

# Une relation hauteur-diamètre tenant compte de l'influence de la station et du climat pour 20 espèces commerciales du Québec



**Mémoire de recherche forestière n° 153**  
**Direction de la recherche forestière**

par

Mathieu Fortin, Sylvain Bernier, Jean-Pierre Saucier et François Labbé



# **Une relation hauteur-diamètre tenant compte de l'influence de la station et du climat pour 20 espèces commerciales du Québec**

Mémoire de recherche forestière n° 153

par

Mathieu FORTIN, ing.f., *Ph. D.*,

Sylvain BERNIER, stat., M. Sc.

Jean-Pierre SAUCIER, ing.f., Dr Sc.,

et

François LABBÉ, ing.f., M. Sc.

## Mandat de la DRF

Le mandat de la Direction de la recherche forestière (DRF) au sein du ministère des Ressources naturelles et de la Faune (MRNF) est de participer activement à l'amélioration de la pratique forestière au Québec. La concrétisation de ce mandat passe par la réalisation de travaux, principalement à long terme et d'envergure provinciale, lesquels intègrent à la fois des préoccupations de recherche fondamentale et de recherche appliquée.

Elle subventionne aussi des recherches universitaires à court ou à moyen terme. Ces recherches, importantes pour le Ministère, sont complémentaires aux travaux de la DRF ou réalisées dans des créneaux où elle ne peut s'impliquer.

Elle contribue à la diffusion de nouvelles connaissances, d'avis et de conseils scientifiques.

## Les mémoires de recherche forestière de la DRF

Depuis 1970, chacun des Mémoires de recherche forestière de la DRF est révisé par un comité *ad hoc* formé d'au moins trois experts indépendants. Cette publication est produite et diffusée à même les budgets de recherche et de développement, comme autant d'étapes essentielles à la réalisation d'un projet ou d'une expérience. Ce document à tirage limité est également disponible dans notre site Internet en format pdf.

Vous pouvez adresser vos demandes à :

Ministère des Ressources naturelles et de la Faune  
Direction de la recherche forestière  
2700, rue Einstein, Québec (Québec)  
Canada, G1P 3W8  
Courriel : [recherche.forestiery@mrnf.gouv.qc.ca](mailto:recherche.forestiery@mrnf.gouv.qc.ca)  
Internet : [www.mrnf.gouv.qc.ca/forets/  
connaissances/recherche](http://www.mrnf.gouv.qc.ca/forets/connaissances/recherche)

© Gouvernement du Québec

On peut citer ce texte en indiquant la référence.

Toutes les publications produites par la Direction de la recherche forestière, du ministère des Ressources naturelles et de la Faune, sont protégées par les dispositions de la Loi sur le droit d'auteur, les lois, les politiques et les règlements du Canada, ainsi que par des accords internationaux. Il est interdit de reproduire, même partiellement, ces publications sans l'obtention préalable d'une permission écrite.

ISBN : 978-2-550-55437-0  
ISBN (PDF) : 978-2-550-55438-7  
F.D.C. 522 (047.3)(714)  
L.C. SD 387.M33

## Notes biographiques



Mathieu Fortin, ingénieur forestier depuis 1998, détient une maîtrise ès sciences forestières (2001) et un grade de *Philosophiæ Doctor* (2006). À l'emploi de la Direction de la recherche forestière depuis 2005, ses travaux portent sur la modélisation de l'accroissement, de la mortalité et du recrutement.



Sylvain Bernier, statisticien diplômé depuis 1989, détient une maîtrise ès sciences en mathématiques, spécialité : statistiques. À l'emploi de la Direction des inventaires forestiers depuis 1997, ses principaux mandats portent sur l'analyse des résultats de l'inventaire forestier décennal. Il agit aussi comme conseiller en statistiques de différents projets.



Jean-Pierre Saucier, ingénieur forestier depuis 1984, détient une maîtrise en aménagement forestier et sylviculture (1992) ainsi qu'un doctorat en sciences forestières de l'École Nationale du Génie Rural des Eaux et des Forêts de Nancy, France (1997). À l'emploi du ministère des Ressources naturelles et de la Faune du Québec depuis 1985, il a participé au développement du système hiérarchique de classification écologique du territoire et a coordonné la réalisation de la série des Guides de reconnaissance des types écologiques ainsi que l'établissement de la productivité des stations. Il occupe le poste de chef du Service de la sylviculture et du rendement des forêts à la Direction de la recherche forestière depuis septembre 2008.



François Labbé, ingénieur forestier depuis 1999, détient une maîtrise ès sciences forestières (2006). À l'emploi de la Direction des inventaires forestiers depuis 2002, l'analyse des résultats de l'inventaire forestier décennal constitue l'un de ses principaux mandats.

# Remerciements

Les auteurs tiennent à remercier M. Georges Blais (ministère des Ressources naturelles et de la Faune du Québec – Direction du soutien aux opérations Faune et Forêts) pour ses commentaires sur la méthodologie employée, M. Jean Noël (ministère des Ressources naturelles et de la Faune du Québec – Direction de la recherche forestière) pour l'illustration de la figure 2, trois réviseurs

anonymes ainsi que l'éditrice associée pour leurs commentaires constructifs, M. Pierre Bélanger, Mmes Sylvie Bourassa et Maripierre Jalbert (ministère des Ressources naturelles et de la Faune du Québec – Direction de la recherche forestière) pour l'édition de l'ouvrage et la mise en page du texte respectivement.



## Résumé

Dans cette étude, nous proposons une relation hauteur-diamètre (HD) pour 20 espèces commerciales du Québec. Le modèle, du type linéaire mixte, a été étalonné à l'aide de données du réseau provincial de placettes permanentes de la Direction des inventaires forestiers. Plusieurs variables locales et régionales, telles que la surface terrière et la température annuelle moyenne, ont été testées dans le modèle. Un effet aléatoire de placette et une structure de covariance linéaire ont également été ajoutés afin de prendre en compte les corrélations spatiales et temporelles. Les résidus normalisés ont été utilisés pour vérifier l'adéquation des éléments de covariance. Une validation croisée (*cross-validation*) a été effectuée afin de tester la capacité prédictive du modèle.

En plus de l'espèce et du diamètre à hauteur de poitrine (dhp, 1,3 m), la surface terrière, la température annuelle moyenne de la période 1971-2000, le drainage, le statut social et l'occurrence d'une perturbation légère se sont avérés des variables significatives dans la prévision de la hauteur d'un arbre. On peut présumer que les inférences statistiques du modèle sont presque sans biais puisque l'analyse des résidus normalisés n'a révélé que de très faibles corrélations persistantes. Par ailleurs, les estimations des biais résultant de la validation croisée étaient très faibles. Comparé à la méthode actuellement en vigueur dans l'inventaire provincial, le modèle de cette étude tient compte de la variabilité climatique et permet ainsi de répondre aux besoins tout aussi bien à l'échelle provinciale que régionale et locale.



## **Abstract**

*In this study, we propose a height-diameter relationship for 20 commercial tree species in Québec. The model, which is of the linear mixed type, was parameterized using data from the provincial network of permanent sample plots of the Direction des inventaires forestiers (Forest Inventory Directorate). Many local and regional variables, such as basal area and mean annual temperature, were tested in the model. A plot random effect and a covariance structure were also included to account for spatial and serial correlations. Normalized residuals were used to check the adequacy of the covariance features. Moreover, a cross-validation was carried out to test the predictive ability of the model.*

*In addition to diameter at breast height (1.3 m) and species, basal area, the 1971–2000 mean annual temperature, plot drainage, social status and the occurrence of disturbance were significant variables used to predict tree height. The statistical inferences could be considered unbiased because the normalized residuals analysis only revealed small, persistent correlations. Moreover, the bias estimates resulting from the cross-validation were small. Compared with the current methodology used in the provincial inventory, the model used in this study accounts for climate variability, and makes it possible to fulfill the needs not only at the provincial level, but also at the regional and local levels.*



# Table des matières

	page
<b>Remerciements</b> .....	iii
<b>Résumé</b> .....	v
<b>Abstract</b> .....	vii
<b>Liste des tableaux</b> .....	xi
<b>Liste des figures</b> .....	xiii
<b>Introduction</b> .....	1
<b>Chapitre premier – Matériel et méthodes</b> .....	3
1.1 Données .....	3
1.2 Sélection du modèle .....	4
1.3 Théorie des modèles mixtes .....	5
1.4 Caractéristiques du modèle.....	6
1.5 Évaluation du modèle.....	8
<b>Chapitre deux – Résultats</b> .....	9
<b>Chapitre trois – Discussion</b> .....	15
3.1 Aspects statistiques.....	15
3.2 Influence des variables.....	16
<b>Conclusion</b> .....	19
<b>Références bibliographiques</b> .....	21



# Liste des tableaux

	page
<b>Tableau 1.</b> Quelques caractéristiques du jeu de données (les variables climatiques sont basées sur les normales climatiques de la période 1971-2000) .....	3
<b>Tableau 2.</b> Moyennes des hauteurs et des dhp observées des 20 espèces du jeu de données .....	4
<b>Tableau 3.</b> Liste des variables de base disponibles pour la modélisation .....	6
<b>Tableau 4.</b> Estimation des paramètres dépendants des espèces dans le modèle (6) .....	11
<b>Tableau 5.</b> Estimation des paramètres indépendants des espèces dans le modèle (6) .....	12
<b>Tableau 6.</b> Estimation des biais et des racines de l'erreur quadratique moyenne (REQM) des prévisions en hauteur .....	14



# Liste des figures

	page
<b>Figure 1.</b> Quelques exemples de relations hauteur-diamètre issus du modèle (1) .....	5
<b>Figure 2.</b> Quatre subdivisions écologiques du territoire retenues pour la modélisation de la relation hauteur-diamètre.....	8
<b>Figure 3.</b> Examen visuel des résidus normalisés .....	9
<b>Figure 4.</b> Corrélations empiriques à l'échelle de l'arbre et du mesurage niché dans la placette.....	9
<b>Figure 5.</b> Estimations et intervalles de confiance à 95 % de la hauteur selon la température annuelle moyenne, de la surface terrière, de la subdivision écologique, de la classe de drainage, de l'indice de statut social et de l'occurrence de perturbation légère sur les prévisions en hauteur (les covariables de référence sont : sapin baumier, dhp de 20 cm, température annuelle moyenne de 0 °C, surface terrière de 20 m <sup>2</sup> ha <sup>-1</sup> , subdivision Centre, drainage mésique, indice de statut social à 0 et aucun perturbation légère) .....	13



# Introduction

La hauteur d'un arbre est l'une des variables dendrométriques de base en foresterie. Plusieurs autres variables indispensables à la prise de décision en aménagement forestier, comme le défilement des arbres et leur volume, la hauteur dominante du peuplement et l'indice de qualité de station, sont dérivées de la hauteur des arbres. Cependant, mesurer la hauteur de chaque arbre est une tâche fastidieuse. En conséquence, la hauteur n'est généralement mesurée que sur un sous-échantillon de tiges dans la plupart des inventaires forestiers. Une relation hauteur-diamètre (HD) est alors étalonnée<sup>1</sup> à partir de ce sous-échantillon et les hauteurs des arbres de la population peuvent ainsi être estimées. La littérature scientifique traitant des relations hauteur-diamètre est très abondante (p.ex. : HUANG *et al.* 2000, SHARMA et ZHANG 2004, CASTEDO DORADO *et al.* 2005, NEWTON et AMPONSAH 2007).

