IFT-7028 Conception et simulation de systèmes intelligents pour l'industrie 4.0

Projet de session : Modélisation du prix des secteurs de récoltes forestières du Québec

Par Maude Beauchemin, Simon Schmiedel,
Gabrielle St-Georges et François-Alexandre Tremblay

À l'attention de Jonathan Gaudreault

Juin 2021

Table des matières

1 Introduction **2**

2 Définition du problème **3**

2.1 Variable expliquée 3

2.2 Variables explicatives 4

2.3 Modèles 4

2.4 Mesure de performance 4

3 Collecte et transformation des données **4**

3.1 Données initiales 6

3.1.1 Nettoyage des données initiales 6

3.1.2 Formation des regroupements 6

3.2 Données additionnelles 7

3.3 Données collectées 7

4 Analyse statistique **8**

4.1 Analyses descriptives 8

4.2 Analyse en composantes principales 8

5 Résultats et performance **11**

5.1 Modèles prédictifs 11

5.2 Importance des variables 16

6 Discussion et limites **17**

7 Conclusion **18**

8 Remerciements **18**

A Annexe **20**

1 Introduction

Au Québec, 465 912 km^2 du territoire est composé de forêts accessibles [1] dont 84% sont de tenure publique sous la responsabilité du Ministère de la Forêt, de la Faune et des Parcs (MFFP). Dans l'optique de préserver ce patrimoine québécois, le MFFP est responsable d'administrer les forêts du domaine de l'état pour les activités de récoltes forestières et les travaux sylvicoles tout en assurant l'aménagement durable du territoire et le respect des lois et règlements en vigueur. La détermination des garanties d'approvisionnement (GA) n'y fait pas exception. Actuellement, le MFFP accorde ces garanties aux compagnies forestières qui oeuvrent dans la transformation du bois à la hauteur de 75% de leurs besoins totaux de production. Pour subvenir au 25% manquant, les compagnies doivent compétitionner, à travers des enchères privées, pour l'acquisition de différents secteurs de récoltes qui se distinguent par plusieurs caractéristiques telles que leur localisation, leur accessibilité et leur volume de différentes essences. Le prix des enchères peut ainsi fortement varier d'un secteur à l'autre. Par exemple, les essences fréquemment retrouvées dans la forêt tempérée du sud de la province sont beaucoup plus variées que celles se situant dans la forêt boréale [1], ce qui a un impact sur sa valeur monétaire.

En plus de garantir une partie de l'approvisionnement, le MFFP, et plus précisément le Bureau de Mise en Marché des Bois (BMMB), doit aussi déterminer le prix de vente des récoltes garanties (par les GA) aux entreprises. Pour réaliser ce mandat, depuis 2015, le BMMB utilise les données de 882 ventes réparties sur les cinq dernières années (2015 à 2019) pour modéliser le problème. Leurs modèles font usage de régressions linéaires multiples.

Actuellement, le BMMB utilise deux équations de régression distinctes pour estimer la valeur des secteurs mis aux enchères, soit une pour les secteurs ayant une proportion de SEPM (sapins, épinettes et pins gris et mélèzes) supérieure à 75% et une autre qui englobe les secteurs restants. À titre d'exemple, les variables explicatives du modèle SEPM et leurs coefficients sont présentés au Tableau 10 en annexe. De plus, une description des différentes variables utilisées dans le modèle est présentée dans le Tableau 1.

Toutefois, des erreurs dans l'estimation de la juste valeur marchande du bois des forêts publiques mènent à des situations non-désirables. D'une part, une sous-estimation du prix de vente génère des pertes monétaires pour l'État puisque les redevances reçues ne représentent pas la juste valeur des bois. D'autre part, une surestimation du prix s'associe à des pertes pour les industriels, ce qui engendre de nombreuses critiques. Il existe ainsi un intérêt à réduire la magnitude des erreurs et le niveau de risque afférent afin d'éviter une baisse générale du prix que les compagnies sont prêtes à payer.

Le nombre important d'enchères depuis la mise en place du BMMB en 2011 laisse entrevoir de nouvelles possibilités d'amélioration des prévisions des prix par le recours à des modèles provenant du domaine de l'apprentissage automatique. Ces modèles, plus puissants, peuvent profiter de grands ensembles de données pour modéliser des relations plus complexes et, conséquemment, augmenter les gains en performance.

Suite à une discussion avec le BMMB le lundi 31 mai 2021, notre équipe a choisi de s'intéresser à la comparaison du potentiel des modèles d'apprentissage supervisés pour la prévision des prix de ventes des secteurs de récoltes forestières aux modèles présentement en place. Dans le cadre de ce projet, nous évaluons la performance des forêts aléatoires, des machines à vecteur de support (SVR) et des

réseaux de neurones en utilisant la régression linéaire multiple comme base de référence. Notre défi est d'autant plus grand puisque les données de nature sensible ne sont pas accessibles et ne peuvent donc pas être utilisées. On compte notamment parmi ces données le nombre de mises et le nombre de fois que le secteur a été mis en vente. À partir de la confirmation du BMMB, nous savons que le nombre de mises est une variable qui influence très fortement le prix de vente d'un secteur.

