

**PROPOSITION D'UN REGROUPEMENT DES
STRATES D'INVENTAIRE POUR LES UNITÉS
DE COMPILATION 064F, 064M ET 064R
GRÂCE À L'UTILISATION DU LOGICIEL
SAS/STAT**

Présenté à :

Le Groupe Optivert inc.

André Carle

Et

CSRE

Raymond Barrette

Par :



Centre Collégial de Transfert
de Technologie en foresterie

Pascal Gauthier, ing.f.

Carl Bergeron, ing.f.

Guy Lessard, ing.f. M.Sc.

TABLE DES MATIÈRES

Table des matières.....	i
Liste des figures	ii
Liste des tableaux.....	ii
Résumé.....	iii
1. Introduction.....	1
2. Méthodologie	2
2.1. Outils d'analyse et choix techniques.....	2
2.2. Variables retenues	4
2.3. Transformation linéaire de l'âge et des proportions de surfaces terrières	5
2.4. Identification du nombre de regroupements à conserver.....	6
2.5. Schéma hiérarchique	7
3. Résultats.....	9
3.1. Nombre de groupes	9
3.2. Analyse statistique	9
3.3. Validation du nombre de regroupements.....	13
4. Discussions	15
5. Conclusion	17
6. Références.....	18

ANNEXES

Annexe 1 - Schéma hiérarchique des regroupements.....	19
Annexe 2 - Résultats détaillés de l'analyse de regroupement	20
Annexe 3 - Correspondance de M ^{me} Hélène Crépeau	21

LISTE DES FIGURES

Figure 1 -	Valeurs du pseudo t^2 en fonction du nombre de groupes (<i>clusters</i>) générés	7
Figure 2 -	Arbre hiérarchique généré par la procédure TREE pour l'analyse faite sur le groupe régulier de l'unité 064F.....	8
Figure 3 -	Évolution du R^2 en fonction du nombre de groupes retenus	14

LISTE DES TABLEAUX

Tableau 1 -	Nombre de groupes retenus par unité de compilation et par ensemble	9
Tableau 2 -	Comparaison des précisions par type de regroupement pour l'unité de compilation 064M.....	10
Tableau 3 -	Comparaison des précisions (%) de la surface terrière par essence et totale pour le regroupement 33 de l'ensemble régulier de l'unité de compilation 064M.....	11
Tableau 4 -	Comparaison des précisions (%) de la surface terrière par essence et totale pour le regroupement 33 de l'ensemble irrégulier de l'unité de compilation 064M.....	12

RÉSUMÉ

Le Centre de Services aux Réseaux d'entreprises (CSRE) - Secteur Forêt a mandaté le Groupe Optivert inc afin de réaliser un calcul de possibilité forestière à l'aide du logiciel Patchworks en collaboration avec l'IQAFF et le CERFO. Un des mandats du CERFO consistait à proposer un regroupement de strates d'inventaire pour les unités de compilation 064R, 064M et 064F afin de diminuer le nombre de strates d'inventaire tout en représentant le maximum de la variabilité des différentes variables caractérisant chaque strate. En conséquence, moins de familles de courbes sont produites et le logiciel Patchworks peut réaliser ses itérations dans un délai raisonnable. Cet objectif est en accord avec la recommandation 5.4 de la **Commission d'étude sur la gestion des forêts publiques du Québec**.

La procédure CLUSTER du logiciel SAS/STAT est toute indiquée pour réaliser des regroupements hiérarchiques (SAS Institute, 2003). La méthode de calcul avec un « beta-flexible » de -0,5 permet ici un juste équilibre dans l'effet des données extrêmes sur les regroupements générés. Les strates sont séparées en deux ensembles (réguliers et irréguliers). Les variables retenues sont : la surface terrière par essence, la surface terrière totale, le diamètre moyen quadratique de la strate, la proportion (%) de la surface terrière par classe de diamètre de 10 cm. Certaines essences ont été regroupées et les essences n'atteignant jamais plus de 2 m²/ha dans un ensemble d'une unité de compilation ont été retirées de l'analyse.

L'analyse de regroupement réalisée est concluante puisqu'il en résulte une diminution de strate de 79 % en moyenne. La précision moyenne sur le volume total est augmentée de 25 % alors que, par essence, elle augmente d'au moins 55 %, et parfois même, de plus de 400 %. Une analyse de la précision au niveau de 2 regroupements de l'unité de compilation 064M, sur la surface terrière totale et par essence a démontré également l'amélioration substantielle de celle-ci.

L'analyse de regroupement se base sur les données issues des placettes et non sur l'appellation de la strate d'inventaire. Elle n'amène donc pas de « perte de diversité ». Un nombre de placettes beaucoup plus élevé a été obtenu par unité de regroupement soit en moyenne 43,2 plutôt que 9,4. Elle implique cependant, pour l'analyste, de nombreux choix et une série d'essais. Ceci pourrait permettre par la suite une approche probabiliste sur chaque regroupement pour gérer la grande variété observée, particulièrement lors de l'analyse des strates pour le choix de production prioritaire, l'élaboration de scénario sylvicole ou la gestion de la dynamique évolutive. Il est important de rappeler que la planification à cette échelle est d'ordre stratégique et non d'ordre opérationnel, qui serait illusoire.

L'évaluation de la variabilité « intra et inter » strate a aussi permis de mettre en lumière la troublante diversité qui est incluse dans une même strate d'inventaire. À cet effet, les causes d'une aussi faible expression de la proportion de la variance devrait être analysées et corrigées le cas échéant (superficie minimale de photo-interprétation, regroupement des strates cartographiques, etc.).

1. INTRODUCTION

Dans la recommandation 5.4 de la **Commission d'étude sur la gestion des forêts publiques du Québec**, il était stipulé que le système d'analyse qui sous-entend le calcul de la possibilité ligneuse devait favoriser un fort regroupement des strates d'inventaire, afin d'augmenter la précision dans l'estimation des volumes de bois. D'autre part, cet exercice de planification est à une échelle stratégique et le nombre de strates d'inventaire est souvent critiqué et d'utilisation incorrecte.

Le Centre de Services aux Réseaux d'entreprises (CSRE) - Secteur Forêt a mandaté le Groupe Optivert inc. afin de réaliser un calcul de possibilité forestière à l'aide du logiciel Patchworks, avec la participation de l'IQAFF et du CERFO. Un des mandats du CERFO consistait à proposer un regroupement de strates d'inventaire pour les unités de compilation 064R, 064M et 064F. Plus spécifiquement, l'objectif de ce travail était donc de développer une méthode afin de diminuer de façon importante le nombre de strates d'inventaire tout en représentant le maximum de la variabilité des différentes variables caractérisant chaque strate. Ceci avait pour but de diminuer le nombre de strates à gérer, de réduire le nombre de familles de courbes à produire et de permettre ainsi au logiciel Patchworks de réaliser ses itérations dans un délai raisonnable.

Comme l'utilisation d'outils statistiques était recherchée, la procédure CLUSTER du logiciel SAS/STAT avec un « beta-flexible » est toute indiquée pour réaliser des regroupements hiérarchiques (SAS Institute, 2003). Elle permet de regrouper avec une intensité variable en fonction des objectifs et de la précision voulue, divers éléments en fonction de variables quantitatives. Les regroupements hiérarchiques permettent de générer d'un seul coup tous les niveaux de regroupement possibles, allant d'un seul groupe jusqu'à tous les éléments soumis. Cette procédure avait été avantageusement utilisée lors des travaux du CERFO pour la **Commission d'étude sur la gestion des forêts publiques du Québec (CERFO, 2004)** ainsi que pour des regroupements en phytosociologie (Lessard, 1991).