Modéliser la hauteur d'un arbre est une opération souvent complexe en raison de la nature des données d'étalonnage. En effet, les mesures de hauteur sont généralement corrélées. Par exemple, les hauteurs mesurées sur un sous-échantillon d'arbres d'une placette donnée peuvent être corrélées dans l'espace (corrélations spatiales), alors que les mesures de hauteur prises sur le même arbre à différents intervalles de temps sont corrélées en série (corrélations temporelles). Depuis deux décennies, l'approche par modèle mixte (effets fixes et effets aléatoires) est devenue courante pour traiter ces problèmes d'autocorrélation (p.ex. : CALAMA et MONTERO 2004, ROBINSON et WYKOFF 2004, SHARMA et PARTON 2007, TRINCADO *et al.* 2007). Un modèle à effets mixtes, ou simplement un modèle mixte, permet de configurer la matrice de variance-covariance des termes d'erreur afin de prendre en compte les corrélations spatiales et temporelles. Lorsque ces corrélations ont été estimées, elles peuvent être utilisées pour améliorer les prévisions si des observations de hauteur du sujet sont déjà disponibles (ROBINSON et WYKOFF 2004, TRINCADO *et al.* 2007).

La forêt publique commerciale de la province de Québec couvre plus de 350 000 km<sup>2</sup> (PARENT et FORTIN 2007). Jusqu'à présent, les relations HD de l'inventaire provincial ont toujours été étalonnées à une échelle régionale pour des espèces particulières. Ce procédé est fastidieux lorsque des prévisions sont nécessaires à l'échelle provinciale parce qu'il exige l'étalonnage d'un grand nombre de relations HD. Par ailleurs, plusieurs espèces deviennent des espèces rares à l'échelle régionale, de sorte que l'étalonnage d'une relation HD pour celles-ci est difficile, voire impossible. Par conséquent, une seule relation HD provinciale constituerait une alternative intéressante puisqu'elle permettrait (i) de réduire le nombre d'espèces rares en les regroupant à l'échelle provinciale et (ii) de rendre le procédé plus cohérent puisqu'il se traduirait par l'étalonnage d'une seule relation HD. Cependant, un tel modèle général devrait prendre en compte les différences régionales d'une certaine façon. En plus des variables du climat et de la station, l'ajout d'un effet aléatoire de placette dans ce modèle permettrait de tenir compte des différences régionales comme l'ont démontré TRINCADO *et al.* (2007).

L'objectif de cette étude était d'étalonner une seule relation HD applicable aux principales espèces commerciales de la forêt publique québécoise. En plus des variables locales comme la surface terrière et la densité de tiges, des variables climatiques ont été introduites dans le modèle afin de tenir compte des différences régionales dans la mesure du possible. À cela, nous avons ajouté un effet aléatoire et une structure de covariance afin d'obtenir des inférences statistiques valables. La relation entre les variables explicatives et la hauteur des tiges ainsi que les enjeux statistiques liés à l'étalonnage de ce genre de modèle sont abordés dans la discussion.

---

1 On utilise souvent à tort le terme calibrage ou encore l'anglicisme calibration pour se référer à l'action d'étalonner un modèle.



# Chapitre premier

## Matériel et méthodes

### 1.1 Données

Au début des années 1970, le Service des inventaires forestiers du ministère des Terres et Forêts du Québec a établi un réseau provincial de placettes permanentes dans le but de suivre l'évolution de la forêt publique (MINISTÈRE DES RESSOURCES NATURELLES, DE LA FAUNE ET DES PARCS DU QUÉBEC 2003). Ce réseau comprend 12 001 placettes distribuées aléatoirement sur l'ensemble de la forêt commerciale. Chaque placette a une superficie de 400 m<sup>2</sup> (0,04 ha) et tous les arbres dont le diamètre à hauteur de poitrine (dhp, 1,3 m) est égal ou supérieur à 9,1 cm ont été numérotés. L'espèce et le dhp au millimètre près de chacun des arbres ainsi identifiés ont été notés.

Un sous-échantillon comprenant en moyenne cinq arbres vivants sur pied a été choisi de façon aléatoire pour la mesure de la hauteur lors de l'établissement des placettes. La hauteur de ces cinq arbres a été mesurée à nouveau lors des mesurages subséquents si ceux-ci étaient toujours vivants. En cas de mortalité, un autre arbre vivant sur pied a été sélectionné aléatoirement pour remplacer l'arbre mort (MINISTÈRE DES RESSOURCES NATURELLES ET DE LA FAUNE DU QUÉBEC 2007).

En plus des variables allométriques usuelles (hauteur et dhp), deux autres sources d'information sont disponibles à l'échelle de la placette. D'abord, un inventaire écologique a été effectué dans la vaste majorité des placettes durant les

deux dernières décennies. Cet inventaire fournit de l'information à propos du drainage, du sol et du type écologique de la placette, des variables qui sont présumées être constantes dans le temps. Ensuite, durant les années 1990, le Service canadien des forêts a mis au point un logiciel nommé BioSIM. Ce logiciel permet d'estimer les conditions climatiques de la plupart des localisations géographiques du Québec méridional (RÉGNIERE *et al.* 1995; RÉGNIERE et SAINT-AMANT 2007, 2008). BioSIM utilise les données des quatre stations climatiques les plus proches pour simuler des variables comme la température et les précipitations annuelles moyennes à un point géographique donné. Dans le cas présent, nous avons utilisé les données des normales climatiques de la période 1971-2000 pour imputer des valeurs de degrés-jours (> 5 °C), de température annuelle moyenne et de précipitations annuelles moyennes à chacune des placettes.

Comme nous pouvions présumer que les variables écologiques et climatiques ont un effet sur la relation HD, nous avons rejeté les placettes pour lesquelles ces informations n'étaient pas disponibles. Par exemple, certaines placettes ont été abandonnées dans les années 1980, de sorte que l'inventaire écologique n'a pas été effectué dans celles-ci. La base de données utilisée dans cette étude comprend donc 11 215 placettes. Quelques caractéristiques de ce jeu de données sont présentées au tableau 1.

**Tableau 1. Quelques caractéristiques du jeu de données (les variables climatiques sont basées sur les normales climatiques de la période 1971–2000)**

Caractéristique	n <sup>†</sup>	min	moyenne	max
Degrés-jours (°C·an <sup>-1</sup> ) <sup>††</sup>	11 215	579	1310	2162
Température annuelle moyenne (°C)	11 215	-4,7	1,4	6,8
Précipitations annuelles moyennes (mm)	11 215	641	1064	1698
Surface terrière (m <sup>2</sup> ha <sup>-1</sup> ) <sup>†††</sup>				
période 1970–1979	9175	0,2	18,1	54,0
période 1980–1989	6033	0,2	17,7	60,0
période 1990–1999	9978	0,2	18,4	64,1
période 2000–2006	5664	0,2	20,3	66,9

<sup>†</sup> n : nombre d'observations; <sup>††</sup> Au-dessus de 5 °C; <sup>†††</sup> Surface terrière de toutes les tiges de 9,1 cm et plus au dhp.

Le jeu de données comprenait 41 espèces arborescentes à l'origine. Cependant, la moitié de celles-ci étaient peu abondantes non seulement à l'échelle régionale, mais également à l'échelle provinciale. Comme l'objectif de cette étude était d'étalonner un modèle à l'échelle provinciale, nous n'avons conservé que les espèces pour lesquelles le jeu de données comptait au moins 500 observations. Vingt espèces satisfaisaient ce critère pour un total de 144 262 observations. La moyenne et l'étendue des hauteurs et des dhp observés sont présentées au tableau 2 pour chaque espèce.

## 1.2 Sélection du modèle

Afin de faciliter la compréhension de la structure de données, définissons  $i$ ,  $j$  et  $k$  comme les indices de la placette, de l'arbre et du mesurage, respectivement, de sorte que  $i = 1, 2, \dots, 11215$ ,  $j = 1, 2, \dots, p_i$  et  $k = 1, 2, \dots, q_{ij}$ . De nombreux modèles de

relation HD ont été proposés dans la littérature scientifique (LÓPEZ SÁNCHEZ *et al.* 2003) et la plupart d'entre eux ont une forme non linéaire. Bien que cette non-linéarité puisse être désirable afin d'assurer un comportement logique du modèle, elle rend l'analyse beaucoup plus complexe lorsque les corrélations doivent être prises en compte. Les données autocorrélées sont généralement traitées en ajoutant des effets aléatoires ou une structure de covariance. Les propriétés des estimateurs basés sur le maximum de vraisemblance des modèles non linéaires mixtes sont méconnues (VONESH et CARTER 1992, PINHEIRO et BATES 2000, p. 323). Dans ce cas d'étude, nous avons préféré un modèle linéaire mixte parce que l'estimateur statistique permettant de l'étalonner est reconnu pour être asymptotiquement convergent et sans biais (PINHEIRO et BATES 2000, p. 81).

**Tableau 2. Moyennes des hauteurs et des dhp observées des 20 espèces du jeu de données (l'étendue des observations figure entre parenthèses)**

$s^{\dagger}$	Espèce	$n^{\ddagger}$	hauteur (m)	dhp (cm)
1	Bouleau jaune	6 259	15,8 (5,0-30,8)	25,1 (9,1-79,6)
2	Bouleau à papier	14 719	13,6 (4,7-30,5)	17,2 (9,1-68,5)
3	Chêne rouge	897	15,6 (5,2-30,8)	23,5 (9,1-67,0)
4	Épinette blanche	6 494	12,9 (3,7-32,5)	20,7 (9,1-73,8)
5	Épinette noire	35 424	11,4 (3,7-26,3)	14,5 (9,1-47,8)
6	Épinette rouge	1 208	12,5 (4,7-31,3)	18,1 (9,1-53,1)
7	Érable rouge	7 811	14,7 (4,3-30,2)	17,6 (9,1-62,9)
8	Érable à sucre	13 252	17,0 (5,4-31,9)	21,9 (9,1-79,3)
9	Frêne noir	917	15,2 (4,6-30,1)	18,3 (9,1-57,5)
10	Hêtre à grandes feuilles	2 524	16,7 (5,2-31,6)	21,4 (9,1-79,7)
11	Mélèze laricin	621	13,1 (4,2-29,3)	16,2 (9,1-41,1)
12	Ostryer de virginie	635	13,3 (5,5-22,3)	13,7 (9,1-35,2)
13	Peuplier à grandes dents	1 005	18,6 (6,7-35,0)	22,4 (9,1-63,1)
14	Peuplier faux-tremble	6 886	16,1 (5,2-32,9)	19,2 (9,1-65,8)
15	Pin blanc	1 318	17,0 (4,3-34,5)	27,5 (9,1-91,0)
16	Pin gris	6 005	12,9 (3,7-29,3)	15,5 (9,1-52,6)
17	Pruche du Canada	1 117	13,8 (4,6-31,9)	24,5 (9,1-76,0)
18	Sapin baumier	32 741	11,3 (3,4-35,0)	15,4 (9,1-50,5)
19	Thuya occidental	3 864	11,3 (3,4-22,9)	21,1 (9,1-80,6)
20	Tilleul d'Amérique	565	18,0 (5,7-33,8)	22,7 (9,2-54,9)

$\dagger s$  : indice des espèces,  $\ddagger n$  : nombre d'observations.

