



ISIMA – INSTITUT SUPÉRIEUR D’INFORMATIQUE, DE MODÉLISATION ET DE LEURS APPLICATIONS
RUE DE LA CHEBARDE, 63178 AUBIÈRE

INRAE (LISC ET UMR SILVA)
9, AVENUE BLAISE PASCAL, 63178 AUBIÈRE
CENTRE INRAE GRAND EST, SITE DE NANCY, RUE D’AMANCE, 54280 CHAMPENOUX

Développement d’un outil d’aide à la gestion forestière multi-critères dans le contexte du changement climatique

Rapport du stage ZZ2 - F4

Élève:
Loubna TALEB

Tuteurs:
Marion JOURDAN
Jean-Baptiste PICHANCOURT
Jean-Denis MATHIAS

Référent ISIMA:
Frédéric RENAUT

Soutenue le MARDI 02 JUILLET 2024

REMERCIEMENTS

À l'issue de ce stage, je tiens à remercier mes tuteurs pour leur accompagnement, leur patience et leur bienveillance.

Merci à Marion pour tes explications claires lors de nos réunions, le temps que tu as pris pour tes explications détaillées.

Merci à Jean-Baptiste pour le temps que tu m'as consacré, ton aide précieuse et tes explications enrichissantes chaque fois que j'en avais besoin.

Merci à Jean Denis pour tes conseils avisés, ton soutien constant et tes encouragements.

Merci à toute l'équipe de LISC. Nos discussions pendant les pauses café ont été enrichissantes et bénéfiques pour mon développement personnel et ma réflexion.

Enfin, merci à mon tuteur académique Frédéric pour votre accompagnement régulier toutes les deux semaines. Votre soutien a été vraiment bénéfique et enrichissant pour moi.

Résumé

La gestion durable des forêts dans le contexte du changement climatique est un défi complexe en raison des changements écologiques imprévisibles, des événements climatiques extrêmes et des limitations technologiques. Cette étude vise à relever ces défis en développant un **outil d'aide à la décision** pour la gestion forestière, garantissant le maintien des **services écosystémiques** fournis par les forêts pour les générations futures.

Le projet s'appuie sur les travaux antérieurs de Michelle Malara (2023) et consiste à coupler la **théorie du la viabilité** avec la **modélisation mécaniste** de la dynamique forestière à long terme. Dans cette étude, j'ai travaillé sur un modèle mécaniste individu-centré, **ForCEEPS**, implémenté en **Java** pour simuler la dynamique forestière, en mettant l'accent sur la **biodiversité** comme nouvel aspect de gestion. L'algorithme génétique a été redéfini pour intégrer l'espace d'exploration élargi des individus viables. Les résultats orientent les gestionnaires forestiers dans le choix des stratégies de gestion les plus adaptées en fonction des données climatiques et de l'évolution de l'état des forêts, assurant ainsi une gestion durable des forêts respectant à la fois les contraintes écologiques et économiques.

À ce jour, l'algorithme génétique a réussi à fournir des **itinéraires sylvicoles viables** et non viables, malgré la haute dimensionnalité du problème posé. J'ai ensuite réalisé une analyse de ces itinéraires pour caractériser les forêts. Il reste à établir le lien entre les itinéraires sylvicoles fournis par l'algorithme génétique et l'état des forêts pour faciliter la compréhension des gestionnaires forestiers.

Mots-clés : gestion durable, théorie du contrôle viable, algorithme génétique, ForCEEPS, Java, changement climatique.

Abstract

Sustainable forest management in the context of climate change is a complex challenge due to unpredictable ecological changes, extreme climatic events, and technological limitations. This study aims to address these challenges by developing a **decision support tool** for forest management, ensuring the maintenance of the **ecosystem services** provided by forests for future generations.

The project builds on the previous work of Michelle Malara (2023) and involves coupling **viable theory** with the **mechanistic modeling** of long-term forest dynamics. In this study, I worked on an individual-based mechanistic model, **ForCEEPS**, implemented in **Java** to simulate forest dynamics, focusing on **biodiversity** as a new management aspect. The genetic algorithm was redefined to integrate the expanded exploration space of viable individuals. The results guide forest managers in choosing the most suitable management strategies based on climatic data and the evolving state of forests, thus ensuring sustainable forest management that respects both ecological and economic constraints.

To date, the genetic algorithm has successfully provided **viable and non-viable silvicultural pathways**, despite the high dimensionality of the posed problem. I then conducted an analysis of these pathways to characterize the forests. The next step is to establish the link between the silvicultural pathways provided by the genetic algorithm and the state of the forests to facilitate understanding for forest managers.

Keywords: sustainable management, viable control theory, genetic algorithm, ForCEEPS, Java, climate change.

Contents

1	Introduction	0
1.1	Présentation de l'entreprise	0
1.2	Contexte scientifique et problématique	0
1.3	Objectif du stage	1
1.4	Plan du rapport	1
2	Le modèle de trouées ForCEEPS	2
2.1	Introduction	2
2.2	Modèle environnemental	2
2.2.1	Données site	3
2.2.2	Variables thermiques	3
2.2.3	Bilan hydrique	3
2.3	Modèle des plantes	4
2.3.1	Établissement des semis	4
2.3.2	Croissance des arbres	5
2.3.2.1	Croissance optimale	5
2.3.2.2	Croissance réelle	5
2.3.3	Mortalité des arbres	5
2.4	Variables de sortie d'une simulation en ForCEEPS	6
2.4.1	Variables à l'échelle individuelle	6
2.4.2	Variables à l'échelle populationnelle	7
3	Application de la théorie de la viabilité à la gestion sylvicole : Introduction de la diversité des espèces dans les pratiques de gestion	8
3.1	La théorie de la viabilité	8
3.1.1	Optimisation avec contraintes vs théorie de viabilité	8
3.2	Poser le problème de contrôle pour la gestion sylvicole	9
3.2.1	Système étudié	9
3.2.2	Contrôle : Gestion sylvicole	9
3.2.2.1	Variables de contrôle	9
3.2.2.2	Caractéristiques de l'intervention	10
3.2.3	Les contraintes écologiques et économiques	12
3.3	Résumé du problème de contrôle pour la gestion forestière	14
4	Méthode de résolution du problème de viabilité pour la gestion forestière	15
4.1	Résolution par algorithme génétique	15
4.2	Évaluation de la Combinatoire du Problème	15
4.3	Une simulation en ForCEEPS	17
4.4	Définition d'un individu en GA	17
4.5	Opérations réalisées pour une population en GA	18
4.5.1	Création d'un individu aléatoire	18
4.5.2	Calcul de la fitness	18
4.5.3	Nettoyage de la population	19
4.5.4	Sélection	19
4.5.5	Croisement des parents	20
4.5.6	Mutation après croisement	20
4.5.7	Exploitation du meilleur	20
4.5.8	Ajout d'individus aléatoires	21
4.5.9	Critère d'arrêt de l'algorithme génétique	21

4.6	Parallélisation des étapes de l'algorithme génétique	21
4.7	Fixation des paramètres de l'algorithme génétique	23
4.8	Analyse de sensibilité pour le nombre de parcelles	23
4.9	Méthodes pour analyser les résultats	23
4.9.1	Analyse en Composantes Principales (ACP)	24
4.9.2	Classification Hiérarchique Agglomérative (CHA)	24
4.9.3	Classification K-Means	24
5	État de la forêt	26
5.1	Définition des variables d'état de la forêt	26
5.2	Méthode 1: Le contrôle et l'état de la forêt	27
5.2.1	Résolution par moindres carrés	28
5.2.2	Régression Linéaire avec Ridge	28
5.3	Méthode 2 : Trouver la relation avec l'ACP et K-means	29
5.3.1	ACP et K-means	29
5.3.2	Relation entre l'état de la forêt et le contrôle appliqué	29
5.4	L'effet du controle sur l'Etat et la viabilité de système	30
5.4.1	Matrice de Transition	30
5.4.2	L'effet sur la non-viabilité : $V(E(t + \theta_s - 1)) = f(V(E(t_c - 1)), u(t_c))$	31
5.4.3	L'effet du controle sur l'Etat $E(t) : E(t + \theta_s - 1) = f(E(t_c - 1), u(t_c))$	31
6	Résultats et analyse	32
6.1	Choix de données	32
6.2	Données d'entrée et configuration des simulations	32
6.3	Choix du nombre de parcelles	32
6.4	Analyse de sensibilité pour l'algorithme génétique	34
6.5	Résultats de l'ensemble des individus viables	36
6.6	Résultats de l'ACP et La classification	36
6.6.1	Matrice de corrélation	36
6.6.2	Analyse en composantes principales	37
6.6.3	Analyse Hiérarchique Ascendante et K-means	38
6.6.4	Analyse des résultats de l'ACP et de la classification basée sur les contrôles effectués	39
6.7	État des forêts et le contrôle appliqué	43
6.7.1	Résultats des moindres carrés	43
6.7.2	Résultats de K-means pour lier état et contrôle	45
6.7.3	Résultats de la matrice de transition : l'effet sur la viabilité	46
6.7.4	Résultats des matrices de transition : l'effet sur l'état de la forêt	47
6.7.4.1	Surface terrière	47
6.7.4.2	Pielou	48
6.7.4.3	Gini	49
7	Discussion	50
7.1	Objectifs de la composition sylvicole	50
7.1.1	Intégration de la composition de contrôle dans la gestion sylvicole	50
7.1.2	Exploitation du meilleur	50
7.1.3	Parallélisation	50
7.2	Résultats de la classification et des itinéraires de gestion	51
7.3	Discussion méthodologique	51
8	Conclusion	53
9	Annexes	54
9.1	Modèle de croissance	54
9.2	Illustration de la coupe	55

1 Introduction

1.1 Présentation de l'entreprise

L'INRAE (Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement) est un organisme public français de recherche, créé le 1er janvier 2020 par la fusion de l'INRA (Institut National de la Recherche Agronomique) et de l'IRSTEA (Institut National de Recherche en Sciences et Technologies pour l'Agriculture et l'Environnement). Sa mission est de produire des connaissances scientifiques et techniques dans les domaines de l'agriculture, de l'alimentation et de l'environnement, afin de répondre aux grands enjeux sociétaux tels que la sécurité alimentaire, le changement climatique, la gestion durable des ressources naturelles, la transition agroécologique et la protection de la biodiversité. En résumé, l'INRAE joue un rôle crucial dans le développement de solutions innovantes pour une agriculture durable, une alimentation saine et une gestion responsable des ressources naturelles.

Mon stage, à caractère interdisciplinaire, se déroule au sein de deux unités de recherche de l'INRAE :

- L'UMR SILVA à Nancy, dirigée par Damien Bonal. Cette unité regroupe des membres d'AgroParisTech, de l'INRAE et de l'Université de Lorraine, et mène des recherches pluridisciplinaires sur le bois, les arbres et les écosystèmes forestiers. Elle vise à répondre aux questions de la société, notamment celles des gestionnaires forestiers, concernant l'adaptation des écosystèmes forestiers aux changements globaux et les services qu'ils fournissent, tels que la production de bois et leur contribution à l'atténuation du changement climatique.
- L'UR LISC de l'INRAE à Clermont-Ferrand, dirigée par Jean-Denis Mathias. Cette unité est spécialisée dans l'étude et la modélisation des systèmes dynamiques complexes. L'UR LISC développe des approches quantitatives pour analyser et gérer durablement les systèmes socio-écologiques, qui comprennent les interactions entre les sociétés humaines et leur environnement naturel.

1.2 Contexte scientifique et problématique

Les écosystèmes forestiers subissent des effets néfastes croissants dus au changement climatique, impactant leur santé, leur biodiversité et leur capacité à fournir des services écosystémiques essentiels. Parmi ces effets, on observe une sécheresse accrue qui conduit à l'augmentation des incendies. Tout ceci est lié à l'augmentation de la température causée par une augmentation des gaz à effet de serre, ce qui induit la propagation des parasites et des maladies au sein des forêts. Le changement climatique entraîne également une perte de biodiversité, avec la disparition d'espèces végétales et animales adaptées à des conditions climatiques spécifiques. Cette perte de biodiversité fragilise les écosystèmes forestiers, réduisant leur résilience face aux perturbations.

Maintenir dans le temps l'approvisionnement de ces écosystèmes est donc un enjeu crucial, notamment pour la régulation du climat, la protection des sols et la production de bois. Il est donc important de tester de nouvelles pratiques de gestion forestière qui permettent de préserver la santé et la résilience des forêts face aux changements climatiques. En effet, Jourdan et al. [1] ont démontré en 2021 que la conversion d'un peuplement monospécifique vers un peuplement mixte pourrait maintenir une productivité forestière plus élevée à long terme, renforçant ainsi la résilience des forêts face aux perturbations climatiques.

Dans ce contexte, j'ai utilisé un modèle forestier mécaniste individu-centré, ForCEEPS, développé par Morin et al. en 2013 [2]. Ce modèle simule l'évolution de chaque arbre en interaction avec son environnement, prenant en compte des facteurs tels que la croissance, la mortalité, l'intervention humaine, la compétition lumineuse et les conditions climatiques. Il permet de représenter la complexité de la dynamique forestière et d'évaluer l'impact des interventions humaines sous divers scénarios climatiques futurs, tels que la coupe des arbres sur un pas de temps défini.

La problématique abordée concerne la recherche de stratégies sylvicoles viables qui prennent en considération le changement climatique tout en respectant les contraintes liées à la santé de la forêt et à la productivité du bois pour des aspects économiques. Pour répondre à cette problématique, nous nous sommes inspirés de la théorie

du contrôle viable développée par le mathématicien Jean-Pierre Aubin [3]. Cette approche vise à identifier des contrôles viables permettant de maintenir le système dynamique étudié dans un état satisfaisant, en respectant les contraintes préalablement définies pour l'évolution du système. Cette approche fonctionne bien lorsque le nombre de contraintes et de variables d'état est limité. Par exemple, en pratique, Deffuant et al. [4] ont montré que cette méthode peut être appliquée jusqu'à neuf variables du système étudié.

Le défi se pose lors de l'application à des modèles informatiques comportant un grand nombre de variables, tel que notre modèle ForCEEPS pour simuler l'évolution dynamique complexe de la forêt.

La recherche de stratégies sylvicoles viables est un processus computationnellement coûteux en raison de l'augmentation de la dimensionnalité, et il est réalisable uniquement pour des systèmes triviaux. En 2023, Tonda et al [5]. ont utilisé la programmation génétique en considérant le problème de contrôle comme un problème d'optimisation en grande dimension pour découvrir des lois de contrôle pour les problèmes de viabilité sous forme analytique. Cette méthode de découverte des itinéraires sylvicoles viables a été mise en œuvre avec succès lors des travaux antérieurs de Malara et al. en 2023 [6], abordant ce problème de contrôle avec trois types de variables de contrôle à rechercher : la période de coupe, la méthode de coupe et la quantité d'arbres à laisser après la coupe.

1.3 Objectif du stage

Sur la base des travaux de Malara et al., l'objectif de mon stage a été double: le premier est d'intégrer la dimension de la biodiversité dans le cadre du contrôle viable appliqué au modèle forestier individu-centré ForCEEPS. Plus spécifiquement, il s'agit d'ajouter à trois variables de contrôle présentées dans le paragraphe précédent, une quatrième variable de contrôle qui modélise la proportion des espèces laissées après chaque coupe forestière.

Mon deuxième objectif est d'analyser ces itinéraires sylvicoles durables afin de comprendre leur impact à long terme sur l'évolution des écosystèmes forestiers, en tenant compte des effets du changement climatique. Cette approche permettra aussi de caractériser les forêts en fonction des pratiques sylvicoles mises en œuvre.

1.4 Plan du rapport

Tout d'abord, je présenterai le modèle de trouée forestière ForCEEPS en détaillant les différentes entrées du modèle telles que les données du site étudié et les données climatiques. Ensuite, je décrirai les différentes dynamiques d'un arbre, en commençant par l'établissement des semis, la croissance d'un arbre et la mortalité des arbres. Enfin, je présenterai les variables de sortie à l'échelle individuelle et à l'échelle populationnelle.²

Ensuite, j'aborderai le problème de contrôle viable en introduisant le nouveau contrôle de la diversité des espèces. Je commencerai par une définition générale de la théorie de viabilité introduite par Aubin. Ensuite, je poserai le problème de viabilité pour la gestion forestière en définissant le système étudié, le contrôle, ses variables et ses caractéristiques qui expliquent la méthode adoptée pour couper les arbres. Je définirai ensuite les contraintes économiques et écologiques qui limitent l'espace de notre problème.³

Après avoir posé le problème, je définirai la méthode de résolution qui utilise les algorithmes génétiques pour trouver la gestion forestière. Je commencerai par définir la combinatoire du problème pour déterminer si nous pouvons résoudre le problème avec les algorithmes génétiques. Ensuite, je présenterai les différentes étapes des algorithmes génétiques et les analyses que j'ai effectuées pour fixer les paramètres des algorithmes génétiques.⁴

J'ai ensuite réalisé une analyse de données pour comprendre les itinéraires viables fournis par les algorithmes génétiques. J'ai commencé par une analyse en composantes principales (ACP) pour visualiser les données et réduire la dimensionnalité des paramètres à analyser. Ensuite, j'ai utilisé une classification ascendante hiérarchique (HAC) comparée à une classification K-means pour classer les forêts et observer les liens entre les variables des contraintes caractérisant la forêt et son évolution. Enfin, j'ai examiné l'évolution de ces paramètres au fil du temps pour comprendre l'impact des contrôles appliqués sur son évolution.⁶

Dans la partie discussion, j'ai analysé les implications des résultats pour les pratiques de gestion forestière, présenté les défis et les limites rencontrés, ainsi que les perspectives de recherche future nécessaires pour soutenir les gestionnaires dans la prise de décisions éclairées pour la gestion des forêts.⁷

2 Le modèle de trouées ForCEEPS

2.1 Introduction

ForCEEPS (Forest Community Ecology and Ecosystem Processes) est un modèle de dynamique forestière introduit en 2013 par Morin et al. [2], basé sur le modèle FORclim (Bugmann 1996, Didion et al. 2009) [7]. Ce modèle est basé sur les modèles JABOWA (Botkin et al. 1972). Il simule la dynamique et la succession forestières en prenant en compte les effets des contraintes abiotiques (climat et propriétés du sol) et biotiques (compétition pour la lumière) sur l'établissement, la croissance et la survie des arbres dans de petites parcelles de terrain.

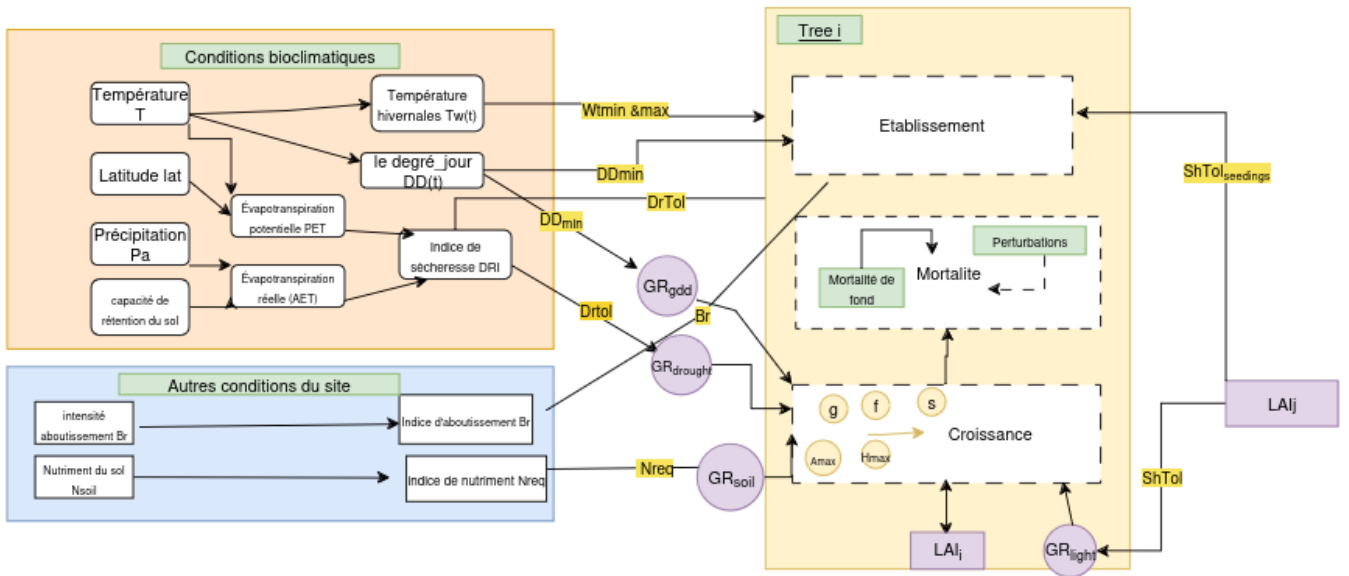


Figure 2.1: Schéma de la dynamique des processus écologiques modélisés dans ForCEEPS, représenté en me basant sur les annexes des travaux de Morin et al [2]. Les éléments sont codés par couleur : les cercles orange représentent les paramètres liés aux espèces et influençant la croissance, les écritures en jaunes représentent les paramètres de réponse aux facteurs de pilotage, et les cercles violets représentent les facteurs externes influençant la croissance et l'établissement. Les cadres noirs avec une flèche sortante indiquent les entrées de chaque équation provenant des conditions abiotiques. Les cadres avec une flèche entrante montrent les variables environnementales qui impactent les trois dynamiques principales du modèle ForCEEPS : établissement, croissance et mortalité, représentées par des cadres en pointillés. Les flèches vertes dans le cadre de la mortalité illustrent les différents effets qui influencent ce processus. Les différentes variables de réduction de croissance (GR_*) sont détaillées par la suite.

2.2 Modèle environnemental

Le modèle environnemental représente les contraintes extérieures qui influent sur l'évolution des arbres. En s'appuyant sur le cadre bleu des autres conditions du site et le cadre orange des conditions bioclimatiques, ces facteurs peuvent être regroupés en trois catégories : les caractéristiques du site étudié telles que l'azote, la latitude et l'intensité d'abrouissement ; ensuite, les variables thermiques basées sur les données climatiques comme les précipitations et la température ; enfin, le bilan hydrique, qui peut être analysé grâce à la capacité du sol à retenir l'eau, incluant également les données climatiques et de latitude.

2.2.1 Données site

Les caractéristiques du site peuvent être définies selon 3 paramètres. Tout d'abord, l'azote disponible dans le sol (**Nsoil**) joue un rôle crucial en caractérisant la fertilité du sol, ce qui influence directement la croissance des plantes. Ensuite, la latitude (**Lat**) est prise en compte pour calculer la durée des jours, un élément essentiel qui impacte le bilan hydrique et donc la disponibilité en eau pour les plantes. Enfin, l'intensité d'abrouissement (**BrI**) est un facteur à considérer, représentant l'impact des animaux se nourrissant de jeunes pousses, ce qui peut avoir des conséquences significatives sur la régénération des plantations forestières.

2.2.2 Variables thermiques

Les variables thermiques sont calculées à partir des données climatiques mensuelles :

- **Température hivernale (Tw)** : Moyenne des températures de décembre, janvier et février.

$$Tw(t) = \frac{T_{\text{décembre}}(t) + T_{\text{janvier}}(t) + T_{\text{février}}(t)}{3} \quad (2.1)$$

- **Degrés-jours (DD)** : Calculés comme suit :

$$DD(t) = \sum_{m=1}^{12} \max(0, T_m(t) - 5.5) \times 30.5 \times k_{\text{corr}} \quad (2.2)$$

- **Coefficient de correction (k_{corr})** : $k_{\text{corr}} = \begin{cases} 8.52 \times 10^{0.165 \times T_m} & \text{si } T_m \leq 5.5 \\ 187.2 \times 10^{0.0908 \times T_m} & \text{si } 5.5 < T_m \leq 15.5 \\ -31.8 + 2.377 \times T_m & \text{si } T_m > 15.5 \end{cases} \quad (2.3)$

Ces variables sont utilisées pour évaluer l'impact des températures sur la croissance et l'établissement des arbres.

2.2.3 Bilan hydrique

Le bilan hydrique évalue l'eau disponible pour les arbres en utilisant les précipitations mensuelles (W_m), les températures mensuelles (T_m) et la durée journalière (DL). Ces variables permettent d'évaluer l'évapotranspiration potentielle mensuelle (PET_m) et l'évaporation du sol (Dm). Elles permettent ensuite de quantifier les perturbations que subiront les arbres à cause du manque d'eau.

- **Indice de chaleur (HtIA, HtIB)** :

$$HtIA(t) = \sum_{m=1}^{12} \max(0, c_1 \times T_m(t))^{c_2} \quad (2.4)$$

$$HtIB(t) = c_3 \times HtIA(t)^3 + c_4 \times HtIA(t)^2 + c_5 \times HtIA(t) + c_6 \quad (2.5)$$

où $(c_i)_{i=1,\dots,6}$ sont des constantes.

- **Durée journalière (DL)** :

$$DL_m = c_A + c_B \times Lat \quad (2.6)$$

où c_A et c_B sont des constantes mensuelles.

- **Évapotranspiration potentielle mensuelle (PET)** :

$$PET_m(t) = c_p \times DL_m \times \left(\frac{10 \times HtIA(t)}{\max(0, T_m(t))} \right)^{HtIB(t)} \quad (2.7)$$

où c_p est une constante, souvent égale à 1.6.

- **Eau interceptée par le couvert végétal (Wout)** :

$$Wout_m(t) = \min(c_{ip} \times W_m(t), PET_m(t)) \quad (2.8)$$

où c_{ip} est une constante représentant la proportion moyenne d'eau interceptée avant d'arriver au sol.

- **Eau infiltrée dans le sol (Win)** :

$$Win_m(t) = W_m(t) - Wout_m(t) \quad (2.9)$$

- **Évaporation du sol (D) :**

$$D_m(t) = \left(1 + \frac{\log(F_w(t) + 1)}{20}\right) \times PET_m(t) - Wout_m(t) \quad (2.10)$$

où $F_w(t)$ est la quantité totale de feuillage sur la parcelle.

2.3 Modèle des plantes

ForCEEPS est un modèle individu-centré de dynamique du peuplement. Cela le distingue des autres modèles de trouées, qui supposent que tous les arbres de la même espèce et de la même époque se comportent de la même manière, dans l'intérêt de l'efficacité de la simulation. Cependant, ForCEEPS est entièrement basé sur les individus, ce qui pourrait permettre de prendre en compte la variabilité inter-individuelle. Le principal développement de ForCEEPS par Morin et al., par rapport au schéma classique des modèles de trouées, est la mise en œuvre d'un nouveau module de compétition pour la lumière où les longueurs de couronnes individuelles sont explicitement représentées dans l'espace vertical du couvert.

2.3.1 Établissement des semis

L'établissement des semis est modélisé en fonction des réponses spécifiques des espèces à l'indice de sécheresse (DrI), aux degrés-jours (GDD), à la température hivernale (TW), à l'intensité de broutage (Br), et à la disponibilité de la lumière (Lav). Cet établissement se fait avec un diamètre à hauteur de poitrine (DHP) de 1 cm. La probabilité de succès d'établissement (P_{est}^S) est calculée comme suit :

$$P_{est}^S = P_{TW} \times P_{GDD} \times P_{Dr} \times P_{Br} \times P_{LA} \times c_{est} \quad (2.11)$$

P_{TW} est la probabilité que le semis puisse s'établir en fonction de la température hivernale TW , et $WtminS$ et $WtmaxS$, respectivement les températures hivernales minimales et maximales mensuelles tolérées pour la régénération de l'espèce S , avec :

$$P_{TW} = \begin{cases} 1 & \text{si } WtminS \leq TW \leq WtmaxS \\ 0 & \text{autrement} \end{cases}$$

P_{GDD} est la probabilité que le semis puisse s'établir en fonction des GDD de l'année et de la somme minimale annuelle ou saisonnière des degrés-jours requis $DDminS$, l'établissement étant supposé impossible lorsque la somme annuelle des degrés-jours de croissance ne correspond pas aux exigences en degrés-jours de l'espèce arboricole :

$$P_{GDD} = \begin{cases} 1 & \text{si } GDD \geq DDminS \\ 0 & \text{si } GDD < DDminS \end{cases}$$

P_{Dr} est la probabilité que le semis puisse s'établir en fonction du stress hydrique de l'année et de la tolérance spécifique à la sécheresse de l'espèce $DrTolS$, avec :

$$P_{Dr} = \begin{cases} 1 & \text{si } DrI \leq DrTolS \\ 0 & \text{si } DrI > DrTolS \end{cases}$$

P_{Br} est la probabilité que le semis puisse s'établir en fonction de l'intensité de broutage du site Br et de la susceptibilité au broutage des semis de l'espèce S , déterminant ensemble la susceptibilité réelle des semis de l'espèce S sur le site $Bactual$, avec :

$$\begin{cases} \text{Si } BrS = 1, & Bactual = Br \times 4 \\ \text{Si } BrS = 2, & Bactual = Br \times 2 \\ \text{Si } BrS = 3, & Bactual = Br \\ \text{Si } BrS = 4, & Bactual = Br \times 0.5 \\ \text{Si } BrS = 5, & Bactual = Br \times 0.25 \end{cases}$$

Ensuite, P_{Br} est calculé comme suit :

$$P_{Br} = \begin{cases} 1 & \text{si } Bactual \leq k_{Br} \\ 0 & \text{si } Bactual > k_{Br} \end{cases}$$

où k_{Br} est un nombre aléatoire compris entre 0 et 1.

P_{LA} est la probabilité que le semis puisse s'établir en fonction de la disponibilité de la lumière au sol (Lav) et de la tolérance à l'ombre spécifique aux semis de l'espèce $ShTol_seedlingS$, avec :

$$P_{LA} = \begin{cases} 1 & \text{si } Lav \geq ShTol_seedlingS \\ 0 & \text{si } Lav < ShTol_seedlingS \end{cases}$$

où

$$Lav = e^{-k \times LA}$$

LA étant la surface foliaire cumulée de tous les arbres de la parcelle.

Le coefficient c_{est} est un facteur aléatoire compris entre 0 et 1 pour introduire de la stochastique dans la régénération des semis.

Enfin, l'établissement des semis d'une espèce donnée S est possible si

$$P_{est}^S \geq k_{est}$$

où k_{est} est un nombre aléatoire compris entre 0 et 1.

2.3.2 Croissance des arbres

2.3.2.1 Croissance optimale

La croissance annuelle des arbres est modélisée par l'incrément du diamètre à hauteur de poitrine (D). L'incrément de diamètre potentiel (ΔD_{opt}) est calculé comme suit (Moore 1989) :

$$\Delta D_{opt} = \frac{g_s}{D(1 - \frac{H}{H_{max}})^{2H_{max} - b_s \cdot e^{c_s \cdot D} \cdot (c_s \cdot D + 2)}} \quad (2.12)$$

où D est le diamètre à hauteur de poitrine, H est la hauteur de l'arbre, g_s est le taux de croissance maximal de l'espèce s , H_{max} est la hauteur maximale atteignable par l'espèce s , et b_s et c_s sont des paramètres spécifiques à l'espèce (avec $b_s = H_{max} - 137$; et $c_s = \frac{s_s}{b_s}$) ; s_s est un paramètre allométrique spécifique à l'espèce reliant la hauteur de l'arbre H et le diamètre D comme suit (Bugmann 1996) :

$$H = a + (H_{max} - a) \times (1 - e^{-\frac{s_s \cdot D}{H_{max} - a}})$$

avec $a = 1.37$ m (c'est-à-dire la hauteur de poitrine).