Dans ce rapport, nous définirons d'abord le problème et ses composantes principales. La deuxième section porte sur la collecte et le traitement de l'information afin de bâtir les ensembles de données. Nous présentons ensuite des analyses statistiques descriptives ainsi que des analyses par composantes principales (ACP). Enfin, les résultats et la performance des modèles prédictifs sont évalués et discutés.

Tableau 1 – Description des variables utilisées dans le modèle SEPM du BMMB

Variable	Unité	Description
Nombre de mises	Unité	Nombre de mises conformes reçues sur le secteur vendu.
Revenus nets usine	$/m^3	Revenus nets moyens (qualité B, C et M) du modèle usine calculés pour chaque secteur à l'aide de la distribution de tiges, selon les prix PRIBEC trois mois précédant le mois de publication de l'appel d'offres.
Indice de prix SEPM	Indice en base 100	Prix des produits finis moyen 6 derniers mois divisé par le prix moyen 2010 multiplié par 100.
Coûts de transport aux cinq usines et aux marchés	$/m^3	Coûts de transport moyens pondérés aux 5 usines de sciage de SEPM les plus proches, plus le coût de transport moyen des 5 usines aux marchés.
Proportion SEPM trituration	%	Proportion des volumes SEPM de qualité trituration.
Nombre de fois que le secteur a été mis en vente	N^{bre} discret	Nombre de fois que le secteur a été affiché avant d'être vendu.
Proportion SEPM	%	Proportion de SEPM (bois d'œuvre et trituration) sur le volume total dans le secteur.
Autres coûts	$/m^3	Autres coûts pondérés selon les volumes de récolte du secteur : SOPFIM, SOPFEU, hébergement ou camp et administration.
Coûts de chemin	$/m^3	Coûts de construction de chemin, de réfection, d'entretien et des infrastructures extraordinaires (ex. : pont) diminués des sommes reçues dans le cadre du programme de remboursement des chemins multiressources.
Herfindahl	Indice entre 0 et 10	Indice de concentration calculé en utilisant les volumes des permis d'usines résineux d'une région administrative. Plus l'indice est proche de 0, plus la concentration est faible (marché compétitif). À l'inverse, un indice de 10 indique un marché concentré (ex. : monopole ou monopsone).
Territoire structuré	Indice entre 0 et 7	Indice marquant la présence d'une ou de plusieurs structures d'organisation du territoire forestier.
Vol/ha	m^3/ha	Volume à l'hectare du secteur.
Difficulté terrain	%	Moyenne pondérée des pourcentages de correction de chaque machine utilisée pour la récolte. Cette correction est en fonction de la pente, du drainage et de la rugosité.
Éloignement des utilisateurs de copeaux	km	Distance moyenne aux 5 usines utilisatrices de copeaux SEPM les plus proches.
Distance de transport à la 3e, 4e et 5e usine	km	Distance moyenne pondérée aux usines de sciage SEPM, et ce, pour les 3e, 4e et 5e usines les plus proches.
Carie	%	Altération du bois par les champignons qui en modifie le poids, la couleur, la texture et la résistance.

2 Définition du problème

Notre équipe s'intéresse à l'élaboration d'un modèle prédictif qui permet de déterminer les prix des ventes des secteurs publics. Cette section définit les principaux aspects du problème.

2.1 Variable expliquée

La variable expliquée choisie est celle de la valeur monétaire des secteurs par mètre cube de bois désiré par l'acheteur (i.e. $\$/m^3$). Cette quantité est couramment utilisée dans le domaine forestier

et se trouve d'ailleurs à être la variable utilisée par le BMMB dans leurs modèles. À la suite d'une recommandation de M. Gaudreault, nous avons tenté d'exprimer le problème sous forme de ratio entre le prix de vente et l'index des prix du bois, tous les deux standardisés, afin d'améliorer la performance des algorithmes. Cette approche ne semblait toutefois pas améliorer leur apprentissage ni leur performance. Ces expériences additionnelles ont confirmé le choix de notre définition initiale de la variable expliquée.

2.2 Variables explicatives

Deux ensembles de variables explicatives sont utilisés pour entraîner les modèles. Un premier jeu de données comportant 65 variables s'intitule l'ensemble *complet*. Le deuxième, une version épurée nommée *minimaliste*, compte les 18 variables explicatives jugées les plus pertinentes. Le Tableau 2 rapporte l'ensemble des variables. La dernière colonne intitulée **Min.** permet d'indiquer celles qui sont incluses dans la base de données *minimaliste*.