Le modèle choisi peut être exprimé de la façon suivante :

$$h_{ijk} = 1,3 + A_{ijk} \ln(dhp_{ijk} + 1) + B_{ijk} (\ln(dhp_{ijk} + 1))^2 + \varepsilon_{ijk} \quad (1)$$

où  $h_{ijk}$  est la hauteur de l'arbre  $j$  dans la placette  $i$  lors du mesurage  $k$  (m),  $dhp_{ijk}$  est le diamètre à hauteur de poitrine de ce même arbre (cm),  $\ln(\bullet)$  est le logarithme naturel de l'argument,  $A_{ijk}$  et  $B_{ijk}$  sont les paramètres généraux de l'équation et  $\varepsilon_{ijk}$  est le terme d'erreur aléatoire. La valeur de 1 est ajoutée au  $dhp$  pour s'assurer du comportement cohérent du modèle et de sa convergence vers 1,3 m lorsque le  $dhp$  tend vers 0. Les paramètres généraux  $A_{ijk}$  et  $B_{ijk}$  sont en fait des fonctions linéaires de variables explicatives tant à l'échelle de l'arbre que du mesurage et de la placette. Nous présumons que les paramètres de ces fonctions linéaires qui remplacent  $A_{ijk}$  et  $B_{ijk}$  peuvent être généralisés à la population. Des exemples de prévisions de hauteur du modèle (1) sont présentés à la figure 1.

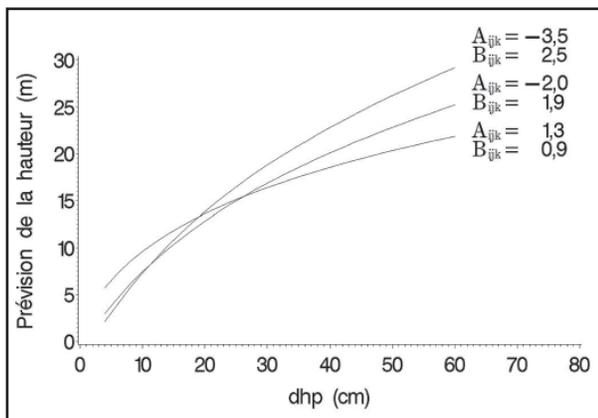


Figure 1. Quelques exemples de relations hauteur-diamètre issus du modèle (1).

### 1.3 Théorie des modèles mixtes

Les données de placettes permanentes sont souvent caractérisées par des corrélations spatiales et temporelles, lesquelles sont induites par la mesure de plusieurs individus à l'intérieur de la même superficie (la placette) et par la mesure répétée de la hauteur dans le temps pour un même individu. Les estimateurs par la méthode traditionnelle des moindres carrés fournissent tout de même des estimations de paramètre sans biais lorsqu'ils sont appliqués à des données autocorrélées (VANCLAY 1994, p. 119). Cependant, les variances de ces estimations de paramètre sont biaisées et les variables explicatives ne peuvent être choisies selon leur niveau de signification réel (LEMAY 1990, GREGOIRE et al. 1995).

Les modèles mixtes comportent à la fois des effets fixes et des effets aléatoires. L'ajout d'effets aléatoires ou d'une structure de covariance fait en sorte que le modèle ne repose pas sur le postulat d'indépendance des erreurs. La théorie des modèles mixtes est décrite par LITTELL et al. (2006) et PINHEIRO et BATES (2000). Un modèle linéaire mixte peut être exprimé sous forme matricielle de la façon suivante :

$$y_i = X_i \beta + Z_i u_i + \varepsilon_i \quad (2)$$

où  $y_i$  est le vecteur contenant les variables dépendantes du sujet  $i$ ,  $X_i$  est la matrice des variables explicatives du sujet  $i$ ,  $\beta$  est un vecteur de paramètres inconnus,  $Z_i$  est la matrice qui conditionne les effets aléatoires,  $u_i$  est le vecteur d'effets aléatoires, lequel est distribué selon une loi normale multivariée de moyenne  $0$  et de variance-covariance  $G$ , c'est-à-dire  $u_i \sim N(0, G)$ ; et  $\varepsilon_i$  est un vecteur de termes d'erreur résiduelle, lequel suit une loi normale multivariée de moyenne  $0$  et de variance-covariance  $R_i$ , c'est-à-dire  $\varepsilon_i \sim N(0, R_i)$ .

À partir du modèle (2), le terme d'erreur complet peut être défini comme  $Z_i u_i + \varepsilon_i$  et, selon un postulat d'indépendance entre les vecteurs  $u_i$  et  $\varepsilon_i$  sa variance peut être dérivée comme suit :

$$\text{var}(y_i - X_i \beta) = Z_i G Z_i^T + R_i \equiv V_i \quad (3)$$

où  $T$  indique la transposition d'un vecteur ou d'une matrice. L'équation (3) démontre qu'il existe deux façons de configurer la matrice de variance-covariance des termes d'erreur complets ( $V_i$ ) : par le terme  $Z_i G Z_i^T$  qui représente les effets aléatoires ou par la matrice  $R_i$  dans laquelle on peut définir une structure de covariance.

Une fois le modèle étalonné, l'ajustement de la matrice  $V_i$  peut être vérifié à l'aide de résidus dits normalisés. Le vecteur des résidus normalisés du sujet  $i$  ( $r_{\text{norm},i}$ ) s'obtient à l'aide de l'équation suivante (PINHEIRO et BATES 2000, p. 239) :

$$r_{\text{norm},i} = (\hat{V}_i^{-1/2})^T (y_i - X_i \hat{\beta}) \quad (4)$$

où l'accent circonflexe indique l'estimation de la matrice ou du vecteur par la méthode du maximum de vraisemblance. Si les éléments de covariance sont adéquats, les éléments du vecteur  $r_{\text{norm},i}$  devraient être indépendants les uns des autres et

distribués selon une loi normale de moyenne 0 et de variance 1 (SAS INSTITUTE 2003).

Les postulats de normalité et d'homogénéité des variances peuvent être vérifiés visuellement à l'aide de graphiques. Des corrélations empiriques doivent être calculées à partir des résidus normalisés pour vérifier le postulat d'indépendance (FORTIN *et al.* 2008). Dans le cas présent, une corrélation empirique intra-arbre peut être calculée comme suit :

$$\hat{\rho}_{\text{arbre}} = \frac{\sum_{i=1}^{11215} \sum_{j=1}^{p_i} \sum_{k=1}^{q_{ij}-1} \sum_{k'=k+1}^{q_{ij}} r_{\text{norm},ijk} \cdot r_{\text{norm},ijk'}}{N} \quad (5)$$

$$\frac{\sum_{i=1}^{11215} \sum_{j=1}^{p_i} \sum_{k=1}^{q_{ij}} r_{\text{norm},ijk}^2}{\sum_{i=1}^{11215} \sum_{j=1}^{p_i} q_{ij}}$$

où N est le nombre de paires de résidus utilisées dans le calcul de la somme et  $r_{\text{norm},ijk}$  sont les éléments du vecteur de résidus normalisés ( $r_{\text{norm},i}$ ). En suivant le même raisonnement, des corrélations empiriques peuvent être calculées pour chaque niveau de regroupement de la base de données ou en fonction d'une quelconque covariable.

### 1.4 Caractéristiques du modèle

Compte tenu de l'information disponible, plusieurs variables pouvaient être testées dans le modèle. La liste des variables de base disponibles pour la

modélisation apparaît au tableau 3. Comme les corrélations entre certaines de ces variables étaient élevées, celles-ci n'ont pu être testées simultanément dans le modèle. La température annuelle moyenne et les degrés-jours en sont un exemple. Par ailleurs, des analyses préliminaires ont permis de regrouper certaines classes de variables catégoriques et de créer des indices qui intégraient des variables continues.

L'ajout des effets fixes et des éléments de covariance dans le modèle repose généralement sur l'analyse des résidus et les critères d'information comme le critère d'information d'Akaike et le critère d'information bayésien (CIA et CIB, PINHEIRO et BATES 2000, p. 83). Des analyses préliminaires ont été effectuées au moyen d'un estimateur par la méthode des moindres carrés afin d'identifier les variables explicatives les plus susceptibles d'être significatives. À partir de ce modèle de base contenant seulement quelques variables, nous avons testé différentes structures de covariance et différentes configurations d'effets aléatoires. Lorsque l'analyse des résidus normalisés s'est avérée concluante, des variables explicatives additionnelles ont été intégrées une à une dans le modèle et sélectionnées en fonction des critères d'information CIA et CIB (processus de sélection dit « FORWARD »). Pour s'assurer que l'ajuste-

**Tableau 3. Liste des variables de base disponibles pour la modélisation**

Variables continues	Variables catégoriques
Échelle de l'arbre	
dhp (cm)	espèce (20 classes)
surface terrière des tiges dont le dhp est plus grand que le sujet (m <sup>2</sup> ha <sup>-1</sup> )	groupe d'espèces (2 classes)
Échelle de la placette	
densité de tiges (tiges·ha <sup>-1</sup> ) <sup>†</sup>	drainage (4 classes)
surface terrière (m <sup>2</sup> ha <sup>-1</sup> ) <sup>†</sup>	perturbation (3 classes)
diamètre moyen quadratique (cm)	région écologique (47 classes)
latitude (degrés)	type de dépôt (10 classes)
longitude (degrés)	épaisseur du dépôt (3 classes)
altitude (m)	
degrés-jours (°C·an <sup>-1</sup> ) <sup>††</sup>	
température annuelle moyenne (°C)	
précipitations annuelles moyennes (mm)	

<sup>†</sup> Calculé à partir des tiges de 9,1 cm et plus au dhp; <sup>††</sup> Au-dessus de 5 °C.