2.3.2.2 Croissance réelle

Ensuite, l'incrément de diamètre réalisé ΔD est calculé en modifiant ΔD_{opt} en fonction des facteurs de réduction de croissance abiotiques ou biotiques (tous les facteurs sont bornés entre 0 et 1) comme suit (Bugmann 1996, Didion et al. 2009) :

$$\Delta D = \Delta D_{opt} \times \sqrt[3]{GR_{light} \times GR_{gdd} \times GR_{drought} \times GR_{soil}} \quad (2.13)$$

où GR_{light} est le facteur de réduction de croissance lié à la disponibilité de la lumière pour l'arbre, GR_{gdd} est le facteur de réduction de croissance lié aux températures de la saison de croissance du site (GDD), $GR_{drought}$ est le facteur de réduction de croissance lié à l'indice de sécheresse (DrI), et GR_{soil} est le facteur de réduction de croissance lié au contenu en nutriments du sol (Nsoil). Le calcul exact de chaque facteur de réduction est détaillé dans l'annexe 9.

2.3.3 Mortalité des arbres

La mortalité des arbres est modélisée de façon partiellement stochastique car la probabilité de mortalité ($P_{mort,i}$) de l'arbre i dépend de deux composants : une mortalité de fond (P_0) et une mortalité liée à la croissance (P_g) :

$$P_{mort,i} = \max(P_0, P_g) \quad (2.14)$$

P_0 est calculée comme suit :

$$P_0 = \begin{cases} 0 & \text{si } m_0 < p_{dist} \\ 1 & \text{si } m_0 \geq p_{dist} \end{cases} \quad (2.15)$$

où m_0 est un nombre aléatoire compris entre 0 et 1, p_{dist} étant une constante représentant la probabilité de mortalité de fond ou de perturbation. La mortalité de fond P_0 est donc purement stochastique et dépend de la longévité maximale des espèces, et est mise en œuvre pour prendre en compte les événements de mortalité induits par des perturbations à petite échelle (par exemple, une attaque de pathogène en phase endémique), mais les perturbations à grande échelle (par exemple, les chablis, les incendies) peuvent être prises en compte en augmentant le taux de mortalité de fond.

P_g est la probabilité de mortalité liée au stress, calculée comme suit :

$$P_g = P_{\text{age}} + (1 - P_{\text{age}}) \times P_{\text{stress},i} \quad (2.16)$$

La probabilité de mortalité liée à l'âge (P_{age}) dépend de l'âge maximal de l'espèce A_{max}^S et peut être définie comme suit :

$$P_{\text{age}} = \frac{k_{\text{death}}}{A_{\text{max}}^S} \quad (2.17)$$

où k_{death} est une constante et A_{max}^S est l'âge maximum de l'espèce S .

La probabilité de mortalité liée au stress (P_{stress}) dépend de la croissance de l'arbre :

$$P_{\text{stress},i} = \begin{cases} p_{\text{slowgr}} & \text{si } \text{SlowGr}_i \geq k_{\text{SlowGrTime}} \\ 0 & \text{autrement} \end{cases} \quad (2.18)$$

où p_{slowgr} est une constante représentant la probabilité de mortalité pour les arbres stressés, SlowGr_i est un indice correspondant au nombre d'années consécutives pendant lesquelles la croissance de l'arbre i est inférieure au seuil t_{slowgr} , et $k_{\text{SlowGrTime}}$ est une constante représentant le nombre d'années de croissance lente après lequel l'arbre est considéré comme stressé.

ForCEEPS permet de calculer différentes variables à l'échelle individuelle pour voir l'évolution d'un arbre sur $[t_0; t_{\text{fin}}]$.

2.4 Variables de sortie d'une simulation en ForCEEPS

Une simulation en ForCEEPS, couvrant une période $[t_0 : t_{\text{fin}}]$, permet de calculer des variables forestières à l'échelle individuelle (arbre) et populationnelle. Ces calculs sont réalisés à partir des propriétés de tous les arbres de la placette modélisée, des données climatiques et des informations spécifiques au site étudié. Les mécanismes décrits précédemment permettent de déterminer les variables suivantes :

2.4.1 Variables à l'échelle individuelle

Les variables calculées pour chaque arbre sont :

- **Volume de l'arbre** : Le volume d'un arbre est calculé en additionnant le volume de la tige à 1,37 m, modélisée comme un cylindre, et celui de la couronne, modélisée comme un cône :

$$V_{\text{cylindre}}(t) = \pi \times \left(\frac{D_i(t)}{2} \right)^2 \times 1.37$$

$$V_{\text{cône}}(t) = \frac{\pi}{3} \times \left(\frac{D_i(t)}{2} \right)^2 \times (H(t) - 1.37)$$

D'où :

$$V_i(t) = V_{\text{cylindre}}(t) + V_{\text{cône}}(t)$$

- **Croissance réelle** : La croissance réelle est influencée par divers facteurs externes, comme détaillé en annexe.9
- **Surface foliaire (foliageArea) (t)** : La surface foliaire de l'arbre est définie en annexe.9
- **Disponibilité lumineuse (lightAvailability) (t)** : La lumière disponible pour chaque arbre est également détaillée en annexe 9 dans le modèle de croissance.
- **Surface terrière ($G_i(t)$)** : La surface terrière d'un arbre est la surface de la section transversale de cet arbre à une hauteur de 1,30 m. Elle est exprimée comme suit :

$$G_i(t) = \pi \times \left(\frac{D_i(t)}{2} \right)^2$$

- **Quantité de bois** ($B_i(t)$) : Estimée à partir du diamètre à hauteur de poitrine (1,3 m) :

$$B_i(t) = 0.12 \times D_i(t)^{2.4}$$

- **Poids du feuillage** ($F_{w_i}(t)$) : Dépend du diamètre de la tige à hauteur de poitrine D et de la taille de la couronne cs :

$$F_{w_i}(t) = f_s'' \times cs \times D_i(t)^{a_s}$$

2.4.2 Variables à l'échelle populationnelle

Les variables calculées pour la population d'arbres sont :

- **Population d'arbres** ($\text{pop}(t)$) : L'ensemble des arbres présents durant l'année $t + 1$ est calculé en ajoutant les nouveaux arbres établis $E(t)$ et en enlevant ceux qui sont morts $M(t)$:

$$\text{pop}(t + 1) = E(t) \cup \text{pop}(t) \setminus M(t)$$

- **Surface terrière** ($G(t)$) : La somme des surfaces des sections transversales des troncs d'arbres, mesurées à 1,3 m au-dessus du sol, est normalisée à un hectare en divisant par la surface échantillonnée S_u :

$$G(t) = \frac{\sum_{i \in \text{pop}(t)} G_i(t)}{S_u}$$

- **Volume total de bois** ($V(t)$) : Le volume total de bois est calculé à partir du volume de la couronne (cône) et du volume de la tige (cylindre à 1,37 m), pour chaque arbre $i \in \text{pop}(t)$, normalisé à un hectare :

$$V(t) = \frac{\sum_{i \in \text{pop}(t)} V_i(t)}{S_u}$$

- **Biomasse du bois totale** ($B(t)$) : La somme de la biomasse des arbres, calculée en fonction du diamètre de la tige à hauteur de poitrine D :

$$B(t) = \frac{\sum_{i \in \text{pop}(t)} B_i(t)}{S_u}$$

- **Poids total du feuillage** ($F_w(t)$) : La somme des poids du feuillage de tous les arbres :

$$F_w(t) = \frac{\sum_{i \in \text{pop}(t)} F_{w_i}(t)}{S_u}$$

Cette quantité est utilisée dans le calcul du bilan hydrique.

3 Application de la théorie de la viabilité à la gestion sylvicole : Introduction de la diversité des espèces dans les pratiques de gestion

3.1 La théorie de la viabilité

3.1.1 Optimisation avec contraintes vs théorie de viabilité

Dans le domaine de la recherche opérationnelle, deux approches principales sont couramment utilisées pour résoudre des problèmes complexes : l'optimisation avec contraintes et la théorie de la viabilité. L'optimisation avec contraintes vise à trouver le maximum ou le minimum d'une fonction objectif tout en respectant certaines contraintes sur les variables de décision. Par exemple, on peut chercher à maximiser le profit d'une entreprise ou à minimiser la production de bois dans une forêt, tout en respectant certaines contraintes. En revanche, la théorie de la viabilité ne cherche pas à optimiser une fonction objectif, mais plutôt à identifier toutes les solutions possibles qui permettent à un système de survivre ou de rester viable face à des perturbations. Cet ensemble de solutions viables est appelé le noyau de viabilité.

- **Optimisation avec contraintes :**

Trouver x qui minimise la fonction objectif $f(x)$ sous les contraintes :

$$\begin{cases} g_i(x) \leq 0 & \text{pour } i = 1, \dots, m, \\ h_j(x) = 0 & \text{pour } j = 1, \dots, p. \end{cases}$$

- **La théorie de la viabilité :**

Étant donné un système dynamique décrit par l'équation différentielle :

$$x(t+1) = f(x(t), u(t)),$$

où :

- $x \in \mathbb{R}^n$ est le vecteur d'état décrivant l'état courant du système,
- $u \in \mathbb{R}^p$ est le vecteur de contrôle contenant les variables qui peuvent être changées pour contrôler le système dynamique,
- $f : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$ décrit l'évolution du système au cours du temps.

L'espace de contraintes K détermine le noyau de viabilité, défini comme :

$$\text{Viab}(u(t)) = \{u(t) \in \mathbb{R}^p \mid x(t+1) = f(x(t), u(t)) \in K \text{ pour tout } t \geq 0\}.$$

Déterminer des stratégies de gestion sylvicole qui maintiennent la forêt dans un état stable sur le long terme, tout en respectant les contraintes imposées par les gestionnaires forestiers, constitue un défi dynamique. Ce défi nécessite une connaissance approfondie de l'évolution de la forêt au fil du temps afin de déterminer quel contrôle viable nous permet de maintenir l'évolution de la forêt dans l'espace des contraintes. C'est pour cette raison qu'il est judicieux de résoudre ce problème en utilisant la théorie de la viabilité. Cette théorie ne cherche pas un unique contrôle optimal, mais plutôt un ensemble de stratégies de gestion viables. Par conséquent, je vais par la suite formuler le problème de contrôle en détaillant le système dynamique, les variables de contrôle et les contraintes qui définissent les états souhaitables. [[8] , [9]]

3.2 Poser le problème de contrôle pour la gestion sylvicole

3.2.1 Système étudié

Soit $\text{pop}(t) = 1, 2, \dots, N(t)$ l'ensemble des arbres présents dans la forêt à l'instant t , où $N(t)$ représente le nombre total d'arbres. L'état du système à l'instant t est décrit par le vecteur d'état $\mathbf{x}(t) = [\mathbf{D}(t) \ \mathbf{c}(t)]$, où :

- $\mathbf{D}(t) = [D_i(t)]$ $i \in \text{pop}(t)$ est le vecteur des diamètres des arbres,
- $\mathbf{c}(t) = [c_i(t)]$ $i \in \text{pop}(t)$ est le vecteur des proportions de la couronne des arbres.

L'évolution temporelle du système est décrite par la relation d'état suivante :

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t)) \text{ pour } t \in [0, T]$$

où T représente le temps final de la simulation et \mathbf{f} est la fonction d'évolution du système, qui dépend de l'état courant $\mathbf{x}(t)$. Les équations décrivant la fonction \mathbf{f} ont été présentées dans le chapitre 1 grâce à la dynamique en ForCEEPS.

3.2.2 Contrôle : Gestion sylvicole

Dans ForCEEPS, il est possible d'appliquer des éclaircies sous la présence de plusieurs paramètres, que nous allons détailler par la suite. Ainsi, pour bien présenter l'évolution de la forêt au cours du temps, il est nécessaire d'introduire le contrôle $\mathbf{u}(t)$.

L'évolution temporelle du système est décrite par la relation d'état suivante :

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \quad (3.1)$$

où $\mathbf{f} : \mathbf{x}(t) \times \mathbf{u}(t) \rightarrow \mathbf{x}(t+1)$ est la fonction d'évolution du système, qui dépend de l'état courant $\mathbf{x}(t)$ et du vecteur de contrôle $\mathbf{u}(t)$.

Pour bien présenter la population à l'instant $t+1$, il est nécessaire d'ajouter l'ensemble des arbres coupés $\mathcal{C}(t)$ aux autres ensembles car les interventions réalisées correspondent à des coupes d'arbres.

$$\text{pop}(t+1) = \mathcal{E}(t) \cup \text{Pop}(t) \setminus \mathcal{M}(t) \cup \mathcal{C}(t), \quad (3.2)$$

où $\mathcal{E}(t)$ l'ensemble des arbres nouvellement établis, $\mathcal{M}(t)$ l'ensemble des arbres morts.

3.2.2.1 Variables de contrôle

Selon ForCEEPS, Les gestionnaires forestiers peuvent travailler avec 4 paramètres pour contrôler la forêt :

1. θ : la période de l'intervention, qui peut être définie sur l'ensemble

$$V_\theta = \{5, 10, 15, 20\} \text{ (exprimée en années).}$$

2. t_p : le type d'éclaircie, qui indique la méthode de coupe des arbres selon les valeurs

$$V_{t_p} = \{0, 1, 0.5\}$$

0 correspond à une coupe décroissante en termes de circonférence de l'arbre, 1 à une coupe croissante, et 0.5 à une coupe aléatoire.

3. G_{obj} : la surface terrière restante après l'intervention sous forme des proportions et définie par l'ensemble

$$V_{G_{obj}} = \{0.5, 0.6, 0.7, 0.8, 0.9\}$$

4. C_f : la composition cible des espèces après l'intervention, c'est-à-dire le pourcentage de chaque espèce cible après éclaircie, qui peut être présenté sous forme d'un n-uplet ou n est le nombre des espèces choisis pour la simulation.

$$V_{C_f} = \left\{ (C_{fi}) \text{ où } \forall i \in [1, n], C_{fi} \in \{0, 20, 40, 60, 80, 100\} \text{ et } \sum_{i=1}^n C_{fi} = 100. \right\}$$

Donc ,Puisque les interventions n'ont pas lieu tous les ans, on définit par la suite T_{inv} qui représente les années d'intervention grâce aux périodes des nb_{inv} interventions. On pose θ_k le nombre d'années entre la k -ième intervention et celle qui la précède, et tel que $\theta_k \in \{5, 10, 15, 20\}$:

$$T_{\text{inv}} = \left\{ \sum_{k=1}^l \theta_k \mid \forall l = 1, \dots, nb_{\text{inv}} \right\}$$

Soit $t_{\text{inv}} \in T_{\text{inv}}$, on définit donc la variable $v(t_{\text{inv}})$ qui repèsente le controle appliqué au système pendant t_{inv} :

$$v(t_{\text{inv}}) = (tp(t_{\text{inv}}), G_{\text{obj}}(t_{\text{inv}}), C_f(t_{\text{inv}})) \in V,$$

tels que $V = V_{t_p} \times V_{G_{\text{obj}}} \times V_{C_f}$ On peut donc définir le contrôle $u(t)$ comme ceci :

$$u : [0, T] \longrightarrow (\{5, 10, 15, 20\} \times V)^{nb_{\text{inv}}} = U$$

$$t \longrightarrow (\theta_k, v(t))_{k=1, \dots, nb_{\text{inv}}}$$

Par exemple, si $nb_{\text{inv}} = 2$, pour $\theta_1 = 5$ et $\theta_2 = 5$, alors $u(t) = [(5, v(t_1)), (5, v(t_2))]$.

On conclut donc :

$$u(t) = \begin{cases} v(t) = (tp(t), G_{\text{obj}}(t), C_f(t)) & \text{pour } t \in T_{\text{inv}}, \\ \emptyset & \text{sinon.} \end{cases}$$

3.2.2.2 Caractéristiques de l'intervention

Maintenant que les variables de contrôle sont définies, il est nécessaire de caractériser l'intervention d'éclaircie appliquée au système pendant l'année $t_{\text{inv}} \in T_{\text{inv}}$. Cela signifie qu'il faut bien caractériser l'ensemble des arbres coupés $\mathcal{C}(t_{\text{inv}})$ pour l'année t_{inv} .

Après avoir analysé le type d'intervention implémenté par les modélisateurs de ForCEEPS, j'ai choisi de suivre cette approche en raison de la démonstration sous forme d'exemple détaillé en annexe 9

Pour une intervention en forêt, l'année doit d'abord être une année $t \in T_{\text{inv}}$ d'intervention, en raison de la dynamique forestière (contre-exemple : si l'on coupe une forêt tous les ans avec un faible taux de croissance, on risque de déboiser la forêt).

Deuxièmement, il faut que la surface terrière objective soit positive et supérieure à la surface totale de la forêt à l'instant t , et il faut aussi exprimer la surface terrière en m^2 si elle était initialement sous forme de pourcentage, donc $G_{\text{obj}}(t) > 0$, $G_{\text{obj}}(t) \leq G(t)$ et $G_{\text{obj}}(t) = G(t) \times G_{\text{obj}}\%$.

Ensuite, lorsque ces conditions sont assurées, on attribue à chaque arbre un score calculé selon le type d'éclaircie t_p :

$$\alpha_i(t) = \begin{cases} \frac{1+C_{\text{max}}(t)-C_i(t)}{1+C_{\text{max}}(t)-C_{\text{min}}(t)} & \text{si } t_p = 0, \\ \frac{1+C_{\text{max}}(t)+C_i(t)}{1+C_{\text{max}}(t)-C_{\text{min}}(t)} & \text{si } t_p = 1, \\ P_{\text{uni}} & \text{si } t_p = 0.5. \end{cases}$$

Pour chaque espèce esp , on calcule le score des arbres de cette espèce et on trie cet ensemble dans l'ordre croissant des scores $\alpha_i(t)$. On peut définir l'ensemble $\text{Tri}_{\text{esp}}(t)$ comme l'ensemble des arbres triés restants pour l'espèce esp , où $\text{pop}_{\text{esp}}(t)$ est l'ensemble des arbres de l'espèce esp et $C_{\text{esp}}(t)$ est l'ensemble des arbres coupés pour l'espèce esp :

$$\text{Tri}_{\text{esp}}(t) = \{i \in \text{pop}_{\text{esp}}(t) \setminus C_{\text{esp}}(t) \mid \forall (i, j) \in \text{pop}_{\text{esp}}(t) \setminus C_{\text{esp}}(t), \alpha_i(t) \leq \alpha_j(t)\}.$$

On peut montrer d'après ces figures que le score correspond bien au type d'éclaircie choisi.

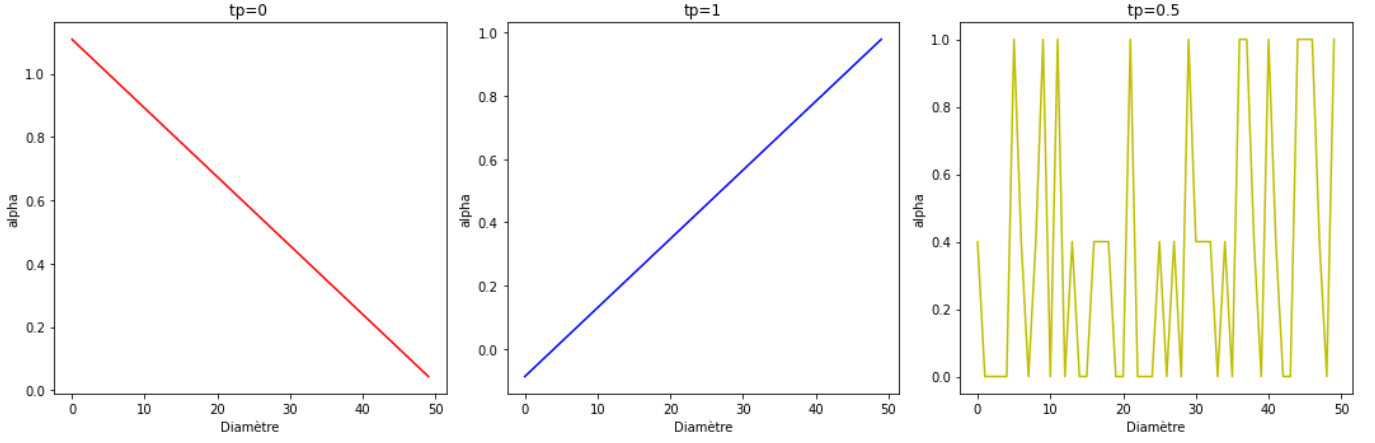


Figure 3.1: Score α_i en fonction du diamètre D_i , pour une éclaircie de plus haut au plus petit ($t_p = 0$, rouge), aléatoire ($t_p = 0.5$, jaune) et de plus petit au plus grand ($t_p = 1$, bleu).

Quand les arbres de chaque espèce sont triés dans l'ordre croissant de score, on crée l'ensemble K_{esp} qui prend le premier arbre de $\text{Tri}_{\text{esp}}(t)$ pour chaque espèce (trié selon l'ordre alphabétique de l'espèce). Selon le modèle de ForCEEPS, les arbres sont identifiés par un ID, donc on peut définir :

$$K_{\text{esp}} = \{k_j \mid \forall j \in \text{Esp où Esp est l'ensemble des espèces choisies}\}.$$

On peut ensuite définir l'ensemble des proportions possibles durant la coupe comme ceci :

$$C_{\text{réel}}^k = \left\{ \left(\frac{G_{\text{esp}}^k}{G(t)} \right) \mid \forall \text{esp} \in \text{Esp}, G_{\text{esp}}^k(t) = G_{\text{rest-esp}}(t) - G_a^k(t) \right\}$$

où

$$G_a^k(t) = \begin{cases} \text{surface de l'arbre } k & \text{si } k \in \text{Tri}_{\text{esp}}, \\ 0 & \text{sinon,} \end{cases}$$

et

$$G_{\text{rest-esp}}(t) = G_{0_{\text{esp}}}(t) - \sum_{i \in C_{\text{esp}}(t)} G_i(t).$$

où $G_{0_{\text{esp}}}(t)$ est la surface forestière initiale de l'espèce esp à l'année t .

Pour choisir l'arbre à couper, on prend l'arbre de l'ensemble K_{esp} qui minimise la différence entre la composition réelle et la composition souhaitée :

$$k = \underset{k \in K_{\text{esp}} \setminus C(t)}{\text{argmin}} \|C_{\text{réel}}^k(t) - C_f(t)\|.$$

Ensuite, les arbres sont coupés jusqu'à ce que la surface terrière restante passe en dessous de $G_{\text{obj}}(t)$, c'est-à-dire on coupe si et seulement si :

$$\sum_{i \in C(t)} G_i(t) < G(t) - G_{\text{obj}}(t)$$

et il est possible que la surface terrière restante passe légèrement en dessous de la surface espérée, donc on peut écrire :

$$\exists j \in C(t) \quad \sum_{i \in C(t) \setminus \{j\}} G_i(t) > G(t) - G_{\text{obj}}(t).$$

Ainsi, on peut caractériser l'ensemble des arbres coupés $C(t)$ pendant l'intervention $t \in T_{\text{inv}}$ comme suit :

$$C(t) = \begin{cases} \{i \in \text{pop}_{\text{esp}} \mid i \in 3.2.2.2 \text{ et } 3.2.2.2\} & \text{si } t \in T_{\text{inv}} \text{ et } 3.2.2.2, \\ \emptyset & \text{sinon.} \end{cases}$$

Voici un organigramme qui résume la procédure de l'intervention que j'ai adopté pour ce projet implémenté en JAVA en se basant sur les méthodes déjà utilisé en ForCEEPS

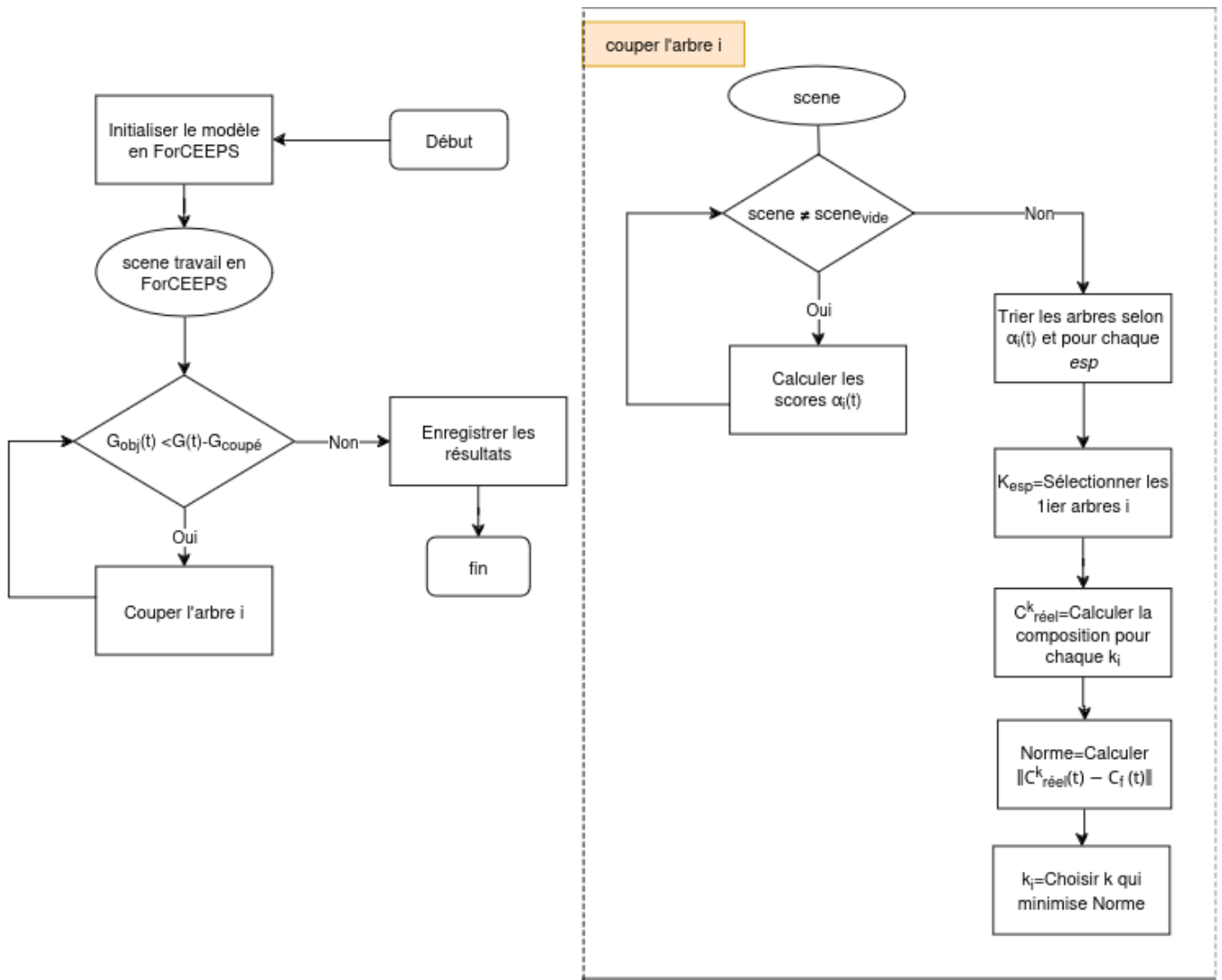


Figure 3.2: Représentation du fonctionnement de l'algorithme d'intervention implémenté en Java (**à gauche**) : Le test en **losange** indique la condition selon laquelle la coupe est effectuée uniquement si la surface terrière objective est inférieure à la surface restante lors de l'éclaircie. Détail du processus de la coupe (**à droite**) : La scène correspond à l'ensemble des arbres avec les différentes caractéristiques implémentées dans ForCEEPS (mortalité, croissance, régénération). Le test en **losange** indique la condition pour calculer le score de tous les arbres de la scène au début de l'éclaircie. Le dernier cadre permet de choisir l'arbre i qui minimise l'écart entre la composition actuelle et la composition cible, ce qui nous ramène ensuite au schéma global pour effectuer le test sur $G_{obj}(t)$. L'enregistrement des résultats se fait dans **3 fichiers de sortie** regroupant l'ensemble des variables à l'échelle individuelle et populationnelle fournies par le modèle ForCEEPS.

3.2.3 Les contraintes écologiques et économiques

Après avoir défini le système étudié ($x(t)$) et le contrôle ($u(t)$) qui gère l'évolution du système, on arrive à la définition de l'espace des contraintes (K) qui nous permet de déterminer les états souhaitables ou non. Il faut donc définir les critères de satisfaction qui sont identifiés par l'IGN (Institut Géographique National) et selon la littérature.

Commençons par la surface terrière annuelle ($G(t)$). Les gestionnaires veulent maintenir une certaine densité d'arbres pour les services écosystémiques, ils ne cherchent pas une coupe rase qui perturberait le fonctionnement de la flore et la faune, donc il faut avoir au moins chaque année une surface terrière de $10 \text{ m}^2/\text{ha}$ dans la forêt. Selon les gestionnaires, pour bien comparer la surface terrière mesurée sur toutes les placettes, il faut :

1. Multiplier la surface terrière mesurée sur chaque placette par la surface de la placette (S_p) pour obtenir la surface terrière totale sur cette placette.
2. Ensuite, diviser cette surface terrière totale par 10 000 pour la convertir en surface terrière par hectare. Donc, on peut écrire :

$$\forall t = 1, \dots, T \quad G(t) = \sum_{\forall i \in \text{pop}(t)} \frac{\pi(D_i^2)}{4} \times \frac{1}{S_p} > 10 \times \frac{S_p}{10000}$$

Ensuite, pour approvisionner la filière économique, il est nécessaire d'avoir un minimum de production de bois. D'après l'IGN, le critère est de 3.1 m³/ha/an lors de la simulation. Et pour avoir un résultat qui soit conforme à la réalité où les forêts font plusieurs hectares et regroupent des zones de différents stades de développement, on a décidé de regarder la production de bois totale (B_{tot}) en m³ à la fin de la simulation à T et de la comparer avec la production de bois minimale qui doit être multipliée par (S_p) et divisée par 10 000 pour la convertir en ha et multipliée aussi par T pour parcourir la production de bois minimale à la fin de la simulation.

$$B_{\text{tot}}(t) = \sum_{t=1}^T \sum_{i \in C(t)} V_i(t) > 3.1 \times \frac{S_p}{10000} \times T$$

où C(t) est l'ensemble du bois coupé durant toute la simulation, $V_i(t)$ est le volume du bois calculé d'après ce qui est présenté ci-dessus et S_p est la surface de la taille de la placette étudiée.

