loi normale pour passer d'un encadrement de 99,75 % à 95 % de notre loi par 1,542.

7 – 3 – D : Transformation de la moyenne et des bornes pour la création de répliques de synthèse

L'ensemble des valeurs calculées étant censé représenter un intervalle de confiance de 95 % pour générer en plus de la moyenne deux répliques informatifs de l'écart-type on divise par 1.8 l'écart à la moyenne de tous les seuils précédemment calculés.

#### 7- 4 : Limites de la méthode 2

D'abord, la Simplification méthodologique 7 a réindexé de manière artificielle les simulations de pente maximum au-dessus de la moyenne. Cela était dû à :

- 3- La Simplification méthodologique 3 qui a généré une mauvaise translation des courbes pour minimiser les effets compensatoires de la variabilité de la pente entre les répliques.
- 4- La Simplification méthodologique 5 de lissage des bornes qui a fait passer la borne supérieure en-dessous de la moyenne créée

Les perspectives pour résoudre ces problèmes seraient :

- 6- D'utiliser un algorithme plus aboutit pour la création des centres de gravité en prenant en compte le point d'annulation de la dérivée première pour éviter un biais lié au choix de la plage d'abscisse de création des images des courbes.
- 7- De mieux comprendre les comportement univariés sur l'établissement d'un intervalle de confiance entre la moyenne et le réplica : Simplification méthodologique 4
- 8- D'améliorer le lissage des courbes (Simplification méthodologique 5) en prenant notamment en compte (dans l'algorithme) la moyenne en tous points de la génération des bornes pour contrôler l'écart avec la moyenne.

Par ailleurs, deux éléments pouvaient être réalisés pour améliorer la création des simulations une fois les bornes translattées générées en choisissant plus intelligemment les points utilisés pour les régressions destinés à simuler les pentes seuils max et min :

- 9- Améliorer par le choix des coupures destinées à faire passer les simulations de pente minimale et maximale entre les bornes : Voir Simplification méthodologique 6
- 10- En créant aléatoirement des points au-dessus et au-dessous des bornes sur des faibles plages d'intervalles. Puis :
  - a. Réaliser plusieurs régressions (pour chaque génération aléatoire de points) pour les deux simulations de pente maximale et minimale
  - b. Prendre la moyenne de la valeur maximale des dérivées (de chaque équation de simulation créée) pour établir le seuil de pente maximale et minimale

D'autres améliorations et d'autres méthodologies pourraient être envisagées.

## 7 – 5 : Classement des génotypes par cluster

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5
D1 2022 Mons	ADVISOR ARKEOS CHEVIGNON EXTASE	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN		ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2023 VLB	RUBISKO	ADVISOR ARKEOS CHEVIGNON		ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN	ARKEOS EXTASE OBIWAN
Normal 2022 Mons	ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS EXTASE RUBISKO
Normal 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
Normal 2023 Mons	OBIWAN	ARKEOS RUBISKO	ADVISOR	CHEVIGNON	EXTASE
Normal 2023 VLB	RUBISKO				ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN
Normal 2023 Thizay	ADVISOR CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN
X80 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		

Figure 42 Tableau de comparaison entre environnements des ensembles de variétés réalisant des différences significativement différentes classés du plus grand au plus petit pour le Drop NDVI

Ce tableau représente les ensembles de variété admettant des différences significatives entre les environnements (lignes). Les colonnes correspondent au nombre maximum d'ensembles formant des différences significatives sur un environnement pour la valeur de Drop NDVI. Les environnements ayant admis moins d'ensembles après le test de post-hoc ont été réalignés suivant la Simplification méthodologique 1, (page .19).