ment du modèle final était adéquat et qu'aucune tendance n'eût été oubliée, des graphiques des résidus normalisés en fonction de toutes les variables explicatives disponibles ont aussi été produits. Le modèle final issu de ce processus de sélection peut être exprimé comme suit :

$$h_{ijk} = 1,3 + (\beta_{1,s} + \beta_{2,s}ST_{ik} + \beta_3T_i + \beta_{4,d} + \beta_{5,e} + \beta_{6,f} + \beta_7ISS_{ijk}) \ln(dhp_{ijk} + 1) + (\beta_{8,s} + \beta_{9,s}ST_{ik} + \beta_{10}ISS_{ijk} + b_{is}) (\ln(dhp_{ijk} + 1))^2 + \varepsilon_{ijk} \quad (6)$$

où  $ST_{ik}$  est la surface terrière de tous les arbres dont le dhp est égal ou supérieur à 9,1 cm dans la placette  $i$  durant le mesurage  $k$  ( $m^2ha^{-1}$ ),  $T_i$  est la température annuelle moyenne de la placette  $i$  pour la période 1971–2000 ( $^{\circ}C$ ),  $s$  est l'indice des espèces, de sorte que  $s = 1, 2, \dots, 20$ ,  $d$  est l'indice des classes de drainage, de sorte que  $d = 1, 2, 3, 4$ ,  $e$  est un indice de subdivision écologique, de sorte que  $e = 1, 2, 3, 4$ ,  $f$  est un indice de type de perturbation légère, de sorte que  $f = 1, 2, 3$ ,  $ISS_{ijk}$  est un indice de statut social,  $\beta_{1,s}$ ,  $\beta_{2,s}$ ,  $\beta_3$ ,  $\beta_{4,d}$ ,  $\beta_{5,e}$ ,  $\beta_{6,f}$ ,  $\beta_7$ ,  $\beta_{8,s}$ ,  $\beta_{9,s}$  et  $\beta_{10}$  sont les paramètres des effets fixes, c'est-à-dire les éléments de  $\beta$ ,  $b_{is}$  est un effet aléatoire de placette qui varie selon l'espèce, et  $\varepsilon_{ijk}$  est l'erreur résiduelle. Il est à noter que l'effet aléatoire de placette  $b_{is}$  ne s'applique que sur le carré du logarithme du dhp.

Dans le modèle (6), le drainage de la placette est exprimé comme une variable catégorique. Les quatre classes de drainage retenues ici sont : xérique ( $d = 1$ ), mésique ( $d = 2$ ), subhydrique ( $d = 3$ ) et hydrique ( $d = 4$ ). La subdivision écologique représente un regroupement de régions écologiques reconnues dans le système de classification écologique provinciale (GRONDIN *et al.* 1996, SAUCIER *et al.* 1998). Les essais préliminaires ont démontré que quatre subdivisions pouvaient être distinguées : l'Ouest ( $e = 1$ ), le Centre ( $e = 2$ ), le Nord-Est ( $e = 3$ ) et le Sud-Est ( $e = 4$ ) (Figure 2). Le type de perturbation est une variable catégorique à trois classes : les perturbations anthropiques ( $f = 1$ ), les perturbations naturelles comme les chablis partiels et les épidémies légères ( $f = 2$ ), et l'absence de perturbation ( $f = 3$ ). Ces types de perturbation font référence aux codes de perturbation légère ou d'intervention partielle notés lors de l'inventaire en forêt (MINISTÈRE DES RESSOURCES NATURELLES ET DE LA FAUNE DU QUÉBEC 2007). Ces perturbations moyennes ou interventions partielles sont notées lorsqu'elles ont entraîné l'élimination de 25 à 75 % de la surface

terrière depuis le dernier mesurage. L'indice de statut social  $ISS_{ijk}$  qui tient compte du statut hiérarchique de l'arbre par rapport à ses voisins a été calculé comme suit :

$$ISS_{ijk} = (dhp_{ijk} - dm_{q_{ik}})^2 \quad (7)$$

où  $dm_{q_{ik}}$  est le diamètre moyen quadratique des tiges de la placette  $i$  lors du mesurage  $k$  (cm). La tendance quadratique exprimée dans l'équation (7) était nécessaire parce que les résidus normalisés ont démontré que les arbres dont le dhp est plus gros ou plus petit que le diamètre moyen quadratique ont tendance à être plus courts à un dhp comparable.

L'effet aléatoire de placette compris dans le modèle (6) est distribué selon une loi normale de moyenne 0 et de variance  $G_s$ , c'est-à-dire  $b_{is} \sim N(0, G_s)$ . Nous présumons que la variance  $G_s$  diffère selon l'espèce. Le vecteur des termes d'erreur résiduelle intra-arbre ( $\varepsilon_{ij}$ ) est également distribué selon une loi normale multivariée de moyenne  $\mathbf{0}$  et de covariance-covariance  $\mathbf{R}_{ij}$ , c'est-à-dire  $\varepsilon_{ij} \sim N(\mathbf{0}, \mathbf{R}_{ij})$ . Des essais préliminaires ont démontré qu'une structure de covariance dite linéaire (LITTELL *et al.* 2006) à l'intérieur de la matrice  $\mathbf{R}_{ij}$  permettait d'obtenir le meilleur ajustement du modèle. Cette structure comporte deux paramètres ( $\gamma$  et  $\delta$ ) et nécessite le vecteur des années des mesurages ( $t_{ij}$ ) pour l'arbre  $j$  dans la placette  $i$  afin d'estimer la covariance entre les termes d'erreur résiduelle :

$$\text{Cov}(\varepsilon_{ijk}, \varepsilon_{ijk'}) = \begin{cases} \gamma_h (1 - \delta_h |t_{ijk} - t_{ijk'}|) & \text{si } \delta_h |t_{ijk} - t_{ijk'}| < 1 \\ 0 & \text{si } \delta_h |t_{ijk} - t_{ijk'}| \geq 1 \end{cases} \quad (8)$$

où  $h$  est l'indice du groupe d'espèces de sorte que  $h = 1$  pour le groupe des espèces caducifoliées et  $h = 2$  pour les espèces de conifères. Les paramètres  $\gamma$  et  $\delta$  représentent la variance et la corrélation, respectivement. La structure de covariance (8) présume que les corrélations déclinent de manière constante dans le temps. Les corrélations négatives sont remplacées par des corrélations nulles (LITTELL *et al.* 2006, p. 441). Par ailleurs, nous présumons que les paramètres de la structure de covariance (8) sont différents selon le groupe d'espèces.

Le modèle a été étalonné à l'aide de la procédure MIXED disponible dans le progiciel SAS System. L'option VCIRY a été sélectionnée pour obtenir les résidus normalisés. L'énoncé ESTIMATE a également été utilisé afin d'obtenir des prévisions de cas

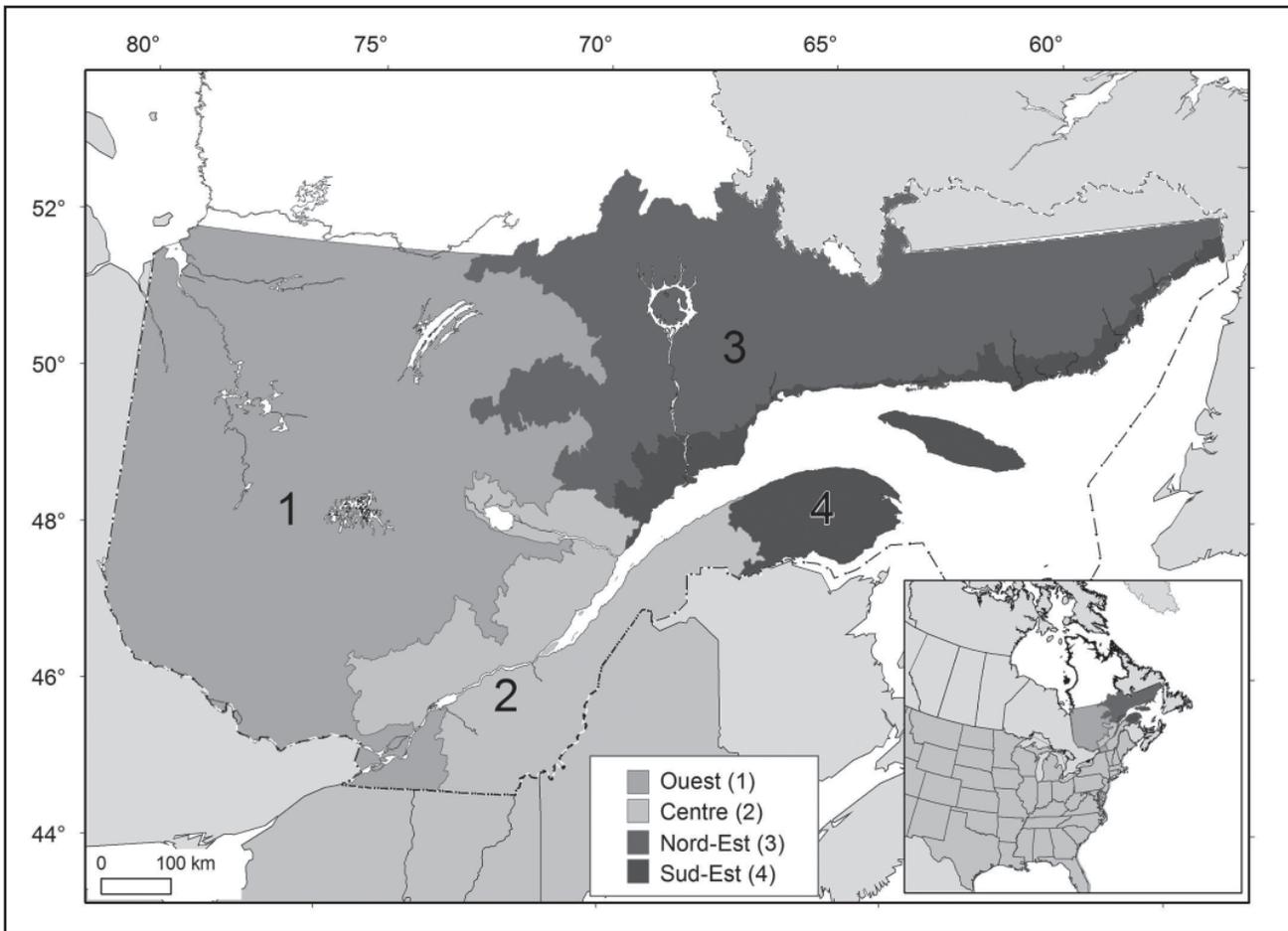


Figure 2. Quatre subdivisions écologiques du territoire retenues pour la modélisation de la relation hauteur-diamètre.

particuliers, ce qui nous a permis de vérifier visuellement les relations entre les variables explicatives et la hauteur moyenne des tiges. Les calculs à la base de ces prévisions sont semblables à ceux utilisés pour calculer des estimations par la méthode des moindres carrés (SAS Institute 2003).