Pour les trois dernières contraintes, les valeurs sont moyennées sur 5 ans parce que ces contraintes sont plus souples par rapport à la surface terrière qui ne doit jamais être en dessous de 10 m²/ha, et pour des raisons de la méthode choisie pour résoudre le problème qui sera détaillée dans le chapitre suivant, ces valeurs seront vérifiées sur $t=5, 10, \dots, T$.

1. Commençons par la diversité structurelle. Cette diversité est mesurée par le coefficient de Gini . Cette métrique est comprise entre 0 et 1. Des valeurs trop élevées ou trop basses peuvent indiquer des déséquilibres structurels qui peuvent affecter la

santé de la forêt et sa capacité à résister aux perturbations. C'est pour cette raison qu'il est préférable de choisir l'indice de Gini sur l'intervalle [0,25 ; 0,75] .

$$\text{Gini}(t) = \frac{\sum_{i,j \in \text{pop}(t)} |D_i(t) - D_j(t)|}{2 * \text{card}(\text{pop}(t)) * \sum_{i \in \text{pop}(t)} D_i(t)}$$

Pour chaque période de 5 ans, il faut :

$$m_{\text{Gini}}(t) = \frac{1}{5} \sum_{t=tc-4}^{tc} \text{Gini}(t) \in [0, 25; 0, 75]$$

2. Pour la résilience de la forêt face aux perturbations [1] et pour la richesse de biodiversité, il existe un nombre important de métriques pour présenter ceci comme l'indice de Shannon $H = -\sum_{i=1}^S p_i \log p_i$ où S est le nombre d'espèces et p_i est la proportion de l'espèce i . Cette métrique permet d'évaluer à la fois le nombre d'espèces et leur abondance relative, et est utilisée dans les travaux de Montgolfier en 2024 [10] pour montrer comment la biodiversité peut être un outil pour maintenir l'écosystème dans un état désirable []. Elle a pris 20% pour représenter des espèces minoritaires dans la population, se traduisant par un indice de Shannon égal ou supérieur à 0,72. Et pour l'uniformité verticale, elle a voulu avoir une forêt d'âges inégaux pour assurer que la couche la plus riche comprenne moins de 80% des tiges, ce qui est traduit par un indice vertical de Shannon supérieur ou égal à 0,92. Pour notre travail, on a choisi d'indiquer 2 espèces au minimum comme critère pour la biodiversité sur la placette étudiée. Donc après chaque période de 5 ans :

$$\text{il faut } m_{\text{esp}}(t) = \frac{1}{5} \sum_{t=tc-5}^{tc} \text{nb}_{\text{Esp}}(t) \geq 2 \text{ tel que } \text{nbs}(t) = \sum s \in \text{Esp} \text{ si } \text{nbs}(t) > 0 \text{ et } 0 \text{ sinon}$$

avec $\text{nb}_s(t)$ le nombre de l'espèce s et Esp l'ensemble des espèces pouvant être présentes dans la simulation.

3. Finalement, pour bien caractériser la santé de la forêt, il fallait évaluer le taux de mortalité chaque 5 ans et qui ne doit pas être supérieur à 10%. Soit le taux de mortalité :

$$\tau_M(t) = \frac{|M(t)|}{|M(t) + \text{pop}(t)|} \text{ donc il faut } m_M(t) = \frac{1}{5} \sum_{t=tc-5}^{tc} \tau_M(t) \geq 0.1$$

où $M(t)$ est l'ensemble des arbres morts et $\text{pop}(t)$ est l'ensemble des arbres toujours vivants.

Donc, puisque nous avons défini toutes les contraintes écologiques et économiques qui caractérisent l'espace des gestions sylvicoles admissibles, nous pouvons maintenant écrire :

$$K = \{x = \{D_i, cs_i\} \forall i \in \text{pop}(t) \mid 3.2.3; 2, 1, 3, 2\}$$

3.3 Résumé du problème de contrôle pour la gestion forestière

Maintenant que toutes les variables de notre problème de viabilité sont définies, voici un résumé du problème et de ce que nous cherchons :

1. **le Système dynamique:** L'état du système à l'instant t est décrit par le vecteur d'état

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{D}(t) & \mathbf{cs}(t) \end{bmatrix}, \text{ où :}$$

$$\mathbf{D}(t) = [D_i(t)]_{i \in \text{Pop}(t)}, \quad \mathbf{cs}(t) = [cs_i(t)]_{i \in \text{Pop}(t)}$$

L'évolution du système est donnée par :

$$\mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))$$

2. **Variables de contrôle:** Les gestionnaires forestiers peuvent utiliser les variables de contrôle suivantes :

$$u(t) = (\theta_k, t_p(t), G_{\text{obj}}(t), C_f(t))$$

Avec :

$$\theta_k \in \{5, 10, 15, 20\},$$

$$t_p \in \{0, 0.5, 1\},$$

$$G_{\text{obj}} \in \{0.5, 0.6, 0.7, 0.8, 0.9\},$$

$$C_f \in \{(C_{fi}) \mid \forall i \in [1, n], C_{fi} \in \{0, 20, 40, 60, 80, 100\}, \sum_{i=1}^n C_{fi} = 100\}$$

3. **Contraintes:** L'espace des contraintes K est défini comme :

$$K = \{\mathbf{x} = \{D_i, c_i\} \forall i \in \text{pop}(t) \mid \text{respect des contraintes de } G(t), B_{\text{tot}}(t), m_{\text{Gini}}(t), m_{\text{esp}}(t), m_M(t)\}$$

Le noyau de viabilité est défini comme suit :

$$\text{Viab}_K(u(t)) = \{u(t) \in U \mid \mathbf{x}(t+1) = \mathbf{f}(\mathbf{x}(t), u(t)) \in K \text{ pour tout } t \in [0, T]\}$$

Nous cherchons donc à trouver un ensemble A tel que $\mathbf{A} \subseteq \text{Viab}_K(u(t))$, où A approche autant que possible de $\text{Viab}_K(u(t))$.

4 Méthode de résolution du problème de viabilité pour la gestion forestière

4.1 Résolution par algorithme génétique

En 2023, Michelle Malara et al ont fusionné un modèle dynamique avec la théorie de la viabilité pour sélectionner une approche algorithmique déjà établie. Ils ont adopté l'algorithme génétique, une méthode heuristique initiée à partir d'une population initiale aléatoire de contrôles, utilisée pour explorer l'espace des contrôles viables, noté $Viab(u(t))$, tout en respectant les contraintes K . Cette méthode s'inspire du processus de sélection naturelle en biologie et implique des techniques telles que la mutation, la sélection naturelle, et le croisement génétique pour aborder les problèmes.

Un algorithme génétique commence avec un ensemble de solutions potentielles, appelé "population". Les solutions de cette population sont évaluées à l'aide d'une métrique appelée "fitness", et plus cette métrique est élevée, plus la solution est efficace pour résoudre le problème. Les meilleures solutions sont alors sélectionnées pour la reproduction, engendrant ainsi une nouvelle génération de solutions. Ce processus est répété sur plusieurs générations, dans l'espoir d'améliorer les solutions au fil du temps. L'objectif est de trouver la meilleure solution possible ou une solution suffisamment bonne pour un problème spécifique.

Avant de commencer, il est nécessaire d'évaluer la dimension de l'espace à explorer et le nombre de combinaisons possibles puisque j'ai ajouté aux contrôles préexistants une autre variable pour l'espace, ce qui est la composante objective $C_f(t)$. D'après la littérature, l'algorithme génétique a pu explorer un espace de 50 dimensions et de 10^{64} combinaisons dans le cas du problème classique du voyageur de commerce (TSP). C'est pour cette raison qu'il est essentiel de définir la combinatoire de notre problème et les simplifications apportées pour rendre l'exploration possible pour l'analyse des résultats subséquents. [[11],[12],[13],[14],[15],[16]]

4.2 Évaluation de la Combinatoire du Problème

La période choisie pour chercher les contrôles viables est de 80 ans. Selon l'intervalle V_θ , il est possible d'effectuer un contrôle tous les 5 ans, 10 ans, ou 20 ans. Sur une période de 80 ans, le nombre maximal d'interventions possibles serait de 16, dans le cas où un contrôle serait effectué tous les 5 ans. Ainsi, le contrôle $u(t)$ sur 80 ans peut prendre 16 valeurs possibles au total.

$$u(t) = (\theta_k, tp_k(t), G_{objk}(t), C_{fk}(t))_{k=1, \dots, 16}$$

Étant donné que notre problème est chronologique, cela signifie que si nous choisissons de faire une intervention après 20 ans, nous devons laisser la forêt évoluer pendant ces 20 années. Nous ne pouvons donc pas effectuer de coupe pendant ces 20 ans.

Malara et al. en 2023 ont fixé les années de contrôle antérieures à 0, ce qui signifie qu'on ne coupe pas. Par exemple, si $\theta_k = 10$ ans, on ne coupe pas pendant t-5 ans, donc $u(t_{k-1}) = (0, 0, 0, 0)$. De même, si $\theta_k = 15$ ans, on ne coupe pas pendant t-10 ans, et si $\theta_k = 20$ ans, on ne coupe pas pendant t-20 ans.

Ainsi, nous ne pouvons pas évaluer la combinatoire de notre problème avec la méthode classique qui consiste à multiplier les cardinaux de chaque intervalle, car les années de non-coupe dépendent de la valeur de θ_k .

Par exemple, si on ne prend pas l'échelle temporelle en considération, on pourrait utiliser cette méthode :

Commençons par déterminer la cardinalité de chaque variable

$$\begin{aligned}\theta_k &\in \{5, 10, 15, 20\}, \\ t_p &\in \{0, 0.5, 1\}, \\ G_{obj} &\in \{0.5, 0.6, 0.7, 0.8, 0.9\}, \\ C_f &\in \{(C_{fi}) \mid \forall i \in [1, n], C_{fi} \in \{0, 20, 40, 60, 80, 100\}, \sum_{i=1}^n C_{fi} = 100\}\end{aligned}$$

Il est clair que θ est composé de 4 valeurs, t_p de 3 valeurs, et G_{obj} de 5 valeurs. Cependant, la complexité de calculer l'ensemble de variation de C_f s'accroît du fait de la contrainte selon laquelle la somme des C_{fi} doit être égale à 100. Pour aborder ce problème, j'ai initialement calculé la combinatoire sans prendre en compte cette condition, comptant ensuite uniquement les quintuplets dont la somme est égale à 100. J'ai choisi de faire le calcul au début pour $n=5$ car on a espéré au début de traiter 5 espèces au maximum dans la forets.

Soit $L = \{0, 20, 40, 60, 80, 100\}$ l'ensemble des valeurs possibles pour chaque variable. Nous définissons alors V_{C_f} (sans la condition) comme l'ensemble des tuples $(C_{f1}, C_{f2}, C_{f3}, C_{f4}, C_{f5})$ où chaque $C_{fi} \in L$. Ainsi, le cardinal de V_{C_f} (sans la condition) est $|L|^5 = 6^5 = 7776$. À l'aide d'un script en Python, nous trouvons que le nombre de combinaisons possibles respectant la condition de la somme égale à 100 est de 126.

Ainsi, pour 16 interventions, on trouve que le nombre de combinaisons possibles pour tout le problème sans compter les zéros est $(3 \times 4 \times 5 \times 126)^{16} = 2.8 \times 10^{76}$, mais pour déterminer les vraies combinaisons possibles, il fallait adopter une autre méthode. Pour ce faire, j'ai créé un graphe qui prend en considération le décalage temporel défini ci-dessus.

Définition du graphe : Soit S_k les nœuds de notre graphe et en se basant sur l'explication définie ci-dessus qui détermine l'enchaînement de notre contrôle $u(t)$, on peut écrire

1. si $S_k = 5$ pas de condition
2. si $S_k = 10$, il faut que $S_{k-1} = 0$
3. si $S_k = 15$, il faut que $S_{k-2} = 0$ et $S_{k-1} = 0$
4. si $S_k = 20$, il faut que $S_{k-1} = 0, S_{k-2} = 0, S_{k-3} = 0$

donc on crée **10 nœuds** et les arcs sont orientés selon cette logique, cela signifie qu'un arc sortant représente une série temporelle possible dans le contrôle.

Après avoir dessiné le graphe, j'ai déterminé la matrice de transition P_{ij} telle que pour chaque arc $i \rightarrow j$ possible, on fixe la case ij à 1 sinon à 0. En utilisant l'**algorithme de parcours en largeur**, nous avons trouvé tous les chemins possibles de longueur égale à 16, en partant d'un état initial. Dans notre cas, il y a 4 états initiaux : l'état 5 et les états zéro 0 (coloriés en orange pour les états 10, 15 et 20). J'ai implémenté tout ceci en Python et voici le lien vers le script sur GitHub : https://github.com/lotaleb/Forceps_code_Inrae

Revenons donc à la définition de l'intervalle du contrôle :

$$U = \{(\theta, t_p, G_{obj}, C_f) \in (V_{t_p}, V_{G_{obj}}, V_{C_f})^{nb} \text{ tel que } nb \text{ est le nombre de périodes dans le contrôle}\}$$

Donc, pour chaque chemin possible, on calcule le nombre de périodes présentes $\{5, 10, 15, 20\}$ et on les combine avec les autres variables. Par exemple, pour un chemin possible sur 80 ans on trouve qu'on a 5 périodes possibles : 5- 20- 20- 20- 10. Ainsi, on peut dire que le nombre de combinaisons possibles pour tout ce chemin est $(3 \times 4 \times 1001)^5$ et ainsi de suite.

On peut écrire la formule suivante :

$$\text{Comb}_{\text{total}} = \sum_{\text{path}_i \in \{5, 10, 15, 20\}} \sum$$

En implémentant tout ceci, nous trouvons que le nombre de chemins possibles, en prenant en considération la chronologie du contrôle, est 1.07×10^{53} . Pour la dimension, puisque nous utilisons 8 bits pour chaque période sur 80 années (1 bit pour chaque variable du contrôle, soit θ , t_p , G_{obj} , et 5 bits pour la composition) et 16 interventions au total, nous avons exactement $16 \times 8 = 128$ bits.

Donc, puisque le nombre de combinaisons possibles est très grand, il était nécessaire de faire des réductions de variables dès le début avant de modifier l'algorithme génétique implémenté par Michelle. Nous avons décidé de travailler avec trois espèces, ce qui réduit la dimension du problème à $16 \times 6 = 96$ bits et la combinaison à 3×10^{40} .

Maintenant que la combinatoire calculée est faisable à explorer, je vais pouvoir définir les différentes étapes et métriques pour l'algorithme génétique.

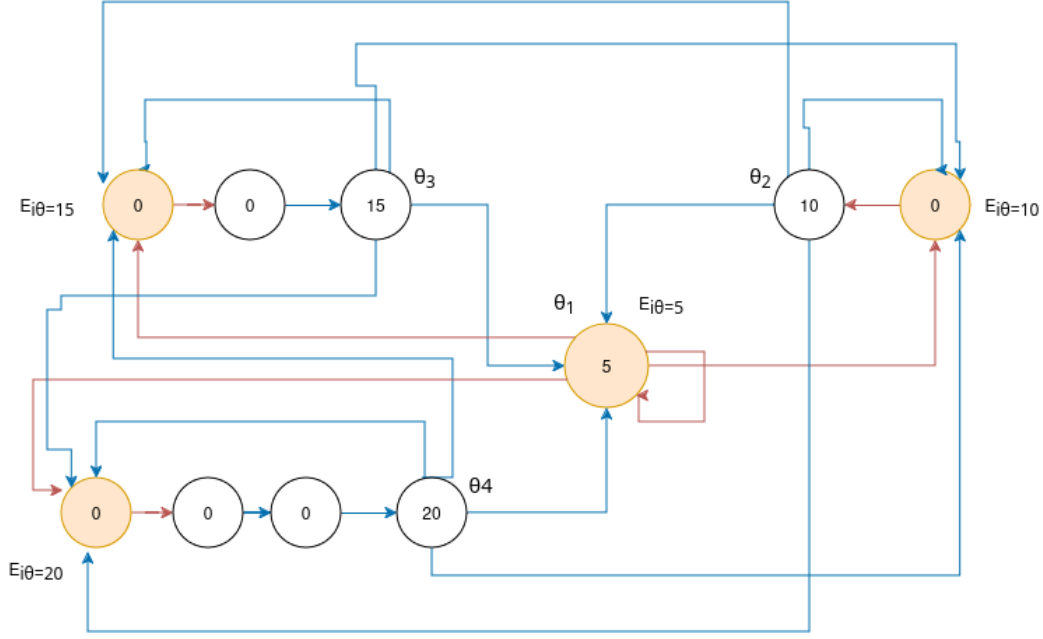


Figure 4.1: Le graphe permet de déterminer la combinatoire du problème : les cercles oranges représentent les états initiaux possibles pour chaque chemin (identifiés par $E_{i\theta}$), les autres cercles représentent les périodes de θ , et les arcs oranges représentent les transitions possibles d'un état initial vers d'autres états. Les arcs bleus représentent les transitions possibles entre les états au milieu de la séquence.

4.3 Une simulation en ForCEEPS

Avant de détailler les différentes étapes implémentées pour l'algorithme génétique (GA), il est essentiel de comprendre ce qu'est une simulation avec ForCEEPS afin de saisir le fonctionnement du GA. Comme expliqué dans les chapitres 1 et 2, l'évolution des forêts est régulée par plusieurs processus, notamment la mortalité, la croissance, la régénération, et également par l'intervention humaine selon les pratiques de coupe.

Ainsi, une simulation en ForCEEPS peut être décrite comme l'ensemble des mécanismes définis par le modèle et de l'intervention d'un gestionnaire. Pour notre étude de la viabilité, nous cherchons à déterminer les différentes stratégies de gestion forestière applicables sur une période totale de 80 ans. Cela implique d'intégrer un contrôle adapté dans la simulation ForCEEPS.

Cette clarification nous permet de mieux visualiser et structurer les différentes étapes que nous définirons par la suite pour le GA.

4.4 Définition d'un individu en GA

Comme introduit précédemment, un algorithme génétique (GA) se compose principalement de populations qui évoluent à chaque itération. Chaque population comprend un certain nombre d'individus, noté n_{ind} , qui subissent des modifications d'une itération à l'autre. Dans notre cas, un individu représente une stratégie de contrôle $u \in U$, simulée sur une durée de 80 ans. Nous désignons cet individu par le terme de "génom" Y^j , qui est constitué de gènes conformément à la définition du GA.

Chaque intervention forestière est représentée par un ensemble de 6 gènes :

- Un gène pour la période θ ,
- Deuxième gène pour le type d'éclaircie tp ,
- Le troisième gène pour la surface terrière objective,
- Les gènes 4 à 6 définissent la composition objective souhaitée après chaque éclaircie.

Ainsi, le génome est constitué de $16 \times 6 = 96$ bits. La structure d'un génome peut être définie comme suit :

$$Y^j = (\theta_k^j, tp_k^j, G_{obj}^k, C_{f1}^j, C_{f2}^j, C_{f3}^j)_{k=1, \dots, 16}$$

où Y^j appartient à l'espace produit des valeurs possibles pour chaque gène, soit :

$$Y^j \in (V_\theta \times V_{tp} \times V_{G_{obj}} \times V_{C_f})^{16}$$

Voici un exemple d'un génome présenté pour deux séquences (Le génome complet est présenté sur 16 séquences)

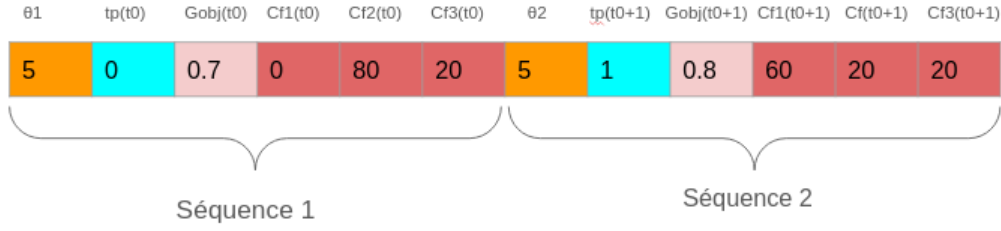


Figure 4.2: deux premières séquences d'un génome, représentées chacune sur 6 bits, sont décrites ci-dessous. Pour chaque séquence, la première valeur en orange représente θ , la seconde en bleu tp , et la troisième en rose clair G_{obj} . Les trois valeurs suivantes en rose foncé définissent la composition cible des espèces après éclaircie. Cette structure est identique pour les 14 séquences suivantes.

4.5 Opérations réalisées pour une population en GA

4.5.1 Création d'un individu aléatoire

Au début de l'algorithme génétique (GA), une population initiale de n_{init} individus est choisie aléatoirement. Ce choix est principalement basé sur la période θ . On commence par sélectionner θ de manière aléatoire et, pour chaque période définie, les gènes antérieurs sont fixés à 0. Par exemple, si $\theta = 10$, les 6 gènes précédents sont nuls; pour $\theta = 15$ et $\theta = 20$, cela correspond respectivement à 12 et 18 gènes nuls. Après avoir placé θ à l'indice approprié, tp et G_{obj} sont tirés aléatoirement et placés sur les deux gènes suivant θ . Pour la composition, un fichier contenant toutes les combinaisons possibles pour trois gènes est utilisé. Un triplet est donc sélectionné aléatoirement à partir de ce fichier et placé dans les trois gènes suivants. Ce processus est répété pour remplir les séquences du génome suivants.

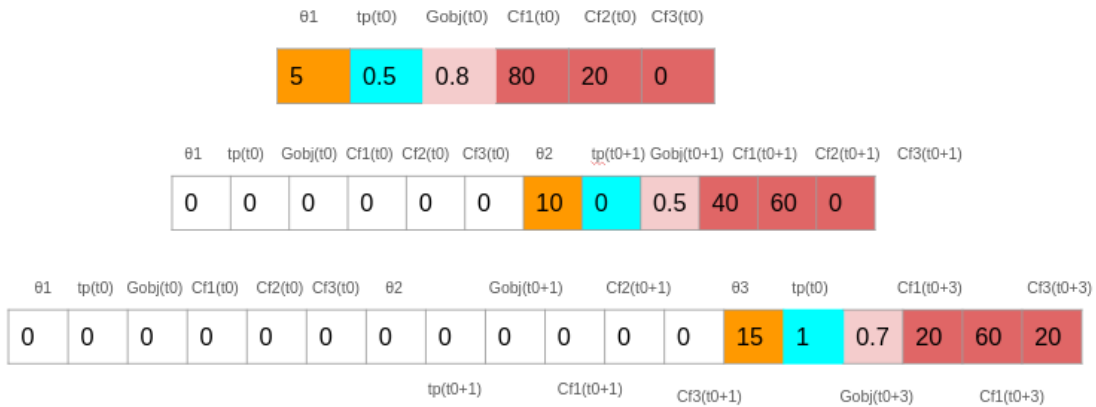


Figure 4.3: 3 types de séquences dans un individu en GA. La première séquence est pour une période de 5, donc il n'y a pas de conditions sur les gènes antérieurs. La deuxième séquence est pour une période de 10, donc une séquence de 6 gènes nuls. Pour une période de 15, il y a 12 gènes nuls, et pour $\theta = 20$, c'est 18 gènes nuls, bien que je ne l'aie pas présentée ici.

4.5.2 Calcul de la fitness

Après avoir généré un individu aléatoire, nous abordons l'étape du calcul de la fitness, ou de la fonction objective, qui va déterminer si cet individu est viable ou non. Pour ce faire, il faut se rappeler des contraintes que nous avons fixées au chapitre 3.5. Donc à partir de ces contraintes, la métrique utilisée est la somme des distances aux contraintes tout au long de la simulation. Étant donné que les contraintes n'ont pas les mêmes dimensions temporelles, parce que nous examinons la surface terrière chaque année, la production de bois à la fin de la simulation, l'indice de Gini, le taux de mortalité, et le nombre d'espèces après 5 ans d'évolution, il est nécessaire

de normaliser toutes les contraintes en fonction de leur apparition. De plus, puisque les unités diffèrent entre chaque grandeur, il est nécessaire de diviser par leur écart-type. On définit donc la fitness comme suit :

La fonction d'évaluation Φ est définie par :

$$\Phi(u_j) = \begin{cases} \phi(u_j) & \text{si } u_j \in Viab_K(x_0) \\ 0 & \text{sinon.} \end{cases}$$

où $Viab_K(x_0)$ représente l'ensemble de contrôle admissible à partir de l'état initial x_0 .

La fonction $\phi(u_j)$ est calculée comme suit :

$$\phi(u_j) = \frac{B_p - 3.1}{B_{pmax} - 3.1} + \frac{1}{80} \sum_{t=1}^{80} \left(\frac{G(t) - 10}{G_{max} - 10} \right) + \frac{1}{16} \sum_{t=1}^{16} \frac{|m_{Gini}(t_c) - 0.5|}{0.5} + \frac{m_{nbEsp}(t_c) - 2}{nbEsp_{max} - 2} + \frac{0.25 - m_{\tau M}(t_c)}{0.25}$$

Les paramètres B_{pmax} , G_{max} , et $nbEsp_{max}$ désignent respectivement la production maximale en mètres cubes par hectare, la surface terrière maximale en mètres carrés, et le nombre maximal d'espèces qui peuvent être atteintes durant la simulation. La simulation est réalisée sur plusieurs définitions dans la forêt et le calcul de la fitness s'effectue sur tous les patches. Il est donc nécessaire de déterminer quand un individu est considéré non viable. Cela implique de vérifier si un individu respecte les contraintes au niveau de chaque patch, un processus que l'on nomme aversion au risque. Cette méthode consiste à évaluer la conformité aux contraintes de chaque patch individuellement:

$$u^j \in Viab_k(x_0) \Leftrightarrow \forall p, \forall t, \quad x_p(t+1) = f_p(x_p(t), u(t)) \in K$$

Une autre approche consiste à examiner les contraintes sur l'ensemble des patches. Ainsi, on calcule $B_{tot}(t)$, $G(t)$, $m_{Gini}(t)$, $\tau_M(t)$, et $m_{nbEsp}(t)$ pour l'ensemble des patches et on vérifie si la moyenne de chaque indicateur respecte les contraintes pour chaque intervalle temporel choisi. Cela peut s'écrire comme suit:

$$u^j \in Viab_k(x_0) \Leftrightarrow \forall p, \forall t, \quad \frac{1}{P} \sum_{p=1}^P x_p(t+1) = f_p(x_p(t), u(t)) \in K$$

Il est ainsi possible d'adopter un schéma qui peut occasionnellement sortir des contraintes. Ce choix, moins contraignant que l'aversion au risque, est justifié par le fait que ForCEEPS est un modèle et ne reflète pas parfaitement la réalité. Lors de la mise en œuvre, décrite dans la section 4, la version retenue est celle qui utilise la moyenne.

4.5.3 Nettoyage de la population

À la suite du calcul de la fitness pour chaque individu, si un individu ne respecte pas les contraintes durant la simulation, sa fitness est réinitialisée à zéro, indiquant ainsi sa non-viabilité. Cet individu est donc retiré de la population courante. En revanche, les individus qui respectent les contraintes sont conservés et ajoutés à l'ensemble *viables* et à la population courante.

4.5.4 Sélection

Nous abordons la phase de sélection des parents pour la population suivante. Ce processus se divise en deux étapes. Tout d'abord, nous conservons n_{save} des meilleurs individus comme parents pour la prochaine génération. Ensuite, pour les $n_{ind} - n_{save}$ individus restants, la sélection s'effectue selon une probabilité P_{save_ind} définie par :

$$P_{save_ind} = \frac{\text{fitness_ind}}{\text{best_fitness}}$$

Si cette probabilité est supérieure à un nombre aléatoire tiré dans l'intervalle $[0,1]$, l'individu est considéré comme parent. Dans le cas contraire, l'individu n'est pas sélectionné.



Figure 4.4: Illustration de la sélection dans une population de 10 individus : Les 3 individus oranges, ayant les meilleures performances, sont automatiquement retenus comme parents pour la prochaine génération. Les individus en bleu sont sélectionnés selon une probabilité basée sur leur fitness relative.

4.5.5 Croisement des parents

Après avoir sélectionné et mélangé aléatoirement les parents, nous procédons à la création de deux enfants pour chaque paire de parents. Le croisement s'effectue point par point pour les trois premiers gènes (θ , tp , et G_{obj}). Pour les trois gènes suivants, le croisement se fait par groupe pour maintenir l'équilibre des compositions. En effet, un croisement individuel pour ces gènes pourrait résulter en une somme de composition soit supérieure à 100, soit inférieure à 100, ce qui contredirait les conditions d'application des interventions décrites au chapitre 3.

Pour les gènes nuls, aucun croisement n'est effectué afin d'éviter de compromettre la structure du génome. Lorsqu'une séquence de gènes est nulle chez les deux parents, les enfants héritent de ces séquences nulles. Si un parent possède une séquence de gènes nuls, le premier enfant hérite de cette séquence, et le second enfant hérite de la séquence du deuxième parent, et vice versa.

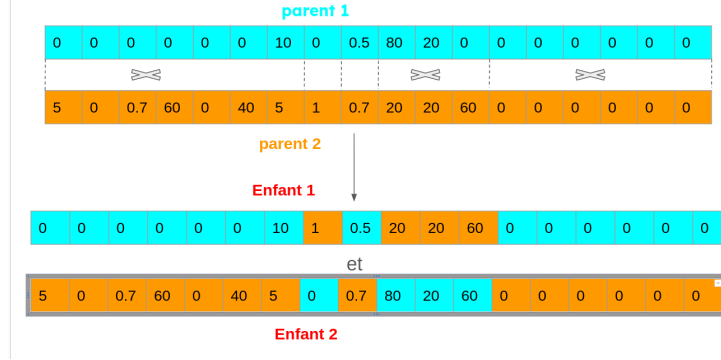


Figure 4.5: Exemple de crossover dans un algorithme génétique : Les gènes issus du parent 1 sont en bleus, ceux du parent 2 en orange et les points de croisement sont en pointillés.

4.5.6 Mutation après croisement

Après la création des enfants par croisement, une mutation est appliquée aléatoirement à chaque enfant. Cette mutation peut survenir sur n'importe quel gène, avec une probabilité P_{mut} . Toutefois, en raison de la complexité du problème, nous avons décidé de ne pas muter le gène de la période (θ), car cela nécessiterait de modifier les séquences nulles antérieures. Ainsi, les mutations sont appliquées aléatoirement sur les gènes tp , G_{obj} , et Cf . Pour chaque gène, une nouvelle valeur est choisie aléatoirement dans son ensemble de définition. Cette étape crée de la diversité parmi les descendants et aide à mieux explorer l'espace de contrôle viable.

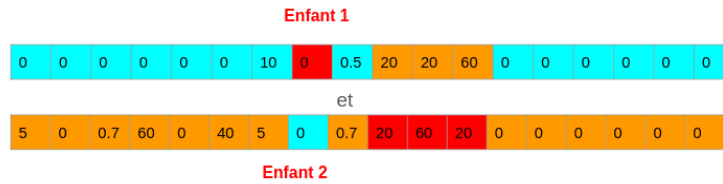


Figure 4.6: Illustration de mutation aléatoire appliquée sur les descendants. Les zones rouges indiquent les mutations survenant dans les génomes des enfants, affectant aléatoirement la composition et le type d'éclaircie tp .