2.3 Modèles

Pour ce problème de régression, nous sélectionnons quatre techniques d'apprentissage automatique : les régressions linéaires multiples, les forêts aléatoires, les réseaux de neurones et les machines à vecteurs de support (SVR). Ces techniques, qui sont orthogonales, connues et très utilisées en pratique [2], constituent à notre avis des choix pertinents pour explorer le potentiel de modèles prédictifs. Compte tenu que le BMMB utilise également des régressions linéaires, notre modèle de régression linéaire sert de base de comparaison aux trois autres.

2.4 Mesure de performance

Le BMMB utilise présentement les mesures du R^2 et du R^2 ajusté à titre d'unité de mesure de la performance. Celles-ci facilitent la comparaison de différents modèles, issus de problème de régression, à partir du pourcentage de variance expliquée. Dans le cadre de notre rapport, nous rapportons ces indicateurs à des fins de comparaison. Par ailleurs, puisque l'erreur absolue moyenne (MAE) est un indicateur facile à interpréter, nous retenons également cette mesure de performance à des fins descriptives.

3 Collecte et transformation des données

Afin de mettre en place la base de données *complète* comportant 65 variables explicatives, différentes sources sont utilisées. La base de données publique du BMMB (disponible ici) constitue notre point de départ. Celle-ci contient 2614 enregistrements de secteurs mis en ventes aux enchères et 201 colonnes rapportant diverses informations autant géographiques (ex. coordonnées, territoire) que biophysiques (ex. volume de bois répertorié par essence). Nous ajoutons ensuite à la base de données des variables additionnelles de nature économique [3] [4], routière [5] [6], et dendrométrique (BMMB). Cette base de données comporte également la variable explicative du prix de vente d'un secteur. La colonne source du Tableau 2 présente en détail nos sources.