### 1.5 Évaluation du modèle

Une validation croisée a été effectuée afin de vérifier la fiabilité et la justesse des prévisions du modèle. Les placettes ont été d'abord divisées en dix groupes de façon aléatoire. Le modèle a ensuite été étalonné dix fois en omettant chacun des dix groupes successivement. Des prévisions ont été générées après chaque étalonnage pour le groupe de placettes n'ayant pas été inclus dans le processus. Une fois les dix étalonnages terminés, des estimations du biais et de la racine de l'erreur quadratique moyenne (REQM) ont été calculées pour chacune des espèces comme suit :

$$\text{biais} = \frac{\sum_{i=1}^{11215} \sum_{j=1}^{p_i} \sum_{k=1}^{q_{ij}} (h_{ijk} - \hat{h}_{ijk})}{\sum_{i=1}^{11215} \sum_{j=1}^{p_i} q_{ij}} \quad (9a)$$

$$\text{REQM} = \sqrt{\frac{\sum_{i=1}^{11215} \sum_{j=1}^{p_i} \sum_{k=1}^{q_{ij}} (h_{ijk} - \hat{h}_{ijk})^2}{\sum_{i=1}^{11215} \sum_{j=1}^{p_i} q_{ij}}} \quad (9b)$$

où  $\hat{h}_{ijk}$  est la prévision de la hauteur (m), obtenue en fixant l'effet aléatoire et l'erreur résiduelle à 0, c'est-à-dire  $b_{is}$  et  $\varepsilon_{ijk} = 0$ . Des valeurs relatives sont obtenues en divisant les résultats des équations (9a) et (9b) par la moyenne des hauteurs observées. Lorsque les résultats de la validation croisée se sont avérés satisfaisants, le modèle a été étalonné avec toutes les données disponibles.

# Chapitre deux

## Résultats

Les résidus normalisés n'ont affiché aucune divergence par rapport aux postulats de normalité et d'homogénéité des variances (Figure 3). Des corrélations empiriques ont été calculées à l'échelle de l'arbre, du mesurage et de la placette. Des graphiques des corrélations intra-arbre et intra-mesurage en fonction du temps entre les mesurages et des différences par rapport à la classe de diamètre ont été également produits (Figure 4). De manière générale, la corrélation intra-arbre était aussi faible que 0,04. Cependant, des corrélations positives persistantes ont pu être décelées. En effet, les erreurs prévisionnelles pour deux mesures de hauteur prises sur le même arbre à 25 ans d'intervalle tendent à être corrélées à 0,20, même si

une structure de covariance linéaire a été ajoutée au modèle (Figure 4a). À l'échelle du mesurage (niché dans la placette), une corrélation persistante de 0,14 a pu être observée. Cette corrélation peut s'expliquer par le fait que les erreurs prévisionnelles des arbres de dhp semblables ont tendance à être significativement corrélées jusqu'à 0,20 (Figure 4b). La corrélation moyenne à l'échelle de la placette était de 0,03.

Toutes les composantes fixes du modèle étaient hautement significatives avec des probabilités inférieures à 0,0001. Les estimations des paramètres du modèle (6) sont présentées dans les tableaux 4 et 5.

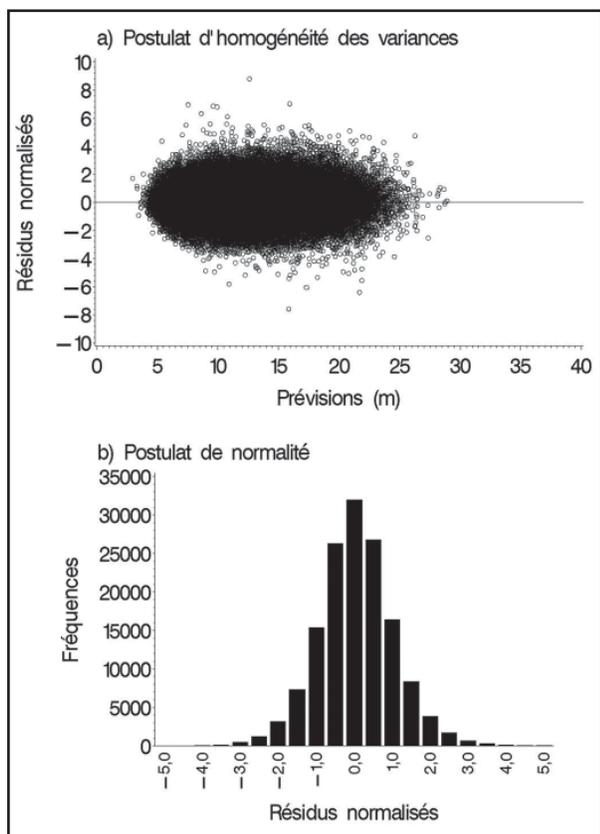


Figure 3. Examen visuel des résidus normalisés.

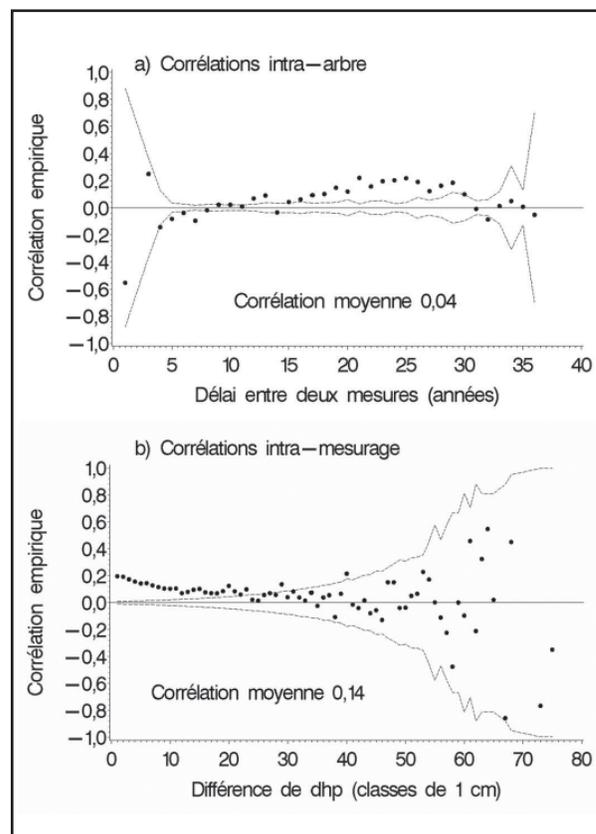


Figure 4. Corrélations empiriques à l'échelle de l'arbre et du mesurage niché dans la placette (les lignes pointillées représentent l'intervalle de confiance à 95 % sous l'hypothèse nulle que la corrélation est nulle).

Les estimations des paramètres de covariance présentées au tableau 5 fournissent une information additionnelle sur le comportement du modèle. L'estimation de  $\gamma$  est en fait la variance de l'erreur prévisionnelle conditionnelle à l'effet aléatoire. La comparaison entre les deux éléments du vecteur  $\hat{\gamma}$  démontre que la variance est plus élevée pour les espèces caducifoliées. La corrélation entre deux mesures du même arbre tend également à diminuer plus rapidement dans le temps pour les espèces caducifoliées. La valeur du premier élément du vecteur  $\hat{\delta}$  est en effet plus élevée que celle du second. À partir de la structure de covariance (8), les corrélations entre deux mesures prises à un intervalle de 10 ans sont estimées à 0,55 ( $1 - 0,04468 \cdot 10$ ) et à 0,74 ( $1 - 0,02620 \cdot 10$ ) pour les espèces caducifoliées et les conifères, respectivement.

Des graphiques ont été produits en utilisant un arbre de référence afin d'illustrer les relations entre les différentes variables explicatives et la prévision en hauteur. L'arbre de référence sélectionné ici possède les caractéristiques suivantes : sapin baumier, dhp de 20 cm, indice de statut social de 0, placette de  $20 \text{ m}^2\text{ha}^{-1}$  de surface terrière, drainage mésique, température annuelle moyenne de  $0^\circ\text{C}$ , localisation dans la subdivision écologique Centre et aucune perturbation légère répertoriée. Ces caractéristiques ont été choisies parce qu'elles correspondent aux conditions moyennes du jeu de données et parce que le sapin baumier est l'une des espèces dont l'aire de distribution est la plus étendue. Pour illustrer la relation entre les variables explicatives et la prévision de la hauteur (Figure 5), nous faisons varier la valeur de chacune d'elles à tour de rôle tout en conservant les autres caractéristiques de référence constantes.

En somme, nous constatons que la hauteur moyenne augmente en fonction de la température annuelle moyenne (Figure 5a). Les différences peuvent dépasser 2 m entre des températures annuelles

moyennes de  $-4^\circ\text{C}$  et  $+4^\circ\text{C}$ . L'augmentation de la surface terrière de la placette induit également une augmentation de la hauteur moyenne. Dans l'exemple présenté à la figure 5b, la différence est d'environ 2,5 m lorsque la surface terrière passe de 10 à  $40 \text{ m}^2$ . Il importe de noter que la relation entre la hauteur et la surface terrière est propre aux espèces. En conséquence, l'ampleur de la relation peut changer pour une autre espèce.

En ce qui concerne la subdivision écologique, les arbres évoluant dans la zone Ouest ont tendance à être plus grands pour un même dhp que ceux des autres subdivisions (Figure 5c). À l'opposé, les arbres de la subdivision Sud-Est sont plus courts en général. La relation entre la classe de drainage et la hauteur se traduit par des tiges généralement plus petites sur les stations xériques (Figure 5d). Par ailleurs, les arbres dont le dhp se rapproche du diamètre moyen quadratique ont tendance à être plus grands en moyenne (Figure 5e). En d'autres termes, un sapin baumier de 20 cm de dhp en association avec un grand nombre de tiges de 10 cm ou de 30 cm est plus petit en moyenne qu'un sapin semblable évoluant parmi des tiges de 20 cm. Les perturbations légères ont également une influence sur la relation HD. Si nous supposons que toutes les autres variables explicatives sont constantes, les tiges évoluant dans des placettes perturbées sont significativement plus grandes en moyenne (Figure 5f).

Les résultats de la validation croisée sont présentés au tableau 6. La plupart des biais sont inférieurs à 0,20 m en valeur absolue et 1,5 % en valeur relative. En fait, le pin blanc est l'espèce affichant le plus grand biais absolu et relatif. Les hauteurs des tiges de cette espèce sont sous-estimées de 0,30 m en moyenne, ce qui représente 1,9 % de la valeur moyenne observée. Les racines de l'erreur quadratique moyenne des estimations excèdent largement leurs biais respectifs, avec des valeurs variant de 1,5 à 3,1 m (15,0 % et 19,8 %).