4.5.7 Exploitation du meilleur

Dans cette étape, nous sélectionnons le meilleur individu de la population courante, c'est-à-dire l'individu ayant la meilleure fitness. Pour chaque gène du génome, à l'exception des périodes, nous créons de nouveaux individus avec une mutation appliquée sur un seul gène à la fois. Pour la composition, nous appliquons des mutations deux à deux. Cependant, cette fois-ci, la mutation n'est pas appliquée en choisissant un nombre aléatoire dans l'ensemble de définition du gène, mais en exploitant le meilleur individu pour explorer localement son environnement et voir si nous pouvons obtenir un individu amélioré.

Pour réaliser cette exploitation locale, nous mutons vers le voisin le plus proche :

1. Pour tp : Si $tp = 0$, nous évaluons la fitness de l'individu muté à 1. Si c'est 0, nous mutons à 1 et si c'est 0.5, nous examinons deux individus mutés à 1 et 0.
2. Pour G_{obj} : Si $G_{obj} < 0.8$, nous mutons à $G_{obj} + 0.1$. Si $G_{obj} > 0.8$, nous mutons à $G_{obj} - 0.1$.

3. Pour Cf : Nous mutons deux gènes à la fois de +10 et -10, mais seulement si les deux valeurs successives sont > 10 et < 100 . Dans ce cas, nous ajoutons +10 à l'un et soustrayons -10 à l'autre. Sinon, nous examinons les deux paires suivantes.

Pour chaque nouvel individu créé, nous évaluons sa fitness et la stockons si elle est meilleure que celle du meilleur individu exploré, et nous enregistrons l'individu dans l'ensemble des viables s'il est viable.

À la fin de l'itération, nous ajoutons à la population suivante le meilleur individu créé par cette étape.

5	0	0.5	80	20	0
---	---	-----	----	----	---

5	1	0.5	80	20	0
---	---	-----	----	----	---

+0.1

5	0	0.6	80	20	0
---	---	-----	----	----	---

-10 +10

5	0	0.5	70	30	0
---	---	-----	----	----	---

-10 +10

5	0	0.5	80	10	10
---	---	-----	----	----	----

Figure 4.7: Les mutations sur le meilleur individus effectuées sont illustrées en rouge, indiquant les ajustements précis sur des gènes spécifiques pour améliorer la fitness globale de l'individu.

4.5.8 Ajout d'individus aléatoires

Afin d'explorer efficacement l'espace de recherche, nous ajoutons n_{rand} individus aléatoires à la population suivante. Ces individus sont créés suivant les principes détaillés dans la section dédiée à la création d'individus aléatoires. Cette méthode permet d'introduire de la diversité génétique dans la population, ce qui est essentiel pour éviter la convergence prématurée vers des optima locaux et pour augmenter la probabilité de découvrir de nouvelles solutions optimales.

4.5.9 Critère d'arrêt de l'algorithme génétique

Les populations sont générées itérativement dans l'algorithme génétique, et le processus de création s'arrête lorsque l'un des deux critères suivants est atteint :

- Les meilleures valeurs de fitness n'évoluent pas après cinq générations successives, ce qui indique une convergence potentielle de l'algorithme.
- Un total de 10,000 itérations a été effectué sur l'ensemble de l'algorithme génétique, ce qui constitue une limite fixe pour terminer le processus.

Ces critères d'arrêt sont conçus pour garantir que l'algorithme ne tourne pas indéfiniment tout en permettant suffisamment de temps pour une exploration approfondie de l'espace de solutions.

4.6 Parallélisation des étapes de l'algorithme génétique

Les étapes de l'algorithme génétique (GA) sont généralement exécutées de manière séquentielle, c'est-à-dire que chaque étape est réalisée l'une après l'autre sur un seul processeur, comme illustré dans l'organigramme présenté ci-dessous. Cependant, compte tenu que la durée de simulation d'un individu peut être longue et que l'exploration de l'ensemble de l'espace de recherche est chronophage — avec un lancement de l'algorithme génétique nécessitant environ 9 secondes, auxquelles s'ajoutent 12 secondes pour l'enregistrement des données — nous avons opté pour une approche visant à accélérer le processus et à rendre les tests plus efficaces.

Pour cela, nous avons décidé de paralléliser la simulation des individus à chaque étape de l'algorithme. La première étape consiste à générer tous les individus aléatoirement, sans immédiatement calculer leur fitness. Une fois tous les individus aléatoires générés, les simulations sont lancées simultanément sur plusieurs processeurs actifs en parallèle.

Ensuite, les individus sont sélectionnés comme parents pour la génération suivante. Le calcul de la fitness des enfants issus du croisement (crossover) est également effectué en parallèle. Pour l'évaluation du meilleur individu de la population courante, nous générons tous les individus mutés et lançons les simulations de manière parallèle pour chacun d'eux. De même, pour les individus générés aléatoirement, nous produisons plusieurs génomes aléatoires et calculons leur fitness en parallèle sur les processeurs actifs.

Cette approche parallèle permet non seulement d'accélérer le processus global de l'algorithme génétique mais aussi d'améliorer l'efficacité des tests en exploitant mieux les ressources matérielles disponibles.

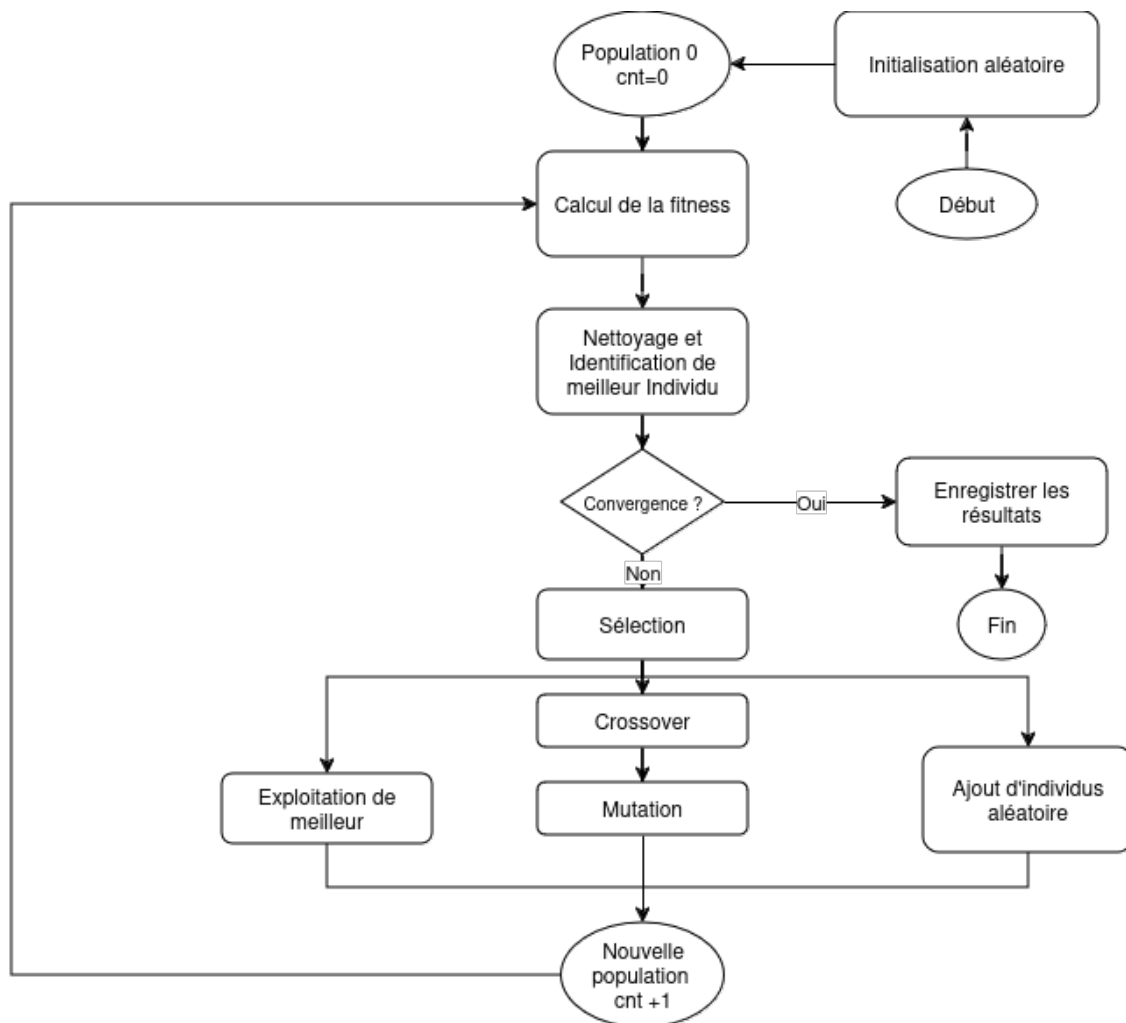


Figure 4.8: Schéma séquentiel de l'algorithme génétique détaillant les étapes du processus depuis l'initialisation jusqu'à l'achèvement d'une génération.

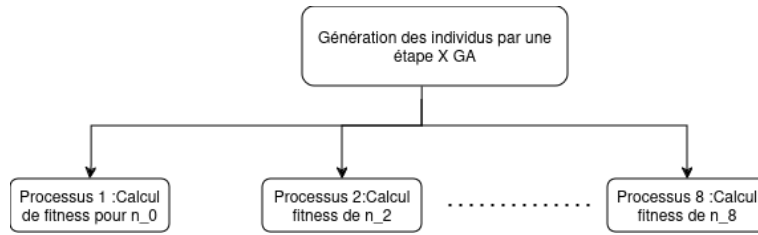


Figure 4.9: Illustration de la parallélisation appliquée après chaque génération d'individus pour différentes étapes de l'algorithme génétique (GA). Les étapes **X** : comprennent la génération de la population initiale, le crossover, l'exploitation du meilleur individu, et l'ajout d'individus aléatoires. Le terme n_i représente le nombre d'individus traités par chaque processeur. Les points indiquent d'autres calculs effectués en parallèle sur les processeurs.

4.7 Fixation des paramètres de l'algorithme génétique

Avant d'implémenter l'algorithme génétique, il est essentiel de l'ajuster afin qu'il explore de manière représentative l'ensemble des solutions possibles. À cet effet, j'ai réalisé une analyse de sensibilité pour évaluer l'impact de la variation des paramètres sur les performances de l'algorithme. Quatre paramètres clés ont été modifiés :

- n_{init} : le nombre d'individus générés aléatoirement lors de l'initialisation pour constituer la première génération,
- n_{save} : le nombre des meilleurs individus conservés automatiquement lors de la sélection,
- P_{mut} : la probabilité qu'un gène mute lors des croisements,
- n_{rand} : le nombre d'individus aléatoires ajoutés à chaque génération.

Nous avons testé entre deux et quatre valeurs différentes pour chacun de ces paramètres. Pour minimiser les perturbations causées par les variations climatiques, nous avons utilisé des données climatiques historiques de la ville de Berne de 1950 à 2030. Le climat de Berne est classé comme Cfb selon la classification de Köppen et Geiger, ce qui indique un climat océanique tempéré avec des étés doux, des hivers modérés et des précipitations importantes tout au long de l'année.

Afin de déterminer les paramètres les plus efficaces, nous avons analysé quatre variables principales : la combinaison produisant la plus grande valeur de fitness, le nombre d'individus explorés à chaque génération, le temps d'exécution pour évaluer l'efficacité des tests, et la durée du processus pour s'assurer que le nombre de générations n'est pas trop restreint. Cette approche nous permet d'éviter de conclure prématurément à un optimum local et de garantir une exploration adéquate de l'espace des solutions.

4.8 Analyse de sensibilité pour le nombre de parcelles

Lors du lancement des simulations, j'ai observé que l'augmentation du nombre de parcelles dans une scène forestière entraînait un allongement significatif du temps de calcul. Pour cette raison, j'ai décidé d'analyser l'impact du nombre de parcelles sur les performances de l'algorithme génétique. Cette analyse a été réalisée en fixant les paramètres de l'algorithme génétique décrits précédemment et en utilisant les mêmes données climatiques historiques de Berne que celles employées pour l'analyse de sensibilité des paramètres de l'AG.

Pour conduire cette analyse, j'ai testé des configurations comprenant entre 10 et 30 parcelles. J'ai examiné les caractéristiques des contraintes mentionnées dans les fichiers de sortie pour chaque individu viable généré par l'algorithme. Plus précisément, j'ai calculé la moyenne de la surface terrière $G(t)$, la valeur finale de la production de bois $B_{\text{tot}}(t)$ – puisque le bois est comptabilisé de manière cumulative –, ainsi que le coefficient de Gini $Gini(t)$, le taux de mortalité $\tau_M(t)$ et le nombre d'espèces pour toutes les années et sur toutes les parcelles. Cela est nécessaire car les données dans un fichier de sortie sont présentées par parcelle et par année.

Ensuite, j'ai évalué les variables fournies par l'algorithme génétique, telles que la meilleure fitness évoluant au fil du temps, le nombre d'individus générés et le temps de simulation pour confirmer que le temps de simulation augmente avec le nombre de parcelles utilisées.

4.9 Méthodes pour analyser les résultats

Afin d'analyser efficacement les résultats, j'ai d'abord procédé à la visualisation des caractéristiques des contraintes générées, présentes dans les fichiers de sortie, en examinant les évolutions temporelles pour les meilleurs

individus viables ainsi que pour les individus non viables. Cette visualisation aide à comprendre comment les différentes contraintes évoluent au fil du temps et leur impact sur la viabilité des individus générés par l'algorithme génétique.

Pour évaluer la densité des individus viables produits par chaque instance de l'algorithme génétique, j'ai analysé l'évolution des populations d'individus viables et non viables. Cette analyse consiste à calculer la moyenne de tous les paramètres des contraintes sur toutes les années pour chaque taille initiale de population dans l'algorithme génétique.

Étant donné la grande dimensionnalité des fichiers de sortie, j'ai effectué une Analyse en Composantes Principales (ACP) pour réduire la dimensionnalité des données et les représenter en 2D. Cette réduction dimensionnelle, grâce à la décorrélation des données, permet une visualisation plus claire et plus facile à interpréter.

Ensuite, pour classifier les individus viables en fonction des caractéristiques des contraintes, j'ai réalisé une classification non supervisée utilisant à la fois l'analyse hiérarchique ascendante et les k-means. Cette approche de clustering aide à identifier comment différentes caractéristiques des contraintes influencent les stratégies de gestion sylvicole et à regrouper les individus qui présentent des comportements similaires en termes de respect des contraintes.

4.9.1 Analyse en Composantes Principales (ACP)

L'Analyse en Composantes Principales (ACP) est une technique statistique utilisée pour réduire la dimensionnalité de grands ensembles de données en transformant un grand ensemble de variables en un plus petit qui contient encore la majeure partie de l'information de l'ensemble large. Dans le contexte de nos données, l'ACP a été utilisée pour identifier les composantes principales qui expliquent le plus de variance au sein de notre jeu de données, ce qui aide à comprendre la structure sous-jacente des données. J'ai commencé par la standardisation des données pour assurer une comparaison équitable entre les variables. Ensuite, j'ai calculé la matrice de covariance pour identifier les relations linéaires entre les variables. La décomposition de cette matrice en valeurs propres et vecteurs propres révèle les directions principales le long desquelles les données varient le plus. Les principales composantes sont sélectionnées en fonction de leur capacité à expliquer une part significative de la variance totale des données.

4.9.2 Classification Hiérarchique Agglomérative (CHA)

La Classification Hiérarchique Agglomérative (CHA) est une méthode d'analyse de clusters visant à construire une hiérarchie de groupes. Dans cette étude, la CHA a été utilisée pour classifier les individus viables selon les caractéristiques des contraintes observées à travers différents patches et années. Cette approche commence par considérer chaque individu comme un cluster indépendant, puis les fusionne progressivement en fonction de la similarité de leurs caractéristiques. Cette similarité est mesurée par la distance de Ward, une méthode particulièrement adaptée aux analyses hiérarchiques car elle vise à minimiser la somme des carrés des distances entre tous les points d'un même cluster. La distance de Ward entre deux clusters A et B est calculée selon la formule suivante :

$$d(A, B) = \sqrt{\frac{|A| \times |B|}{|A| + |B|} \times \|\bar{x}_A - \bar{x}_B\|^2}$$

où $|A|$ et $|B|$ représentent les tailles des clusters A et B , et \bar{x}_A et \bar{x}_B sont les centres de gravité de A et B respectivement. Cette méthode est essentielle pour garantir que la fusion des clusters à chaque étape agglomérative minimise l'augmentation de la variance totale intra-cluster, ce qui facilite l'identification des regroupements naturels au sein des données. Les résultats sont présentés sous forme de dendrogramme, qui permet de déterminer le nombre de classes en choisissant un niveau de coupe interceptant des branches relativement longues.

4.9.3 Classification K-Means

K-Means est un algorithme de classification populaire connu pour sa simplicité et son efficacité. L'algorithme partitionne l'ensemble de données en K classe distincts et non chevauchants en minimisant la somme des carrés intra-cluster. Dans notre application, K-Means a été appliqué pour grouper les individus en clusters basés sur leurs caractéristiques des contraintes. Le nombre optimal de clusters a été déterminé en utilisant la méthode du coude, qui implique de tracer la variation expliquée en fonction du nombre de clusters et de sélectionner le "coude" de la courbe comme le nombre de clusters optimaux. Cette méthode est efficace pour identifier des groupes homogènes d'individus qui présentent des caractéristiques similaires et pour évaluer la distribution globale des traits génétiques à travers la population.

J'ai calculé ensuite pour chaque méthode de classification, l'indice de silhouette afin de déterminer quelle

méthode est la plus précise pour classifier les individus viables. L'indice de silhouette est une mesure de la qualité du clustering qui évalue à quel point chaque objet est bien classé. Cet indice est calculé pour chaque échantillon et est défini comme suit :

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

où $a(i)$ est la distance moyenne entre l'échantillon i et tous les autres points dans le même cluster, et $b(i)$ est la distance moyenne entre l'échantillon i et tous les points dans le cluster le plus proche dont i n'est pas membre. L'indice peut varier entre -1 et 1, où une valeur élevée indique que les classes sont bien séparées et que chaque point est bien placé dans son propre classe. Une valeur proche de zéro suggère que les classes chevauchent, tandis qu'une valeur négative indique que les points pourraient avoir été assignés à de meilleurs classes.

Ces méthodes analytiques ont été mises en œuvre en utilisant les bibliothèques scientifiques de Python, y compris Scikit-Learn pour l'ACP et la classification, Pandas pour la manipulation des données, et Matplotlib et Seaborn pour la visualisation.

5 État de la forêt

5.1 Définition des variables d'état de la forêt

Pour définir les variables d'état d'une forêt, il est important de déterminer les paramètres qui caractérisent la forêt à une année donnée t à la fois à l'échelle de la population et à l'échelle des individus.

Soit $E(t)$ l'état de la forêt à l'année t . Pour bien comprendre l'état d'une forêt, il est essentiel de prendre en compte les différentes dimensions de son écosystème. Nous choisissons de les regrouper en trois groupes.

Commençons par les variables populationnelles, qui se rapportent à la structure et à la dynamique des populations d'organismes vivant dans la forêt, et plus précisément des arbres. Nous pouvons travailler avec la surface terrière totale $G_{\text{tot}}(t)$ ou bien la surface terrière totale pour chaque espèce. Ensuite, le nombre des tiges d'arbres $N_{\text{tree_tot}}(t)$ et à l'échelle des espèces $N_{\text{tree_esp}}(t)$.

Pour la diversité structurelle, nous pouvons utiliser l'écart-type des diamètres ou des hauteurs, ou encore le coefficient de Gini. Comme ces trois variables sont fortement corrélées, comme montré dans la figure ci-dessus, il est judicieux de travailler avec $Gini(t)$.

Pour la diversité du peuplement, il existe différents descripteurs statistiques pour caractériser la diversité des peuplements. Commençons par la richesse spécifique S , qui représente le nombre d'espèces dans un peuplement, et la diversité spécifique, qui prend en compte à la fois le nombre d'espèces et la distribution des individus au sein de ces espèces. Les indices d'équitabilité traduisent la répartition des abondances d'espèces dans le peuplement.

L'indice le plus couramment employé est l'indice de Shannon H' , aussi appelé indice de Shannon-Wiener. Il se calcule par la formule:

$$H' = - \sum_{j=1}^S p_i \log_2(p_i)$$

où p_i est l'abondance proportionnelle ou le pourcentage d'abondance d'une espèce présente ($p_i = \frac{n_i}{N}$), n_i est le nombre d'individus d'une espèce présente, N est le nombre total d'individus dénombrés toutes espèces confondues, et S est le nombre total d'espèces présentes.

Si le peuplement est homogène (constitué d'une seule et même espèce), alors l'indice $H' = 0$. Plus nous avons d'espèces différentes, plus la valeur augmente de façon logarithmique. Il est courant de voir des valeurs de H' comprises entre 1 et 5 pour tenir compte de la diversité spécifique. La valeur maximale $H_{\text{max}} = \log_2(S)$ correspond à un peuplement hétérogène où tous les individus de toutes les espèces sont répartis de manière égale. L'indice H' varie donc entre ces deux limites.

Pour faciliter l'interprétation, l'indice de Shannon est toujours accompagné de l'indice d'équitabilité de Pielou $E = \frac{H'}{H_{\text{max}}}$. Cet indice varie entre 0 et 1. S'il tend vers 1, alors les espèces présentes dans le peuplement ont des abondances identiques. S'il tend vers 0, alors nous observons un déséquilibre où une seule espèce domine tout le peuplement. L'apparition d'une espèce invasive ou des pics d'opportunistes entraînent une baisse significative conjointe de H' et E .

Puis, il y a l'indice de Simpson, introduit par le statisticien Edward H. Simpson, qui mesure la probabilité que deux individus pris au hasard appartiennent au même groupe. Plusieurs formes de l'indice existent dans la littérature scientifique, c'est pourquoi nous utiliserons la formule originale de Simpson en 1949 $P(w) = \sum p_i^2$. Deux formules en découlent selon l'échantillon : pour les tirages sans remise (indice L), sinon pour les tirages avec remise (indice λ):

$$\lambda = \sum_{i=1}^S p_i^2$$
$$L = \frac{\sum n_i(n_i - 1)}{N(N - 1)}$$

où p_i est la proportion d'individus de l'espèce i ($p_i = \frac{n_i}{N}$), n_i est le nombre d'individus de l'espèce i , N est le nombre total d'individus, et S est le nombre total d'espèces présentes.

Plus cet indice est proche de 1, plus le peuplement est homogène.

Il y a deux approches pour cet indice : soit un indice de diversité plus intuitif pour la lecture avec la formule $D = 1 - L$ ou $D = 1 - \lambda$. Lorsque $D = 0$, une seule espèce est présente dans le peuplement. Si toutes les espèces ont la même probabilité $p_i = 1/S$ de présence, alors $D = 1 - \frac{1}{S}$. Enfin, la valeur $D = 1$ apparaîtrait dans le cas de figure où un nombre infini d'espèces sont présentes, mais toutes de probabilité quasi nulle.

Une autre formulation de l'indice de Simpson stipule que $D = \frac{1}{\lambda}$ (indice de diversité statistique). Il devient ainsi possible de considérer que si $E_D = \frac{D}{D_{\max}}$ alors $D_{\max} = S$. Soit l'équation $E_D = \frac{D}{S}$. Le choix de la formule est donc crucial pour comparer les résultats entre eux.

L'indice d'équitabilité de Simpson permet de traduire la dominance d'une espèce (E_D tend vers 0) ou la codominance de plusieurs espèces (E_D tend vers 1).

Pour la physiologie de la forêt, nous avons choisi de considérer l'indice de sécheresse moyen par espèce $Drought_{\text{index_mean}}$ pour représenter les changements climatiques, ainsi que la lumière accessible dans chaque strate $light_{\text{availability_mean}}$ pour indiquer la compétition lumineuse entre les espèces.

Nous avons également décidé de prendre en compte l'historique de l'état de la forêt à l'année t par rapport à la dernière coupe pour observer comment la forêt a évolué après cette coupe. Nous pouvons l'appeler Δ_{recup} , qui représente la pente d'une variable d'état, par exemple $G_{\text{tot}}(t)$, entre la dernière éclaircie et l'année actuelle.

Donc, voici un récapitulatif pour représenter l'état de la forêt :

$$E(t) = \left\{ (V_{\text{pop}}(t), V_{\phi}(t), V_{\text{recup}}(t)) \forall t \in [0, T] \right\}$$

tel que

$$V_{\text{pop}}(t) = \left\{ (G_{\text{tot}}(t), N_{\text{tree_tot}}(t), I_{\text{diversité}}(t), Gini(t)) \forall t \in [1, T] \right\}$$

$$V_{\phi}(t) = \left\{ (drought_index_mean(t), light_availability_mean(t)) \forall t \in [0, T] \right\}$$

$$V_{\text{recup}}(t) = \Delta_{\text{recup}}(t) = \frac{G(t) - G(t - \theta(t))}{\theta(t)}$$

tel que $\theta(t)$ est la période entre les deux coupes.

5.2 Méthode 1: Le contrôle et l'état de la forêt

Après avoir obtenu tous les itinéraires sylvicoles par l'algorithme génétique, les gestionnaires ont besoin de savoir quel contrôle appliquer en partant d'un état de la forêt. Il est donc nécessaire d'exprimer ce contrôle en fonction de l'état de la forêt, c'est-à-dire de trouver la fonction f qui lie le contrôle U et l'état E . Ainsi, on peut écrire :

$$u(t_c) = f(e(t_c - 1)).$$

Matriciellement, on peut définir $U = [u_{ij}]$, où U est la matrice de toutes les variables de contrôle pour chaque simulation d'un itinéraire sylvicole. Ici, i représente la ligne des variables de contrôle $[\theta, G_{\text{obj}}, t_p, C_{f1}, C_{f2}, C_{f3}]$, et j représente les observations obtenues pour chaque itinéraire sylvicole, avec j prenant des valeurs de 0 à 9305 pour le jeu de données considéré.

On définit également la matrice $E = [e_{ij}]$, qui est la matrice des variables d'état pour chaque simulation. Ici, i représente :

- BA : Surface terrière totale de la forêt avant coupe, c'est-à-dire à l'année $t_c - 1$.
- N_{tree} : Nombre total d'arbres à $t_c - 1$.
- Pielou : Indice d'équirépartition des espèces à l'année $t_c - 1$.
- drought_index : Indice de sécheresse moyenne totale pour l'ensemble des espèces.
- light_mean_index : Indice de lumière moyenne totale pour l'ensemble des espèces.
- Gini : Coefficient de diversité structurelle pour l'ensemble des arbres.

L'objectif est donc de trouver la matrice M telle que :

$$U = ME,$$

5.2.1 Résolution par moindres carrés

Pour résoudre ce problème, nous cherchons à déterminer la matrice M et le biais b . Nous pouvons utiliser la méthode des moindres carrés pour minimiser la norme suivante :

$$\|U - ME\|$$

Cependant, lorsque le nombre de lignes est inférieur au nombre de colonnes dans la matrice E , E n'est pas inversible. Pour contourner ce problème, nous avons divisé les matrices en blocs de taille 6×6 . Ainsi, nous avons 6 lignes et 9305 colonnes, ce qui génère 1550 blocs de matrice M .

La fonction `lstsq` de la bibliothèque Python est conçue pour résoudre le problème pour $U = EM$, et non pour $U = ME$. Afin de résoudre le problème pour $U = ME$, nous devons reformuler le problème.

Pour ce faire, nous définissons une nouvelle matrice M' telle que :

$$U = EM'$$

nous obtenons donc :

$$EM' = ME$$

Si E est carrée et inversible (ce qui est le cas ici car toutes les matrices ont un déterminant non nul), nous pouvons isoler M' en multipliant par l'inverse de E :

$$M' = E^{-1}ME$$

Ensuite, la matrice M est obtenue par :

$$M = EM'E^{-1}$$

Cela nous permet de résoudre le problème de moindres carrés tout en utilisant les blocs pour gérer la taille des matrices et la bibliothèque `lstsq` pour ajuster les paramètres dans les limites de ce cadre.

5.2.2 Régression Linéaire avec Ridge

Pour résoudre le problème de régression linéaire avec régularisation Ridge, nous cherchons à déterminer la matrice M et le biais b dans le modèle $U = ME + b$. La fonction objectif à minimiser est :

$$J(M, b) = \|U - ME - b \cdot \mathbf{1}\|_2^2 + \lambda \|M\|_2^2$$

où :

- b est le biais ,
- λ est le paramètre de régularisation.

En incluant le biais b dans le modèle, nous réécrivons U comme suit :

$$U = ME + b \cdot \mathbf{1}$$

où $\mathbf{1}$ est un vecteur colonne de dimension 6 avec toutes les valeurs égales à 1.

La fonction de coût Ridge devient :

$$J(M, b) = \|U - ME - b \cdot \mathbf{1}\|_2^2 + \lambda \|M\|_2^2$$

Pour résoudre ce problème, nous reformulons le problème en termes matriciels. La formulation matricielle de la fonction de coût peut être exprimée comme suit :

$$\begin{bmatrix} E^T E + \lambda I & E^T \mathbf{1} \\ \mathbf{1}^T E & \mathbf{1}^T \mathbf{1} \end{bmatrix} \begin{bmatrix} M \\ b \end{bmatrix} = \begin{bmatrix} E^T U \\ \mathbf{1}^T U \end{bmatrix}$$

où I est la matrice identité de taille appropriée à M , et $E^T \mathbf{1}$ et $\mathbf{1}^T E$ représentent les produits scalaires entre E et $\mathbf{1}$.

5.3 Méthode 2 : Trouver la relation avec l'ACP et K-means

5.3.1 ACP et K-means

Pour explorer la relation entre l'état de la forêt E et le contrôle U , nous avons combiné l'Analyse en Composantes Principales (ACP) avec l'algorithme K-means. Cette approche vise à réduire la dimensionnalité des données tout en identifiant des clusters naturels. Nous avons commencé par standardiser les données en soustrayant la moyenne et en divisant par l'écart-type de chaque variable. Cette étape est essentielle pour garantir que toutes les variables sont sur une échelle comparable avant de procéder à l'ACP.

L'ACP est ensuite appliquée pour réduire la dimensionnalité des données tout en conservant une grande partie de la variance. Pour cette analyse, nous avons sélectionné les variables des ensembles $V_{\text{pop}}(t)$ et $V_{\phi}(t)$. La matrice des données standardisées \mathbf{X} est alors construite à partir de ces variables. La première étape consiste à calculer la matrice de covariance, définie par :

$$\mathbf{C} = \frac{1}{n-1} \mathbf{X}^T \mathbf{X}$$

où n représente le nombre d'observations. Cette matrice de covariance est ensuite décomposée en valeurs propres et vecteurs propres :

$$\mathbf{C} = \mathbf{P} \mathbf{\Lambda} \mathbf{P}^T$$

où \mathbf{P} est la matrice des vecteurs propres et $\mathbf{\Lambda}$ est la matrice diagonale des valeurs propres. Les données sont projetées sur les composantes principales en utilisant les vecteurs propres :

$$\mathbf{Z} = \mathbf{X} \mathbf{P}$$

Les composantes principales sont alors classées par ordre décroissant de la variance expliquée, calculée comme :

$$\text{Variance expliquée} = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}$$

où λ_i est la valeur propre de la i -ème composante et p est le nombre total de composantes.