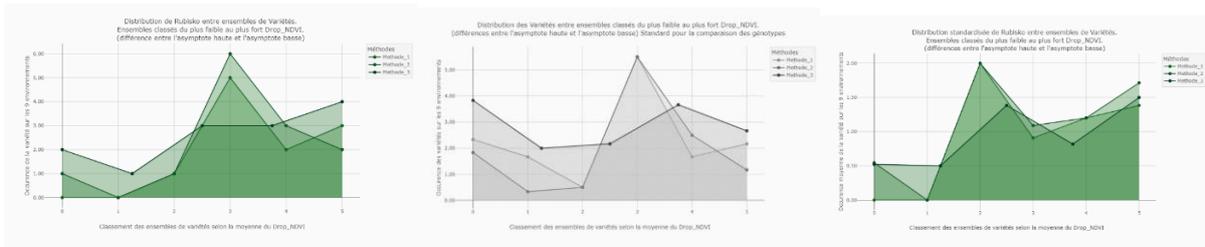


Figure 43 Graphique de densité d'apparition de la variété rubisko selon la position dans le classement des ensembles sur la valeur de Drop NDVI

Graphiques de densité d'apparition de la variété rubisko selon la position dans le classement des ensembles sur la valeur de Drop NDVI environnements avec moins d'ensemble que le nombre maximal. Les environnements ayant admis moins d'ensembles après le test de post-hoc ont été réaligné suivant la Simplification méthodologique 1, (page .19).

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5
D1 2022 Mons	A		ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2023 VLB	ARKEOS OBIWAN	RUBISKO		ADVISOR CHEVIGNON	EXTASE
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO				EXTASE
Normal 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
Normal 2023 Mons	ARKEOS OBIWAN	RUBISKO	CHEVIGNON EXTASE	ADVISOR CHEVIGNON EXTASE	ADVISOR EXTASE
Normal 2023 VLB	EXTASE OBIWAN	ARKEOS RUBISKO	ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON RUBISKO	ADVISOR CHEVIGNON
Normal 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			CHEVIGNON EXTASE
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO				EXTASE

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	Ensemble_6
D1 2022 Mons	B	ARKEOS CHEVIGNON	EXTASE RUBISKO	EXTASE OBIWAN RUBISKO	ADVISOR OBIWAN RUBISKO	ADVISOR OBIWAN
D1 2022 Thizay		ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO				ARKEOS
D1 2023 VLB	ARKEOS OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO			ADVISOR CHEVIGNON EXTASE RUBISKO	ADVISOR CHEVIGNON EXTASE
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE
Normal 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	EXTASE	ADVISOR
Normal 2023 VLB	ADVISOR ARKEOS OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
Normal 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO				CHEVIGNON	EXTASE
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE

Environnement	faible pente	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	forte pente
D1 2022 Mons	C	ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS EXTASE OBIWAN RUBISKO	ADVISOR EXTASE OBIWAN RUBISKO
D1 2022 Thizay		ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO				ARKEOS
D1 2023 VLB	ARKEOS OBIWAN	RUBISKO			ADVISOR CHEVIGNON	EXTASE
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE
Normal 2022 Thizay	ARKEOS EXTASE OBIWAN	ADVISOR ARKEOS EXTASE OBIWAN			ADVISOR ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON RUBISKO
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	EXTASE	ADVISOR
Normal 2023 VLB	EXTASE OBIWAN	ARKEOS CHEVIGNON RUBISKO			ADVISOR ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON
Normal 2023 Thizay	ADVISOR ARKEOS OBIWAN RUBISKO					CHEVIGNON EXTASE
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO					EXTASE

Figure 44 : Tableaux d'ensemble de variétés significativement différents sur leur valeur de pente Obiwan vs Advisor

Tableaux représentant les clusters issus d'un tukey-test, de la plus faible valeur de pente de sénescence à gauche, à la valeur de sénescence la plus forte à droite. Les cases noires représentent une absence de cluster, les cases blanches des clusters comprenant les deux variétés Obiwan et Advisor, les cases violettes la variété Obiwan sans Advisor et les cases jaunes l'inverse sans qu'il n'y ait sur la ligne une seule case comprenant les deux variétés (distinction des deux variétés significatives). Tableau A : réalisée sur la méthode de récupération des données numéro 1. Tableau B : réalisée sur la méthode de récupération des données numéro 2. Tableau C : combinaison des deux méthodes