**Tableau 4. Estimation des paramètres dépendants des espèces dans le modèle (6)**

$s^{\dagger}$	Espèce	$\hat{\beta}_{1,s}$	$\hat{\beta}_{2,s}$	$\hat{\beta}_{8,s}$	$\hat{\beta}_{9,s}$	$\hat{G}_s$
1	Bouleau jaune	0,7655	0,06611	0,8253	-0,01352	0,02422
2	Bouleau à papier	0,2034	0,08593	0,9994	-0,01918	0,03284
3	Chêne rouge	-0,7520	0,08372	1,2325	-0,01717	0,04103
4	Épinette blanche	-3,1046	-0,00420	1,9371	0,00994	0,03023
5	Épinette noire	-2,9239	0,04592	2,1216	-0,00741	0,02120
6	Épinette rouge	-1,9719	-0,01763	1,6137	0,01351	0,02172
7	Érable rouge	0,3860	0,07946	0,9997	-0,01670	0,03053
8	Érable à sucre	0,7308	0,07261	1,0134	-0,01629	0,02959
9	Frêne noir	-1,6474	0,12740	1,7443	-0,03367	0,04285
10	Hêtre à grandes feuilles	-0,9384	0,08412	1,4856	-0,01825	0,02760
11	Mélèze laricin	-2,3083	0,07413	1,8788	-0,00984	0,03119
12	Ostryer de Virginie	0,6756	0,10900	0,8731	-0,02860	0,03395
13	Peuplier à grandes dents	-0,0827	0,07650	1,2626	-0,01257	0,03695
14	Peuplier faux-tremble	-0,7888	0,09821	1,4957	-0,02082	0,02778
15	Pin blanc	-3,9134	0,07463	2,1652	-0,01360	0,04945
16	Pin gris	-1,9179	0,13390	1,5895	-0,02363	0,02864
17	Pruche du Canada	-3,5560	0,04974	1,9743	-0,01073	0,01861
18	Sapin baumier	-2,6584	0,02748	1,8726	0,00055	0,01918
19	Thuya occidental	-1,3181	0,02282	1,1661	-0,00292	0,00619
20	Tilleul d'Amérique	-1,4742	0,04758	1,6586	-0,00444	0,04267

$\dagger$  : indice des espèces

**Tableau 5. Estimation des paramètres indépendants des espèces dans le modèle (6)**

	Paramètre	Catégorie	Estimation
<b>Composantes fixes</b>			
Température annuelle moyenne pour la période 1971–2000	$\beta_3$		0,08828
Classe de drainage	$\beta_{4,1}$	Xérique	-0,1117
	$\beta_{4,2}$	Mésique	0,03237
	$\beta_{4,3}$	Subhydrique	-0,00277
	$\beta_{4,4}$	Hydrique	0
Subdivision écologique	$\beta_{5,1}$	Ouest	0,6008
	$\beta_{5,2}$	Centre	0,3266
	$\beta_{5,3}$	Nord-Est	0,3319
	$\beta_{5,4}$	Sud-Est	0
Occurrence de perturbation légère	$\beta_{6,1}$	Anthropique	0,03563
	$\beta_{6,2}$	Naturelle	0,02191
	$\beta_{6,3}$	Aucune	0
Indice de statut social	$\beta_7$		-0,00235
Indice de statut social	$\beta_{10}$		0,000483
<b>Éléments de covariance</b>			
Variance dans la matrice $\mathbf{R}_{ij}$	$\gamma_1$	Caducifoliés	3,1148
	$\gamma_2$	Conifères	2,0370
Corrélation dans la matrice $\mathbf{R}_{ij}$	$\delta_1$	Caducifoliés	0,04468
	$\delta_2$	Conifères	0,02620

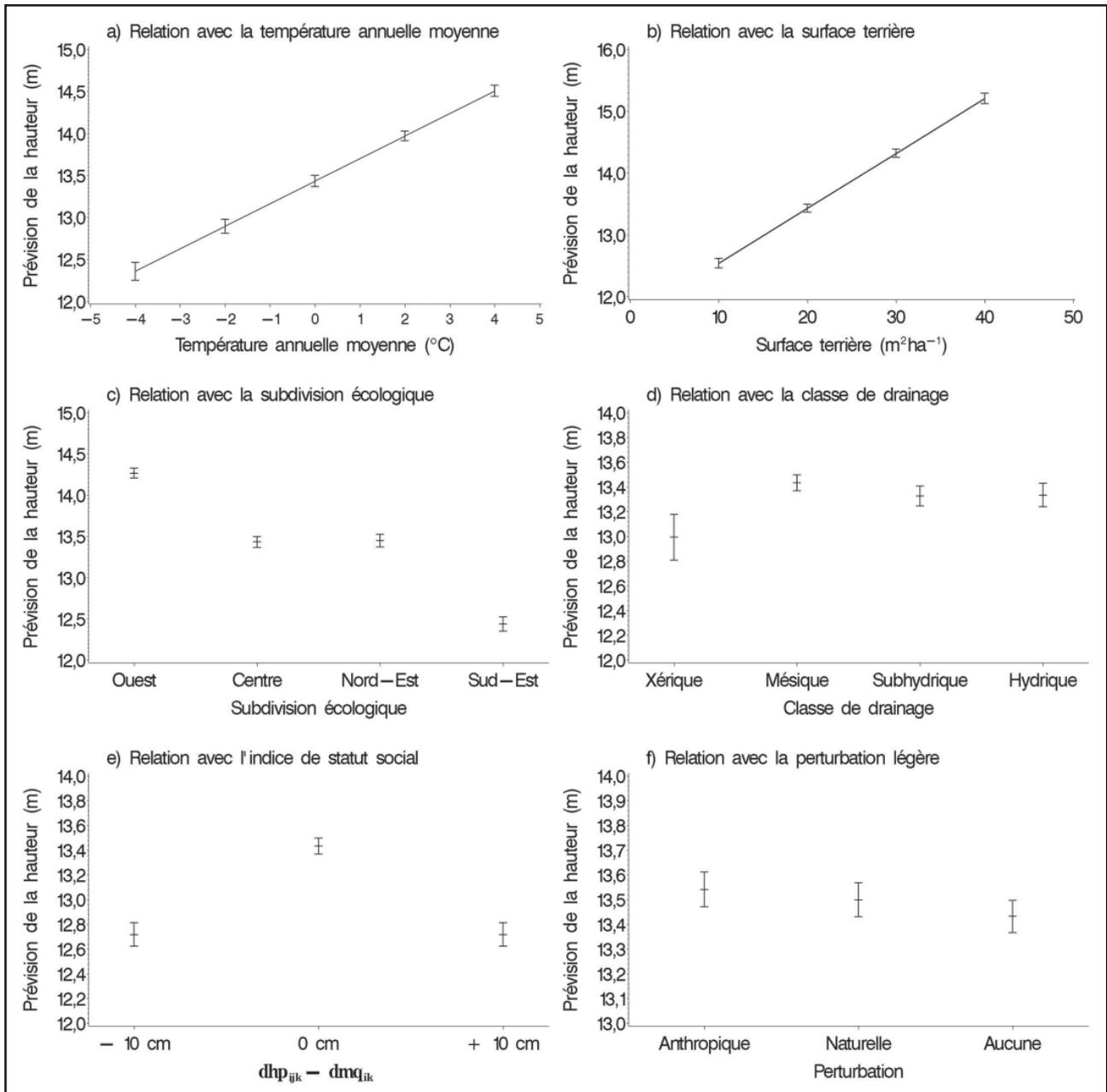


Figure 5. Estimations et intervalles de confiance à 95 % de la hauteur selon la température annuelle moyenne, de la surface terrière, de la subdivision écologique, de la classe de drainage, de l'indice de statut social et de l'occurrence de perturbation légère sur les prévisions en hauteur (les covariables de référence sont : sapin baumier, dhp de 20 cm, température annuelle moyenne de 0 °C, surface terrière de 20 m²ha⁻¹, subdivision Centre, drainage mésique, indice de statut social à 0 et aucun perturbation légère).

**Tableau 6. Estimation des biais et des racines de l'erreur quadratique moyenne (REQM) des prévisions en hauteur**

s †	Espèce	Biais		REQM	
		(m)	(%)	(m)	(%)
1	Bouleau jaune	0,05	0,4	2,4	16,3
2	Bouleau à papier	0,15	1,2	2,2	18,2
3	Chêne rouge	-0,19	-1,4	2,7	18,8
4	Épinette blanche	-0,16	-1,4	2,3	19,8
5	Épinette noire	0,07	0,7	1,7	17,1
6	Épinette rouge	-0,08	-0,7	2,0	17,8
7	Érable rouge	0,05	0,4	2,2	16,8
8	Érable à sucre	0,21	1,3	2,4	15,6
9	Frêne noir	-0,01	-0,0	2,4	17,2
10	Hêtre à grandes feuilles	0,20	1,3	2,6	17,1
11	Mélèze laricin	0,03	0,3	2,1	17,9
12	Ostryer de Virginie	-0,03	-0,3	2,1	17,4
13	Peuplier à grandes dents	0,18	1,0	2,6	15,2
14	Peuplier faux-tremble	0,16	1,1	2,2	15,0
15	Pin blanc	0,30	1,9	3,1	19,6
16	Pin gris	0,07	0,6	1,9	16,4
17	Pruche du Canada	0,02	0,2	2,1	17,0
18	Sapin baumier	-0,06	-0,6	1,8	18,4
19	Thuja occidental	-0,03	-0,4	1,5	15,0
20	Tilleul d'Amérique	-0,02	-0,1	2,6	15,4

†s : indice des espèces

# Chapitre trois

## Discussion

### 3.1 Aspects statistiques

Dans cette étude, nous avons réussi à étalonner une relation HD provinciale applicable aux 20 espèces commerciales les plus abondantes de la forêt publique québécoise. Ce modèle comprend des variables qui caractérisent la station, telles la température annuelle moyenne, la surface terrière, la subdivision écologique, la classe de drainage et l'occurrence de perturbation légère. Ce modèle tient compte des corrélations spatiales et temporelles par l'ajout d'un effet aléatoire de placette et d'une structure de covariance linéaire. Par ailleurs, la validation croisée révèle que les biais sont relativement faibles, ce qui contribue à accroître le degré de confiance envers les prévisions de ce modèle.

La littérature scientifique abonde d'ouvrages traitant de la modélisation des relations HD (p.ex. : HUANG *et al.* 2000, LÓPEZ SÁNCHEZ *et al.* 2003, ROBINSON et WYKOFF 2004, SHARMA et PARTON 2007, TRINCADO *et al.* 2007, MENG *et al.* 2008). Cependant, à notre connaissance, très peu d'études ont porté sur autant d'espèces et sur la relation avec les variables climatiques tout en mettant l'accent sur les enjeux statistiques comme nous l'avons fait ici. MENG *et al.* (2008) est en fait la seule étude qui fait état de l'influence d'une variable climatique, soit la vitesse du vent. Du point de vue statistique, ROBINSON et WYKOFF (2004), SHARMA et PARTON (2007) et TRINCADO *et al.* (2007) ont effectué un travail similaire, en utilisant une approche basée sur les modèles mixtes pour étalonner les relations HD de plusieurs espèces à l'échelle régionale. Cependant, aucune de ces études n'aborde l'examen des corrélations persistantes pour confirmer l'adéquation des éléments de covariance comme nous l'avons fait.