Après avoir réduit la dimensionnalité des données à deux composantes principales, nous avons appliqué l'algorithme de classification K-means pour regrouper les observations en clusters. L'algorithme K-means partitionne les observations en k clusters en suivant plusieurs étapes. D'abord, les centres des clusters sont initialisés de manière aléatoire. Ensuite, chaque observation est assignée au cluster dont le centre est le plus proche, ce qui est formulé mathématiquement par :

$$c_i = \arg \min_j \|\mathbf{x}_i - \mu_j\|^2$$

où \mathbf{x}_i est un vecteur d'observation et μ_j est le centre du j -ème cluster. Les centres des clusters sont ensuite mis à jour en calculant la moyenne des observations assignées à chaque cluster :

$$\mu_j = \frac{1}{|C_j|} \sum_{i \in C_j} \mathbf{x}_i$$

où $|C_j|$ est le nombre d'observations dans le j -ème cluster et C_j est l'ensemble des indices des observations appartenant au j -ème cluster. L'algorithme converge lorsque les centres des clusters ne changent plus ou lorsque le nombre d'itérations atteint une limite prédéfinie. Cette approche permet d'identifier des structures sous-jacentes dans les données et de regrouper les observations en clusters pertinents.

5.3.2 Relation entre l'état de la forêt et le contrôle appliqué

L'objectif de cette étude est de déterminer le contrôle à appliquer à partir de l'état de la forêt à l'année t , noté $e(t)$, sans considérer l'effet direct, car nous savons déjà que, selon l'algorithme, l'état de la forêt est viable et respecte toutes les contraintes. Soit $e(t)$ l'état de la forêt à l'année t . Le contrôle appliqué à cet état est noté $u(t)$. Nous cherchons donc une fonction f telle que :

$$u(t) = f(e(t-1)).$$

Définissons l'ensemble des années tous les 5 ans comme :

$$T_{5 \text{ ans}} = \{t_0 + 5i \mid i \in [1, 15]\}.$$

Soit \mathcal{E} l'ensemble des états de la forêt aux années de l'ensemble $T_{5 \text{ ans}}$:

$$\mathcal{E} = \{E(t_c - 1) \mid t_c \in T_{5 \text{ ans}}\}.$$

L'ensemble \mathcal{E} est divisé en 3 groupes à l'aide de K-means et de l'ACP, ce qui permet d'écrire :

$$\mathcal{E} = (E_1(t_c - 1) \cup E_2(t_c - 1) \cup E_3(t_c - 1)),$$

où $E_1(t_c - 1)$, $E_2(t_c - 1)$, et $E_3(t_c - 1)$ représentent les groupes identifiés.

Nous cherchons donc une fonction f qui relie l'état de la forêt au contrôle appliqué, de sorte que :

$$\mathcal{U} : (\{5, 10, 15, 20\} \times V)^{\text{nb}_{\text{inv}}} \longrightarrow E_1 \cup E_2 \cup E_3 = \mathcal{E}$$

$$u(t) = (\theta_k, v(t_c))_{k=1, \dots, \text{nb}_{\text{inv}}} \longrightarrow (e_1(t_c - 1), e_2(t_c - 1), e_3(t_c - 1))$$

Soit $\mathcal{P} = \{5, 10, 15, 20\}$ l'ensemble des périodes de coupe. Pour chaque classe C_i , le nombre total de configurations possibles est donné par :

$$N_{\text{config}} = n_{\theta} \times n_{\text{périodes}} \times n_{G_{\text{obj}}} \times n_{C_{\text{fi}}} = 4 \times 5 \times 3 \times 121.$$

Nous avons examiné la densité de chaque variable de contrôle u dans chaque classe C_i pour chaque période $p \in \mathcal{P}$:

$$\rho_{u, C_i, p} = \frac{1}{N_{C_i, p}} \sum_{j=1}^{N_{C_i, p}} \delta(u - u_j),$$

où $N_{C_i, p}$ est le nombre total d'observations dans la classe C_i pour la période p et $\delta(u - u_j)$ est une fonction indicatrice qui vaut 1 si $u = u_j$ et 0 sinon.

5.4 L'effet du controle sur l'Etat et la viabilité de système

Après avoir examiné les différents aspects, nous nous sommes posés la question suivante : quel est l'effet du contrôle sur l'état de la forêt et sur la viabilité du système ? Pour répondre à cette question, nous avons calculé des matrices de transition qui illustrent la probabilité de passer d'un état à un autre, en présence ou en absence d'un contrôle. En observant la probabilité de transition d'un état à un autre sous l'effet d'un contrôle u à l'année de la coupe $t_c \in T_{5 \text{ ans}}$, nous pouvons mieux comprendre l'impact de ce contrôle.

5.4.1 Matrice de Transition

Une matrice de transition est utilisée pour décrire les probabilités de transition d'un état à un autre dans un système dynamique. Soit $E(t)$ l'état de la forêt à l'année t et $u(t)$ le contrôle appliqué. L'ensemble des états possibles de la forêt est noté S , et $P_{ij}(u, t)$ représente la probabilité de transition de l'état i à l'état j sous l'effet d'un contrôle u à l'année $t_c - 1$.

Ainsi, la matrice de transition $\mathbf{P}(u, t)$ est une matrice carrée de dimension $|S| \times |S|$, où chaque élément $P_{ij}(u, t)$ indique la probabilité de transition de l'état i à l'état j . La matrice peut être écrite comme suit :

$$\mathbf{P}(u, t) = \begin{bmatrix} P_{11}(u, t) & P_{12}(u, t) & \cdots & P_{1|S|}(u, t) \\ P_{21}(u, t) & P_{22}(u, t) & \cdots & P_{2|S|}(u, t) \\ \vdots & \vdots & \ddots & \vdots \\ P_{|S|1}(u, t) & P_{|S|2}(u, t) & \cdots & P_{|S||S|}(u, t) \end{bmatrix}$$

où :

1. **Probabilités de Transition** : Chaque élément de la matrice $\mathbf{P}(u, t)$ est une probabilité, donc $0 \leq P_{ij}(u, t) \leq 1$.
2. **Somme des Probabilités** : La somme des probabilités de transition de chaque état i vers tous les autres états doit être égale à 1 :

$$\sum_{j=1}^{|S|} P_{ij}(u, t) = 1 \quad \text{pour tout } i.$$

Pour calculer les probabilités de transition $P_{ij}(u, t)$, nous utilisons les fréquences observées des transitions dans les simulations effectuées tous les 5 ans. Soit $N_{ij}(u, t)$ le nombre de transitions observées de l'état i à l'état j sous le contrôle u à l'année t , et $N_i(u, t)$ le nombre total de transitions observées à partir de l'état i . La probabilité de transition est alors estimée par :

$$P_{ij}(u, t) = \frac{N_{ij}(u, t)}{N_i(u, t)}.$$

5.4.2 L'effet sur la non-viabilité : $V(E(t + \theta_s - 1)) = f(V(E(t_c - 1)), u(t_c))$

Pour calculer la probabilité de transition d'un état viable à un état non viable en fonction du contrôle appliqué $u(t)$ ou sans contrôle (c'est-à-dire sans coupe), j'ai d'abord traduit le génome en une séquence de 0 et 1, où 1 correspond à une coupe et 0 à l'absence de coupe. Ensuite, j'ai examiné la viabilité de l'état de la forêt avant la coupe à $t_c - 1$ (où t_c est l'année de la coupe) ainsi que la viabilité de l'état de la forêt après la coupe, c'est-à-dire $V(E(t + \theta_s - 1))$.

La viabilité de l'état de la forêt est définie comme suit :

$$V(E(t)) = \begin{cases} 1 & \text{si } E(t) \text{ respecte toutes les contraintes,} \\ 0 & \text{sinon.} \end{cases}$$

Cette analyse est effectuée sur le long terme. Sur le court terme, j'ai examiné l'état de la forêt immédiatement après la coupe, c'est-à-dire à l'année t_c , afin d'observer l'effet direct.

Pour relier cela au contrôle appliqué $u(t)$, j'ai calculé la matrice de transition par groupe de variables de contrôle entre $G_{\text{obj}}(t)$ et $\theta_s(t)$, correspondant à la période suivante. Cela permet de déterminer que si l'on se trouve à l'année t_c et que l'on applique un contrôle avec la variable de contrôle $G_{\text{obj}}(t)$, sachant que la prochaine éclaircie sera après θ_s , alors il y a une plus grande probabilité de rester viable avec ce contrôle qu'avec un autre contrôle.

De plus, pour examiner plus en détail les raisons de la non-viabilité, j'ai classé les variables d'état en trois classes avec k-means. Pour chaque classe, j'ai recalculé cette matrice de transition pour les combinaisons entre les différentes variables de contrôle. J'ai également pu établir ces matrices de transition avec des intervalles sur les variables d'état, notamment entre la surface terrière, le coefficient de Gini, et l'indice de Pielou. Cela permet de déterminer la probabilité de transition vers un autre intervalle, étant donné ce contrôle et l'appartenance à une classe spécifique.

5.4.3 L'effet du controle sur l'Etat $E(t)$: $E(t + \theta_s - 1) = f(E(t_c - 1), u(t_c))$

Dans l'espace des individus viables trouvé avec l'algorithme génétique, j'ai examiné l'état de la forêt avant la coupe à $t_c - 1$ ainsi que les états à $t + 4$ et à $t + \theta_s - 1$ après coupe. J'ai choisi d'analyser l'ensemble des variables d'état en les divisant en plusieurs intervalles, d'abord avec coupe et sans coupe, puis, au sein des scénarios de coupe, en observant la matrice de transition entre les combinaisons des variables de contrôle $G_{\text{obj}}(t)$ et θ_s , ainsi qu'entre t_p et θ_s et entre les Cf_i et θ_s .

Pour cela, j'ai calculé les matrices de transition P définies par :

$$P_{ij} = \mathbb{P}(E(t + \theta_s - 1) = j \mid E(t_c - 1) = i, u(t) \text{ appliqué}),$$

où $E(t)$ représente l'état de la forêt à l'année t , i et j sont des états spécifiques, et $u(t)$ est le contrôle appliqué.

De plus, comme pour la méthode de l'effet sur la non-viabilité, j'ai classifié les variables d'état en trois classes à l'aide de k-means. Pour chaque classe, j'ai recalculé cette matrice de transition en fonction des différents contrôles appliqués. Cela permet d'analyser plus précisément les effets des contrôles sur la transition entre états dans chaque classe.

6 Résultats et analyse

6.1 Choix de données

6.2 Données d'entrée et configuration des simulations

Les données d'entrée utilisées pour nos simulations sont les suivantes : la surface de la placette étudiée est de 1000 m^2 , avec une latitude de 43.93. La teneur en nutriments du sol est mesurée à $N_{\text{soil}} = 100$ et l'indice d'abrutissement, représentant la présence d'animaux, est de $BrI = 0.2$. Le point de départ des simulations, x_0 , a été établi grâce à un inventaire que j'ai créé en utilisant des données aléatoires, tout en fixant des limites supérieures et inférieures pour certaines variables : le diamètre des arbres varie de 5 cm à 50 cm et l'âge de 1 an à 100 ans. J'ai également sélectionné trois types d'espèces pour l'étude : le Pin sylvestre (*Pinus sylvestris*), le Tilleul à petites feuilles (*Tilia cordata*) et le Chêne pédonculé (*Quercus petraea*). Les simulations initiales ont été menées en utilisant un climat historique stable, couvrant la période de 1950 à 2030, comme illustré dans les figures ci-dessous.

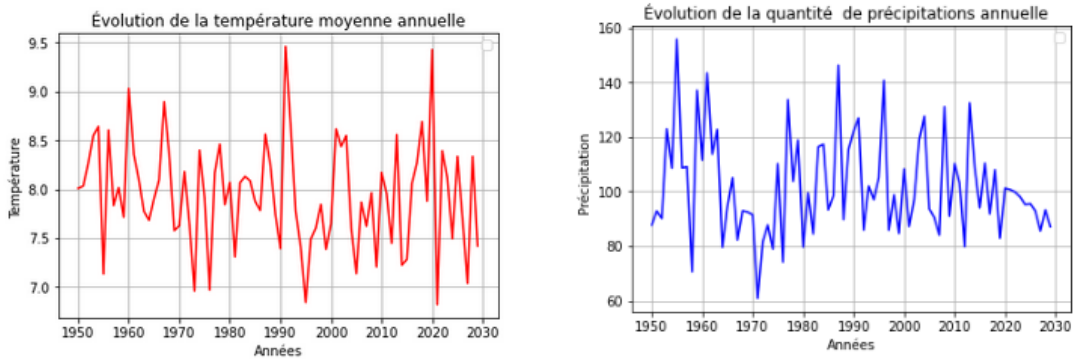


Figure 6.1: Évolution de la température annuelle et de la somme des précipitations annuelles de 1950 à 2030

6.3 Choix du nombre de parcelles

Comme nous l'avons déjà précisé, le but de cette analyse est de déterminer si travailler avec seulement 10 patches, au lieu de 30, affecte nos résultats ultérieurs. Pour cette analyse, nous avons fixé les paramètres suivants pour l'algorithme génétique (GA) : $n_{\text{init}} = 4$, $n_{\text{rand}} = 0$, $n_{\text{save}} = 3$ et $P_{\text{mut}} = 0.1$. Nous avons constaté qu'il n'y a pas de différence notable concernant la moyenne des caractéristiques de tous les individus viables. Cependant, il est à noter que la valeur de la fitness est plus élevée pour 30 patches que pour 10. On peut expliquer cela par le fait qu'il y a 100 arbres pour chaque parcelle et selon le calcul de la fitness, la surface terrière, par exemple, augmente avec le nombre d'arbres et de parcelles, ce qui affecte également l'indice de Gini. Pour les autres caractéristiques des contraintes, cela dépend de la gestion effectuée et du climat. Pour le nombre d'espèces et le taux de mortalité, j'ai bien remarqué que c'est le même pour les deux, car j'utilise le même climat et les mêmes espèces. Quant aux histogrammes, cela est dû au fait que la moyenne est calculée sur toutes les années. En ce qui concerne le temps de simulation, nous observons que le temps nécessaire pour les 30 patches est quatre fois plus élevé que pour les 10 patches. Pour cette raison, nous avons décidé de réaliser les tests subséquents avec 10 patches plutôt qu'avec 30.

Nombre de parcelles	Durée en temps	Individus viables	Nombre de générations
10	459.9	17	10
30	1837.6	25	14

Table 6.1: Comparaison entre 10 et 30 patches pour les paramètres de l'algorithme génétique.

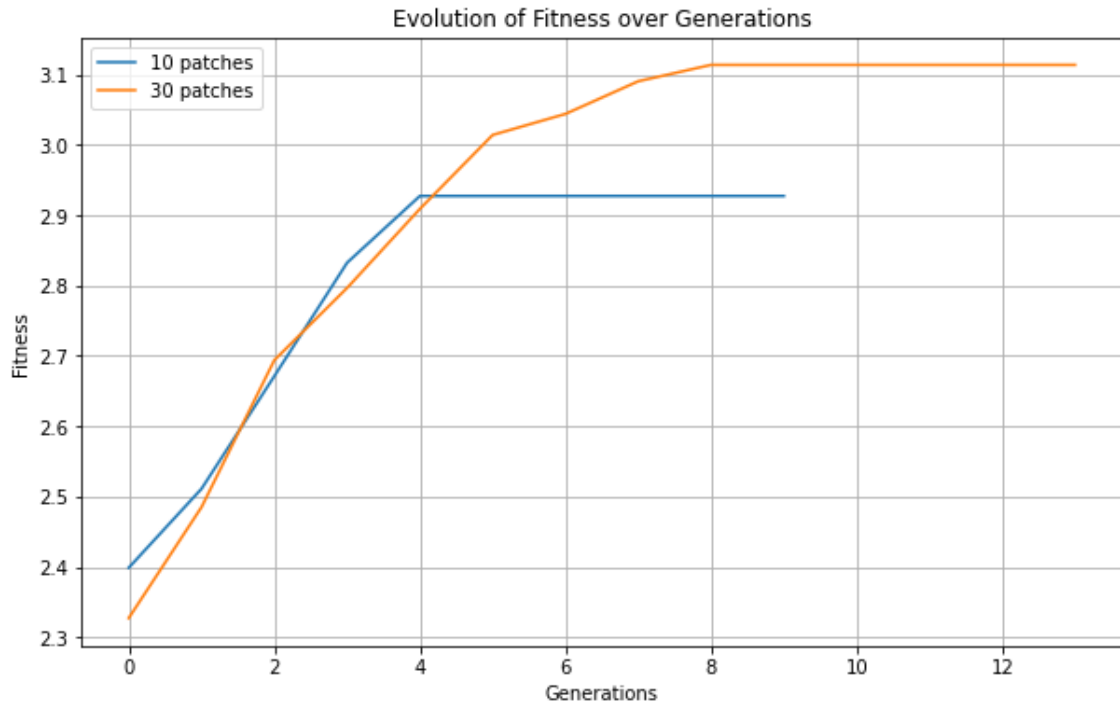


Figure 6.2: Comparaison de l'évolution du meilleur fitness au cours des générations entre 10 et 30 patches pour un seul réplicat .

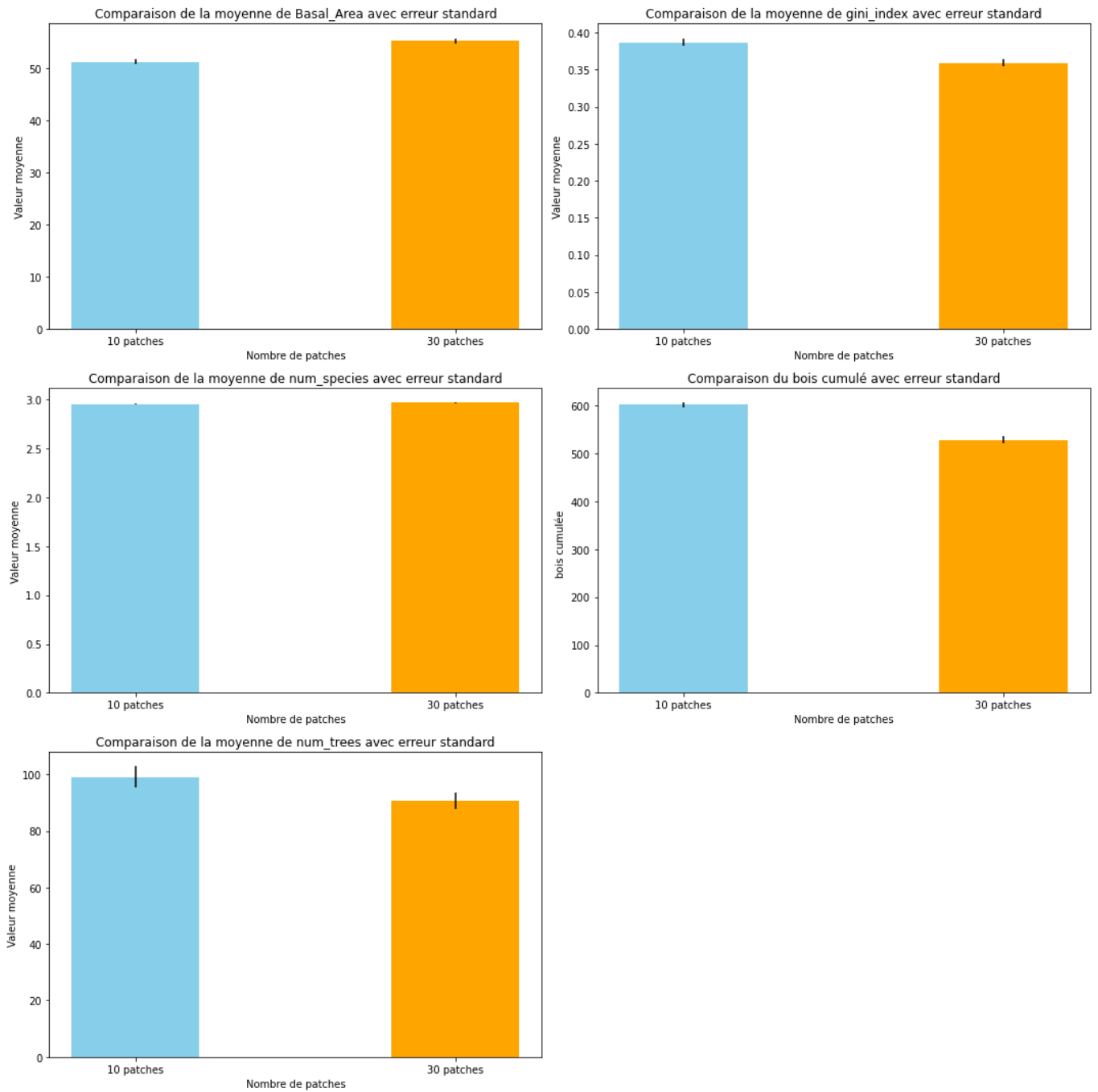


Figure 6.3: Histogrammes illustrant la différence entre les moyennes des caractéristiques sur 80 années pour 10 et 30 patches. La moyenne a été calculée pour chaque caractéristique, sauf pour le bois récolté où la dernière valeur a été utilisée pour représenter le bois cumulé sur 80 années et Les moyennes ont été calculées en fonction du nombre de parcelles (10 et 30).

6.4 Analyse de sensibilité pour l'algorithme génétique

L'objectif de cette analyse est de déterminer les paramètres de l'algorithme génétique, compte tenu de l'augmentation de la dimensionnalité de l'espace à explorer. Pour cela, nous avons testé toutes les combinaisons possibles des paramètres suivants dans les intervalles définis : $n_{init} = [50, 100]$, $n_{rand} = [5, 10]$, $n_{save} = [3, 5]$, et $P_{mut} = [0, 0.05, 0.1]$.

La combinaison retenue est celle qui permet une exploration la plus efficace, c'est-à-dire celle qui conduit aux meilleures valeurs de fitness, au nombre de générations le plus optimal et au plus grand nombre d'individus viables. Les paramètres sélectionnés sont : $n_{init} = 100$, $n_{rand} = 10$, $n_{save} = 5$ et $P_{mut} = 0.05$. Cette combinaison a été identifiée comme la plus performante pour poursuivre les analyses suivantes.

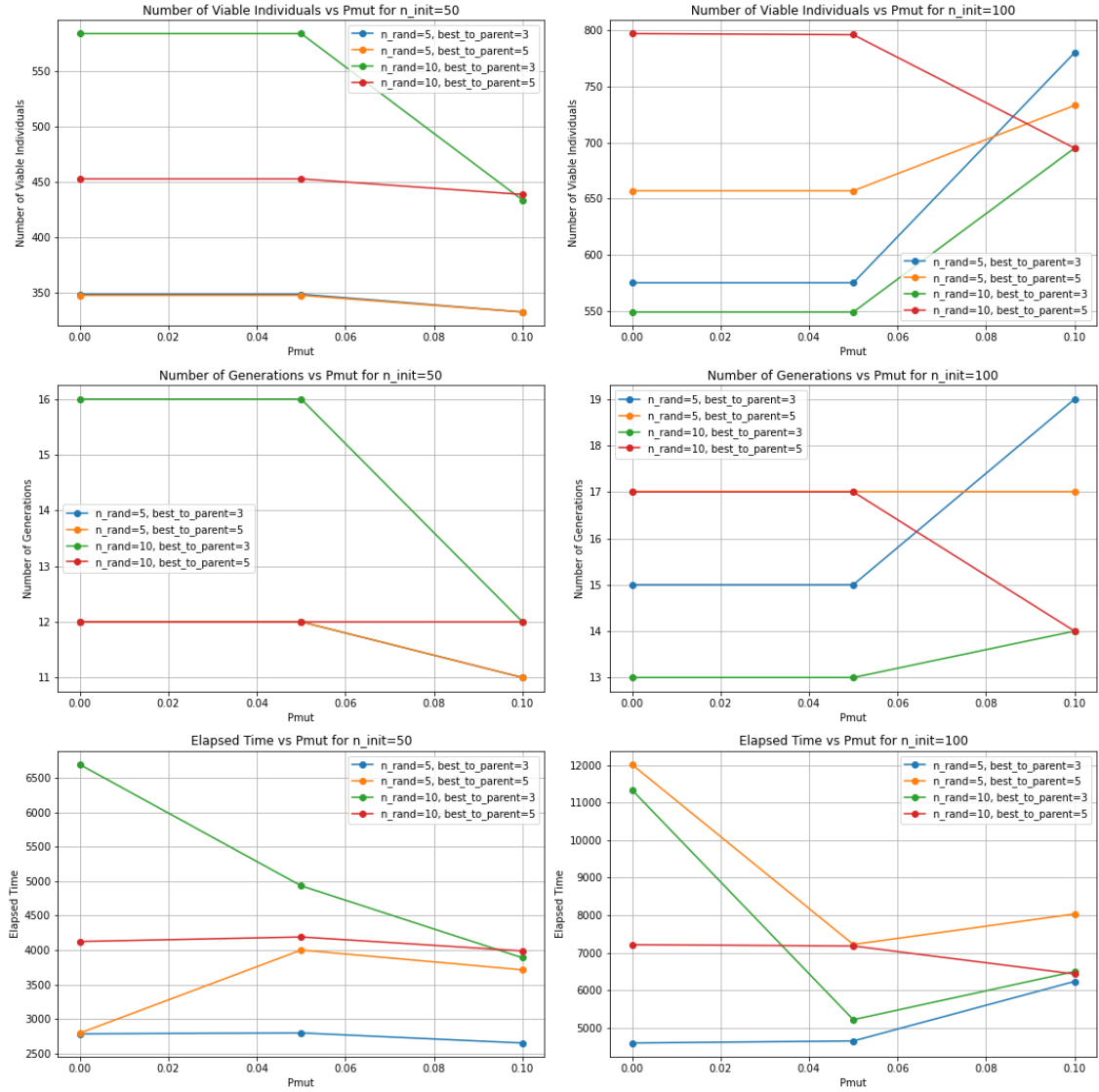


Figure 6.4: Résultat de l'analyse de sensibilité des paramètres de l'algorithme génétique avec un climat historique. Nombre de générations, nombre d'individus viables calculés lors du GA en fonction de P_{mut} pour différentes configurations de n_{save} et n_{rand} et pour $n_{init} = 100$ (à droite) et $n_{init} = 50$ (à gauche).

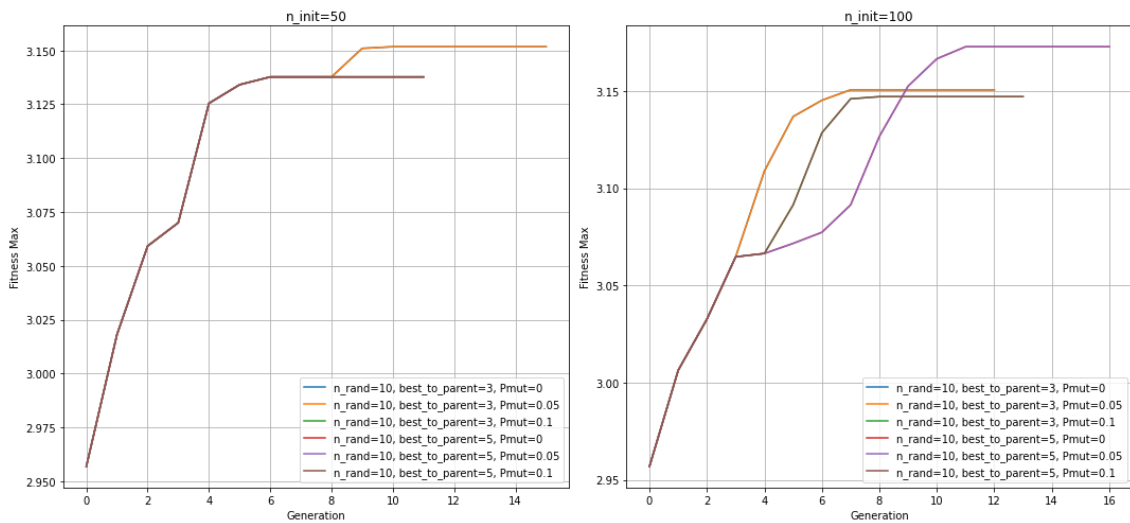


Figure 6.5: Évolution de la fitness maximale par génération selon différents paramètres de l'algorithme génétique pour un seul réplicat.

6.5 Résultats de l'ensemble des individus viables

Après avoir établi les paramètres de perturbation pour l'algorithme génétique et le nombre de parcelles, j'ai présenté en trois dimensions la distribution des individus viables et non viables, en calculant la moyenne sur toutes les années et sur tous les patchs pour la surface terrière, l'indice de Gini et le bois récolté pour chaque individu fourni par l'algorithme génétique entre deux nombres d'individus de la population initiale. Je l'ai présenté entre 500 et 1000 individus.

On peut bien remarquer que certains individus non viables se chevauchent avec des individus viables parce que les individus peuvent sortir de l'espace des contraintes s'il y a au moins une variable parmi la surface terrière, l'indice de Gini, le bois récolté, le taux de mortalité ou le nombre d'espèces qui ne respectent pas les contraintes déjà définies en moyenne sur l'ensemble des patchs. Ainsi, on peut trouver des individus non viables qui respectent l'indice de Gini et la surface terrière, mais qui n'ont pas pu atteindre la valeur annuelle de bois récolté de $3,1 \text{ m}^3/\text{ha}/\text{an}$.

On peut également constater que l'augmentation de la population initiale augmente le nombre d'individus viables trouvés. En effet, on remarque que l'ensemble pour $n_{\text{init}} = 500$ est plus discontinu par rapport à $n_{\text{init}} = 1000$, qui est plus dense en termes d'individus viables trouvés. Donc on peut dire qu'il sera préférable de travailler avec 1000 individus comme population initiale pour mieux explorer l'espace de recherche.