Tableau 2 – Variables explicatives

Variable	**Unités**	**Type**	**Description**	**Source**	**Calculs réalisés**	**Min.**
Territoire	NA	Catégorielle	Territoire où se trouve le secteur	BMMB	NA	X
UA	NA	Catégorielle	Unité d'aménagement où est situé le secteur	BMMB	NA	X
Superficie	ha	Numérique continue	Superficie du secteur	BMMB	X NA	X
Densite volumique	m^3/ha	Numérique continue	Volume total de bois disponible par hectare	BMMB	$\frac{\text{Volume total}}{\text{Superficie}}$	X
Vol total	m^3	Numérique continue	Volume total de bois disponible	BMMB + Calcul	Vol. obligatoire + Vol. optionnel	X
Vol obligatoire total	m^3	Numérique continue	Volume obligatoire total de bois disponible	BMMB	NA	
Volume optionnel total	m^3	Numérique continue	Volume optionnel total de bois disponible	BMMB	NA	
Annee	Année	Numérique discrète	Année de la date fin de la vente	BMMB	NA	
Latitude	NA	Numérique continue	Latitude du secteur	BMMB	NA	X
Longitude	NA	Numérique continue	Longitude du secteur	BMMB	NA	X
Dist vol oiseau 1	NA	Numérique continue	Distance entre le point de récolte et l'usine de production la plus proche à vol d'oiseau	Répertoire des usines de production du Québec + Calcul	Distance euclidienne	
Dist vol oiseau 2	NA	Numérique continue	Distance entre le point de récolte et la deuxième usine de production la plus proche à vol d'oiseau	Répertoire des usines de production du Québec + Calcul	Distance euclidienne	
RandomLenghts	NA	Numérique continue	Index américain du prix du bois	RandomLenghts	NA	
Index de prix forestier canadien	NA	Numérique continue	Index de prix forestier canadien	Banque du Canada	NA	X
SEPM pourcent	%	Numérique continue	Proportion du volume de bois de type SEPM par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Somme SEPM}}{\text{Volume total}}$	X
Autre res pourcent	%	Numérique continue	Proportion du volume de bois de type Autres résineux par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Somme Autres res.}}{\text{Volume total}}$	X
Bois franc pourcent	%	Numérique continue	Proportion du volume de bois de type bois franc par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Somme Bois franc}}{\text{Volume total}}$	X
BP pourcent	%	Numérique continue	Proportion du volume de bois de type bouleau-peuplier par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Somme Bouleau-Peuplier}}{\text{Volume total}}$	X
SEPM Sciage	m^3	Numérique continue	Volume de bois de sciage de type SEPM dans le secteur	BMMB	NA	
SEPM Trituration	m^3	Numérique continue	Volume de bois de trituration de type SEPM dans le secteur	BMMB	NA	
Autres res Sciage	m^3	Numérique continue	Volume de bois de sciage de type Autres résineux dans le secteur	BMMB	NA	
Autres res Trituration	m^3	Numérique continue	Volume de bois de trituration de type Autres résineux dans le secteur	BMMB	NA	
Bois franc Sciage	m^3	Numérique continue	Volume de bois de sciage de type bois franc dans le secteur	BMMB	NA	
Bois franc Trituration	m^3	Numérique continue	Volume de bois de trituration de type bois franc dans le secteur	BMMB	NA	
BP Sciage	m^3	Numérique continue	Volume de bois de sciage de type bouleau-peuplier dans le secteur	BMMB	NA	
BP Trituration	m^3	Numérique continue	Volume de bois de trituration de type bouleau-peuplier dans le secteur	BMMB	NA	
Sciage pourcent	%	Numérique continue	Proportion du volume de bois de sciage par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Somme sciage}}{\text{Volume total}}$	X
Trituration pourcent	%	Numérique continue	Proportion du volume de bois de trituration par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Somme trituration}}{\text{Volume total}}$	
SEPM Sciage pourcent	%	Numérique continue	Proportion du volume de bois de sciage de type SEPM par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{SEPM sciage}}{\text{Volume total}}$	
SEPM Trituration pourcent	%	Numérique continue	Proportion du volume de bois de trituration de type SEPM par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{SEPM trituration}}{\text{Volume total}}$	
Autres res Sciage pourcent	%	Numérique continue	Proportion du volume de bois de sciage de type Autres résineux par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Autres res. sciage}}{\text{Volume total}}$	
Autres res Trituration pourcent	%	Numérique continue	Proportion du volume de bois de trituration de type Autres résineux par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Autres res. trituration}}{\text{Volume total}}$	
Bois franc Sciage pourcent	%	Numérique continue	Proportion du volume de bois de sciage de type Autres résineux par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Bois franc sciage}}{\text{Volume total}}$	
Bois franc Trituration pourcent	%	Numérique continue	Proportion du volume de bois de trituration de type bois franc par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Bois franc trituration}}{\text{Volume total}}$	
BP Sciage pourcent	%	Numérique continue	Proportion du volume de bois de sciage de type bouleau-peuplier par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Bouleau-Peuplier sciage}}{\text{Volume total}}$	
BP Trituration pourcent	%	Numérique continue	Proportion du volume de bois de trituration de type bouleau-peuplier par rapport au volume total de bois dans le secteur	BMMB + Calcul	$\frac{\text{Bouleau-Peuplier trituration}}{\text{Volume total}}$	
Vol par tige moyen	dm^3	Numérique continue		BMMB + Calcul		X
Vol par tige moyen SEPM	dm^3	Numérique continue	Volume moyen des tiges du bois de type SEPM	BMMB	NA	
Vol par tige moyen Autres res	dm^3	Numérique continue	Volume moyen des tiges du bois de type Autres résineux	BMMB	NA	
Vol par tige moyen Bois franc	dm^3	Numérique continue	Volume moyen des tiges du bois de type Bois franc	BMMB	NA	
Vol par tige moyen moyen BP	dm^3	Numérique continue	Volume moyen des tiges du bois de type Bouleau-Peuplier	BMMB	NA	
Superficie CP	ha	Numérique continue	Superficie en coupe partielle	BMMB	NA	
Superficie CT	ha	Numérique continue	Superficie en coupe totale	BMMB	NA	
Coupe partielle pourcent	%	Numérique continue	Pourcentage en coupe partielle	BMMB + Calcul	$\frac{\text{Superficie coupe partielle}}{\text{Superficie totale}}$	
Coupe totale pourcent	%	Numérique continue	Pourcentage en coupe totale	BMMB + Calcul	$\frac{\text{Superficie coupe totale}}{\text{Superficie totale}}$	X
Secteur contrainte	NA	Catégorielle	Contraintes spéciales dans le secteur	BMMB	NA	
Distance SEPM	km	Numérique continue	Distance moyenne des 5 usines SEPM les plus proches	BMMB	NA	
Distance Peuplier	km	Numérique continue	Distance moyenne des 5 usines de peuplier les plus proches	BMMB	NA	
Distance FT	km	Numérique continue	Distance moyenne des 5 usines de feuillus trituration les plus proches	BMMB	NA	
Distance FS	km	Numérique continue	Distance moyenne des 5 usines de feuillus sciage les plus proches	BMMB	NA	
Distance Pins	km	Numérique continue	Distance moyenne des 5 usines de pins les plus proches	BMMB	NA	
Vol desire global	m^3	Numérique continue	Volume total désiré par l'acheteur	BMMB	NA	
Vol desire SEPM	m^3	Numérique continue	Volume de SEPM désiré par l'acheteur	BMMB	NA	
Vol desire Autres res	m^3	Numérique continue	Volume d'autres résineux désiré par l'acheteur	BMMB	NA	
Vol desire Peuplier	m^3	Numérique continue	Volume de peupliers désiré par l'acheteur	BMMB	NA	
Vol desire Autres feuillus	m^3	Numérique continue	Volume d'autres feuillus désiré par l'acheteur	BMMB	NA	
Vol desire SEPM pourcent	m^3	Numérique continue	Pourcentage de SEPM désiré par l'acheteur	BMMB + Calcul	$\frac{\text{Volume désiré SEPM}}{\text{Volume total}}$	
Vol desire Autres res pourcent	m^3	Numérique continue	Pourcentage d'autres résineux désiré par l'acheteur	BMMB + Calcul	$\frac{\text{Volume désiré Autres rés.}}{\text{Volume total}}$	
Vol desire Peuplier pourcent	m^3	Numérique continue	Pourcentage de peupliers désiré par l'acheteur	BMMB + Calcul	$\frac{\text{Volume désiré Bouleau-Peuplier}}{\text{Volume total}}$	
Vol desire Autres feuillus pourcent	m^3	Numérique continue	Pourcentage d'autres feuillus désiré par l'acheteur	BMMB + Calcul	$\frac{\text{Volume désiré Autres feuillus}}{\text{Volume total}}$	
Distance moyenne des usines les plus proches	km	Numérique continue	Distance moyenne aux deux usines les plus proches	API de Google maps	Moyenne	X
Chemin a implanter	km	Numérique continue	Longueur de chemins à implanter	BMMB	NA	
Chemin a reparer	km	Numérique continue	Longueur de chemins à réparer	BMMB	NA	
Densite chemin a implanter	km/m^3	Numérique continue	Longueur de chemins à implanter par m^3 de bois	BMMB + Calcul	$\frac{\text{KM IM}}{\text{Volume total}}$	X
Densite chemin a reparer	km/m^3	Numérique continue	Longueur de chemins à réparer par m^3 de bois	BMMB + Calcul	$\frac{\text{KM RE}}{\text{Volume total}}$	X