Le présent travail présente la démarche utilisée et les résultats obtenus. La précision obtenue sur les essences tant au niveau général par strates d'inventaire est discutée ainsi que la validation du nombre de regroupement.

2. MÉTHODOLOGIE

2.1. OUTILS D'ANALYSE ET CHOIX TECHNIQUES

Le regroupement à l'aide d'outils statistiques se base sur les données issues des placettes échantillons et non sur l'appellation de la strate d'inventaire. Il a été convenu qu'un regroupement devait amalgamer des strates d'inventaire similaires sous un angle dendrométrique et sans égard à la qualité de station. En effet, le regroupement sur la base des types écologiques présentaient **dans ce cas** plusieurs difficultés. Cependant, bien que certaines strates d'inventaire possédant des qualités de station différentes soient placées dans un même regroupement, leurs évolutions dans Patchworks pourront diverger grâce à l'utilisation de plus d'une famille de courbes. De plus, dans le cas d'un regroupement contenant des strates non perturbées et des strates ayant fait l'objet de coupe partielle mais dont l'année de sondage diffère d'une strate à l'autre, il a été convenu de procéder de la façon suivante : le regroupement est maintenu mais le point d'embranchement sur la courbe d'évolution moyenne est ajusté pour chaque strate d'inventaire touchée par ce groupe en fonction de la surface terrière prévue en 2008. Ceci est possible puisque le logiciel Patchworks peut permettre un positionnement différent sur la famille de courbes moyennes pour chaque polygone touché par cette dernière.

Comme nous l'avons mentionné, la procédure CLUSTER du logiciel SAS/STAT est toute indiquée pour réaliser des regroupements hiérarchiques (SAS Institute, 2003). Il importe avec cette procédure de sélectionner une méthode de calcul parmi les onze proposées. Plusieurs essais ont été effectués. La méthode WARD a un biais important au niveau de la dimension de groupes formés, ayant ainsi une forte tendance à créer des groupes constitués d'un nombre semblable d'éléments. Aussi, elle est très sensible aux valeurs extrêmes. L'option TRIM permet d'atténuer ce biais, mais pour cela, elle exclut de l'analyse les valeurs identifiées comme extrêmes. Les essais effectués ont démontré que beaucoup de strates étaient exclues de l'analyse avec cette option. Il s'agit donc d'une option peu pratique puisque nous voulions analyser toutes les strates soumises. La méthode FLEXIBLE, de son côté, apparaît bien adaptée pour les cas de regroupement de strates. Il importe de définir avec cette méthode un facteur nommé « beta ». Plus la valeur définie est élevée, plus l'analyse effectuée sera sensible aux valeurs extrêmes, créant des groupes spécifiques pour chaque cas particulier de valeurs extrêmes. Les valeurs du facteur suggérées sont celles comprises entre 0 et -1 et la valeur définie par défaut est de -0,25 (SAS Institute, 2003). Les essais effectués ont permis de constater que le compromis idéal pour

les données analysées était lorsque la valeur du « beta » égale -0,5 puisqu'elle permettait ainsi un juste équilibre dans l'effet des données extrêmes sur les regroupements générés.

Les premiers essais effectués ont permis de constater que l'influence de chaque variable sur les regroupements effectués dépendait pour beaucoup de son échelle et de sa variance. Le ratio entre la moyenne et la variance a aussi une grande influence. Ainsi, ce qui contribue à augmenter l'influence d'une variable sur les regroupements formés sont des moyennes élevées, des variances élevées et des ratios entre la moyenne et la variance faibles. Concrètement, une variable avec une moyenne de 10 et une variance de 20 aura plus d'influence qu'une variable ayant une moyenne de 20 et une variance de 5. Des techniques sont prévues dans le logiciel SAS/STAT pour atténuer ces distorsions. La procédure ACECLUS transforme les variables en autant de variables canoniques. Les essais effectués ont donné des résultats mitigés et l'importance relative des variables ne semble pas être modifiée. Une autre solution consiste à standardiser les variables avec une moyenne de 0 et un écart type de 1 (option STD de la procédure CLUSTER). Dans ce cas, d'autres distorsions importantes apparaissent, les données extrêmes prenant beaucoup plus de poids. Elles ne semblent donc pas appropriées pour les besoins de l'étude.

Outre ces méthodes programmées, une simple transformation linéaire peut être utilisée en ajustant l'échelle des variables qui ont trop de poids. Cette procédure fut utilisée dans les présentes analyses. Elle permet d'éviter des distorsions dues à l'échelle utilisée et de s'assurer que les variables les plus importantes ont un poids suffisant. L'avantage de cette méthode est que l'analyste est libre d'ajuster le poids qu'a chaque variable dans la détermination des regroupements. Or, malgré qu'une telle approche puisse sembler biaisée, elle est en fait absolument nécessaire puisque certaines variables ont plus d'importance que d'autres dans la détermination des strates. Les premiers éléments à tenir compte dans une strate sont la composition en essence et la surface terrière totale, qui indirectement permettent d'estimer le volume total. Puisque la composition en essence est déterminée par la proportion de chaque essence en termes de surface terrière, il importe que les surfaces terrières par essence soient prioritaires dans la détermination des regroupements. Ces variables doivent donc primer sur d'autres comme par exemple le diamètre moyen ou la proportion de la surface terrière par classe de diamètre. Les transformations linéaires retenues sont décrites à la section 2.3.

2.2. VARIABLES RETENUES

Les essais initiaux ont montré qu'il était préférable de distinguer les strates d'inventaire en deux ensembles dans chaque unité de compilation. Un premier ensemble comprenait des strates à structure plutôt régulière et un deuxième, des strates à structure plutôt irrégulière (ou jardinée). Le classement des strates dans un ou l'autre de ces ensembles s'est fait sous la base de l'appellation de la strate d'inventaire, notamment en fonction du type de peuplement, des groupes d'essences et de la classe d'âge. Ainsi, les strates résineuses ou contenant des essences feuillues intolérantes) ont été placées dans l'ensemble régulier alors que les strates contenant plutôt des essences feuillues tolérantes ou semi-tolérantes ont été mises dans l'ensemble irrégulier. Cette distinction permettait d'utiliser des variables mieux adaptées à chaque cas. De plus, le premier ensemble correspond aux strates habituellement simulées dans le modèle par courbes (logiciel Sylva II) alors que le modèle irrégulier comprend les strates généralement simulées dans le modèle par taux.

Pour l'ensemble régulier, les variables retenues étaient :

- La surface terrière par essence
- La surface terrière totale
- Le diamètre moyen quadratique de la strate
- La hauteur moyenne de la strate (pondérée par le volume par essence)
- L'âge moyen de la strate (pondérée par le volume par essence)
- Proportion (%) de la surface terrière par classe de diamètre (10-18, 20-28, 30-38, 40-48 et 50 cm et plus

Il est à noter que l'âge et la hauteur par essence n'ont pas été retenus comme variables puisqu'ils n'étaient pas disponibles pour toutes les strates (données manquantes). En effet, pour que la procédure CLUSTER s'exerce sur toutes les strates, elle exige une valeur à chaque variable pour toutes les strates. Or, il arrive fréquemment qu'une strate contienne une faible quantité d'une essence mais que la hauteur ou l'âge de cette essence soit inconnue.