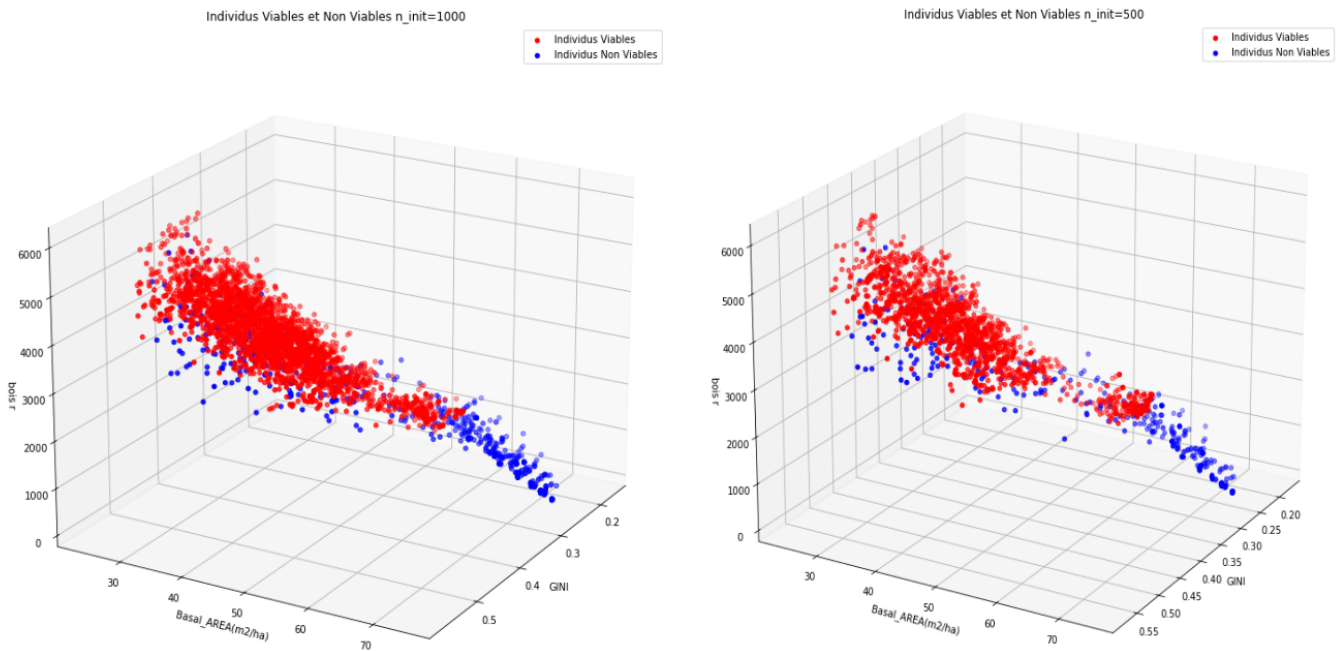


Figure 6.6: Analyse comparative de la viabilité des individus en 3D pour différentes tailles de populations initiales, à droite pour $n_{\text{init}}=500$ et à gauche pour $n_{\text{init}}=1000$

6.6 Résultats de l'ACP et La classification

6.6.1 Matrice de corrélation

La matrice de corrélation entre les variables nous donne un premier aperçu des interactions entre les caractéristiques écologiques et la gestion forestière. Il est notable que la surface terrière présente une forte corrélation négative avec l'indice de Gini et le volume de bois récolté. Cette relation peut s'expliquer par le fait que, lors de coupes intensives, le volume de bois récolté augmente et que cette coupe introduit une diversité structurelle élevée au sein de la forêt. Cette coupe réduit, à son tour, la surface terrière restante, car elle implique souvent l'abattage d'arbres dominants ou structurants, ce qui diminue la densité globale du couvert forestier et augmente la lumière qui atteint le sol. L'augmentation de la lumière favorise la croissance de certaines espèces végétales et accélère le processus de régénération, ce qui peut expliquer pourquoi l'indice de Gini augmente.

D'un autre côté, on peut expliquer cette relation par le fait que des coupes effectuées sur une longue période, comme dans notre cas tous les 20 ans, induisent une faible quantité de bois récolté car moins d'arbres sont

abattus, permettant aux arbres de pousser . Cela augmente la surface terrière et diminue la diversité structurale.

De plus, on remarque que le nombre d'espèces est moins corrélé par rapport aux autres caractéristiques. Cela peut s'expliquer par le fait que nous avons choisi de travailler uniquement avec trois espèces, et que la composition varie entre ces trois espèces sur l'intervalle $\{0, 20, 40, 60, 80, 100\}$. Souvent, l'algorithme génétique a généré des combinaisons mixtes. C'est pour cette raison que, dans la plupart des simulations, nous nous retrouvons avec soit trois espèces, soit deux.

En outre, on remarque aussi que le bois récolté et l'indice de Gini sont fortement corrélés, c'est-à-dire que lorsque l'on récolte du bois, cela augmente la diversité structurale de la forêt et vice versa. On peut expliquer cela par le fait que la récolte de bois peut créer des ouvertures dans la canopée, permettant ainsi à plus de lumière d'atteindre le sol. Cela favorise la croissance de différentes espèces végétales et la régénération d'autres espèces, ce qui entraîne une plus grande diversité structurale.

En revanche, si l'on ne récolte pas trop de bois, cela peut entraîner une compétition pour la lumière entre les espèces végétales dans la forêt. Cette compétition peut limiter la croissance des espèces moins adaptées à des conditions de faible luminosité.

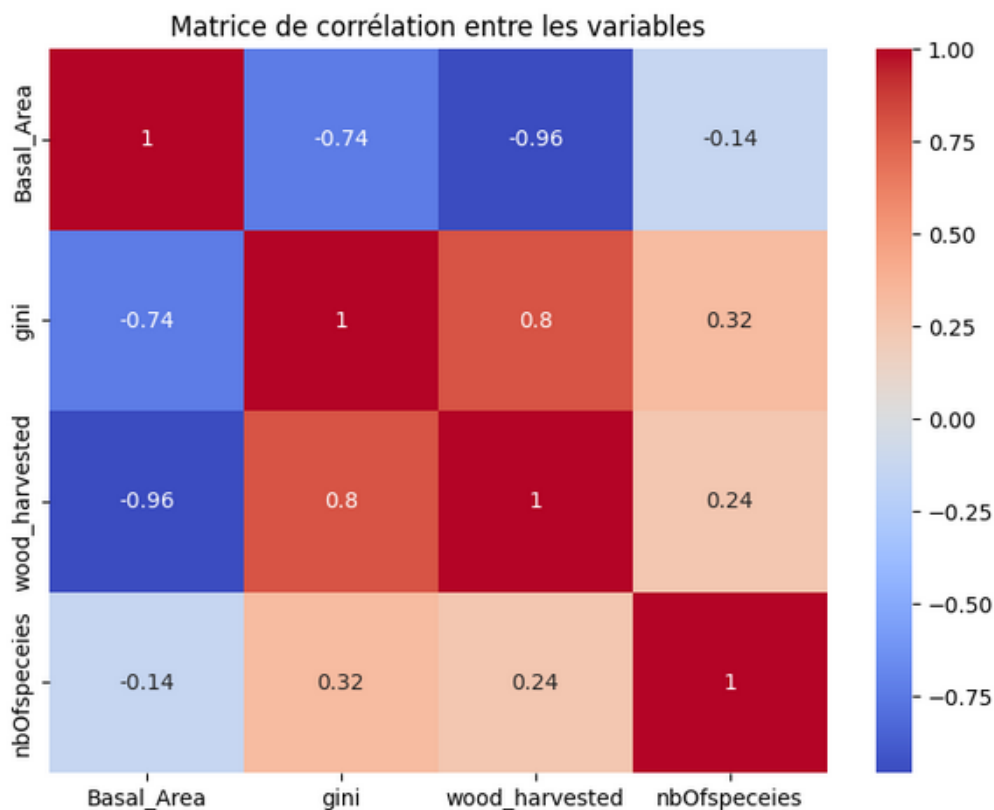


Figure 6.7: Matrice de corrélation entre tous les caractéristiques des contraintes sauf le taux de mortalité car c'est nulle durant toute la simulation.

6.6.2 Analyse en composantes principales

Pour la contribution à la première et à la deuxième composante principale (PC1 et PC2) révèlent des aspects significatifs de la gestion forestière et de la biodiversité.

Pour la **PC1** :

- La *surface terrière* montre une forte contribution négative, indiquant son impact inverse sur la densité forestière dans la composante.
- L'*indice de Gini* et le *volume de bois récolté* présentent des contributions positives, suggérant que des augmentations dans ces variables sont associées à des valeurs élevées de PC1, reflétant des interventions humaines.
- Le *nombre d'espèces* affiche une contribution positive modérée, indiquant une influence relative sur cette composante.

Pour la **PC2** :

- La *surface terrière* contribue négativement de manière la plus significative, soulignant son rôle comme facteur discriminant dans cette composante.
- L'*indice de Gini* et le *volume de bois récolté* ont des contributions négatives moins marquées.
- Le *nombre d'espèces* présente la contribution négative la plus forte, reflétant l'importance de la diversité biologique dans cette composante, potentiellement en réponse aux pratiques de gestion.

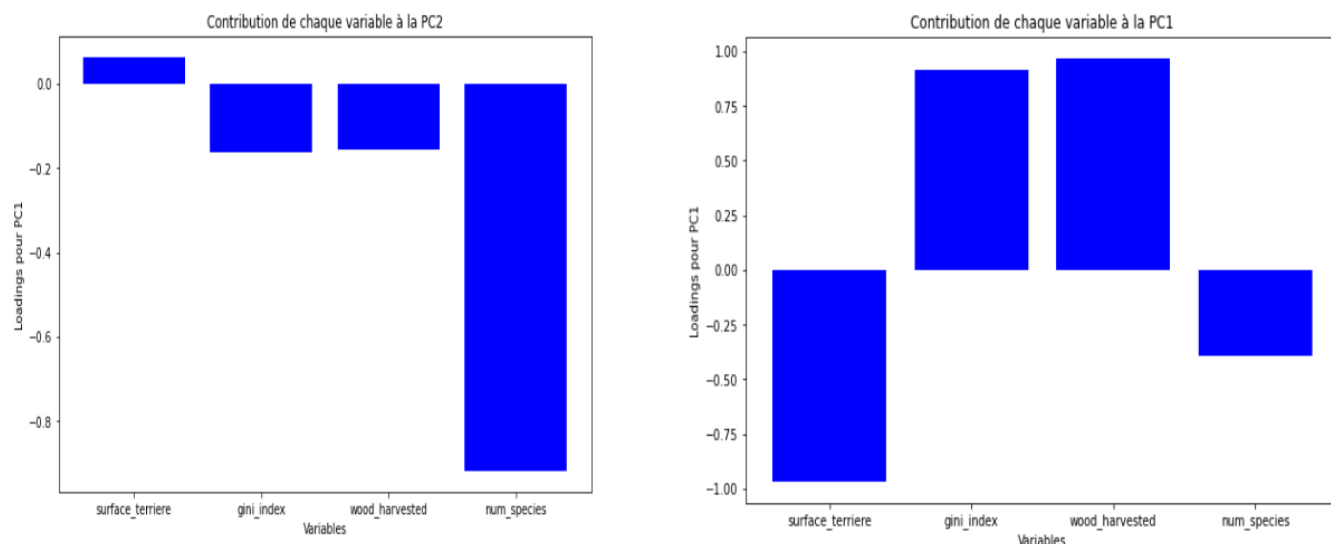


Figure 6.8: Contribution des contraintes pour la première composante principale (à gauche) et la deuxième composante principale (à droite)

6.6.3 Analyse Hiérarchique Ascendante et K-means

L'analyse Hiérarchique Ascendante (HAC) et la méthode K-means ont été utilisées pour identifier des groupes principaux dans les données forestières. Le dendrogramme illustre ces groupes 6.10, distingués par les couleurs rouge, vert, et orange. Les hauteurs des fusions dans le dendrogramme révèlent la distance à laquelle les échantillons sont combinés, avec des écarts significatifs qui indiquent des distinctions nettes entre les clusters. Trois classes principales ont été déterminées, visualisées par une ligne en pointillé sur le graphique. La classification HAC a généré un indice de silhouette de 0.511, signifiant une séparation bien définie entre les clusters. Par ailleurs, l'analyse K-means, utilisant la méthode du coude qui consiste à regarder le graphique la somme des carrés des distances intra-cluster (WCSS) en fonction du nombre de clusters et à identifier le point où l'ajout de clusters supplémentaires n'entraîne plus de diminution significative du WCSS. tout comme un coude. Cette méthode a confirmé que le nombre optimal de clusters est également de trois, avec un indice de silhouette similaire de 0.512.

Les groupes formés par HAC et K-means 6.9 montrent des clusters d'individus partageant des caractéristiques similaires :

- Les **points bleus**, localisés dans le quart inférieur de PC1 et le quart positif de PC2, reflètent des pratiques de coupe intense. Cette localisation indique des zones de gestion forestière intensive, où la réduction de la surface terrière peut être associée à des pratiques de récolte agressives.
- Les **points jaunes**, dispersées principalement vers le centre gauche du graphique, pourraient indiquer des zones où la gestion forestière est plus modérée. Ces échantillons, ayant une diversité spécifique élevée et une surface terrière importante, suggèrent une récolte de bois modérée et une gestion conservatrice.
- Les **points rouges**, situés principalement le long du côté positif de PC1 dans la partie supérieure du graphique, sont caractéristiques de zones avec une surface terrière réduite, une récolte élevée de bois, et un indice de Gini élevé. Ce groupe pourrait refléter des régions de forte exploitation avec une diversité structurelle significative.

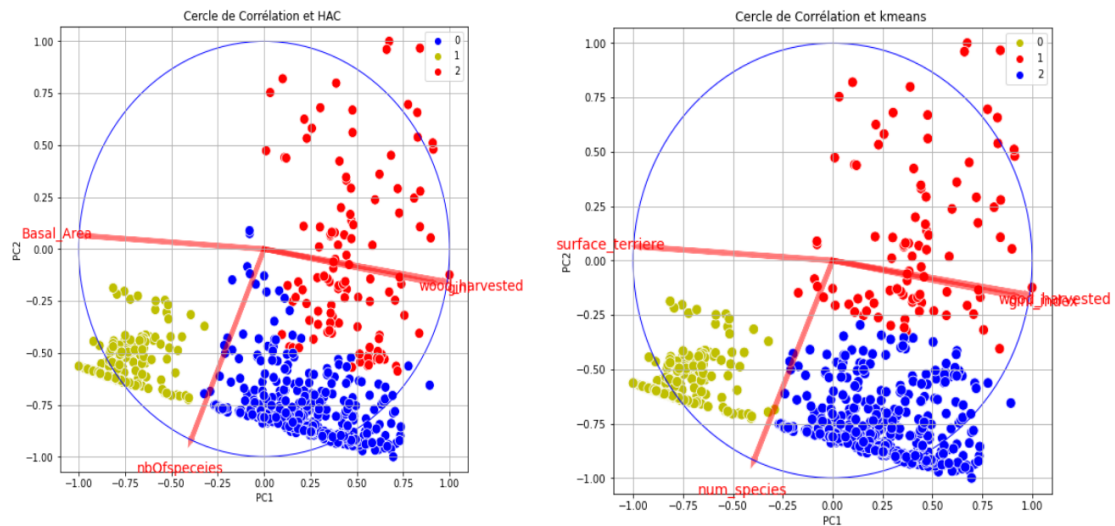


Figure 6.9: la projection des individus et le cercle de corrélation des variables sur les deux principales composantes, avec une classification obtenue via l'Analyse Hiérarchique Ascendante (à droite) et K-means (à gauche). Les différentes couleurs indiquent les classes d'individus, tandis que les flèches illustrent les coefficients de corrélation entre les variables et les composantes principales.

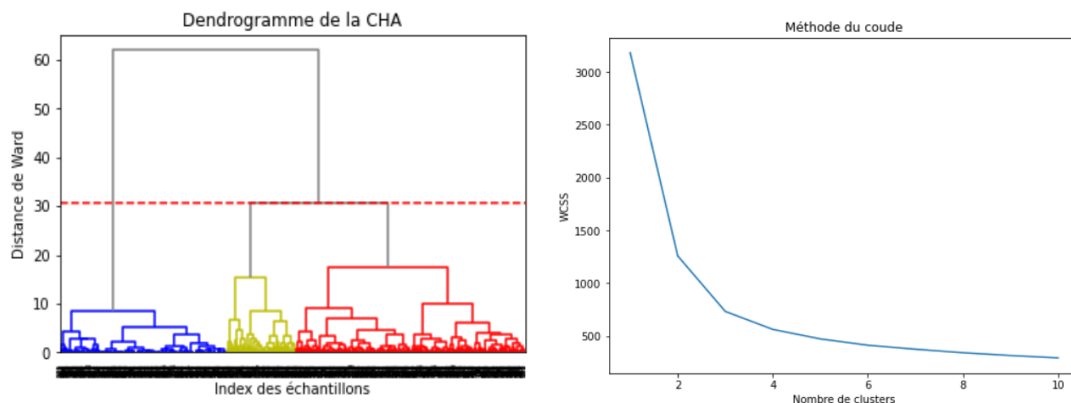


Figure 6.10: Le dendrogramme à gauche illustre les individus sur l'axe des abscisses avec leurs distances hiérarchiques sur l'axe des ordonnées, les lignes pointillées rouges indiquant le niveau de coupe. À droite, le graphique montre la variation du WCSS en fonction du nombre de clusters, illustrant l'application de la méthode du coude pour déterminer le nombre optimal de clusters.

6.6.4 Analyse des résultats de l'ACP et de la classification basée sur les contrôles effectués

Pour bien comprendre et expliquer les résultats de l'ACP et de la classification, j'ai d'abord examiné la moyenne de chaque caractéristique de contrainte avec l'écart type, à l'exception du taux de mortalité (voir Figure 6.11).

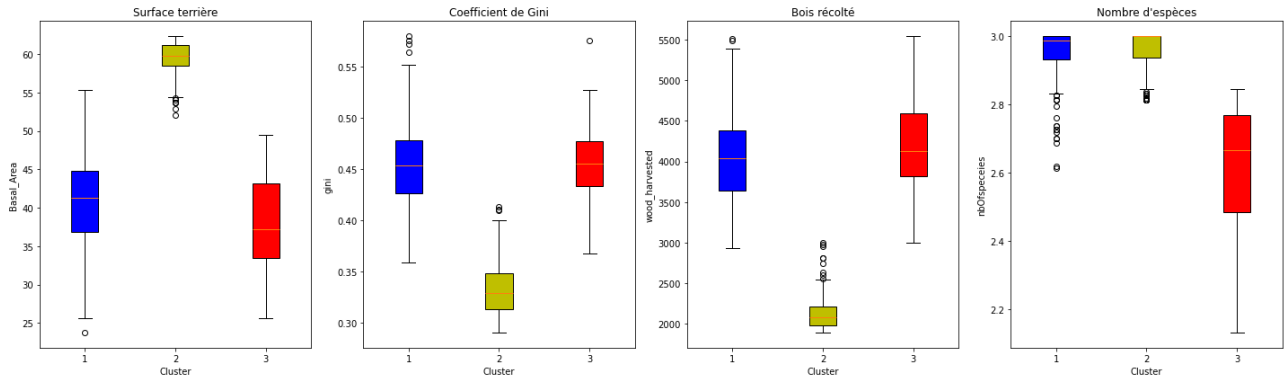


Figure 6.11: la moyenne et l'écart des variables des contraintes pour chaque classe de HAC

J'ai également analysé l'évolution temporelle de chaque caractéristique en calculant la moyenne annuelle sur les 10 parcelles, ainsi que l'écart type entre ces 10 parcelles pour observer les variations (voir Figure 6.12).

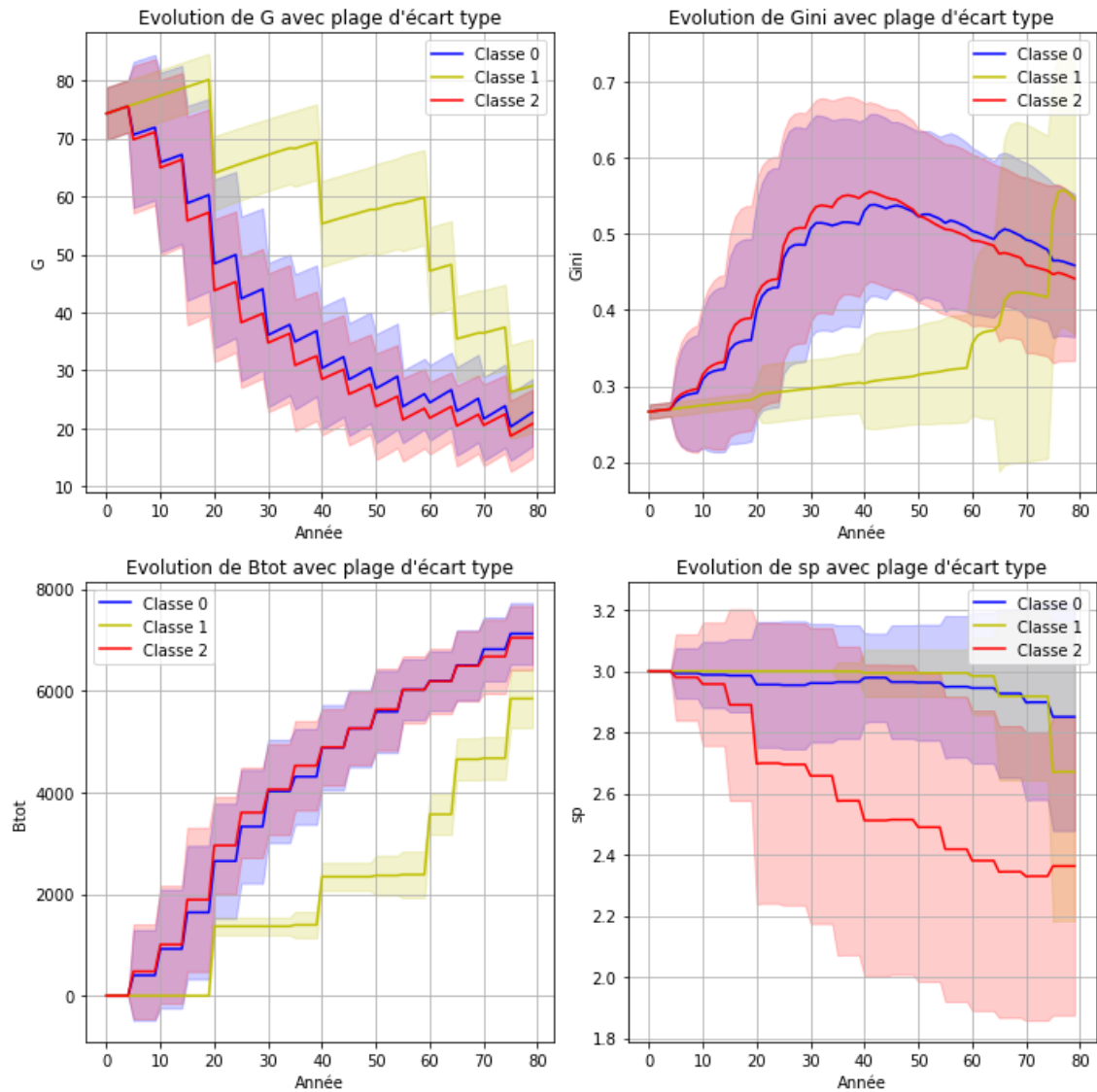


Figure 6.12: Médiane et écart interquartile (0.25-0.75) de l'évolution au cours de la simulation des contraintes, coefficient de Gini, surface terrière, production et nombre d'espèce, pour les différentes classes obtenues lors de la CAH avec la distance de Ward sur les données de ACP sur l'ensemble des contraintes des individus viables identifiés par le GA. Btot : est la production du bois totale (m^3/ha), G:est la surface terrière (m^2/ha) , Gini :est l'indice de Gini, sp :est le nombre des espèces .

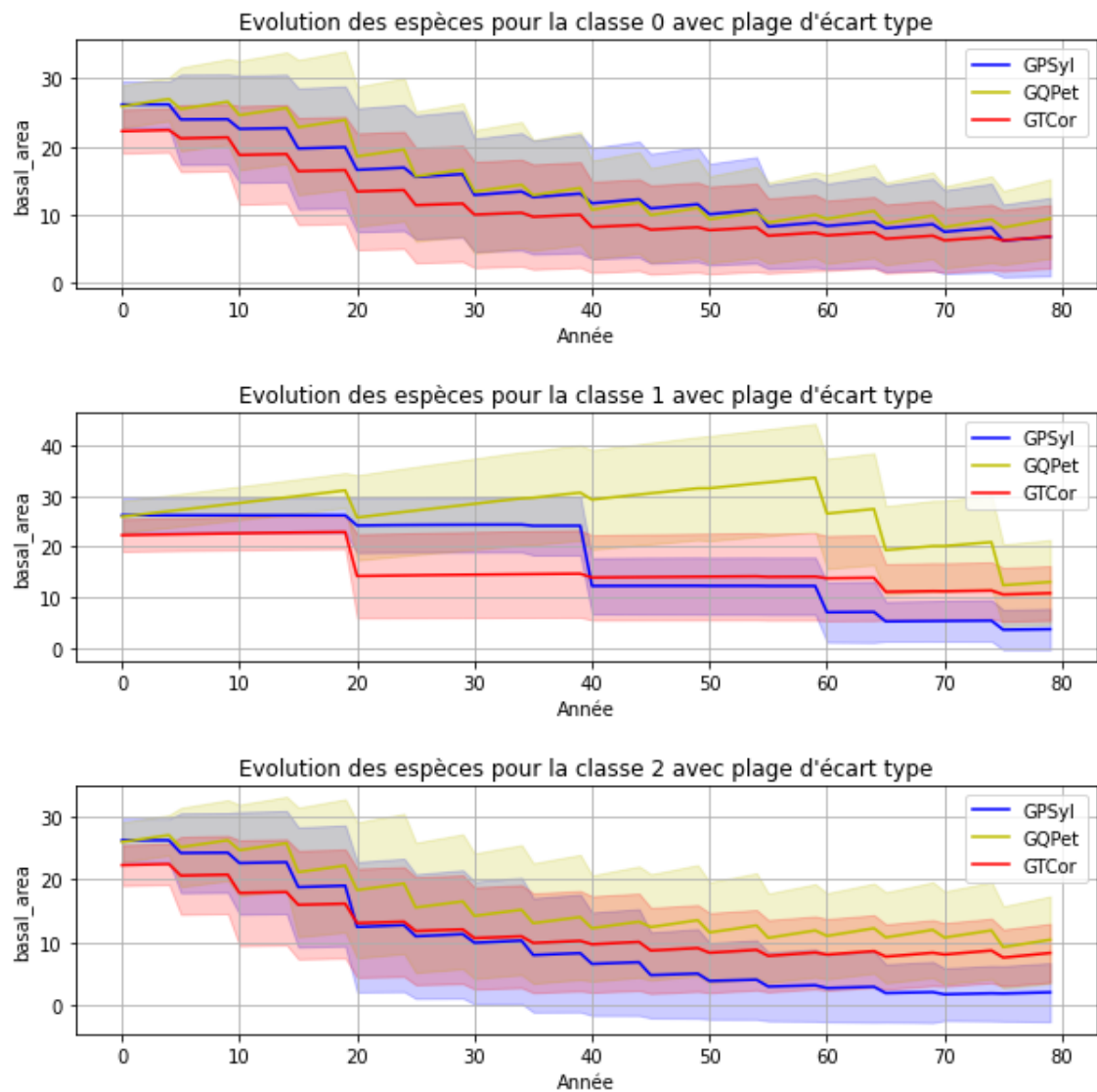


Figure 6.13: Médiane et écart interquartile (0.25-0.75) de l'évolution au cours de la simulation de la surface terrière pour chaque espèce dans chaque classe où : GPSYL : Surface terrière de *Pinus sylvestris*, GQPET : la surface terrière de *Quercus petraea* et GTCOR : la surface terrière de *Tilia cordata*

Pour voir les contrôles appliqués à chaque classe, j'ai étudié la distribution de chaque variable de contrôle pour chaque classe. Pour ce faire, j'ai décomposé le génome expliqué dans le chapitre précédent et regroupé les paramètres communs en groupes pour faciliter l'analyse. Ainsi, pour chaque classe, j'ai calculé le nombre de variables de contrôle, la période, le type d'éclaircie et la surface terrière. En ce qui concerne la composition, puisque le génome fonctionne avec 3 gènes de Cf_i , j'ai calculé l'écart type dans ce trinôme pour observer sa distribution. Par exemple, si l'écart type est significativement éloigné de la moyenne de $100/3$, comme $(80, 20, 0)$, on peut conclure que la composition est déséquilibrée. En revanche, si la distribution est proche de la moyenne de 30, on peut dire que la composition est équilibrée (voir Figure 6.14).

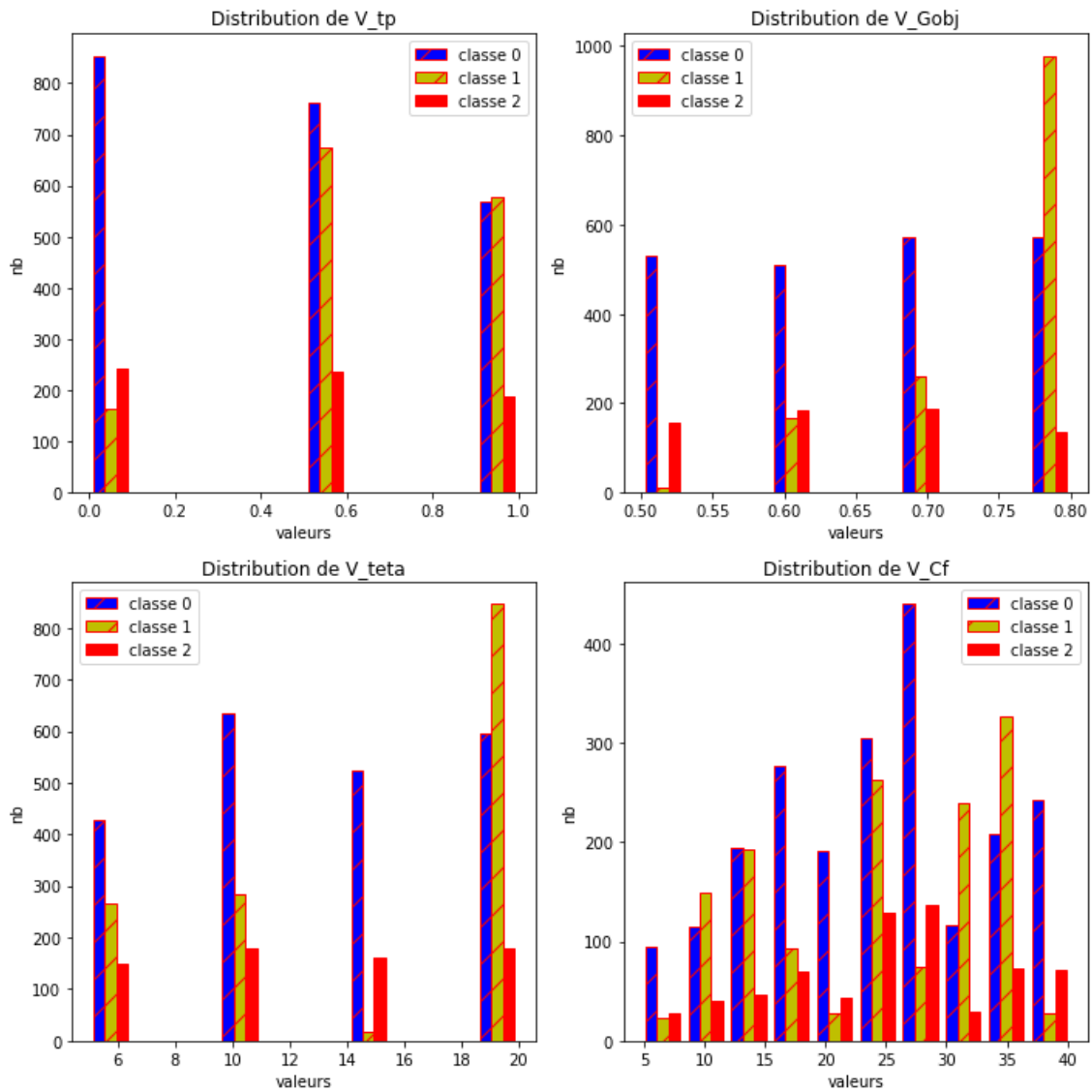


Figure 6.14: Distribution des variables de controle pour chaque classe la periode (V_{θ}), le type d'écclairci (V_{tp}), la surface terrière (V_{Gobj}) et la composition objective (V_{Cf}).