3.1 Données initiales

3.1.1 Nettoyage des données initiales

La première étape du projet consiste à nettoyer la base de données publique du BMMB.

Pour la variable d'«état» de la vente, nous conservons uniquement les enregistrements dont l'état est adjugé puisque les autres types de contrats (ex. annulé, secteur retiré, non adjugé) n'incluent pas de prix de vente. Au total, 871 lignes sont supprimées. Puisque l'«état» de la vente est identique pour toutes les ventes restantes, nous retirons conséquemment cette variable.

Les autres enregistrements n'ayant pas de prix de vente sont ensuite supprimés. Autrement, l'absence de cette valeur rendrait impossible l'approximation de la fonction à partir des algorithmes d'apprentissage choisis. Les enregistrements sans prix de vente sont au nombre de 15 dans la base de données.

Le «type de vente», qui indique essentiellement si le bois est livré, récolté ou encore sur pied, est ensuite nettoyé. Nous décidons de conserver uniquement le type de vente se nommant bois sur pied - mesurage. Il s'agit en effet du type de vente le plus commun et nous souhaitons éviter les enjeux de sous-représentativité associés aux trois autres types de vente. Le retrait de ce type de vente représente 132 enregistrements. Pour la même raison, nous retirons la variable du «type de vente».

La base de données contient trois sortes d'enchères, soient les enchères de base, combinatoires ou combinatoires principales. Pour simplifier le problème, nous décidons de conserver uniquement les observations concernant les enchères de base. Les enchères combinatoires sont en effet plus complexes et difficiles à modéliser et comptent seulement 59 enregistrements.

Nous nous attardons subséquemment à enlever les variables qui n'expliquent pas le prix de vente. À titre indicatif, nous retirons par exemple le «montant de la garantie», le «nom du secteur», le «numéro d'appel d'offre» et le «statut des opérations». Ces variables de nature administrative n'affectent pas la valeur des secteurs.

3.1.2 Formation des regroupements

La base de données contient également les variables suivantes pour chacune des 30 essences de bois, ce qui correspond au total à 180 variables explicatives :

- Volume de bois d'oeuvre obligatoire
- Volume de bois d'oeuvre optionnel
- Prix fixe de bois d'oeuvre par m^3
- Volume de trituration obligatoire
- Volume de trituration optionnel
- Prix fixe de trituration par m^3

Le prix fixe est une manière alternative d'effectuer un achat aux enchères en proposant un prix pour chaque essence se retrouvant sur le secteur d'intérêt plutôt que de proposer un prix global. Nous retirons les variables concernant le prix fixe étant donné que celles-ci sont des sous-divisions du prix de vente du secteur.

La dernière transformation majeure sur la base de données initiale consiste à regrouper les essences dans des catégories où le bois partage des caractéristiques similaires. Selon les principales industries de transformation primaire de bois du Québec, nous regroupons les essences en quatre regroupements distincts : le SEPM (déjà utilisé par le BMMB depuis 2015), le bois franc, le bouleau-peuplier et les autres résineux. Ce choix permet notamment de réduire le nombre de variables et d'éviter de retirer des enregistrements supplémentaires. Les regroupements par essence sont rapportés dans le Tableau 3.