Pour des fins de simplification, certaines essences ont été regroupées telles :

- Les chênes, les frênes, les ormes et les tilleuls
- Les bouleaux blanc et gris
- Les épinettes blanche et rouge
- Les épinettes noires et les mélèzes
- Les peupliers

De plus, les essences n'atteignant jamais plus de 2 m²/ha dans un ensemble d'une unité de compilation ont été retirées de l'analyse.

Pour l'ensemble irrégulier, les variables initialement retenues étaient les mêmes que l'ensemble régulier sans toutefois considérer l'âge et la hauteur totale. Ces variables sont peu utiles dans le cas des structures irrégulières en raison de la forte variabilité de la hauteur et de l'âge au sein des peuplements représentatifs de ces strates et de la trop grande difficulté à sonder tous les individus.

Enfin, il est à noter que la perturbation d'origine ou récente ainsi que le type écologique n'ont pas été retenus pour l'analyse puisque ces éléments seront considérés à une autre échelle de gestion dans Patchworks. Ainsi, un regroupement de strates d'inventaire peut contenir différentes perturbations tout comme posséder diverses stations forestières. Toutefois, leurs caractéristiques dendrométriques lors de l'inventaire sont similaires. Ainsi, bien que certaines strates d'inventaire possédant des qualités de station différentes seront mises dans un même regroupement, leurs évolutions pourront être différentes dans Patchworks.

2.3. TRANSFORMATION LINÉAIRE DE L'ÂGE ET DES PROPORTIONS DE SURFACES TERRIÈRES

Tel que décrit précédemment, une transformation linéaire fut appliquée pour l'âge et les proportions de surface terrière par classe de diamètre exprimées en pourcentage.

Rappelons que le but n'est pas d'uniformiser le poids de chaque variable et qu'il faut au contraire que les surfaces terrières demeurent prioritaires dans la détermination des regroupements. Pour mieux juger de l'influence relative de chaque variable et déterminer la valeur du facteur d'ajustement, s'il y a lieu, une analyse discriminante a été conduite grâce à la procédure

STEPCDISC et de la technique STEPWISE. L'âge total a donc été ajusté par un facteur de 4 alors que le poids des classes a été ajusté en divisant généralement les valeurs par un facteur de 4 (5 dans le cas de l'ensemble 064R régulier).

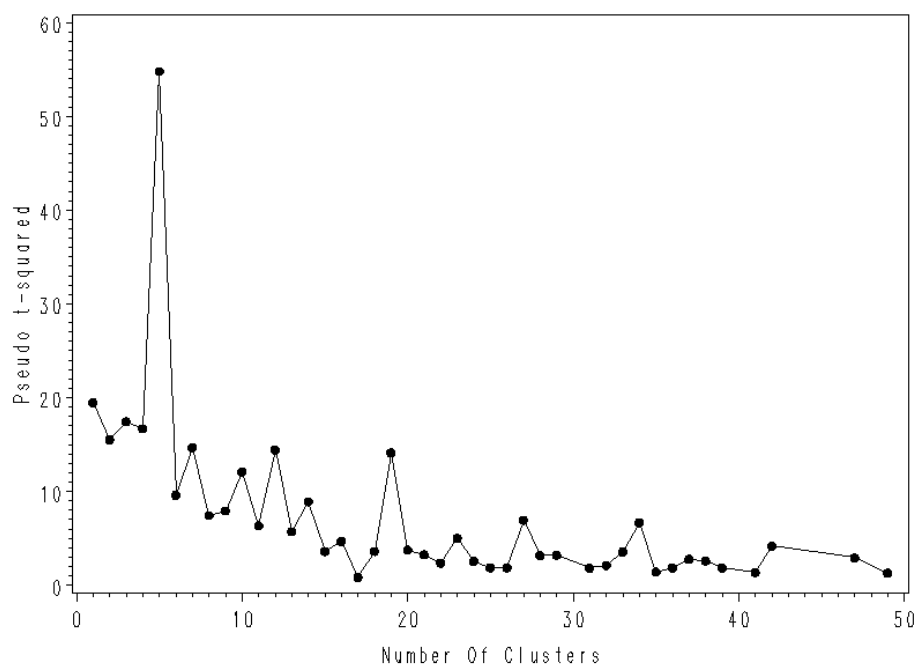
2.4. IDENTIFICATION DU NOMBRE DE REGROUPEMENTS À CONSERVER

Suite à la sélection et l'ajustement des variables, l'analyse de regroupement fut conduite pour chaque unité de compilation et pour chaque ensemble de structure. Puisqu'il s'agit d'une analyse de type hiérarchique, il faut définir le nombre de regroupements qui sera conservé. Or, il n'existe aucune méthode qui soit pleinement satisfaisante pour déterminer précisément le nombre de regroupements à conserver (SAS Institute, 2003). La détermination de ce nombre implique de savoir d'abord le niveau de précision et l'information que doivent contenir les regroupements créés. Pour ce faire, il est utile de calculer le coefficient de détermination (R^2), qui exprime la proportion de la variance qui est expliquée par le niveau de regroupement, pour évaluer le niveau de perte d'information à un niveau donné de regroupement. Ainsi, conserver l'ensemble des strates permet d'expliquer 100 % de la variance ($R^2=1$) et regrouper toutes les strates dans un seul groupe ne permet d'exprimer aucunement la variance ($R^2=0$). Le défi est donc d'identifier l'équilibre idéal entre ces deux extrêmes afin d'exprimer le maximum de variance avec le minimum de groupes.

D'autres outils sont aussi disponibles pour mieux orienter la détermination du nombre de regroupement. Tout d'abord, la procédure CLUSTER permet de générer des statistiques d'analyse. C'est ainsi que nous avons utilisé le *Cubic Clustering Criterion* (CCC), le *Pseudo F* et le *Pseudo t^2* . Ces statistiques ne donnent en aucun cas une réponse unique du nombre de regroupement optimal. Elles donnent des indications générales et permettent d'éliminer les situations extrêmes. Ainsi, le CCC permet principalement de connaître le nombre minimal de groupes à générer pour maintenir une certaine cohérence. Il est recommandé d'avoir un CCC d'au moins de 2 et les valeurs négatives indiquent un nombre clairement insuffisant de regroupement (SAS Institute, 2003). Par exemple, pour l'ensemble régulier de l'unité 064F, un minimum de 6 groupes est suggéré par cette statistique. Pour le *pseudo F*, des valeurs relativement élevées pour un niveau de regroupement indiquent un point d'arrêt possible, sans conclure clairement. Dans la plupart des cas, les valeurs élevées se retrouvent aux extrêmes, soit avec très peu de groupe, autour du nombre suggéré par le CCC, ou avec un très grand nombre de groupe, donnant un R^2 de près de 1. Cette statistique n'apparaissait donc pas très utile avec les présentes données si ce n'est que pour confirmer le nombre minimal de groupes à conserver.

Enfin, le *pseudo t²* doit être analysé en partant du nombre total de strates : en descendant dans la hiérarchie, des bonds locaux de cette statistique proposent des niveaux possibles de regroupement pour le niveau précédent le bond (SAS Institute, 2003). Par exemple, pour le même cas que celui présenté pour le CCC, les valeurs suggèrent entre autres les niveaux de 35, 28, 24, 20, 17, 15, 13, 11, 8 et 6 groupes (figure 1). Le grand nombre de niveaux suggérés implique une hypothèse préalable de définition du niveau de précision voulu car, entre les deux extrêmes (35 et 6 groupes), le R^2 varie de 0,93 à 0,61.