	Stay-Green long			Stay-Green court	
Environnement	ensemble_1	ensemble_2	ensemble_3	ensemble_4	ensemble_5
D1 2022 Mons			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2022 Thizay	EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO
D1 2023 VLB	OBIWAN	ARKEOS		EXTASE RUBISKO	ADVISOR CHEVIGNON
Normal 2022 Mons	EXTASE	CHEVIGNON OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON
Normal 2022 Thizay	CHEVIGNON OBIWAN	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO
Normal 2023 Mons	EXTASE OBIWAN RUBISKO	CHEVIGNON			ADVISOR ARKEOS
Normal 2023 VLB	OBIWAN	EXTASE RUBISKO	CHEVIGNON EXTASE RUBISKO	CHEVIGNON EXTASE	ADVISOR ARKEOS
Normal 2023 Thizay	CHEVIGNON EXTASE	CHEVIGNON EXTASE OBIWAN		ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS OBIWAN RUBISKO
X80 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		

	Stay-Green long			Stay-Green court	
Environnement	ensemble_1	ensemble_2	ensemble_3	ensemble_4	ensemble_5
D1 2022 Mons	CHEVIGNON OBIWAN				ADVISOR ARKEOS EXTASE RUBISKO
D1 2022 Thizay	EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO
D1 2023 VLB	OBIWAN	ARKEOS		EXTASE RUBISKO	ADVISOR CHEVIGNON
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO
Normal 2022 Thizay	OBIWAN				ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO
Normal 2023 Mons	EXTASE OBIWAN RUBISKO	CHEVIGNON			ADVISOR ARKEOS
Normal 2023 VLB	OBIWAN	RUBISKO		CHEVIGNON EXTASE	ADVISOR ARKEOS
Normal 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
X80 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		

	Stay-Green long			Stay-Green court	
Environnement	ensemble_1	ensemble_2	ensemble_3	ensemble_4	ensemble_5
D1 2022 Mons	CHEVIGNON OBIWAN				ADVISOR ARKEOS EXTASE RUBISKO
D1 2022 Thizay	OBIWAN	ARKEOS CHEVIGNON EXTASE RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS CHEVIGNON RUBISKO
D1 2023 VLB	OBIWAN	ARKEOS	RUBISKO	EXTASE	ADVISOR CHEVIGNON
Normal 2022 Mons			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
Normal 2022 Thizay	OBIWAN	ADVISOR CHEVIGNON EXTASE RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS
Normal 2023 Mons	EXTASE OBIWAN	EXTASE OBIWAN RUBISKO	OBIWAN RUBISKO	CHEVIGNON	ADVISOR ARKEOS
Normal 2023 VLB	OBIWAN	RUBISKO		CHEVIGNON EXTASE	ADVISOR ARKEOS
Normal 2023 Thizay	CHEVIGNON EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS OBIWAN RUBISKO
X80 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		

Figure 45: Tableaux d'ensemble de variétés significativement différents sur leur valeur de Stay-green Obiwan vs Advisor

Même chose que la figure précédente mais sur la variable du stay-green. Advisor et Obiwan ont donc des qualités opposées. Obiwan a une pente de sénescence lente mais un bon Stay-Green tandis qu'Advisor a une forte pente de sénescence mais un mauvais Stay-green

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5
D1 2022 Mons			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2023 VLB	ARKEOS OBIWAN	RUBISKO		ADVISOR CHEVIGNON	EXTASE
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN				EXTASE
Normal 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
Normal 2023 Mons	ARKEOS OBIWAN	RUBISKO	CHEVIGNON EXTASE	ADVISOR CHEVIGNON EXTASE	ADVISOR EXTASE
Normal 2023 VLB	EXTASE OBIWAN	ARKEOS RUBISKO	ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON RUBISKO	ADVISOR CHEVIGNON
Normal 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN	ADVISOR ARKEOS CHEVIGNON EXTASE			CHEVIGNON EXTASE
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN				EXTASE