Tableau 3 – Regroupement des essences de bois

Regroupement	**Essences**
SEPM	Sapin beaumier, Épinette sp., Pin gris, Mélèze laricin
Bouleau-Peuplier	Bouleau à papier, Peuplier sp.
Bois franc	Érable sp., Frêne sp., Hêtre à grandes feuilles, Bouleau jaune, Chêne sp., Autres feuillus
Autres résineux	Pin blanc, Pin rouge, Thuya occidental, Pruche de l'est, Autres résineux

3.2 Données additionnelles

Suite à notre rencontre avec le BMMB, ce dernier a accepté de nous fournir un fichier Excel comportant des données additionnelles. En faisant une première analyse, dix enregistrements initiaux sont absents de cette nouvelle base de données. Ils sont donc retirés.

Le volume par tige moyen, qui réfère à la grosseur du tronc des arbres, se retrouve aussi dans le nouveau fichier. Par contre, en croisant la variable avec notre base initiale de données, nous notons que 153 volumes sont manquants. Pour éviter de perdre des données de ventes, nous estimons la valeur à partir de la médiane par secteur et par essence. Cela permet de tenir en compte à la fois les effets biophysiques (ex. climat) et les caractéristiques des arbres.

De plus, nous identifions 124 observations dont le volume moyen des tiges (par regroupement d'essence) contient une valeur supérieure à 0 alors qu'aucun volume de bois n'est vendu, ce qui constitue une incohérence dans les données. En fait, ces observations proviennent de données d'inventaires du MFFP où les essences pour lesquelles il y a une interdiction de récolte n'ont pas été retirées. Pour corriger la situation, nous ramenons à zéro les volumes moyens.

Le BMMB nous a également partagé plusieurs types de données tels que la longueur des chemins à construire et à réparer pour un secteur particulier, la distance en kilomètres des cinq usines de transformation de bois les plus proches par grand type d'industrie ainsi que le volume total désiré par l'industriel pour le secteur. Cette valeur nous a permis de préciser notre variable d'intérêt du prix par mètre cube en modifiant le volume par le volume désiré de l'acheteur.

3.3 Données collectées

Pour calculer la distance de route entre un point de récolte et l'usine de transformation, nous réalisons plusieurs étapes. Tout d'abord, nous faisons une liste des coordonnées géographiques de toutes usines de transformation du Québec [5]. De cette liste, nous conservons seulement les usines qui détiennent des volumes de production annuels de plus de 200 000 mètres cubes. Cela permet de retirer les petites scieries marginales moins propices à traiter les récoltes de bois issues de nos transactions qui viendraient bruiter les distances calculées. Au final, on considère donc 101 usines de transformation.

À partir des coordonnées géographiques des 101 usines, nous choisissons, pour chaque point de récolte, les deux usines avec la plus faible distance euclidienne en terme de latitude et longitude. Les coordonnées des deux usines et celles du point de récolte sont ensuite connectées à l'interface de programmation d'application (API) de Google Maps pour déterminer la distance de route jusqu'aux usines. Malgré tout, il arrive que certains secteurs soient trop loin des routes répertoriées et que l'API ne puisse pas estimer la distance de route. Cette situation survient pour 180 lignes de notre base de données. Dans ces situations, nous calculons la distance en kilomètres à vol d'oiseau entre les deux coordonnées. Nous augmentons ensuite cette valeur de 25%, déterminée empiriquement, pour prendre en compte la non-linéarité des distances. Finalement, nous calculons la distance moyenne entre le secteur et les deux usines les plus proches.

Du côté des variables économiques, les valeurs de l'indice du prix du bois canadien ainsi que de l'indice RandomLengths sont ajoutées aux observations en fonction de la date de la vente. Puisque les deux variables contiennent des informations similaires, une validation supplémentaire est réalisée pour évaluer la similarité des variables au fil du temps. Enfin, puisque les prix de vente remontent à 2011, nous les indexons selon les taux d'indexation publiés par la Banque du Canada [7].

À la suite de toutes ces manipulation, la base de données finale contient 1492 lignes ainsi que 65 et 18 variables pour les base de données complète et minimaliste respectivement.

4 Analyse statistique

4.1 Analyses descriptives

Afin de s'assurer de la validité des données, nous réalisons une analyse descriptive sur l'ensemble de la base de données à l'aide du logiciel R. Tout d'abord, nous vérifions les doublons et les valeurs manquantes. Aucune anomalie n'est constatée.

Les statistiques de base (voir Figure 8 en annexe) sont ensuite produites pour l'ensemble des variables de manière à évaluer rapidement la présence ou non de valeurs aberrantes. Les valeurs minimales, médianes, moyennes et maximales des variables numériques sont cohérentes avec le contexte forestier et aucune irrégularité n'est détectée. Le nombre de modalités et leur uniformité pour les trois variables catégorielles sont également examinés. Aucune action n'est nécessaire à ce sujet.