Figure 1 - Valeurs du pseudo t^2 en fonction du nombre de groupes (*clusters*) générés

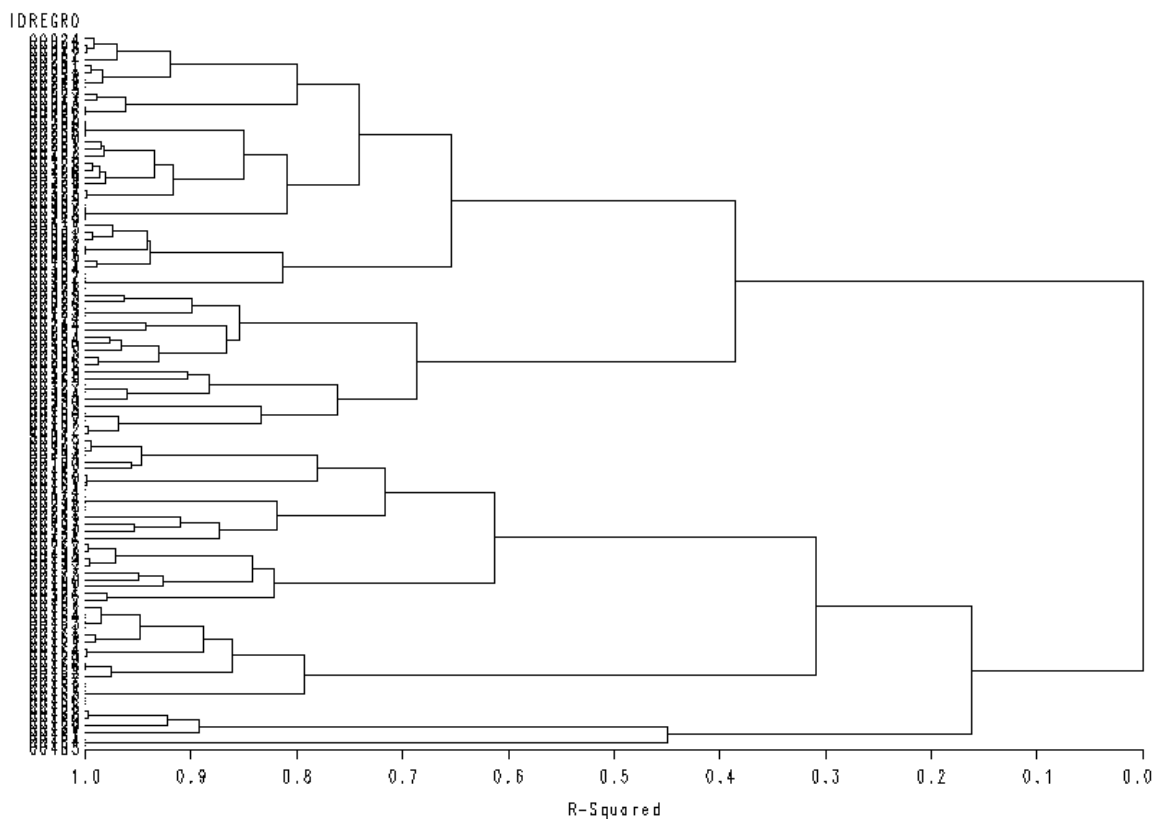


2.5. SCHÉMA HIÉRARCHIQUE

Puisque les regroupements générés par l'analyse sont de type hiérarchique, il est possible de représenter schématiquement l'arbre hiérarchique des regroupements en fonction de différentes variables grâce à la procédure TREE de SAS/STAT. La figure 2 en donne un exemple alors que l'annexe 1 présente sous format numérique les schémas pour tous les ensembles. On y retrouve sur l'axe vertical la liste de toutes les strates soumises à l'analyse (IDREGRO est l'identifiant de chaque strate) et apparaît à droite de cet axe, le schéma de la succession des regroupements. Ces divers niveaux de regroupement sont présentés en fonction du coefficient de détermination (R^2) qu'ils permettent. Ainsi, à mesure que l'on regroupe des strates, une proportion toujours plus

faible de la variance est expliquée. Il est à remarquer cependant que les premiers regroupements n'ont que très peu d'influence sur le R^2 alors que pour des valeurs de R^2 de moins de 0,80, chaque nouveau groupe fait chuter de façon importante la valeur du R^2 . L'observation de cette interaction entre le nombre de groupes formés et le R^2 résultant nous a permis d'orienter le niveau de regroupement. Ainsi, nos divers essais nous ont permis de constater que l'équilibre idéal avec les données analysées se situait avec un R^2 de 0.85 à 0.90, ce qui correspond à environ 20 % du nombre de strates en terme de nombre de groupes formés.

Figure 2 - Arbre hiérarchique généré par la procédure TREE pour l'analyse faite sur le groupe régulier de l'unité 064F



L'utilisation du schéma hiérarchique permet aussi de valider les groupes formés et de faire des corrections s'il y a lieu. Le seuil cible utilisé, soit 20 % du nombre de strates, permet de créer des ensembles homogènes et logiques dans la majorité des cas, mais il importe de faire des ajustements pour quelques cas, en augmentant ou en diminuant le niveau de regroupement. Cette étape de validation et d'ajustement devient donc essentielle pour, d'une part, s'assurer de la justesse des groupes créés et, d'autre part, pour ajuster les groupes face à des facteurs ou des variables secondaires qui pouvaient difficilement être pris en compte dans l'analyse de

regroupement. C'est ainsi que certaines modifications des regroupements ont été faites de manière manuelle. Pour l'ensemble irrégulier, une attention particulière a été portée sur les regroupements possédant une certaine variation de la surface terrière alors que sur l'ensemble régulier, c'est notamment sur les regroupements possédant plus de 2 classes d'âge différentes que la validation a porté. Ceci a eu pour effet de modifier légèrement le nombre de groupes. Par exemple, pour l'ensemble irrégulier de l'unité de compilation 064M, le nombre de groupes initialement fixé était de 40. Après validation, il est passé à 41. À l'opposé, l'ensemble irrégulier de l'unité de compilation 064F est passé de 59 à 51 après validation.

3. RÉSULTATS

3.1. NOMBRE DE GROUPES

Le tableau suivant présente le nombre de groupes finalement retenus pour chaque unité de compilation ainsi que pour chaque ensemble. Il en résulte une diminution de strate de 79 % en moyenne.

Tableau 1 - Nombre de groupes retenus par unité de compilation et par ensemble

Unité de compilation	Ensemble	Nombre initial de strates d'inventaire	Nombre de groupes retenus	Pourcentage réduction (%)
064F	Irrégulier	295	51	83
064F	Régulier	116	28	76
064M	Irrégulier	203	41	80
064M	Régulier	343	68	80
064R	Régulier	163	43	73
	Total	1120	231	79

3.2. ANALYSE STATISTIQUE

Afin d'évaluer l'effet de la diminution du nombre de regroupement sur la précision de l'inventaire, une analyse à l'aide du logiciel SCP a été réalisée pour l'unité de compilation 064M. Le tableau 2 présente la précision moyenne sur le volume total ainsi que pour les principales essences. Ce calcul est réalisé sur les strates ou groupes ayant au moins 1 placette établie, recrutée ou actualisée. Dans le cas des précisions moyennes par essence, les strates ou groupes retenus pour le calcul sont uniquement ceux ayant une présence de l'essence.

Tableau 2 - Comparaison des précisions par type de regroupement pour l'unité de compilation 064M.