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	Ensemble_6
D1 2022 Mons	ARKEOS CHEVIGNON	EXTASE RUBISKO	EXTASE OBIWAN RUBISKO		ADVISOR OBIWAN RUBISKO	ADVISOR OBIWAN
D1 2022 Thizay	ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO					ARKEOS
D1 2023 VLB	ARKEOS OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON OBIWAN			ADVISOR CHEVIGNON EXTASE RUBISKO	ADVISOR CHEVIGNON EXTASE
Normal 2022 Mons	ADVISOR ARKEOS CHEVIGNON OBIWAN					EXTASE
Normal 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	EXTASE	ADVISOR
Normal 2023 VLB			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
Normal 2023 Thizay	ADVISOR ARKEOS OBIWAN				CHEVIGNON	EXTASE
X80 2023 Thizay	ADVISOR ARKEOS CHEVIGNON OBIWAN					EXTASE

	faible pente					forte pente
Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	Ensemble_6
D1 2022 Mons	ARKEOS CHEVIGNON	ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS EXTASE OBIWAN RUBISKO	ADVISOR EXTASE OBIWAN RUBISKO
D1 2022 Thizay	ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO					ARKEOS
D1 2023 VLB	ARKEOS OBIWAN	RUBISKO			ADVISOR CHEVIGNON	EXTASE
Normal 2022 Mons	CHEVIGNON OBIWAN RUBISKO					EXTASE
Normal 2022 Thizay	ARKEOS EXTASE OBIWAN	ADVISOR ARKEOS EXTASE OBIWAN			ADVISOR ARKEOS CHEVIGNON RUBISKO	CHEVIGNON RUBISKO
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	EXTASE	ADVISOR
Normal 2023 VLB	EXTASE OBIWAN	ARKEOS CHEVIGNON RUBISKO			ADVISOR ARKEOS CHEVIGNON RUBISKO	ADVISOR CHEVIGNON
Normal 2023 Thizay	ADVISOR ARKEOS OBIWAN RUBISKO					CHEVIGNON EXTASE
X80 2023 Thizay	CHEVIGNON OBIWAN RUBISKO					EXTASE

Figure 46 Tableau d'ensemble de variétés significativement différents sur leur valeur de pente Arkeos vs Extase

Sur ce tableau de comparaison sur la valeur de pente, on peut voir un exemple de deux variétés avec une très forte interférence génotype-X-environnement avec des différences significatives en faveur d'une pente plus forte pour Arkeos dans deux environnement 'D1 2022 Thizay' et 'Normal 2023 VLB' et pour les autres environnement (à l'exception de D1 2022 Mons et Normal 2022 Thizay qui n'ont pas pu établir de différences entre ces deux génotypes) montrent une pente plus forte pour Extase.

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	
D1 2022 Mons			ADV ISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
D1 2022 Thizay			ADV ISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
D1 2023 VLB	ARKEOS OBIWAN	RUBISKO		ADV ISOR CHEVIGNON	EXTASE	
Normal 2022 Mons	ADV ISOR ARKEOS CHEVIGNON OBIWAN RUBISKO				EXTASE	
Normal 2022 Thizay			ADV ISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
Normal 2023 Mons	ARKEOS OBIWAN	RUBISKO	CHEVIGNON EXTASE	ADV ISOR CHEVIGNON EXTASE	ADV ISOR EXTASE	
Normal 2023 VLB	EXTASE OBIWAN	ARKEOS RUBISKO	ARKEOS CHEVIGNON RUBISKO	ADV ISOR CHEVIGNON RUBISKO	ADV ISOR CHEVIGNON	
Normal 2023 Thizay	ADV ISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADV ISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			CHEVIGNON EXTASE	
X.80 2023 Thizay	ADV ISOR ARKEOS CHEVIGNON OBIWAN RUBISKO				EXTASE	
Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	Ensemble_6
D1 2022 Mons	ARKEOS CHEVIGNON	EXTASE RUBISKO	EXTASE OBIWAN RUBISKO		ADV ISOR OBIWAN RUBISKO	ADV ISOR OBIWAN
D1 2022 Thizay	ADV ISOR CHEVIGNON EXTASE OBIWAN RUBISKO					ARKEOS
D1 2023 VLB	ARKEOS OBIWAN RUBISKO	ADV ISOR ARKEOS CHEVIGNON OBIWAN			ADV ISOR CHEVIGNON EXTASE RUBISKO	ADV ISOR CHEVIGNON EXTASE
Normal 2022 Mons	ADV ISOR ARKEOS CHEVIGNON OBIWAN					EXTASE
Normal 2022 Thizay			ADV ISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
Normal 2023 Mons	OBIWAN	ARKEOS	RUBISKO	CHEVIGNON	EXTASE	ADV ISOR
Normal 2023 VLB			ADV ISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO			
Normal 2023 Thizay	ADV ISOR ARKEOS OBIWAN				CHEVIGNON	EXTASE
X.80 2023 Thizay	ADV ISOR ARKEOS CHEVIGNON OBIWAN					EXTASE