De la même façon, des graphiques de la distribution des variables sont générés. La Figure 9 en annexe présente seulement ceux relatifs aux variables de la base de données *minimaliste* pour alléger le rapport. Les données plus extrêmes sont validées et aucune correction n'est nécessaire. De plus, considérant que les échelles de mesure sont assez diversifiées, la standardisation des variables semble une opération pertinente pour la accélérer la convergence de certains algorithmes.

4.2 Analyse en composantes principales

Dans le contexte où le jeu de données complet comporte beaucoup de variables, soit 65 par rapport à un nombre de lignes relativement limité (moins de 1500), il est intéressant d'explorer la réduction de la dimensionnalité. L'analyse en composantes principales (ACP) permet de trouver une combinaison linéaire des variables qui maximise la variance tout en conservant le maximum d'information sur

les données et leurs interactions. Puisque l'ACP est très sensible à l'échelle de mesure des variables, les logiciels d'analyse appliquent automatiquement une standardisation des données préalablement à la méthode. Dans le cadre de notre rapport, nous utilisons parallèlement le logiciel R et le logiciel JMP ! pour effectuer cette tâche.

Dans un premier temps, pour la base de données *complète*, 51 composantes (aussi appelées dimensions ou axes) sont nécessaires pour chercher 100% de la variation expliquée comme le montre la Figure 1. Toutefois, on remarque que la variance expliquée atteint 80% dès la 11e composante et 95% dès la 24e. Pour sélectionner la quantité adéquate de composantes, Cattell [8] propose de conserver celles qui précèdent le pied de l'éboulis. Sur la base de cette règle, nous retenons et analysons les trois premières composantes qui représentent 52% de la variabilité totale. Le même exercice est réalisé sur la base de données *minimaliste* (Figure 2). Cette fois-ci, la 8e dimension permet d'obtenir 80% de la variance et la combinaison des trois premières totalise environ 52%.

FIGURE 1 – Base de données - *Complet*

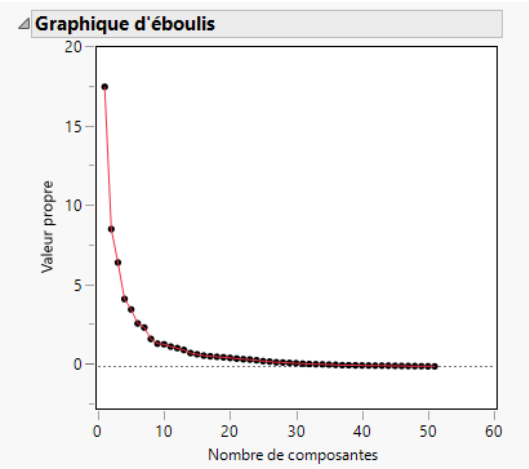
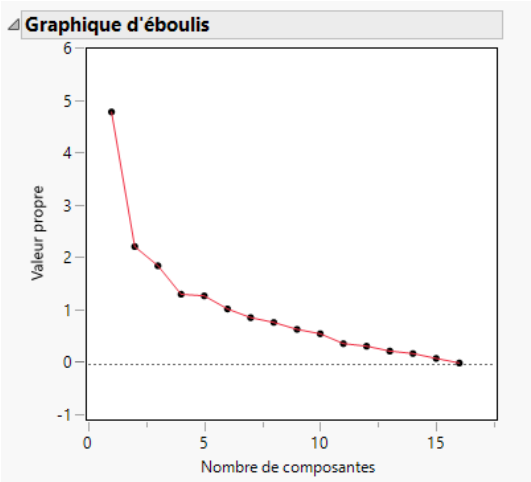
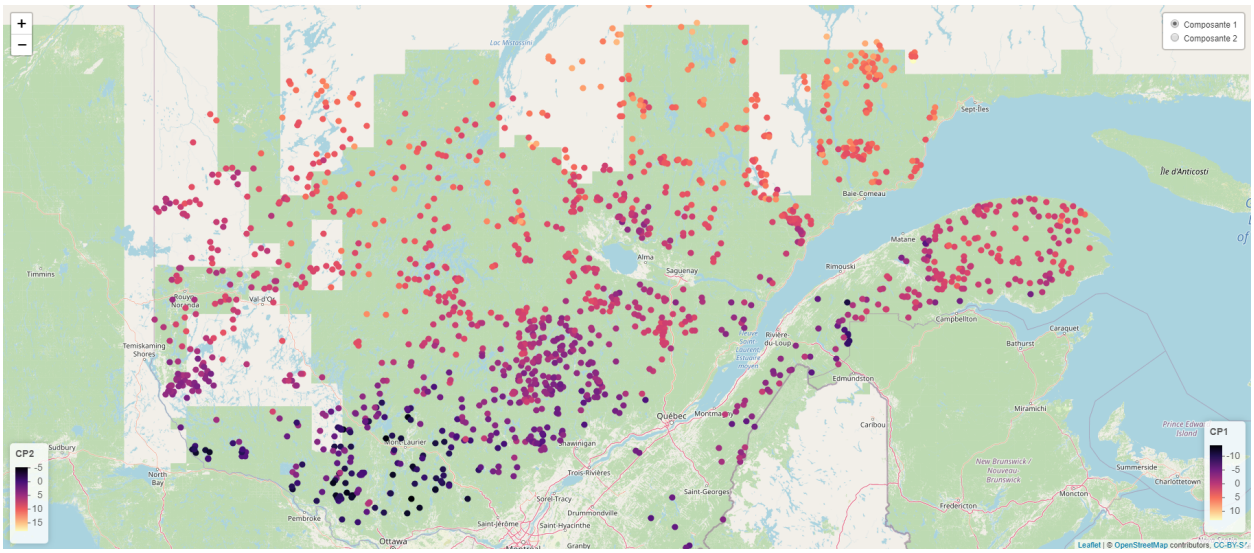