Type de regroupement	Nombre de strates ou groupes	Nombre moyen de placettes	Précision moyenne (%) sur le volume						
			BOJ	BOP	EPN	ERR	ERS	SAB	Total
Original	499	9,4	14	19	16	4	7	18	63
Proposé	109	43,2	22	46	34	21	11	46	79

Il en résulte que le regroupement proposé améliore de façon importante la précision de l'inventaire. La précision moyenne sur le volume total est augmentée de 25 % alors que par essence, elle augmente d'au moins 55 %, et parfois même, de plus de 400 %.

Une analyse de la précision au niveau de 2 regroupements de l'unité de compilation 064M a aussi été réalisée. Les précisions sur la surface terrière totale et par essence des groupements 33 des ensembles régulier (tableau 3) et irrégulier (tableau 4) ont été calculées à l'aide du compilateur SCP (Blouin et al., 2002). Ces tableaux présentent, dans un premier temps, les strates d'inventaire contenues dans un même groupe ainsi que leurs surfaces terrières par essences et totales (champs débutant par ST_). De plus, la précision de ces mêmes éléments est présentée pour chaque strate d'inventaire contenue dans un groupe (champs débutant par PR_) ainsi que pour le groupe lui-même.

Tout comme au niveau de l'unité de compilation, les groupes proposés améliorent sensiblement la précision tant du point de vue de la surface terrière par essence que totale.

Tableau 3 - Comparaison des précisions (%) de la surface terrière par essence et totale pour le regroupement 33 de l'ensemble régulier de l'unité de compilation 064M

idregro	nomregro				sup	STBOJ	PR_BOJ	STBOP	PR_BOP	STEPB	PR_EPB	STEPN	PR_EPN	STERR	PR_ERR	STPET	PR_PET	STSAB	PR_SAB	STTHO	PR_THO	STTOT	PR_TOT	
00051	F	BB1	A 3 50	FE32	323.3	0.0		5.2	0.00	0.6	0.00	1.0	0.00	6.6	0.00	1.7	0.00	1.1	0.00	0.0		18.2	0.16	
00057	F	BBBB	A 3 50	C MJ22	11544.4	2.1	0.26	11.0	0.81	1.7	0.34	1.7	0.03	1.1	0.41	0.0	0.00	2.9	0.50	0.0	0.00	20.6	0.83	
00246	M	BB1R	B 3 70	E MJ10	567.8	2.6	0.00	6.6	0.32	1.2	0.00	0.8	0.00	1.9	0.00	0.5	0.00	4.6	0.11	4.0	0.00	24.3	0.69	
00262	M	BBBBR	B 3 70	B MJ22	2188.9	2.8	0.00	7.7	0.54	1.9	0.23	2.5	0.00	1.2	0.00	0.6	0.00	2.7	0.50	1.1	0.00	21.2	0.74	
00268	M	BBBBB	B 2 70	D MJ22	1029.9	2.9	0.00	7.5	0.60	3.5	0.16	0.7	0.00	0.7	0.00	1.4	0.00	6.0	0.65	0.0		22.8	0.79	
00269	M	BBBBB	B 3 50	B MJ21	410.7	5.0	0.00	4.8	0.00	0.6	0.00	1.8	0.00	1.2	0.00	0.4	0.00	3.4	0.00	0.5	0.00	19.4	0.12	
00270	M	BBBBB	B 3 50	B MJ25	1195.7	2.3	0.24	9.7	0.74	1.6	0.07	1.7	0.00	0.2	0.00	0.0		3.1	0.35	0.0		18.7	0.79	
00271	M	BBBBB	B 3 50	C MJ12	1791.4	2.0	0.00	8.4	0.48	2.0	0.00	0.3	0.00	2.2	0.07	0.0		7.7	0.47	0.0		23.2	0.83	
00272	M	BBBBB	B 3 50	C MJ22	8645.5	3.7	0.41	9.2	0.74	1.4	0.41	1.2	0.00	1.0	0.08	0.8	0.00	7.4	0.52	0.0		24.8	0.88	
00273	M	BBBBB	B 3 50	C MS22	1473.3	2.2	0.00	9.8	0.44	1.5	0.03	0.9	0.00	0.3	0.00	1.0	0.00	5.1	0.37	0.9	0.00	21.7	0.71	
00274	M	BBBBB	A 3 70	MS22	270.8	1.7	0.00	9.5	0.42	2.1	0.20	1.8	0.00	0.2	0.00	0.3	0.00	5.2	0.14	0.0		20.8	0.80	
00275	M	BBBBB	B 3 70	C MJ12	798.9	4.7	0.00	6.8	0.09	0.7	0.00	2.1	0.00	1.7	0.00	1.6	0.00	3.8	0.00	0.0		21.4	0.43	
00276	M	BBBBB	B 3 70	C MJ22	1749.2	2.5	0.30	6.9	0.43	1.1	0.06	0.7	0.00	1.3	0.00	0.0		6.8	0.32	0.0	0.00	19.3	0.75	
00331	M	RBB	B 3 70	B MJ22	1439.4	4.7	0.29	4.3	0.00	2.6	0.00	0.8	0.00	1.0	0.32	0.0		4.9	0.56	0.6	0.00	21.0	0.69	
00337	M	RBB	C 3 70	B RS25	250.7	1.3	0.00	6.5	0.17	1.4	0.09	3.9	0.06	1.2	0.00	0.3	0.00	5.3	0.40	3.4	0.00	23.2	0.81	
00352	M	SBB	B 3 50	B MJ22	2189.7	1.6	0.00	6.2	0.66	1.7	0.18	1.2	0.00	1.0	0.00	0.2	0.00	9.1	0.48	0.1	0.00	21.1	0.73	
00421	M	EL	BBBBR	C 3 70	C MJ22	1842.9	0.9	0.00	6.7	0.40	2.1	0.32	2.1	0.00	3.0	0.17	0.1	0.00	2.2	0.00	0.0		17.8	0.71
00464	M	EL	SBB	C 3 70	MJ22	702.7	1.6	0.00	4.9	0.49	1.3	0.08	1.1	0.00	2.4	0.00	1.6	0.00	6.0	0.57	2.9	0.00	22.0	0.73
07011*	M	CPE 2003	BBBBR	C 3 70	D MJ22	351.4	0.9		6.7		2.1		2.1		3.0		0.1		2.2		0.0		17.8	
Cluster 33 (modèle régulier)					38766.6	2.6	0.75	8.8	0.89	1.7	0.76	1.4	0.59	1.3	0.74	0.4	0.38	5.0	0.84	0.3	0.40	21.6	0.94	

*STRATE ASSOCIÉE

Tableau 4 - Comparaison des précisions (%) de la surface terrière par essence et totale pour le regroupement 33 de l'ensemble irrégulier de l'unité de compilation 064M