Figure 47: Tableau d'ensembles de variétés significativement différents sur leur valeur de pente Arkeos vs Chevignon (Méthode 1 et 2)

Valeurs d'Arkeos et de Chevignon pour la valeur de pente sur le tableau comparatif des variétés. Arkeos est en exemple de valeur faible de pente, tandis que Chevignon de valeur forte. Cas particulier de 'D1 2022 Thizay' où le classement de ces deux variétés s'inverse sur la valeur de pente.

Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	
D1 2022 Mons			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN			
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN			
D1 2023 VLB	EXTASE	ADVISOR CHEVIGNON		RUBISKO	ARKEOS OBIWAN	
Normal 2022 Mons	EXTASE	ARKEOS	ADVISOR CHEVIGNON	ADVISOR CHEVIGNON	OBIWAN RUBISKO	
Normal 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN			
Normal 2023 Mons	EXTASE	ADVISOR CHEVIGNON		RUBISKO	ARKEOS OBIWAN	
Normal 2023 VLB	ADVISOR ARKEOS CHEVIGNON	ADVISOR ARKEOS CHEVIGNON EXTASE		ARKEOS EXTASE OBIWAN RUBISKO	EXTASE OBIWAN RUBISKO	
Normal 2023 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN			
X80 2023 Thizay	EXTASE				ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	
Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	Ensemble_6
D1 2022 Mons	ADVISOR OBIWAN RUBISKO	EXTASE				ARKEOS CHEVIGNON
D1 2022 Thizay	ADVISOR ARKEOS	ADVISOR ARKEOS EXTASE	ADVISOR EXTASE RUBISKO		CHEVIGNON EXTASE OBIWAN RUBISKO	CHEVIGNON OBIWAN RUBISKO
D1 2023 VLB	EXTASE	ADVISOR CHEVIGNON			RUBISKO	ARKEOS OBIWAN
Normal 2022 Mons	EXTASE					ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO
Normal 2022 Thizay	CHEVIGNON	ADVISOR EXTASE OBIWAN RUBISKO			ADVISOR ARKEOS EXTASE OBIWAN RUBISKO	ARKEOS OBIWAN
Normal 2023 Mons	EXTASE	ADVISOR	CHEVIGNON	RUBISKO	ARKEOS	OBIWAN
Normal 2023 VLB	ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO				ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO	ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO
Normal 2023 Thizay	EXTASE				CHEVIGNON	ADVISOR ARKEOS OBIWAN
X80 2023 Thizay	EXTASE	CHEVIGNON			ARKEOS OBIWAN RUBISKO	ADVISOR
Categorie	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5	Ensemble_6
D1 2022 Mons	ADVISOR ARKEOS EXTASE OBIWAN RUBISKO	ADVISOR ARKEOS CHEVIGNON EXTASE				ARKEOS CHEVIGNON
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN			
D1 2023 VLB	EXTASE	ADVISOR CHEVIGNON	RUBISKO		ARKEOS	OBIWAN
Normal 2022 Mons	EXTASE	ARKEOS CHEVIGNON	ADVISOR ARKEOS CHEVIGNON	ADVISOR CHEVIGNON	ADVISOR CHEVIGNON OBIWAN RUBISKO	OBIWAN RUBISKO
Normal 2022 Thizay	CHEVIGNON	ADVISOR EXTASE OBIWAN RUBISKO				ARKEOS
Normal 2023 Mons	EXTASE	ADVISOR	CHEVIGNON	RUBISKO	ARKEOS	OBIWAN
Normal 2023 VLB	ADVISOR CHEVIGNON	ADVISOR ARKEOS CHEVIGNON	ARKEOS CHEVIGNON	ARKEOS EXTASE RUBISKO	EXTASE OBIWAN RUBISKO	EXTASE OBIWAN
Normal 2023 Thizay	CHEVIGNON EXTASE					ADVISOR ARKEOS OBIWAN
X80 2023 Thizay	EXTASE	ARKEOS CHEVIGNON OBIWAN RUBISKO			ADVISOR ARKEOS CHEVIGNON OBIWAN	ADVISOR ARKEOS OBIWAN RUBISKO