Classe bleue (0) : Cette classe, d'après l'ACP et la classification, semble se caractériser par une exploitation intense, ce qui entraîne une faible surface terrière, une récolte de bois et un indice de Gini importants, ainsi qu'un nombre d'espèces élevé. Ainsi, selon les courbes d'évolution temporelle 6.11 et la distribution des variables de contrôle, on peut affirmer que cette classe se caractérise probablement par des coupes rapprochées et une surface terrière restante distribuée de la même manière sur l'ensemble des variables. C'est pourquoi nous récoltons beaucoup de bois et que la surface restante est faible. En ce qui concerne le nombre d'espèces, on remarque que cette variation est en moyenne de trois espèces au fil des années, comme le montre la courbe de distribution où ces coupes sont effectuées de manière équitable, comme le montre également le graphe de la classe 0 pour les 3 espèces 6.13. On constate que, au fil du temps, les espèces n'atteignent pas la valeur nulle, sauf vers la fin, ce qui explique pourquoi la stabilité est maintenue. Pour l'indice de Gini, il serait nécessaire d'examiner la lumière et la croissance des diamètres pour justifier également pourquoi il augmente.

Classe jaune (1) : L'ACP et la classification nous indiquent que cette classe est caractérisée par une diversité spécifique élevée et une surface terrière importante, suggèrent une récolte de bois modérée et une gestion conservatrice.. Nous avons remarqué que cette classe, d'après la projection sur le cercle de corrélation, présente un nombre élevé d'espèces. Cela est visible en calculant la moyenne et l'écart type pour tous les individus viables de cette classe 6.11. Pour comprendre pourquoi, nous pouvons constater que les coupes dans cette classe sont généralement longues 6.14, souvent sur 20 ans. De plus, cette classe se caractérise par une surface terrière objective de 0.8 tirée aléatoirement, ce qui explique pourquoi nous ne récoltons pas beaucoup de bois.

En ce qui concerne l'indice de Gini, nous remarquons qu'il n'évolue pas significativement par rapport aux

autres classes, mais la composition de cette classe reste toujours équilibrée avec 3 espèces. Cela est également visible dans le graphe pour la classe 1 des trois espèces utilisées 6.13. C'est seulement vers la fin que la surface terrière de l'espèce PSYL (le Pin sylvestre) s'annule. Nous observons que la composition est plus ou moins équilibrée au début, mais vers la fin, nous remarquons une coupe plus importante de QPET (le Chêne pédonculé). Cela peut s'expliquer par le fait que le Quercus se développe plus que les autres espèces quand on a moins d'interventions. C'est pourquoi, sur l'ensemble de l'évolution, l'indice de Gini reste faible, augmentant seulement vers la fin.

Classe rouge (2) : L'ACP et la classification pour cette classe nous ont donné des hypothèses selon lesquelles cette classe est caractérisée par une forte exploitation, ce qui augmente l'indice de Gini, le bois récolté et diminue également la surface terrière et le nombre d'espèces. Ainsi, on peut observer que la coupe est effectuée à des périodes rapprochées avec une surface terrière qui varie de manière significative, comme le montrent les histogrammes 6.14 où toutes les variables pour cette classe ont une distribution équilibrée sur l'ensemble des intervalles. C'est pourquoi la surface terrière pour cette classe diminue considérablement au fil du temps, tandis que le bois récolté augmente, de même que l'indice de Gini 6.12.

Cette augmentation de l'indice de Gini peut s'expliquer par le fait que les coupes favorisent différemment la croissance des espèces en fonction de leur tolérance à l'ombre. En effet, l'espèce TCOR, qui est la moins tolérante à l'ombre avec une tolérance de 0.1, a pu mieux s'établir et se développer suite aux ouvertures de la canopée, tandis que les espèces plus tolérantes comme PSYL (0.3) et QPET (0.2) ont été plus affectées par ces perturbations. Cela a entraîné une différenciation de la croissance entre ces espèces, augmentant ainsi l'hétérogénéité de la structure forestière et donc l'indice de Gini.

De plus, la coupe de certaines espèces et pas d'autres, comme le montre la diminution de l'intervalle de surface terrière de PSYL jusqu'à 0 à partir de l'année t_0+40 , explique également la baisse du nombre d'espèces observée au fil du temps. Cependant, il serait nécessaire de tracer l'évolution de la lumière pour mieux comprendre les mécanismes sous-jacents à ces dynamiques.

Voici le tableau des différences entre les classes :

Caractéristique	Classe bleue (0)	Classe jaune (1)	Classe rouge (2)
Exploitation	Forte	Faible	Forte
Surface terrière	Faible	Forte	faible
Récolte de bois	forte	Faible	Forte
Indice de Gini	Forte	Faible	Forte
Nombre d'espèces	Forte	Forte	Faible
Dynamique des coupes	équitables	Longues	équitables
Composition des espèces	Équilibrée	Équilibrée	Croissance différenciée
Lumière	Ouvertures fréquentes, favorisant les espèces moins tolérantes	Coupes espacées, équilibre maintenu	Ouvertures fréquentes, favorisant les espèces moins tolérantes

Table 6.2: Résumé des caractéristiques par classe.

6.7 État des forêts et le contrôle appliqué

6.7.1 Résultats des moindres carrés

Après avoir appliqué la méthode des moindres carrés, j'ai obtenu les matrices de transitions qui peuvent lier le contrôle U à l'état E . Lorsque j'ai calculé la matrice de $U_{\text{prédit}}$, j'ai trouvé que l'erreur quadratique R^2 est de l'ordre de 0.8, sans ajouter de biais, ce qui montre que la méthode des moindres carrés a bien fonctionné pour lier U et E .

Pour trouver les coefficients qui lient exactement U et E , j'ai récupéré la moyenne de chaque coefficient u_{ij} , tel que $u_{ij} = \sum(m_{ik}e_{kj})$. J'ai donc récupéré tous les m_{ik} pour toutes les observations, et j'ai calculé cette moyenne pour toutes les observations :

	e_1	e_2	e_3	e_4	e_5	e_6
θ	-1.152556	108.802633	-20.715674	-120.249888	-10.870553	0.002525
G_{obj}	0.004550	2.718663	-5.229614	-0.930871	-2.172104	0.000268
tp	0.117746	-9.268610	7.175170	7.360230	-0.637560	0.000237
$Cf1$	-10.365458	94.463144	1394.839462	-617.262311	220.080442	0.015139
$Cf2$	-0.616507	626.477518	-2069.465604	-2565.036702	-163.564827	0.032927
$Cf3$	10.759230	-583.333909	733.640776	3273.761763	-42.683566	-0.044356

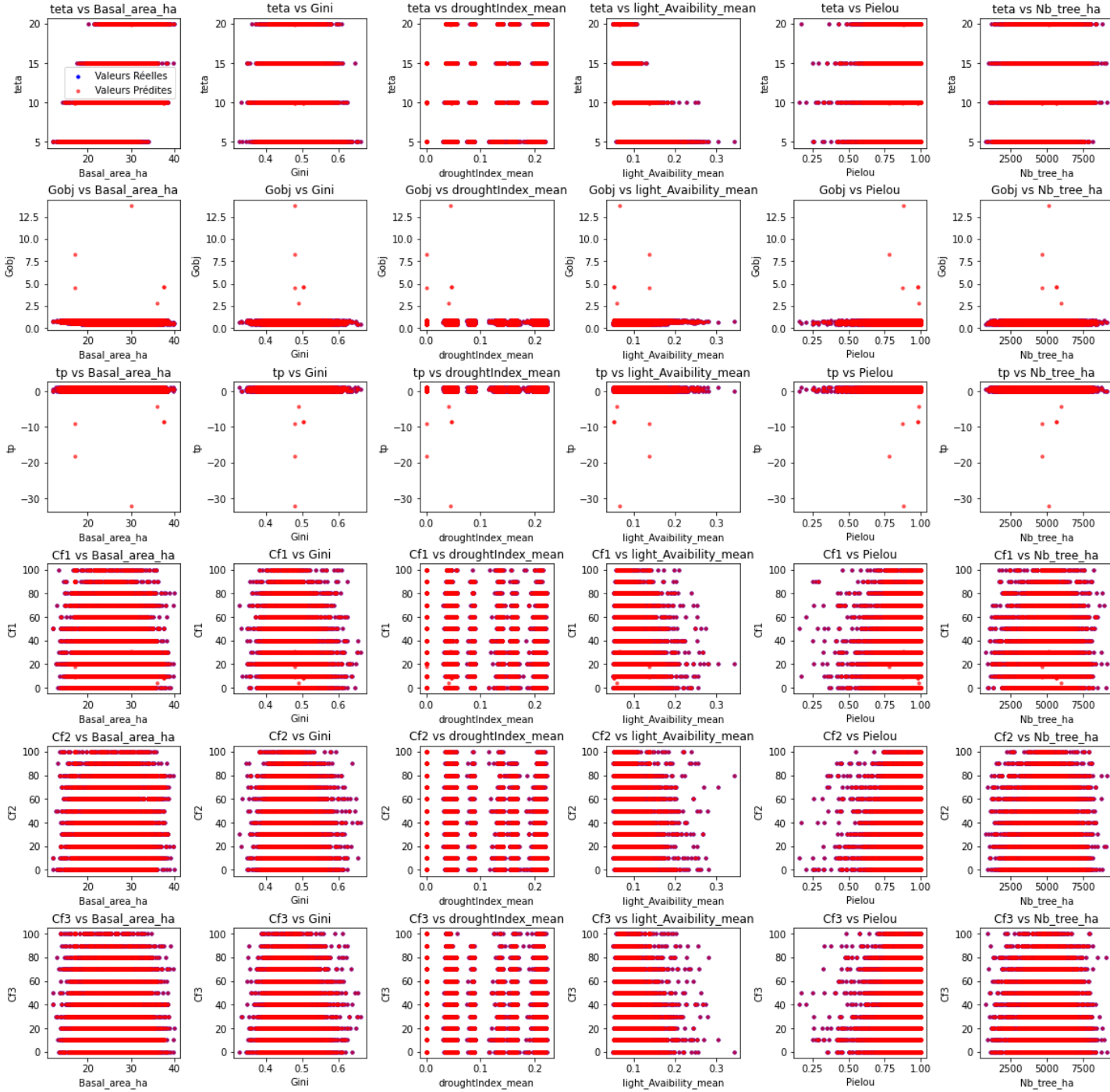


Figure 6.15: Présentation de chaque variable U en fonction de chaque variable E , pour toutes les observations, avec les valeurs prédites par les moindres carrés et les valeurs réelles en bleu.

La dernière étape qui consiste à lier précisément la moyenne en fonction du contrôle en calculant la moyenne pour chaque observation, c'est-à-dire $m_{ij} = \text{mean}(m_{ij_obs}/n_{obs})$. Cependant, cette méthode n'est pas très précise et d'autres approches pourraient mieux améliorer la précision pour l'ensemble des observations.

6.7.2 Résultats de K-means pour lier état et contrôle

Comme présenté dans la section méthodologie, j'ai classifié les variables d'état et examiné les caractéristiques de chaque classe. J'ai trouvé les résultats suivants :

- Cluster bleu** : Les forêts sont densément peuplées en termes de nombre d'arbres, mais avec une faible surface terrière. Elles sont bien diversifiées en termes d'indice de Gini, moins sèches et moins hétérogènes.
- Classe verte** : Les forêts présentent une forte surface terrière en termes d'indice de Pielou, mais sont plus sèches. Ces résultats peuvent s'expliquer par le fait que l'augmentation du nombre d'arbres intensifie la compétition pour les ressources, notamment l'eau.
- Classe orange** : Cette classe se caractérise par une distribution élevée des indices de Pielou et de Gini, mais avec une faible surface terrière, un nombre d'arbres réduit, et un environnement moins sec (autour de zéro).

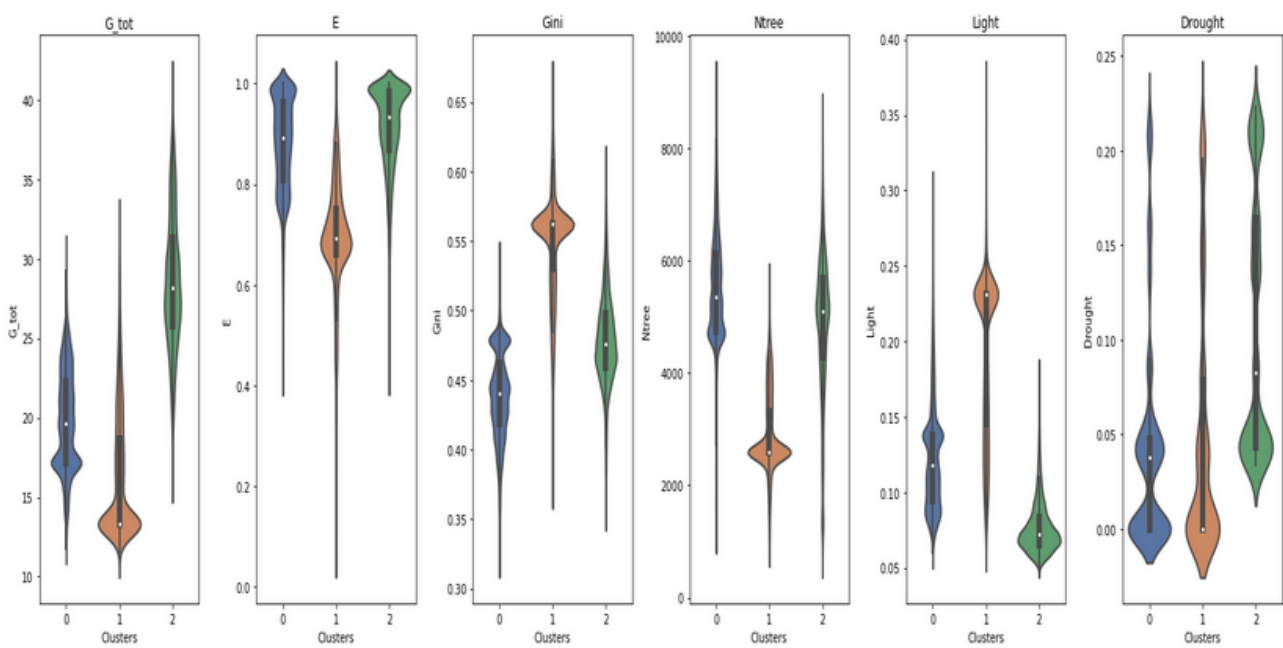


Figure 6.16: Distribution des variables d'état pour chaque classe.

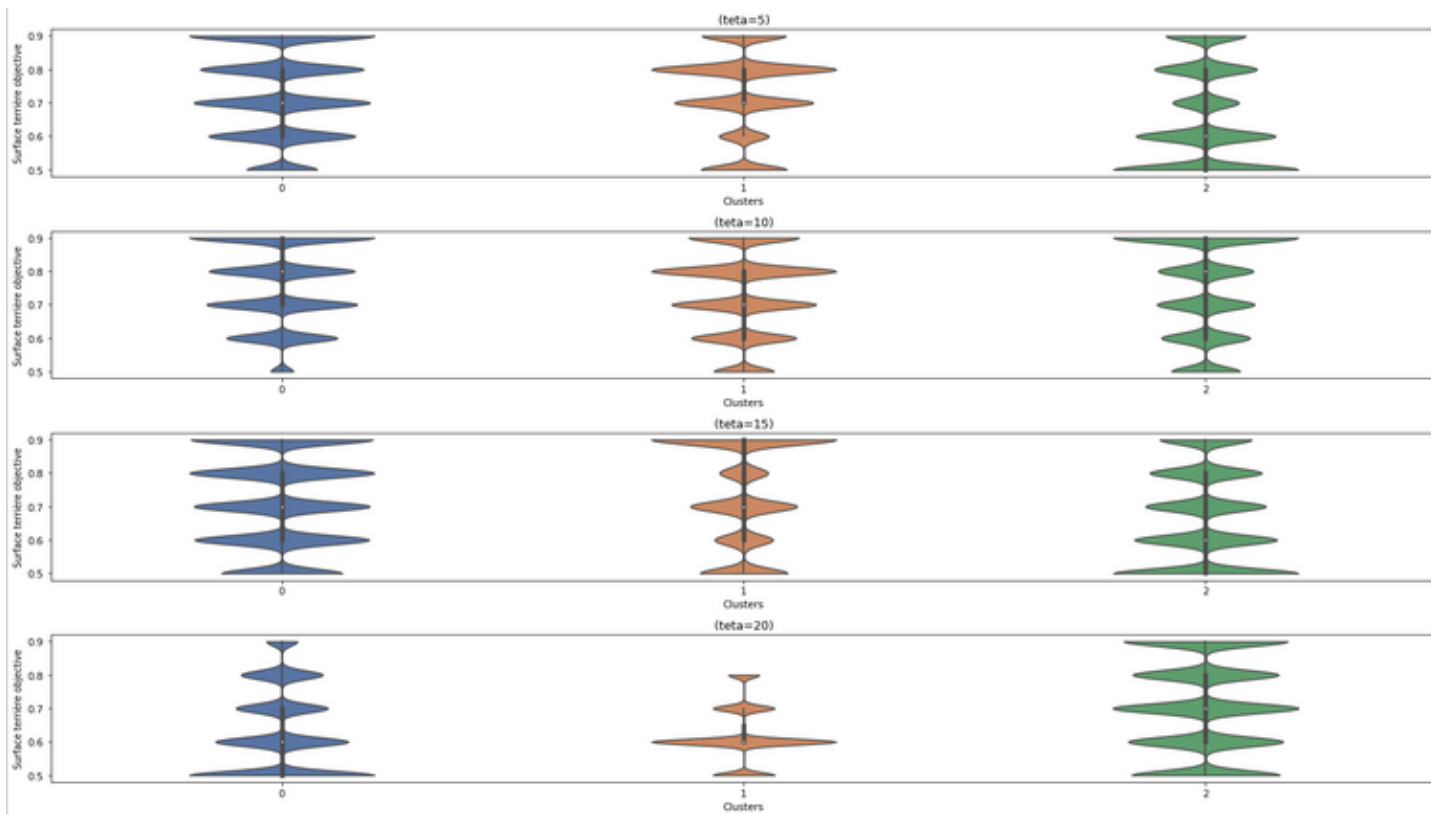


Figure 6.17: Distribution des contrôles de surface terrière objective pour chaque θ et pour chaque classe.

En ce qui concerne les contrôles présentés ci-dessus, on peut noter que pour la classe verte, caractérisée par une forte surface terrière, le contrôle optimisé par l'algorithme génétique (GA) favorise une coupe de 50% si la dernière coupe a eu lieu il y a plus de 5 ans. Pour les clusters rouge et bleu, qui se caractérisent par une faible surface terrière, la coupe est plutôt favorisée pour 20% ou 10% de la forêt.

6.7.3 Résultats de la matrice de transition : l'effet sur la viabilité

Comme présenté dans la section méthodologie, j'ai simulé l'évolution de la forêt, même lorsque celle-ci n'est pas viable, afin de répondre à la question : quel est l'effet sur la forêt après avoir effectué une coupe, ou de ne pas couper du tout, en tenant compte des contraintes qui peuvent apparaître annuellement, comme la surface terrière. Dans la nouvelle version, j'ai modifié cette contrainte pour qu'elle soit testée tous les 5 ans, plutôt qu'annuellement.

Ce que l'on peut observer à partir de la matrice de transition, c'est que lorsque aucune coupe n'est effectuée, la forêt a plus de chances de rester viable. Cela peut s'expliquer par le fait que la forêt continue de croître sans perturbation significative. Bien que la mortalité existe, elle est rare et n'influence que peu l'évolution. En revanche, pour une forêt où une coupe est effectuée, il existe une probabilité de 10% qu'elle devienne non viable, même si l'état initial de la forêt était viable. Il est également à noter qu'une forêt non viable a une probabilité de 0,54 de le rester après une coupe.

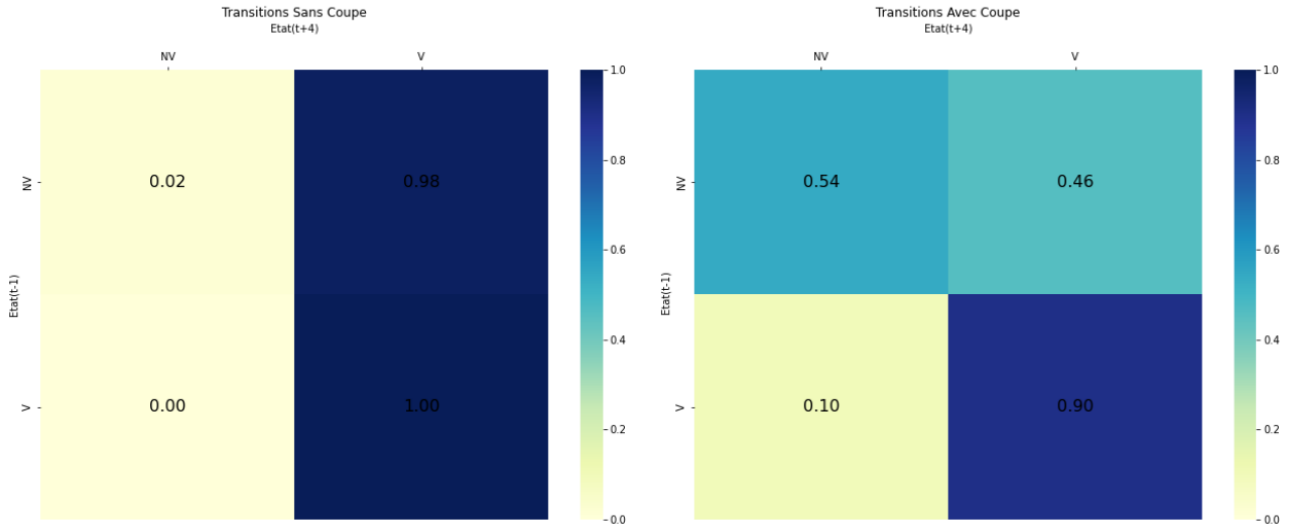


Figure 6.18: Matrice de transition pour la viabilité de l'état de la forêt sans et avec coupe.

Pour mieux comprendre ces 10% de non-viabilité lorsque l'état initial est viable, j'ai analysé le contrôle appliqué en termes de θ et de la surface terrière cible après coupe. J'ai constaté que la cause de cette non-viabilité est due au choix de couper 50% de la forêt après avoir attendu 5 ans sans intervention, ce qui entraîne une probabilité de non-viabilité de 53%. Cette probabilité diminue à mesure que la surface terrière cible augmente, et la forêt devient viable à 100% lorsque l'on choisit de ne couper que 10% de la forêt.

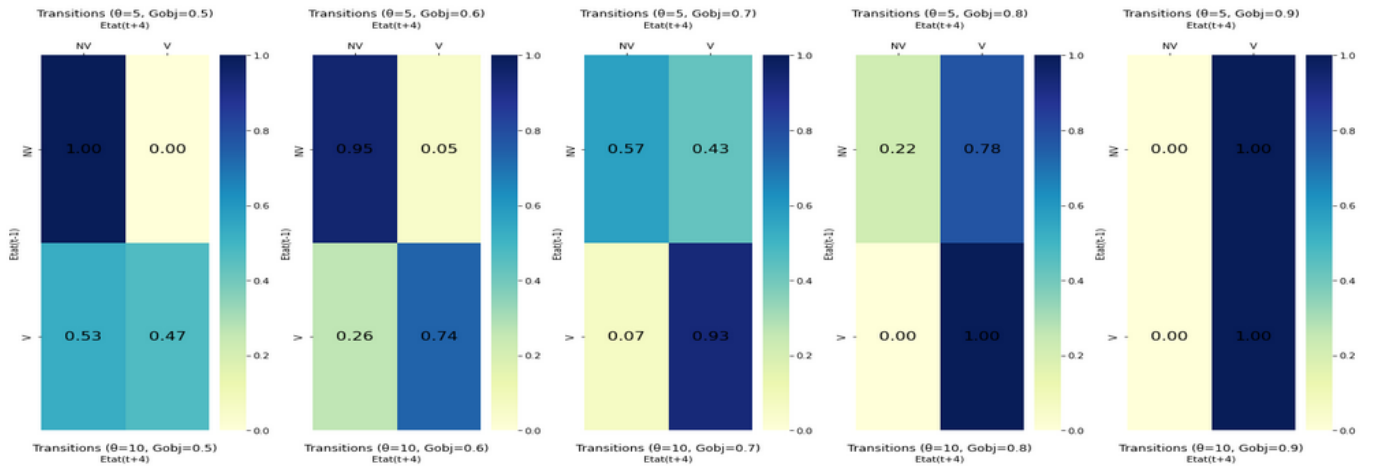


Figure 6.19: Matrice de transition pour la viabilité de l'état de la forêt après coupe sur une période de 5 ans.

6.7.4 Résultats des matrices de transition : l'effet sur l'état de la forêt

6.7.4.1 Surface terrière

Lorsqu'une coupe est réalisée dans la forêt, il y a une plus grande probabilité de transition vers d'autres états forestiers, comparé à la situation où la forêt n'est pas coupée. En revanche, lorsque la forêt n'est pas coupée, la surface terrière reste souvent dans la même gamme. Cela s'explique par la croissance naturelle de la forêt, et par le fait que le contrôle est géré en fonction de la surface terrière objective, ce qui explique pourquoi celle-ci décroît vers des états antérieurs.

On observe également une plus grande probabilité de transition vers des états différents lorsque la surface terrière est faible avant la coupe. Cela peut s'expliquer par le fait que la croissance est favorisée dans ces conditions, car il y a moins d'arbres, donc plus de lumière. En revanche, dans une forêt dense, la coupe impacte davantage la densité, ce qui réduit la probabilité de transition vers des états de faible surface terrière après la coupe. On peut en conclure que la croissance est plus lente dans une forêt dense.

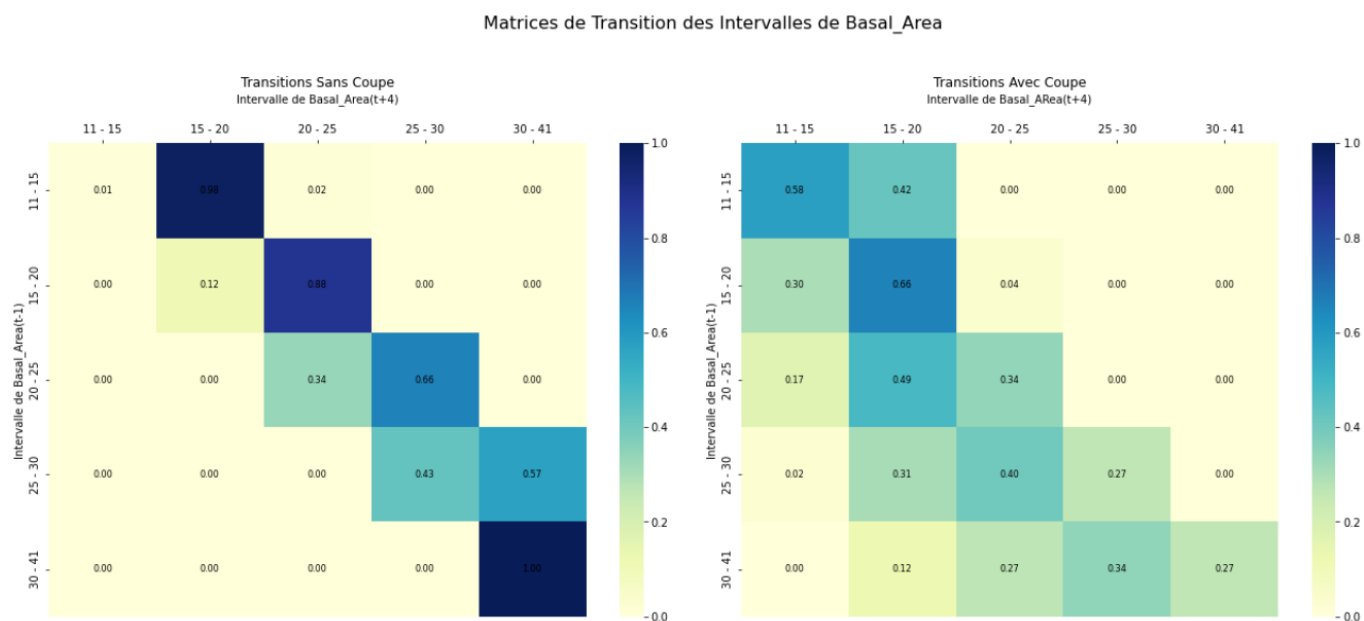


Figure 6.20: Matrice de transition de la surface terrière avant et après coupe.

6.7.4.2 Pielou

Lorsque la forêt est coupée, l'indice de Pielou tend à se déplacer vers des intervalles plus élevés, indiquant que la coupe augmente la diversité après 5 ans d'évolution. Cela peut s'expliquer par le fait que la coupe peut favoriser certaines espèces, ce qui signifie que la composition cible n'est pas équilibrée. De plus, la coupe permet une meilleure pénétration de la lumière au sol, favorisant ainsi la croissance de nouvelles espèces, ce qui crée de l'hétérogénéité, en plus des anciennes espèces qui n'ont pas été coupées lors de l'intervention. En revanche, lorsque la forêt n'est pas coupée, il y a plus de chances que l'indice de diversité reste dans le même intervalle. Cela peut être lié à la compétition pour la lumière. Par exemple, les espèces à croissance rapide et de grande hauteur continuent à croître, créant de l'ombre pour les espèces moins tolérantes à la lumière, ce qui ralentit leur croissance. Ainsi, la vitesse de croissance reste la même pour toutes les espèces, car l'état de la forêt n'a pas été perturbé.