idregro	nomregro	sup	STBOJ	PR_BOJ	STBOP	PR_BOP	STEPB	PR_EPB	STERR	PR_ERR	STERS	PR_ERS	STSAB	PR_SAB	STTHO	PR_THO	STTOT	PR_TOT
00283	M BJ+R B 2 VIN C MJ12 JARDINEE 1.13	14478.4	9.0	0.80	1.1	0.37	2.1	0.54	1.7	0.54	2.7	0.35	5.3	0.70	1.9	0.17	24.4	0.90
00284	M BJ+R B 2 VIN C MJ22 IRREGULIERE 1.01	3963.7	12.1	0.64	0.9	0.00	1.1	0.00	0.7	0.00	1.0	0.00	8.3	0.54	0.0		24.3	0.77
00286	M BJ+R B 2 VIN E MJ10	408.1	11.7	0.78	0.9	0.00	1.7	0.00	1.2	0.00	0.1	0.00	7.2	0.38	2.4	0.00	25.6	0.40
00287	M BJ+R B 3 VIN MJ22 IRREGULIERE 1.12	778.1	9.7	0.00	0.7	0.00	1.7	0.00	1.6	0.00	2.8		5.1	0.20	4.8		27.2	0.43
00290	M BJ+R C 2 VIN C MJ22 IRREGULIERE 1.08	1421.8	11.5	0.54	2.3	0.00	1.9	0.15	0.5	0.00	0.1	0.00	4.8	0.63	2.5	0.00	24.1	0.75
00291	M BJ-R B 3 JIN B MJ25 IRREGULIERE 1.08	513.3	10.4	0.62	0.5	0.00	1.6	0.00	0.4	0.00	0.0		9.3	0.59	0.8	0.00	23.9	0.70
00294	M BJ-R B 3 VIN C MJ12 IRREGULIERE 1.09	1872.3	10.6	0.61	1.5	0.00	0.7	0.00	2.2	0.02	3.7	0.00	5.6	0.46	0.8	0.00	25.0	0.89
00310	M ERR B 2 VIN C MJ12 IRREGULIERE 1.03	1630.3	9.7	0.08	2.2	0.00	1.4	0.00	3.0	0.00	1.7	0.00	4.6	0.00	0.0	0.00	23.1	0.40
00436	M EL BJ+R B 2 VIN C MJ12 IRREGULIERE 1.1	1101.5	11.8	0.28	0.7	0.00	2.7	0.00	1.4	0.00	1.4	0.00	8.8	0.63	0.0		26.9	0.65
00437	M EL BJ+R C 2 VIN C MJ12 IRREGULIERE 1.1	3305.5	9.1	0.65	1.5	0.23	1.4	0.07	1.1	0.19	2.3	0.00	4.6	0.59	2.0	0.00	23.5	0.83
	Cluster 33 (modèle irrégulier)	29473	9.9	0.86	1.3	0.68	1.8	0.69	1.5	0.72	2.2	0.50	5.8	0.86	1.5	0.46	24.4	0.94

3.3. VALIDATION DU NOMBRE DE REGROUPEMENTS

Suite à la présentation des résultats, des demandes divergentes de modifications des groupes ont été formulées. Certaines proposaient d'augmenter significativement leur nombre alors que d'autres, au contraire, avaient pour effet de les diminuer sensiblement. Afin, de justifier ou non la modification du nombre initial de groupes, il a été convenu d'évaluer la variabilité « intra et inter » strates.

L'analyse de regroupement détermine par défaut un R^2 équivalant à 1 lorsque le nombre de groupe égale le nombre d'unité à regrouper (ie au nombre de strate). Or, ceci n'est pas le cas puisque dans les faits, les données issues de la strate d'inventaire constituent une moyenne des échantillons de cette même strate. Ainsi, cette moyenne n'explique pas toute la variance de la strate. Son R^2 n'est donc pas égal à 1 mais inférieur ou égal à 1.

Afin d'évaluer cette variabilité, nous avons vérifié dans quelle mesure la moyenne des échantillons exprime la variance. Pour ce faire, nous avons réalisé le même travail d'analyse de regroupement en utilisant les placettes comme unité de base ($R^2 = 1$) plutôt que la strate d'inventaire. Ceci a été réalisé pour l'ensemble irrégulier de l'unité de compilation 064F. Afin de connaître les seuils issus du nombre de clusters retenus (en fonction du R^2), nous avons fait varier le nombre de groupes de 203 à 2. Ainsi, pour les 203 strates d'inventaire contenues dans l'analyse de regroupement, nous avons calculé le R^2 pour un nombre équivalent de clusters (203), mais cette fois en évaluant le R^2 en fonction des placettes.

Ces travaux ont été confiés à Mme Hélène Crépeau, consultante en statistique du Service de consultation statistique du Département de mathématiques et de statistique de l'Université Laval. La correspondance complète est présentée en annexe 3.

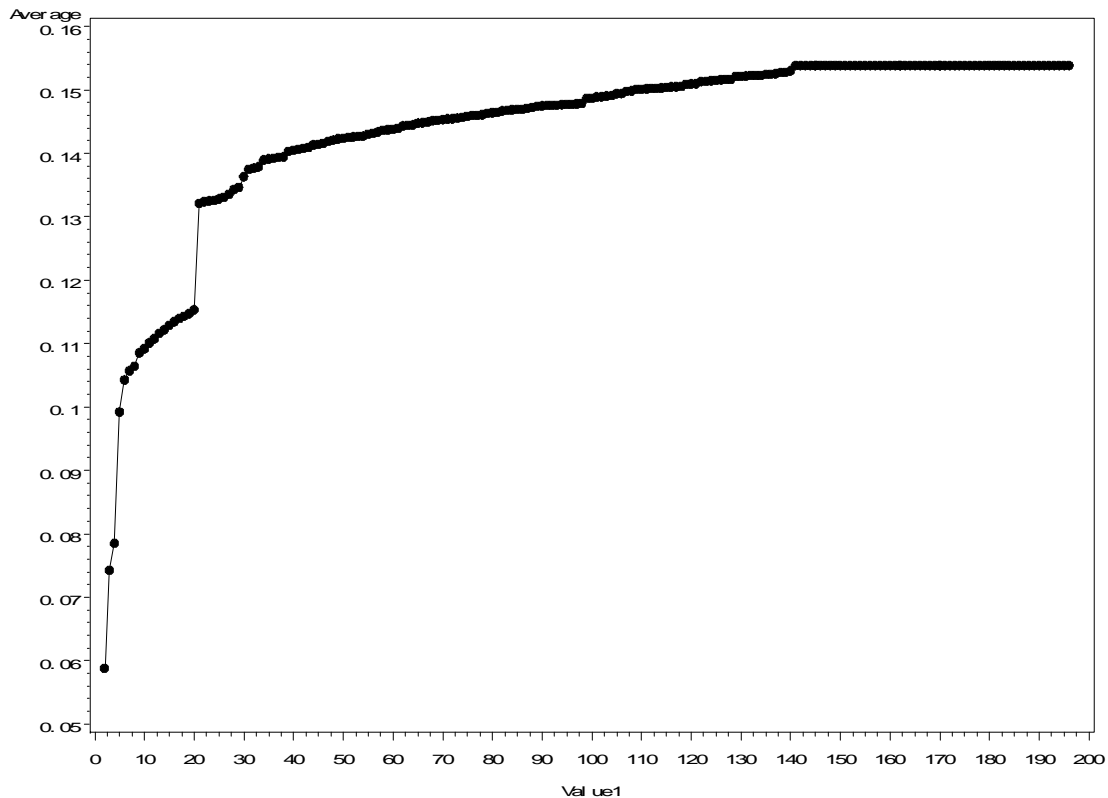
En imposant un regroupement de 203 groupes soit l'équivalent des strates d'inventaires, le R^2 basé sur les placettes comme unité, est de seulement 15,4 % (figure 3). Il y a donc une très grande variabilité d'une placette à l'autre pour une même strate. L'utilisation d'un nombre aussi élevé de strates d'inventaire comportant une telle marge d'erreur est donc questionnable. Il faut donc reconnaître qu'une strate d'inventaire n'est pas homogène et que les données dendrométriques qui s'y rattachent peuvent être sensiblement différentes de ce qu'on pourrait s'attendre, notamment en fonction de son appellation d'inventaire.

Comme le démontre la figure 3, en restreignant le nombre de groupes à 40, le R^2 obtenu est 14,1 %. Ceci est un argument de taille justifiant un fort regroupement puisque la perte en R^2 est faible alors que la diminution du nombre de regroupements est importante. Il s'agit donc d'un compromis intéressant.