Figure 48 Tableau d'ensemble de variétés significativement différents sur leur valeur de score de pente ajustée Arkeos vs Chevignon (méthode 1, 2 et 3)

L'idée d'avoir pris comme exemple Chevignon en tant que valeur de pente haute et Arkeos en tant que valeurs de pente basse se confirme car 4 environnements par rapport à la figure précédent discrimine dans ce sens pour la méthode 3 et aucun environnement discrimine dans l'autre sens.





Environnement	Ensemble_1	Ensemble_2	Ensemble_3	Ensemble_4	Ensemble_5
D1 2022 Mons	CHEVIGNON OBIWAN	ARKEOS CHEVIGNON OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS EXTASE RUBISKO
D1 2022 Thizay			ADVISOR ARKEOS CHEVIGNON EXTASE OBIWAN RUBISKO		
D1 2023 VLB	EXTASE	ADVISOR CHEVIGNON OBIWAN			ARKEOS RUBISKO
Normal 2022 Mons	EXTASE				ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO
Normal 2022 Thizay	CHEVIGNON OBIWAN RUBISKO	ADVISOR CHEVIGNON EXTASE OBIWAN RUBISKO		ADVISOR EXTASE OBIWAN RUBISKO	ARKEOS
Normal 2023 Mons	EXTASE	CHEVIGNON RUBISKO	ADVISOR	OBIWAN	ARKEOS
Normal 2023 VLB	OBIWAN	ADVISOR CHEVIGNON RUBISKO		ADVISOR ARKEOS CHEVIGNON EXTASE RUBISKO	ADVISOR ARKEOS EXTASE
Normal 2023 Thizay	CHEVIGNON EXTASE				ADVISOR ARKEOS OBIWAN RUBISKO
X80 2023 Thizay	EXTASE	ARKEOS CHEVIGNON OBIWAN RUBISKO		ADVISOR ARKEOS CHEVIGNON OBIWAN RUBISKO	ADVISOR ARKEOS OBIWAN RUBISKO

Figure 49 Tableaux de comparaison avec apparition pour 3 variétés et 3 méthodes d'ensembles des variétés significativement différents pour la valeur de score classé du plus fort au plus faible

Exemple de 3 variétés Arkeos, Chevignon et Extase mis en lumière sur les 3 méthodes. Arkeos et Extase semble plus facilement classifiable entre environnements. Arkeos semble avoir globalement un mauvais score par rapport aux autres variétés, tandis qu'Extase semble plutôt avoir un bon score. Il est difficile de comprendre le comportement de Chevignon mais la méthode 3 semble la plus efficace pour distinguer les variétés parmi les 3 méthodes utilisées pour générer les tableaux de comparaisons.