Le modèle de ce cas d'étude a produit des résidus normalisés qui n'affichaient aucune divergence par rapport aux postulats de normalité et d'homogénéité des variances (Figure 3). Pour prendre en compte les corrélations, ce modèle a nécessité non seulement l'ajout d'un effet aléatoire de placette, mais aussi l'ajout d'une structure de covariance intra-arbre. Ce résultat indique qu'un effet aléatoire de placette peut ne pas être suffisant pour tenir compte des corrélations intra-arbre en présence de mesures répétées sur le même arbre.

Comme l'ont mentionné HALL et BAILEY (2001), il devrait être permis à chaque paramètre de base d'un modèle de varier de placette en placette en leur ajoutant un effet aléatoire. Durant l'étalonnage du modèle, nous avons tenté d'attribuer un effet aléatoire de placette à la fois sur les paramètres  $A_{ijk}$  et  $B_{ijk}$ , mais le modèle présentait des problèmes de convergence. En conséquence, le modèle qui ne comprenait qu'un seul effet aléatoire de placette sur le paramètre  $B_{ijk}$  était celui qui présentait le meilleur ajustement. L'absence d'effet aléatoire de placette sur le paramètre  $A_{ijk}$  ne semble pas être un problème dans cette analyse toutefois, puisque l'examen des résidus normalisés révèle une corrélation empirique très basse (de l'ordre de 0,03) à l'échelle de la placette.

Même avec ces éléments de covariance, soit un effet aléatoire de placette et une structure de covariance linéaire, le modèle a produit des résidus normalisés qui affichaient des corrélations faibles, mais significatives à l'échelle de l'arbre et du mesurage (Figure 4). L'ajout d'éléments de covariance additionnels comme un effet aléatoire de mesurage pourrait corriger la situation. Toutefois, ces ajouts complexifient la configuration de la matrice de variance-covariance des termes d'erreur ( $V$ ), de sorte qu'il est difficile d'atteindre un maximum de vraisemblance. Dans ce cas d'étude, nous avons tenté en vain d'étalonner un modèle comprenant un effet aléatoire de placette et de mesurage en plus d'une structure de covariance. D'autres structures de covariance ont également été testées pour corriger les corrélations persistantes à l'échelle de l'arbre, mais sans succès. Bien que l'adéquation des éléments de covariance ne soit pas parfaite, nous pouvons toutefois présumer que les inférences statistiques sont presque sans biais étant donné que les corrélations persistantes sont faibles.

La vaste majorité des modèles de relation HD proposés dans la littérature sont non linéaires et étalonnés comme tels (p.ex. : HUANG *et al.* 2000, CASTEDO DORADO *et al.* 2005, SHARMA et PARTON 2007) ou linéarisés par une transformation logarithmique (p.ex. : ROBINSON et WYKOFF 2004, TRINCADO *et al.* 2007). On présume souvent que les modèles non linéaires sont supérieurs aux modèles linéaires. On reconnaît qu'ils sont plus robustes dans des cas

d'extrapolation au-delà de la plage des données d'étalonnage (VANCLAY 1994, p. 112). Par contre, la maximisation de la fonction de vraisemblance des modèles non linéaires à effets mixtes repose sur des approximations (LINDSTROM et BATES 1988, PINHEIRO et BATES 1995), ce qui rend la convergence plus difficile à atteindre et les inférences statistiques moins fiables lorsque la convergence est atteinte (PINHEIRO et BATES 2000, p. 323).

Le modèle de cette étude diffère des autres modèles du fait qu'il est purement linéaire. Trois raisons ont motivé la sélection d'un tel modèle plutôt qu'un du type non linéaire. D'abord, il a été possible d'étalonner ce modèle avec un effet aléatoire de placette et une structure de covariance, alors que des essais préliminaires ont démontré qu'il était impossible de le faire avec un modèle non linéaire. Ensuite, compte tenu de l'étendue des données à notre disposition, les extrapolations devraient être très près de la plage de données d'étalonnage. Enfin, à un dhp égal ou supérieur à 9,1 cm, le modèle affiche un comportement semblable à l'équation de von Bertalanffy-Richards (aussi appelée équation de Chapman-Richards) (Figure 1). Celle-ci est régulièrement utilisée dans la modélisation de la hauteur. Par conséquent, nous pouvions nous attendre à ce que le modèle génère de prévisions cohérentes même à l'extérieur de la plage des données d'étalonnage. Bien qu'il puisse sembler simpliste de modéliser la relation HD à l'aide d'un modèle linéaire, nous arrivons à la conclusion qu'une telle approche fonctionne bien dans le cas présent.

Il importe de mentionner que ce modèle a été étalonné dans un contexte d'échantillonnage. En conséquence, les sources de variation ne sont pas contrôlées. En d'autres termes, nous présumons que les variables explicatives ont un effet sur la variable dépendante alors que dans les faits, il est possible que l'effet soit dû à une autre variable dont la corrélation avec la variable explicative est étroite. À titre d'exemple, la corrélation entre les degrés-jours et la température annuelle moyenne dans la base de données peut faire en sorte que cette dernière masque l'effet des degrés-jours.

### 3.2 Influence des variables

L'influence des variables de station sur la relation HD a été soulignée par plusieurs auteurs (p.ex. : LÓPEZ SÁNCHEZ *et al.* 2003, SHARMA et PARTON 2007). Les résultats présentés dans cette étude vont dans le même sens que ceux des études précédentes. Dans le cas présent, la surface terrière, la tempéra-

ture annuelle moyenne et la subdivision écologique se sont avérées être des variables dont l'influence est non seulement significative mais importante par rapport à l'amplitude. La surface terrière est une mesure de densité. Lorsque la densité augmente, la compétition réduit l'accroissement diamétral des tiges, ce qui augmente le rapport hauteur-diamètre (CALAMA et MONTERO 2004). Selon l'espèce, la différence en hauteur pour un dhp donné peut être de plusieurs mètres entre des conditions de faible et de forte surface terrière (Figure 5b). Des différences similaires ont pu être observées par SHARMA et ZHANG (2004) et TRINCADO *et al.* (2007). L'augmentation de la température annuelle moyenne induit également une augmentation de la hauteur pour un même dhp. En somme, le rapport hauteur-diamètre augmente sur les stations associées à des températures plus chaudes. Comme la température annuelle moyenne est probablement un facteur d'importance dans la fertilité de la station et peut être liée à la durée de la saison de croissance, ce résultat était attendu.

Le Québec comporte deux gradients écologiques bien connus : du sud vers le nord, la température diminue alors que de l'ouest vers l'est, l'aridité diminue avec l'influence maritime (SAUCIER *et al.* 1998). La subdivision écologique utilisée dans cette étude s'est avérée être une variable explicative dans la modélisation de la relation HD. Même si la température annuelle moyenne de la période 1971-2000 permet de tenir compte du gradient nord-sud, cette variable ne permet pas de distinguer toutes les différences régionales et un zonage écologique devient nécessaire. Ces différences régionales nous ont amenés à diviser le Québec en quatre subdivisions écologiques (Ouest, Centre, Nord-Est et Sud-Est). Nous avons tenté d'introduire d'autres variables climatiques, comme les précipitations annuelles moyennes et les degrés-jours, dans le modèle pour expliquer ces différences régionales, mais aucune de ces variables ne s'est avérée significative. D'une part, la corrélation entre les degrés-jours et la température annuelle moyenne est étroite et d'autre part, la prévision des précipitations annuelles moyennes par BioSIM est moins précise que celle des températures annuelles moyennes (RÉGNIÈRE et SAINT-AMANT 2007). Nous croyons que les différences entre les quatre subdivisions écologiques pourraient être dues à (i) des particularités climatiques qui sont mal évaluées par les variables climatiques dont nous disposons, (ii) une interaction complexe entre les variables climatiques, ou (iii) des caractéristiques de sol et de peuplement différentes. Cette hypothèse devra être testée.

Une relation linéaire entre la température annuelle moyenne et la hauteur dans le modèle, sans aucun effet des précipitations, peut sembler illogique d'un point de vue biologique. Cette relation signifie que la hauteur d'un arbre augmente proportionnellement à la température annuelle moyenne, si tous les autres facteurs sont reconnus comme étant constants. Lors de l'analyse des résidus normalisés, il a été impossible d'identifier une tendance asymptotique dans cette relation. Il importe de mentionner que le gradient de température annuelle moyenne s'étend de -5 °C à +7 °C dans le jeu de données. La perspective aurait pu être différente avec un gradient plus large. Cependant, il faut remarquer que les données de cette étude ne proviennent pas d'un dispositif expérimental factoriel. Parce qu'il s'agit d'un suivi, on ne retrouve pas toutes les espèces tout au long du gradient de température. Par ailleurs, d'autres facteurs, comme la surface terrière, le dhp et l'indice de statut social, sont susceptibles de changer tout au long de ce gradient. En conséquence, il pourrait être difficile d'identifier clairement l'asymptote de la relation entre la température annuelle moyenne et la hauteur des tiges.

Trois autres variables, la classe de drainage, le statut social de l'arbre et l'occurrence de perturbation légère, ont aussi une influence sur la relation HD, mais de moindre amplitude. Parmi ces variables, l'occurrence d'une perturbation légère est

probablement la plus imprécise. La surface terrière des placettes perturbées peut diminuer fortement entre deux mesurages. Parce que la relation HD est définie selon les valeurs de surface terrière, les arbres de cette placette peuvent virtuellement raccourcir. La relation HD étant définie selon les perturbations, cela permet de tenir compte du délai dans la réaction des arbres à la perturbation. Cependant, cette relation est plutôt faible par rapport à l'amplitude (Figure 5f).

Le jeu de données comporte très peu d'information sur la nature de la perturbation et l'année à laquelle elle s'est produite. En fait, la seule information disponible est la suivante : une certaine perturbation légère de nature anthropique ou naturelle s'est produite ou non depuis le dernier mesurage. Comme les mesurages ont été effectués à des intervalles de dix ans en moyenne, cela induit une grande variabilité. Ce manque d'information constitue une limite du modèle. Des détails concernant le type de perturbation et l'année à laquelle elle s'est produite aideraient à mieux définir cette relation entre la perturbation et la hauteur des tiges. En l'absence de telles informations, les prévisions pour des arbres évoluant dans des peuplements sous aménagement intensif pourraient être biaisées et par le fait même, devraient être utilisées avec prudence.