La figure 3 décèle aussi des sauts relativement importants du R^2 lorsque le nombre de groupes passe de 4 à 5 et de 20 à 21. On remarque aussi une légère augmentation du R^2 lorsque l'on passe de 28 à 35 groupes. À partir de 40, l'augmentation du R^2 est régulière pour atteindre un plateau lorsque le nombre de groupes est d'environ 140.

En fonction de cette figure, il apparaît évident qu'il ne serait pas souhaitable de retenir un nombre de regroupement inférieur à 21 et qu'il serait superflu d'en maintenir plus de 40.

Figure 3 - Évolution du R^2 en fonction du nombre de groupes retenus



4. DISCUSSIONS

L'analyse de regroupement se base sur les données issues des placettes et non sur l'appellation de la strate d'inventaire. Dans l'actuel cas, elle utilise des variables quantitatives tels la surface terrière par essence et totale, le diamètre moyen quadratique, l'âge, la hauteur ainsi que le pourcentage de surface terrière par classe de 10 cm. **L'analyse de regroupement n'amène donc pas de « perte de diversité ».** Dans les faits, la seule perte qu'amène l'analyse de regroupement réside dans la diversité des appellations de strate d'inventaire. Pour ne citer qu'un exemple, le regroupement no. 8 de l'unité de compilation 064M (ensemble régulier) regroupe une plantation de PIG avec des peuplements naturels d'épinette noire. Dans les faits, la plantation de PIG est une strate associée. Ces données sont exactement les mêmes que la strate d'épinette noire.

Comme le démontre le tableau 2, l'analyse de regroupement a permis de réduire de façon importante le nombre d'ensemble soit de près de 80 %. L'utilisation des schémas hiérarchiques et les validations d'usage ont permis de maintenir des groupes suffisamment homogènes pour une planification stratégique dans Patchworks. De plus, les analyses statistiques à l'échelle de l'unité de compilation et à l'échelle du regroupement indiquent une amélioration substantielle de la précision sur les volumes par essences de l'unité de compilation, mais également sur les volumes totaux et par essences pour chaque strate d'inventaire.

Le tableau 2 démontre aussi un nombre de placettes beaucoup plus élevé par unité de regroupement soit en moyenne 43,2 plutôt que 9,4. Ceci pourrait permettre une approche probabiliste sur chaque regroupement lors du diagnostic sylvicole (CERFO, 2004) Ainsi, plutôt qu'utiliser des filtres fonctionnant par seuils pour définir la production prioritaire et le traitement à appliquer, il serait possible d'évaluer la probabilité qu'a un groupe de se trouver dans une production prioritaire quelconque ou à être traité d'une certaine façon.

La validation du nombre de groupes par l'évaluation de la variabilité « intra et inter » strate permet de trouver le meilleur compromis. Elle a permis de conclure que pour l'ensemble irrégulier de l'unité de compilation 064M, le nombre initialement fixé (40) était situé à la limite supérieure du nombre de regroupements adéquats. Ce nombre représentait ainsi un bon compromis puisque les valeurs des variables d'un même regroupement étaient similaires comme le démontre l'annexe 2. De plus, il permet de diminuer sensiblement le nombre de strates et respecte les sauts de la figure 3. Toutefois, basé sur la même figure, le nombre optimal serait

plutôt 21. À cet effet, ce nombre équivaut à un R^2 d'environ 0.75 (basé sur l'unité strate d'inventaire), pour l'ensemble irrégulier de l'unité de compilation 064M. Des validations sommaires indiquent que ces regroupements demeurent intéressants pour une planification stratégique, notamment en terme de composition. Toutefois, les strates incluses dans un même regroupement diffèrent parfois en terme de surface terrière totale. Cependant, il faut rappeler que l'objectif de ce travail consistait à identifier des strates d'inventaire pouvant être regroupées afin de générer une courbe décrivant l'évolution de ces dernières dans le temps. Il n'est toutefois pas nécessaire que toutes les strates d'un même groupe soient placées en un même point sur cette courbe puisque le logiciel Patchworks peut permettre un positionnement différent pour chaque polygone touché par une même courbe.

Le nombre de regroupement pourrait même diminuer davantage. En effet, lors d'une planification stratégique avec Patchworks, il est avantageux de minimiser le nombre de regroupement afin de réduire le temps de résolution des solutions optimales pour permettre d'explorer davantage de scénarios.

L'évaluation de la variabilité « intra et inter » strate a aussi permis de mettre en lumière la troublante diversité présente dans une même strate d'inventaire. À cet effet, il serait avantageux d'analyser davantage les causes d'une aussi faible expression de la proportion de la variance. Est-ce la qualité de la photo-interprétation? Est-ce l'utilisation d'une superficie minimale de photo-interprétation? Est-ce le regroupement des strates cartographique en strates d'inventaire ou est-ce que cette piètre performance est inévitable? Quoiqu'il en soit, il apparaît inefficace, peu pertinent et incorrect d'utiliser un nombre élevé de strates d'inventaire pour une planification stratégique dans les conditions actuelles. De plus, il est peu souhaitable d'entreprendre l'élaboration d'un diagnostic sylvicole sans tenir compte de la grande diversité intrinsèque à chaque strate d'inventaire.

5. CONCLUSION

L'analyse de regroupement réalisée est concluante puisqu'elle a atteint l'objectif principal souhaité. Elle a permis de diminuer de façon importante le nombre de strates tout en représentant le maximum de la variabilité des différentes variables caractérisant chaque strate. Les outils utilisés impliquent cependant pour l'analyste de faire de nombreux choix comme les méthodes de calcul et ses réglages, la sélection des variables et leur ajustement ainsi que la détermination d'un seuil de précision souhaité. Cela implique nécessairement une série d'essais et une validation importante afin de s'assurer de la justesse des regroupements réalisés. La représentation schématique a été pour cela fort utile et elle a permis de relier plus facilement l'analyse purement statistique avec le sens pratique que doivent nécessairement avoir les groupes de strates formés.

Il faut mentionner que l'analyse de regroupement se base sur les données issues des placettes et non sur l'appellation de la strate d'inventaire. Dans l'actuel cas, elle utilise des variables quantitatives tels la surface terrière par essence et totale, le diamètre moyen quadratique, l'âge, la hauteur ainsi que le pourcentage de surface terrière par classe de 10 cm. L'analyse de regroupement n'amène donc pas de « perte de diversité ». Dans les faits, la seule perte qu'amène l'analyse de regroupement réside dans la diversité des appellations de strate d'inventaire.

La validation du nombre de groupes par l'évaluation de la variabilité « intra et inter » strate a permis de conclure que pour l'ensemble irrégulier de l'unité de compilation 064M, le nombre initialement fixé (40) était situé dans la limite supérieure du nombre de regroupements adéquats. Dans un cadre de planification stratégique avec Patchworks, plus le nombre de groupes est restreint, plus le temps de résolution est réduit et qu'il devient alors possible d'explorer davantage de scénarios.

L'évaluation de la variabilité « intra et inter » strate a aussi permis de mettre en lumière la troublante diversité qui est incluse dans une même strate d'inventaire. À cet effet, il serait avantageux d'analyser davantage les causes d'une aussi faible expression de la proportion de la variance. L'arrivée du 4^e inventaire décennal permet d'anticiper une nette amélioration. L'analyse de regroupement pourrait alors inclure des variables écologiques.