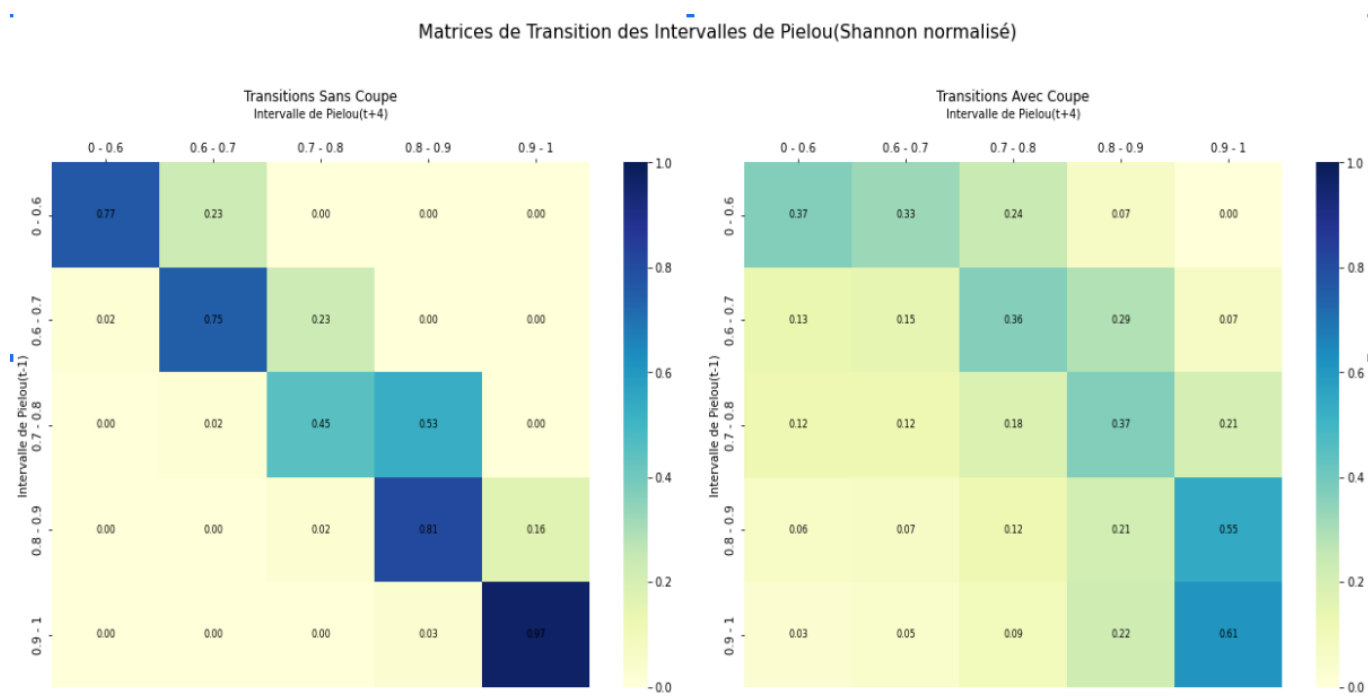


Figure 6.21: Matrice de transition de l'indice de diversité et d'équi-répartition (Pielou) avant et après coupe.

6.7.4.3 Gini

Lorsque la forêt n'est pas exploitée, elle conserve sa diversité structurelle en termes de diamètre. Cela peut s'expliquer par une croissance homogène, même lors de la transition vers un état moins hétérogène, particulièrement lorsque la forêt est déjà très hétérogène. Cette situation peut être observée par la domination de certains arbres, probablement due au fait que les espèces moins tolérantes à l'ombre ne trouvent pas des conditions propices à leur croissance.

En revanche, lorsqu'une coupe est réalisée dans des forêts très hétérogènes, cette intervention impacte la transition vers une forêt moins hétérogène, avec des intervalles de Gini de $[0,45, 0,5]$ et $[0,6, 0,67]$. Cela pourrait s'expliquer par le fait que la coupe a favorisé l'élimination des petits arbres avec un éclaircissement croissant ($tp=1$). À l'inverse, pour des forêts moins hétérogènes, la coupe favorise la transition vers une forêt plus hétérogène, ce qui pourrait être dû à une réduction de la dominance des espèces, grâce à un type d'éclaircie décroissant ($tp=0$) ou même aléatoire, favorisant ainsi la diversité.

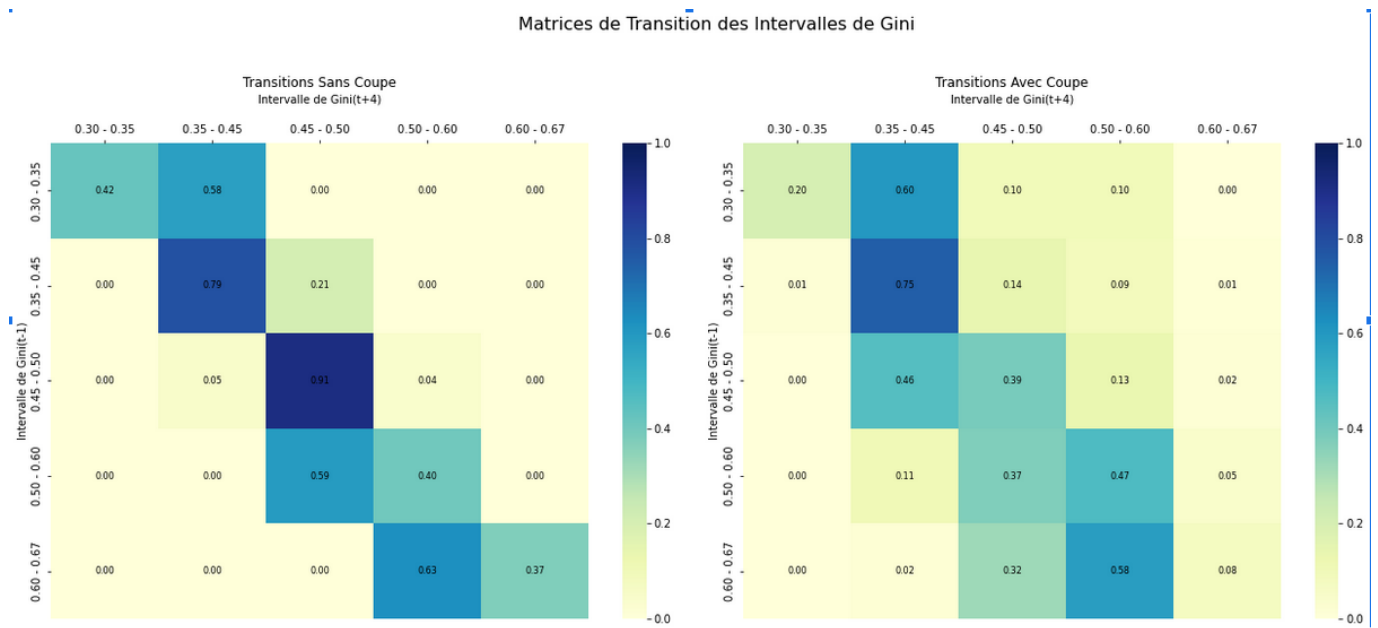


Figure 6.22: Matrice de transition de l'indice de diversité structurelle GINI avant et après coupe.

7 Discussion

7.1 Objectifs de la composition sylvicole

7.1.1 Intégration de la composition de contrôle dans la gestion sylvicole

Le premier objectif de mon stage est d'intégrer le problème de la composition de contrôle afin de déterminer des itinéraires sylvicoles combinant la période d'éclaircie, le type de coupe, la surface terrière visée après éclaircie, et la composition des espèces forestières. Initialement, nous souhaitions travailler avec 5 espèces, mais en raison de la combinatoire élevée de 10^{53} , nous avons décidé de commencer avec 3 espèces, avec un intervalle de tous les 20% .

La façon de coder de l'intervention s'est d'abord concentrée sur les espèces à maintenir après éclaircie, sans tenir compte des autres espèces pouvant s'établir au fil des années. C'est pourquoi les graphiques de présentation du nombre d'espèces montrent un maximum de trois espèces. Pour inclure les autres espèces, il serait nécessaire de les couper également. Une façon de le faire est de les regrouper dans un quatrième groupe et de réappliquer l'algorithme d'intervention, en sélectionnant les arbres parmi ce groupe pour minimiser l'écart entre la composition actuelle et la composition cible, tout en mettant à jour la composition actuelle en tenant compte de ce groupe.

7.1.2 Exploitation du meilleur

Pour exploiter le meilleur individu et obtenir un itinéraire sylvicole optimal proche de la meilleure solution de la génération précédente, j'ai appliqué une mutation de la composition de ± 10 pour explorer le voisinage immédiat. Cette approche visait à ne pas trop s'éloigner de la composition d'origine, tout en respectant la condition que la somme des coefficients de mélange (Cfi) doit être égale à 100. Cependant, cette contrainte impose des changements combinatoires pour les individus descendants. C'est pourquoi il est nécessaire de muter de -20 à +20 pour deux gènes successifs de Cfi.

7.1.3 Parallélisation

Travailler avec un espace de recherche très vaste et une simulation lente pour chaque individu nous a conduits à utiliser la programmation parallèle pour paralléliser les étapes de l'algorithme génétique. L'idée initiale était de paralléliser les simulations des individus sur les processeurs disponibles après chaque ensemble d'individus produite par une étape de l'algorithme génétique (génération initiale, après croisement et mutation, après génération des individus mutés à partir du meilleur individu, et après génération des individus aléatoires).

J'ai effectué des simulations de l'algorithme génétique pour les versions parallèle et séquentielle avec trois valeurs de $n_{ind} = 10, 20$, et 30 , en fixant la graine aléatoire pour garantir la reproductibilité. J'ai constaté que la méthode parallèle est légèrement plus rapide que la méthode séquentielle, avec des rapports de performance allant de 1,04 à 1,07 selon la valeur de n_{ind} .

Cependant, l'utilisation de tous les processeurs pour exécuter l'algorithme génétique en parallèle tend à ralentir l'ordinateur. Cela peut s'expliquer par le fait que lorsque l'algorithme est exécuté sur tous les cœurs, la mémoire vive (RAM) se remplit rapidement, entraînant un transfert des données vers l'espace de swap, qui est une zone de stockage sur le disque dur utilisée par le système d'exploitation.

L'accès au swap étant beaucoup plus lent que l'accès à la RAM, les données de l'algorithme génétique doivent être lues et écrites sur le disque dur. Pour améliorer l'efficacité, j'ai vidé le swap après chaque exécution de l'algorithme pour garantir le bon fonctionnement de l'ordinateur.

J'ai également essayé de n'utiliser que la moitié des processeurs pour réduire la charge globale, mais cela n'a pas bien fonctionné en termes de temps de simulation. Il est donc nécessaire de retravailler la parallélisation des instructions pour réduire le temps de simulation et permettre un nombre plus important de tests de l'algorithme génétique.

7.2 Résultats de la classification et des itinéraires de gestion

L'ACP et la matrice de corrélation nous ont permis de supposer certains liens entre les variables qui caractérisent les contraintes. La production de bois est fortement corrélée avec l'indice de Gini, et ces deux variables sont inversement corrélées avec la surface terrière.

Pour bien comprendre ces relations, on peut supposer que la relation de décorrélation est liée à l'abattage excessif du bois. Cet abattage entraîne une augmentation de la production de bois récoltée et une faible surface terrière. Pour l'indice de Gini, on peut penser que cet abattage entraîne l'augmentation de la lumière captée par le sol, ce qui favorise l'établissement de nouveaux semis. Cela peut donc avoir une influence sur l'augmentation de la diversité structurelle et peut également être influencé par l'abattage des arbres dominants ou structuraux.

Concernant le nombre d'espèces, cette variable était peu corrélée avec les autres variables, car j'ai contraint de travailler avec un maximum de trois espèces. C'est pourquoi on remarque de faibles variations pour cette variable.

Pour comprendre la cause derrière ces corrélations et ces hypothèses, j'ai réalisé une classification de ces itinéraires viables et j'ai identifié trois groupes d'itinéraires de gestion. J'ai bien trouvé que, pour la classe jaune, la décorrélation est liée à l'abattage qui se fait sur de longues périodes et avec un choix de la surface terrière objective de 80% qui dominant pour toute la simulation. On a aussi trouvé sur le graphique lié aux espèces (Fig. 6.13) que les coupes sont effectuées pour les 3 espèces, c'est pour cette raison que le nombre d'espèces reste à 3 pour cette classe dans la plupart des cas. Pour l'indice de Gini, il faudra aussi regarder le nombre d'individus qui entrent dans la simulation pour cette gestion et la croissance en diamètre de chaque arbre. Si le nombre de semis est faible et que la croissance est stable, cela explique pourquoi Gini ne varie pas trop.

Pour la classe rouge et la classe bleue, ce sont des évolutions pour les caractéristiques des contraintes assez similaires au cours du temps, seul le nombre d'individus de GA n'est pas le même pour chaque classe. La classe bleue a le plus grand nombre d'individus par rapport à la classe rouge. Mais la différence entre ces deux classes réside dans la composition des espèces.

Pour la classe rouge, elle se caractérise par une variation avec une composition à 2 espèces après 40 ans. Cela peut s'expliquer soit par le fait que le choix de la composition pour cette classe était dans la plupart des cas en binôme, soit que les interventions effectuées en termes de période et de choix de la surface terrière objective ont favorisé la disparition de l'espèce PSYL.

D'après les effectifs de chaque variable de contrôle, on remarque qu'ils sont distribués équitablement sur tous les intervalles. C'est pourquoi on ne sait pas exactement pourquoi on observe ce résultat pour cette classe.

7.3 Discussion méthodologique

L'objectif de ce stage est d'introduire la composition des espèces dans le problème de contrôle forestier, afin d'analyser les itinéraires sylvicoles issus de l'algorithme génétique. Cela permettra d'établir un lien entre les interventions appliquées aléatoirement sur la forêt et l'évolution de sa composition, facilitant ainsi le choix de gestion pour les gestionnaires forestiers.

Après avoir compris le problème de contrôle posé précédemment, nous avons modifié l'intervention de Forceps pour tenir compte de la composition cible. Cela a permis de calculer la combinatoire du problème et de réduire le nombre d'espèces considérées, reformulant ainsi le problème de contrôle et l'algorithme génétique.

L'objectif est de tester un grand nombre de combinaisons de variables de contrôle afin de trouver un ensemble proche du noyau de viabilité, sans biais. L'analyse des itinéraires sylvicoles avec un climat stable a permis d'identifier des relations entre les variables des contraintes et les itinéraires, ainsi que de tester la validité de la reformulation du problème.

Cependant, il faudra refaire ces analyses avec des données climatiques extrêmes à l'avenir, afin d'orienter la discussion sur les itinéraires sylvicoles dans un contexte de changement climatique extrême. En effet, cela pourrait avoir plusieurs impacts. Tout d'abord, l'utilisation de données climatiques extrêmes, comme celles issues des modèles CNRM et MPI avec les scénarios climatiques RCP 4.5 et RCP 8.5, pourrait augmenter l'aléatoire possible dans les simulations, rendant le noyau de viabilité plus restreint. Ensuite, les simulations avec ces données climatiques extrêmes seraient potentiellement plus courtes, ce qui pourrait réduire les temps de calcul. Enfin, la prise en compte de l'impact du climat sur le taux de mortalité, qui n'a pas été inclus dans les analyses initiales, sera un élément clé à intégrer dans les prochaines analyses. La modification des coefficients de mortalité dans Forceps permettra ainsi d'étudier cet aspect important du problème de contrôle forestier.

L'analyse de l'évolution de la forêt peut être enrichie par l'étude de son espace d'état, c'est-à-dire l'évolution conjointe du diamètre D_i et de la proportion de l'houppier cs_i pour chaque espèce de l'arbre i . Cette approche permettrait d'évaluer l'impact des différents modes de gestion forestière sur la structure et le fonctionnement de l'écosystème. Par exemple, sous un scénario de gestion extensive avec peu d'interventions, on pourrait observer une diversification des diamètres entre les espèces, reflétant une stratification verticale plus marquée de la forêt. À l'inverse, une gestion plus intensive avec des éclaircies régulières pourrait favoriser une homogénéisation des

diamètres et une diminution des proportions d'houpier, indiquant une structure plus uniforme. L'étude de ces paramètres de l'espace d'état, en les mettant en regard des pratiques de gestion, offrirait donc une perspective complémentaire à l'analyse de l'évolution de la forêt.

8 Conclusion

Ce stage, basé sur les travaux de Michelle Malara en 2023 [6], avait pour objectif principal de renforcer la définition des pratiques sylvicoles viables, essentielles pour la préservation de nos écosystèmes, aujourd'hui menacés par le changement climatique.

Grâce au modèle ForCEEPS, centré sur l'individu, qui simule l'évolution des forêts en tenant compte des impacts climatiques et de la faune, des caractéristiques des espèces, des propriétés du sol et de la luminosité, nous avons pu explorer diverses stratégies de gestion. Ce modèle permet de définir les interventions d'un gestionnaire forestier grâce à quatre variables clés : la périodicité des coupes, le type d'éclaircies, l'objectif de surface terrière, et la composition spécifique des espèces visée. Face à la complexité et la combinatoire du problème, l'algorithme génétique a été choisi pour sa capacité à générer des solutions prometteuses pour établir des pratiques de gestion viables.

Durant ce stage, je me suis d'abord familiarisé avec le modèle ForCEEPS, à travers la littérature et le code Java implémenté sur la plateforme Capsis. Après avoir cerné les enjeux, nous avons décidé d'ajouter un quatrième contrôle, permettant de mieux définir un itinéraire sylvicole pour un gestionnaire, en recalibrant la composition des espèces après chaque éclaircie. Cette nouvelle approche a modifié notre manière de concevoir les coupes, en privilégiant la réduction des écarts entre la composition des espèces souhaitée et celle actuellement observée.

Une fois ces ajustements validés, j'ai intégré ce nouveau contrôle dans l'algorithme génétique, ce qui a nécessité de recalculer la combinatoire du problème pour évaluer la faisabilité de cette augmentation des variables et de la complexité. Les modifications apportées à toutes les étapes de l'algorithme génétique ont inclus la création d'individus aléatoires, l'ajustement des croisements et mutations, et l'optimisation de la sélection des meilleurs individus.

En parallèle, j'ai optimisé le calcul de la fitness de l'algorithme génétique pour accélérer le processus, bien que les capacités de mon ordinateur aient limité l'amélioration de la performance. Des tests de sensibilité ont été réalisés pour ajuster les paramètres de l'algorithme et le nombre de parcelles utilisées, s'appuyant sur des données climatiques historiques de la ville de Berne.

En conclusion, une analyse des individus viables et non viables a été effectuée pour estimer la proportion de chacun, et une analyse en composantes principales a servi à identifier les liens entre les variables et les contraintes. Cette analyse a été complétée par une classification hiérarchique ascendante et une méthode K-means pour examiner l'influence de chaque individu sur des paramètres clés tels que la surface terrière, l'indice de Gini, le volume de bois coupé, et le nombre d'espèces. Ces analyses ont permis de caractériser les itinéraires sylvicoles et les contraintes forestières.

Bien que des progrès significatifs aient été réalisés dans la compréhension et le contrôle des dynamiques forestières, des efforts supplémentaires sont nécessaires pour définir les actions politiques adaptées à chaque variable de contrôle, en tenant compte des conditions forestières et des contraintes écologiques. Le mois restant sera donc consacré à la définition de ces variables, en prenant en compte le changement climatique et les dynamiques définies par le modèle Forcepps, pour mieux soutenir les gestionnaires dans la mise en œuvre de pratiques de gestion durable.

9 Annexes

9.1 Modèle de croissance

les facteurs de réduction de croissance sont calculés comme suit :

$$GR_{\text{gdd}} = \max \left(0, 1 - e^{\left(\frac{DD_{\text{min}}}{S} - GDD \right) \times q_{\text{corr}}} \right)$$

avec q_{corr} étant le paramètre de pente de la courbe dans la fonction de croissance en degrés-jours ($q_{\text{corr}} = \frac{1}{750} \text{ } ^\circ\text{C}^{-1} \cdot \text{day}^{-1}$) (Bugmann 1996).

$$GR_{\text{drought}} = \max \left(0, 1 - \frac{DrI}{DrTol_S} \right)$$

$$GR_{\text{soil}} = \max \left(0, 1 - e^{N_{1S} \times (N_{\text{soil}} - N_{2S})} \right)$$

Avec N_{1S} et N_{2S} étant des paramètres spécifiques aux espèces dérivés de $N_{\text{req}}S$

Réduction de croissance due à la compétition pour la lumière:

Comme les arbres ne sont pas spatialement localisés à l'intérieur des parcelles, l'interaction pour la lumière entre les arbres dans ForCEEPS dépend uniquement de la stratification verticale des couronnes des arbres. Dans ForCEEPS, GR_{light} a deux composants :

$$GR_{\text{light}} = GR_{\text{cs}} \times GR_{\text{sh}}$$

avec GR_{cs} représentant le retour de la taille de la couronne sur la croissance de l'arbre, c'est-à-dire, la surface foliaire de l'arbre est positivement liée au taux de croissance de l'arbre. GR_{sh} est le facteur de réduction lié à l'ombrage par les arbres concurrents. Pour chaque arbre i :

$$cl_i = cs_i \times H_i$$

avec H étant la hauteur de l'arbre et cs étant le ratio de la hauteur, qui est lié à l'exposition à la lumière de l'arbre. Pour chaque arbre i , cs varie entre deux valeurs spécifiques à l'espèce qui représentent le cas où l'arbre est complètement ombragé ou en pleine lumière, avec :

$$cs_i = cs_{\text{max}}^S - (cs_{\text{max}}^S - cs_{\text{min}}^S) \times k_{LAI_i}$$

où les valeurs extrêmes cs_{max}^S et cs_{min}^S ont été dérivées de la relation entre le poids frais du feuillage et le DHP décrite par Wehrli et al. (2007) et dépendent du paramètre du type de feuillage f_S , et k_{LAI_i} est le facteur de correction - allant de 0 (pas d'ombrage) à 1 (ombrage complet) - calculé par Didion et al. (2009) comme suit :

$$k_{LAI_i} = \min \left(\left(\frac{LAI_{H_i}}{LAI_{\text{max}}} \right)^2, 1 \right)$$

avec LAI_{H_i} étant l'indice de surface foliaire cumulée entre le sommet de la canopée et le sommet de l'arbre cible (c'est-à-dire entre le sommet de la canopée et la hauteur H), c'est-à-dire la surface foliaire cumulée (en m^2) divisée par la surface de la parcelle (en m^2) et LAI_{max} étant la valeur maximale de la surface foliaire double-face dans une parcelle résultant du point de compensation de la lumière dans les forêts européennes.

Pour une parcelle donnée, l'indice de surface foliaire est calculé pour chaque couche, en sommant l'indice de surface foliaire de toutes les parties de couronnes trouvées dans cette couche. Ainsi, pour la couche x de largeur w , l'indice total de surface foliaire SumLAI_x est :

$$\text{SumLAI}_x = \frac{1}{PA} \times \sum_{i|H_i \geq x-w} LAI_{i,x}$$

avec PA étant la surface de la parcelle, H_i la hauteur de l'arbre i , et $LAI_{i,x}$ la surface foliaire double-face de la partie de la couronne de l'arbre i appartenant à la couche x , calculée comme suit :

$$LAI_{i,x} = r_{i,x} \times LA$$

où $r_{i,x}$ correspond à la fraction de la surface foliaire totale trouvée dans la couche x (en supposant une diminution linéaire de la surface foliaire), et la cohérence entre les niveaux de la couche et de l'arbre entier étant assurée par le fait que :

$$\sum_{x=H_0}^{x=H} r_{i,x} = 1$$

avec x_{H0} et x_H étant la couche incluant la base et le sommet de la couronne de l'arbre i , respectivement, et LA étant la surface foliaire totale double-face d'un arbre, calculée à partir du diamètre de l'arbre D comme suit (Didion et al. 2009) :

$$LA = f'_s \times c_s \times D_s^a$$

Une variable intermédiaire $GR_{sh,i,x}$ est calculée en utilisant la disponibilité en lumière L_{av} dans la couche selon la loi d'extinction de Beer (Botkin et al. 1972, Shugart 1984) pour quantifier l'intensité de l'ombrage dans chaque couche de la couronne de chaque arbre. Pour un arbre i trouvé dans la couche x , la disponibilité en lumière $L_{avi,x}$ est calculée comme suit :

$$L_{avi,x} = r_{i,x} \times e^{-k \times \text{SumLAI}_x}$$

avec $r_{i,x}$ étant le coefficient représentant la diminution de la surface foliaire du sommet à la base de la couronne et k est le coefficient d'extinction de la lumière (avec $k = 0.25$).

Enfin, le facteur de réduction lié à l'ombrage par les arbres concurrents ($GR_{sh,i}$) est calculé pour un arbre i en sommant les $GR_{sh,i,x}$ de toutes les parties de la couronne de l'arbre :

$$GR_{sh,i} = \sum_x GR_{sh,i,x}$$

9.2 Illustration de la coupe

Imaginons une forêt avec deux espèces dominantes, le sapin et le chêne. Voici la surface terrière de chaque espèce en cm^2 :

- Chêne = {60, 50, 30, 10, 5} pour une surface de proportion $\frac{155}{270} \approx 0.574$.
- Sapin = {70, 40, 5} pour une surface de $\frac{115}{270} \approx 0.426$.

Un itinéraire sylvicole pour une année d'intervention:

- $tp = 1$, $G_{\text{objet}} = 100 \text{ dm}^2$ et pour une composition de 80% pour chêne et 20% pour sapin.
- On commence par trier ces arbres selon tp puisque tp est proportionnelle à α , donc on peut dire que les arbres sont déjà triés.
- On coupe de façon croissante de diamètre. On va couper de plus petite au plus haut pour les arbres.
- Selon le processus de la coupe, on a le choix au début entre deux arbres des diamètres $d_{k_{\text{sapin}}} = 5$ et $d_{k_{\text{sapin}}} = 5$. Donc on regarde quel arbre coupé va minimiser la différence entre la composition objective et la composition actuelle.
- On calcule la composition actuelle pour chaque espèce:
 - Si on coupe $k1_{\text{sapin}}$: la composition actuelle pour sapin $\left(\frac{115-5}{270-5}\right) \approx 0.41$ et pour chêne est 0.58.
 - Si on coupe $k1_{\text{chêne}}$: la composition actuelle pour sapin est $\frac{115}{265} \approx 0.43$ et pour chêne est 0.57.
- Donc on évalue lequel va réduire la différence entre la composition actuelle et cible selon la norme euclidienne.
- Pour $K - 1_{\text{sapin}}$ on trouve $\sqrt{(0.41 - 0.8)^2 + (0.58 - 0.2)^2} = 0.54$ et pour $k1_{\text{chêne}}$ on trouve $\sqrt{(0.43 - 0.8)^2 + (0.57 - 0.2)^2} = 0.52$.
- Donc on choisit de couper $k1_{\text{chêne}}$, et la surface terrière restante 265 dm^2 est $> 100 \text{ dm}^2/\text{ha}$.

- On actualise les listes des arbres triés et on retrouve chêne = $\{60, 50, 30, 10\}$ et sapin = $\{70, 40\}$.
- Et maintenant on refait le même calcul mais les arbres testés cette fois-ci est celui de diamètre pour chêne 5dm^2 et pour sapin 40 dm^2 .
- On s'arrête de couper quand la surface terrière restante soit en dessous de la surface terrière cible.

Bibliography

- [1] Marion Jourdan et al. “Managing Mixed Stands Can Mitigate Severe Climate Change Impacts on French Alpine Forests”. In: *Regional Environmental Change* 21.78 (2021), pp. 955–975. DOI: <https://doi.org/10.1007/s10113-021-01805-y>.
- [2] Xavier Morin et al. “Beyond Forest Succession: A Gap Model to Study Ecosystem Functioning and Tree Community Composition Under Climate Change”. In: *Functional Ecology* 35 (2021), pp. 955–975. DOI: 10.1111/1365-2435.13760.
- [3] Jean-Pierre Aubin. “A Survey of Viability Theory”. In: *SIAM Journal on Control and Optimization* 28.4 (1990), pp. 749–788. DOI: 10.1137/0328044.
- [4] Sophie Martin, Guillaume Deffuant, and Justin M. Calabrese. “Defining Resilience Mathematically: From Attractors To Viability”. In: *Book Title (replace with the actual book title)*. First Online: 01 January 2011. Publisher (replace with the actual publisher), 2011, pp. 15–36. DOI: <http://www.springer.com/series/5394>.
- [5] Alberto Tonda et al. “Towards Evolutionary Control Laws for Viability Problems”. In: *GECCO '23: Genetic and Evolutionary Computation Conference*. HAL: hal-04230169. Lisbon, Portugal, July 2023, pp. 1464–1472. DOI: 10.1145/3583131.3590415.
- [6] Michelle Malara. *Analyse de viabilité d’un modèle mécaniste pour l’aide à la décision de gestion : exemple d’un peuplement de chêne*. 2023.
- [7] Harald Bugmann. “On the Ecology of Mountainous Forests in a Changing Climate: A Simulation Study”. Ph.D. Dissertation. ETH Zürich, 1994, pp. 44–93. DOI: 20.500.11850/141625.
- [8] Claire Bernard. “La théorie de la viabilité au service de la modélisation mathématique du développement durable. Application au cas de la forêt humide de Madagascar”. Ph.D. Dissertation. Université Blaise Pascal, 2011.
- [9] Jean-Pierre Aubin, Alexandre M. Bayen, and Patrick Saint-Pierre. *Viability Theory: New Directions*. Springer Science Business Media, 2011. URL: <https://link.springer.com/book/10.1007/978-3-642-16684-6>.
- [10] Clémentine de Montgoffier. *Diversity as a Management Tool for Forest Ecosystem Services*. 2024.
- [11] Ahmad Hassanat et al. “Choosing Mutation and Crossover Ratios for Genetic Algorithms—A Review with a New Dynamic Approach”. In: *Information* 10 (12 2019), p. 390. DOI: <https://www.mdpi.com/2078-2489/10/12/390>.
- [12] Jean-Denis Mathias, Xavier Balandraud, and Michel Grediac. “Applying a Genetic Algorithm to the Optimization of Composite Patches”. In: *Computers Structures* 84 (12 2006), pp. 823–834. DOI: 10.1016/j.compstruc.2005.12.004.
- [13] Michèle Sebag and Marc Schoenauer. “Contrôle d’un algorithme génétique”. In: *Revue d’Intelligence Artificielle* (1996), pp. 389–428. DOI: hal-00111526.
- [14] A.E. Eiben and J.E. Smith. *Introduction to Evolutionary Computing*. Springer Series in Natural Computing. Springer, 2025. URL: <http://www.springer.com/series/4190>.
- [15] David E. Goldberg and Chie Hsiung Kuo. “Genetic Algorithms in Pipeline Optimization”. In: *Journal of Computing in Civil Engineering* 1 (2 1987), pp. 128–141. DOI: [https://ascelibrary.org/doi/abs/10.1061/\(ASCE\)0887-3801\(1987\)1:2\(128\)](https://ascelibrary.org/doi/abs/10.1061/(ASCE)0887-3801(1987)1:2(128)).
- [16] David E. Goldberg. *Genetic Algorithms in Search Optimization and Machine Learning*. Addison-Wesley, 1989, pp. 128–141.