## Conclusion

Dans cette étude, nous avons réussi à étalonner un modèle linéaire mixte pour prévoir la hauteur des arbres des 20 espèces les plus abondantes du Québec méridional. Un effet aléatoire de placette et une structure de covariance linéaire ont été ajoutés au modèle pour tenir compte des corrélations spatiales et temporelles. On peut présumer que les inférences statistiques provenant de ce modèle sont presque sans biais puisque les corrélations persistantes entre les résidus normalisés sont faibles. Par ailleurs, la validation croisée du modèle a démontré que les biais sont faibles à l'échelle provinciale.

En plus de la surface terrière et de l'indice de statut social, le modèle comprend une variable climatique, soit la température annuelle moyenne de la période 1971-2000. Bien que la relation entre cette variable climatique et la hauteur soit importante par rapport à l'amplitude, il semble qu'il ne permette pas de tenir compte de toute la variabilité climatique présente au Québec. Comme les autres variables climatiques disponibles n'étaient pas significatives, une subdivision écologique, représentant une forme de zonage écologique, a dû être ajoutée au modèle.

Une autre variable explicative permet de prendre en compte l'historique récent du peuplement, soit l'occurrence d'une perturbation légère. Cependant, en raison d'informations limitées dans le jeu de données d'étalonnage, la relation entre la perturba-

tion et la hauteur de tiges est plutôt imprécise. De plus amples détails sur la nature de la perturbation et l'année à laquelle elle s'est produite permettraient d'améliorer le modèle.

Ce modèle de relation HD pourrait être utilisé dans l'inventaire provincial. Comparé à la méthode actuellement en vigueur, il permet d'atténuer les problèmes reliés à l'étalonnage de relations locales ou régionales. Les prévisions de hauteur de ce modèle peuvent également être utilisées comme intrants dans le tarif de cubage général afin d'estimer le volume des arbres (FORTIN *et al.* 2007). Par rapport aux relations HD actuellement utilisées, ce modèle donne une meilleure estimation de l'incertitude associée aux prévisions de hauteur. Il serait donc possible de propager l'erreur associée aux prévisions en hauteur dans le tarif de cubage général et de quantifier l'augmentation de la variance des estimations de volume lorsque l'on utilise une prévision en hauteur au lieu d'une hauteur observée. Cette avenue reste à explorer. Par ailleurs, d'autres analyses seront nécessaires (i) pour évaluer l'amélioration des prévisions en hauteur et en volume dans le cas où l'effet aléatoire de la placette est estimé à partir d'observations de hauteur disponibles (CALAMA et MONTERO 2004, TRINCADO *et al.* 2007), et (ii) pour évaluer les prévisions du modèle à l'échelle régionale.



## Références bibliographiques

- CALAMA, R. et G. MONTERO, 2004. *Interregional nonlinear height-diameter model with random coefficients for stone pine in Spain*. Can. J. For. Res. 34 : 150–163.
- CASTEDO DORADO, F., M. BARRIO ANTA, B.R. PARRESOL et J.G. ÁLVAREZ GONZÁLEZ, 2005. *A stochastic height-diameter model for maritime pine ecoregions in Galicia (northwestern Spain)*. Ann. For. Sci. 62 : 455–465.
- FORTIN, M., J. DEBLOIS, S. BERNIER et G. BLAIS, 2007. *Mise au point d'un tarif de cubage général pour les forêts québécoises : une approche pour mieux évaluer l'incertitude associée aux prévisions*. For. Chron. 83 : 754–765.
- FORTIN, M., S. BÉDARD, J. DEBLOIS et S. MEUNIER, 2008. *Accounting for error correlations in diameter increment modelling: a case study applied to northern hardwood stands in Quebec, Canada*. Can. J. For. Res. 38 : 2274–2286.
- GREGOIRE, T.G., O. SCHABENBERGER et J.P. BARRETT, 1995. *Linear modelling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements*. Can. J. For. Res. 25 : 137–156.
- GRONDIN, P., C. ANSSEAU, L. BÉLANGER, J.-F. BERGERON, Y. BERGERON, A. BOUCHARD, J. BRISSON, L. DE GRANDPRÉ, G. GAGNON, C. LAVOIE, G. LESARD, S. PAYETTE, P.J.H. RICHARD, J.-P. SAUCIER, L. SIROIS et L. VASSEUR, 1996. *Écologie forestière*. Dans : Bérard, J.A. et M. Côté (Ed.), Manuel de foresterie, Les Presses de l'Université Laval, Sainte-Foy, Québec, p. 133–279.
- HALL, D.B. et R.L. BAILEY, 2001. *Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models*. For. Sci. 47 : 311–321.
- HUANG, S., D. PRICE et S.J. TITUS, 2000. *Development of ecoregion-based height-diameter models for white spruce in boreal forests*. For. Ecol. Manage. 129 : 125–141.
- LEMAY, V.M., 1990. *MSLS: a linear least squares technique for fitting a simultaneous system of equations with a generalized error structure*. Can. J. For. Res. 20 : 1830–1839.
- LINDSTROM, M.J. et D.M. BATES, 1988. *Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data*. J. Am. Stat. Assoc. 83 : 1014–1022.
- LITTELL, R.C., G.A. MILLIKEN, W.W. STROUP, R.D. WOLFINGER et O. SCHABENBERGER, 2006. *SAS System for mixed models, Second Edition*. SAS Institute Inc., Cary, NC, 814 p.
- LOPEZ SÁNCHEZ, C.A., J. GORGOSO VARELA, F. CASTEDO DORADO, A. ROJO ALBORECA, R. RODRÍGUEZ SOALLEIRO, J.G. ÁLVARIZ GONZÁLEZ et F. SÁNCHEZ RODRÍGUEZ, 2003. *A height-diameter model for Pinus radiata D. Don in Galicia (Northwest Spain)*. Ann. For. Sci. 60 : 237–245.
- MENG, S.X., S. HUANG, V.J. LIEFFERS, T. NUNIFU, et Y. YANG, 2008. *Wind speed and crown class influence the height-diameter relationship of lodgepole pine: nonlinear mixed effects modeling*. For. Ecol. Manage. 256 : 570–577.
- MINISTÈRE DES RESSOURCES NATURELLES ET DE LA FAUNE DU QUÉBEC, 2007. *Normes d'inventaire forestier, Placettes-échantillons permanentes, Édition 2007*. Gouvernement du Québec, ministère des Ressources naturelles et de la Faune, Direction des inventaires forestiers. 246 p.
- MINISTÈRE DES RESSOURCES NATURELLES, DE LA FAUNE ET DES PARCS DU QUÉBEC, 2003. *Croissance de la forêt publique du Québec sous aménagement : évolution mesurée à partir des placettes-échantillons permanentes*. Gouvernement du Québec, ministère des Ressources naturelles, de la Faune et des Parcs. 8 p.
- NEWTON, P.F. et I.G. AMPONSAH, 2007. *Comparative evaluation of five height-diameter models developed for black spruce and jack pine standtypes in terms of goodness-of-fit, lack-of-fit and predictive ability*. For. Ecol. Manage. 247 : 149–166.
- PARENT, B. et C. FORTIN, 2007. *Ressources et industries forestières, Portrait statistique, Édition 2007*. Gouvernement du Québec, ministère des Ressources naturelles et de la Faune, Direction du développement de l'industrie des produits forestiers.

- PINHEIRO, J.C. et D.M. BATES, 1995. *Approximations to the Log-Likelihood function in the nonlinear mixed-effects model*. J. Comp. Graph. Stat. 4 : 12–35.
- PINHEIRO, J.C. et D.M. BATES, 2000. *Mixed-effects models in S and S-PLUS*. Springer Verlag, New York. 528 p.
- RÉGNIÈRE, J. et R. SAINT-AMANT, 2007. *Stochastic simulation of daily air temperature and precipitation from monthly normals in North America north of Mexico*. Int. J. Biometeorol. 51 : 415–430.
- RÉGNIÈRE, J. et R. SAINT-AMANT, 2008. *BioSIM 9 – Manuel de l'utilisateur*. Ressources naturelles Canada, Service canadien des forêts, Centre de foresterie des Laurentides. Rapport d'information LAU-X-134F. 74 p.
- RÉGNIÈRE, J., B. COOKE et V. BERGERON, 1995. *BioSIM : Un instrument informatique d'aide à la décision pour la planification saisonnière de la lutte antiparasitaire – Guide d'utilisation*. Ressources naturelles Canada, Service canadien des forêts, Centre forestier des Laurentides. Rapport d'information LAU-X-116f. 68 p.
- ROBINSON, A.P. et W.R. WYKOFF, 2004. *Imputing missing height measures using a mixed-effects modelling strategy*. Can. J. For. Res. 34 : 2492–2500.
- SAS Institute, 2003. *SAS OnlineDoc 9 [en ligne]*. SAS Institute Inc. Cary, NC. Disponible à l'adresse <http://v9doc.sas.com/sasdoc/>.
- SAUCIER, J.-P., J.-F. BERGERON, P. GRONDIN et A. ROBITAILLE, 1998. *Les régions écologiques du Québec méridional : un des éléments du système hiérarchique de classification écologique du territoire mis au point par le ministère des Ressources naturelles*. L'Aubelle, février-mars 1998. 8 p.
- SHARMA, M. et J. PARTON. 2007. *Height-diameter equations for boreal tree species in Ontario using a mixed-effects modeling approach*. For. Ecol. Manage. 249 : 187–198.
- SHARMA, M. et S.Y. ZHANG. 2004. *Height-diameter models using stand characteristics for Pinus banksiana and Picea mariana*. Scand. J. For. Res. 19 : 442–451.
- TRINCADO, G., C.L. VANDERSCHAAF et H.E. BURKHART. 2007. *Regional mixed-effects height-diameter models for loblolly pine (Pinus taeda L.) plantations*. Eur. J. For. Res. 126 : 253–262.
- VANCLAY, J.K., 1994. *Modelling forest growth and yield: applications to mixed tropical forests*. CAB International, Wallingford, UK. 312 p.
- VONESH, E.F. et R.L. CARTER, 1992. *Mixed-effects nonlinear regression for unbalanced repeated measures*. Biometrics 48 : 1–17.





Les priorités du ministère des Ressources naturelles et de la Faune en matière d'acquisition de connaissances scientifiques ont pour but d'assurer un aménagement durable des forêts pour le bénéfice de la population. L'évaluation de la ressource forestière est donc essentielle pour parvenir à l'aménager de façon adéquate. La hauteur des arbres est une variable indispensable lors de cette évaluation. Toutefois, pour des raisons pratiques, la hauteur ne peut être mesurée sur tous les arbres. On doit donc l'estimer. La Direction de la recherche forestière et la Direction des inventaires forestiers ont travaillé ensemble à la mise au point d'une méthode d'estimation des hauteurs à partir des données du réseau provincial des placettes-échantillons permanentes. Ce mémoire propose un modèle de relation hauteur-diamètre d'envergure provinciale qui permet d'estimer la hauteur des arbres à partir de variables de climat, de station et dendrométriques disponibles dans la plupart des inventaires.