6. RÉFÉRENCES

Blouin, D., E. Gaulin et M. Ouellet, 2002. Compilateur SCP - (Système de calcul de la précision). Centre collégial de transfert de technologie en foresterie (CERFO), Rapport 2002-10. 15 p.

CERFO, 2004. Analyse des problématiques sur les calculs de la possibilité forestière. [Résumé exécutif](#), [résumé](#) et rapport. www.commission-foret.qc.ca

Commission d'étude sur la gestion des forêts publiques du Québec, 2004. Rapport. <http://www.commission-foret.qc.ca/rapportfinal.htm>

Lessard, G, 1991. Groupes socio-écologiques d'espèces pour la sapinière à bouleau blanc de l'est, secteur Forêt Montmorency. Thèse de maîtrise.

SAS Institute Inc., 2003. SAS/STAT User's Guide, Version 9, SAS Institute Inc., Cary, NC.

ANNEXE 1 - SCHÉMA HIÉRARCHIQUE DES REGROUPEMENTS

Voir fichiers powerpoint sur CD

ANNEXE 2 - RÉSULTATS DÉTAILLÉS DE L'ANALYSE DE REGROUPEMENT

Voir fichiers excel sur CD

ANNEXE 3 - CORRESPONDANCE DE M^{ME} HÉLÈNE CRÉPEAU

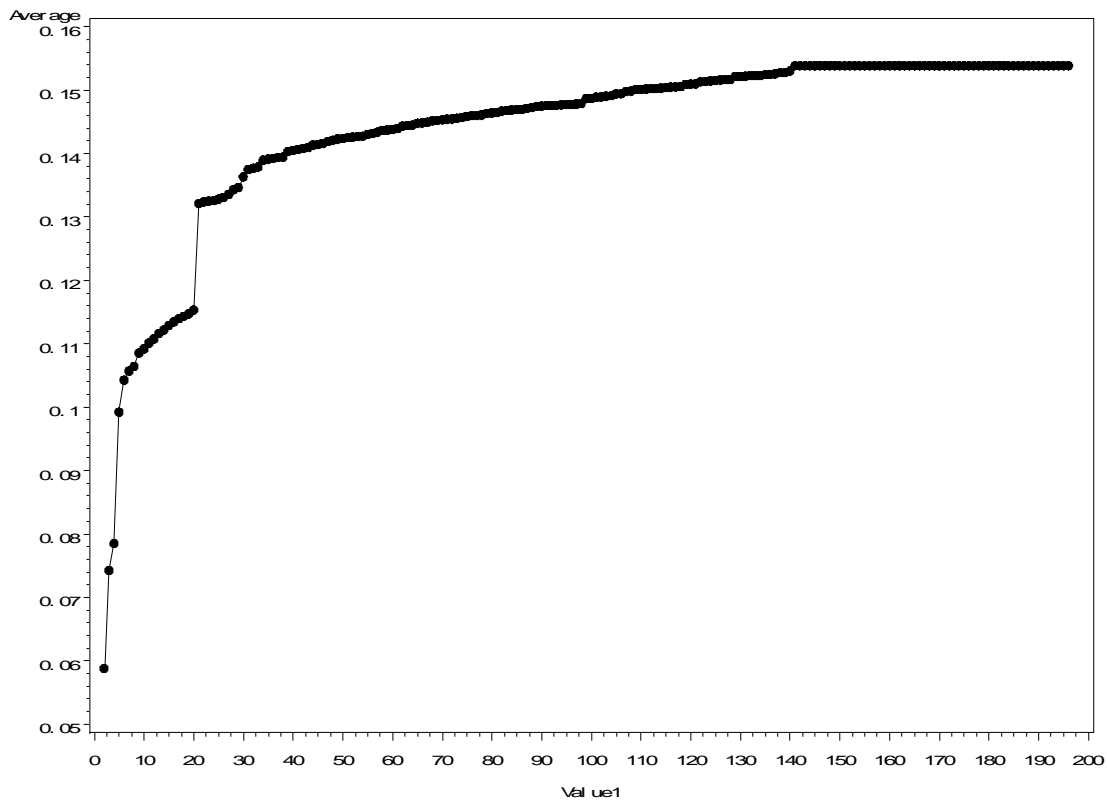
Lundi 10 avril 2006

Pascal Gauthier
Chargé de projet
2424, chemin Sainte-Foy
Sainte-Foy (QC) G1V 1T2

Voici un résumé des analyses effectuées sur les données d'inventaires forestiers que vous m'avez fait parvenir. Deux fichiers de données sont disponibles. Le premier fichier contient les informations (volume, diamètre, qualité etc.) pour chacune des placettes échantillonnées des 203 strates d'inventaire représentant le territoire à l'étude. L'objectif poursuivi est de regrouper les strates d'inventaire par rapport aux variables mesurées sur les placettes échantillonnées dans chacune des strates. L'information des placettes a donc été synthétisée en calculant les moyennes par strates de chacune des variables et que l'on retrouve dans le deuxième fichier.

Une analyse de regroupement hiérarchique (cluster analysis) a été appliquée sur les données du deuxième fichier. Trois critères peuvent être utilisés pour déterminer le nombre de groupes que l'on retrouve dans les données : le « cubic clustering criterion (CCC), le pseudo F (PSF) et le pseudo t2 (PST2) ». Malheureusement, avec nos données, ces critères ne sont pas très concluants pour déterminer le nombre de groupes. En utilisant ces trois critères, on obtient des résultats différents. Par exemple, avec CCC, on a un maximum local à 13 et à 21; avec PSF il y aurait possiblement un maximum à 12 mais très peu marqué et avec PST2 il y aurait plusieurs possibilités i.e. à 2, 3, 5, 6, 9, 11,13, 17 et 21. Pour chaque regroupement, on a aussi une statistique R^2 qui est calculée et qui représente la proportion de la variance qui est expliquée par ce regroupement. Le R^2 est de 100 % lorsque le nombre de groupes est égal au nombre d'observations et 0 % lorsqu'il n'y a qu'un seul groupe. Cependant, puisque l'analyse de regroupement est faite sur des moyennes, le R^2 tel que calculé par la procédure CLUSTER de SAS n'est pas représentatif de nos données par placette. Nous avons donc recalculé le R^2 en utilisant les données par placette pour chacun des regroupements obtenus. Le graphique suivant présente le R^2 en fonction du nombre de cluster.

Graphique du R^2 en fonction du nombre de cluster



Comme vous pouvez le constater, il y a une très grande variabilité entre les parcelles d'une même strate. En effet, le R^2 obtenu si on conserve les 203 strates d'inventaires sans regroupement n'est que de 15.4. D'après ce graphique, on observe une grosse augmentation du R^2 lorsqu'on passe de 4 à 5 groupes (R^2 passe de 7.9 à 9.9) et de 20 à 21 groupes (R^2 passe de 11.5 à 13.2%). D'un point de vue statistique, un regroupement de 21 groupes semblerait un bon choix puisque que le R^2 ne diminue que de 2.2. Par contre, il y a peut-être d'autres considérations qu'il faut examiner afin de faire un choix définitif. Est-ce qu'un nombre de groupes égal à 21 est trop petit pour les objectifs que vous poursuivez? Est-ce qu'un regroupement de 40 groupes qui nous donne un R^2 de 14.1 est plus acceptable d'un point de vue géographique? Quel est l'impact du nombre de cluster sur les analyses finales?

J'espère que les résultats de ces analyses vous aideront dans votre décision. Si vous avez des questions ou commentaires, n'hésitez pas à communiquer avec moi.

Hélène Crépeau
Service de consultation statistique
Département de mathématiques et de statistique
Université Laval