FIGURE 2 – Base de données - *Minimaliste*



Par ailleurs, les valeurs des vecteurs propres représentent les poids accordés aux variables originales dans les composantes principales. Pour la première composante, les observations avec des poids élevés pointent vers le nord de la province, une forte proportion de volume de bois de SEPM ainsi qu'un faible pourcentage de volume désiré des autres feuillus. En visualisant cette composante sur une carte du Québec (Figure 3), il est clair que la représentation des données oppose les secteurs du nord et à ceux du sud.

Figure 3 – Première composante sur la carte du Québec



Pour ce qui est de la deuxième composante, les variables étant plus représentées sont la superficie, le volume total, le volume obligatoire, la superficie en coupe totale, la longueur de chemin à implanter et le volume désiré global. Cette dimension semble ainsi liée à la taille des secteurs, opposant les petits et les grands.

Pour la troisième composante, les enregistrements avec les plus grands poids sont ceux avec de faible volume de bois de bouleau-peuplier et de volume de bois de trituration du même type. Cela paraît cohérent puisque le bouleau et le peuplier sont généralement bien répandus à travers le Québec, peu importe la latitude et la taille du secteur. Par conséquent, pour maximiser la variance, le troisième axe est lié à la quantité de bois de bouleau-peuplier, car ces variables sont moins bien représentées par les deux premières composantes.

Il est important de mentionner que la contribution des variables à la variance des axes est assez bien distribuée et ne dépasse jamais 4% pour le premier axe, 9% pour le deuxième et 12% pour le troisième (Figure 4). Ainsi, les variables mentionnées ci-haut contribuent davantage à la variabilité de chaque composante, mais ne l'expliquent pas complètement.

Par ailleurs, la première composante explique une bonne proportion de la variabilité, selon la qualité de sa représentation. La prudence est donc de mise lors de son interprétation. Cependant, l'ajout des territoires en tant que variable catégorielle supplémentaire permet de constater qu'ils sont particulièrement bien représentés par le premier axe, ce qui vient renforcer la représentation géographique des données qui en est fait.

FIGURE 4 – Contribution partielle de la base de données *Complète*

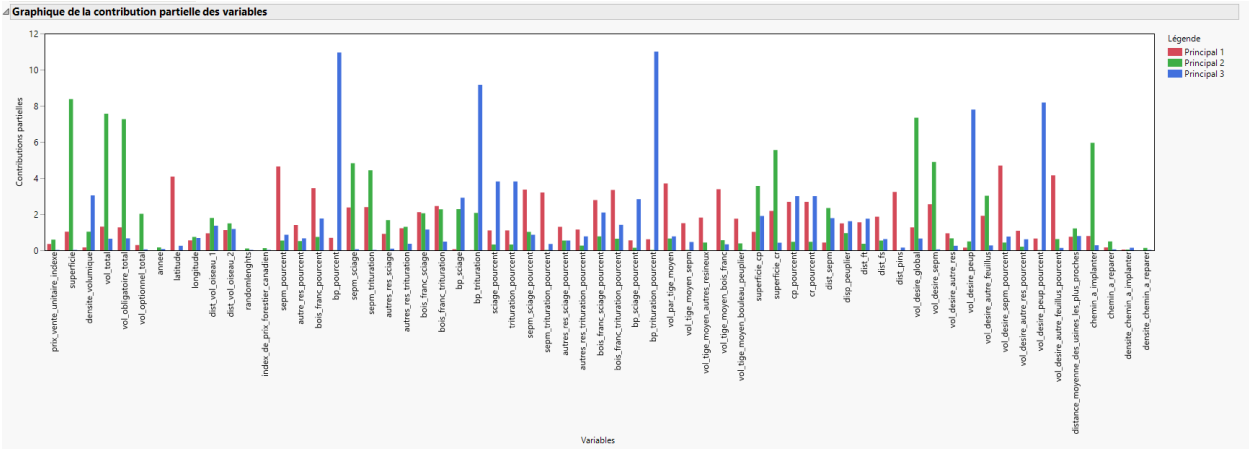
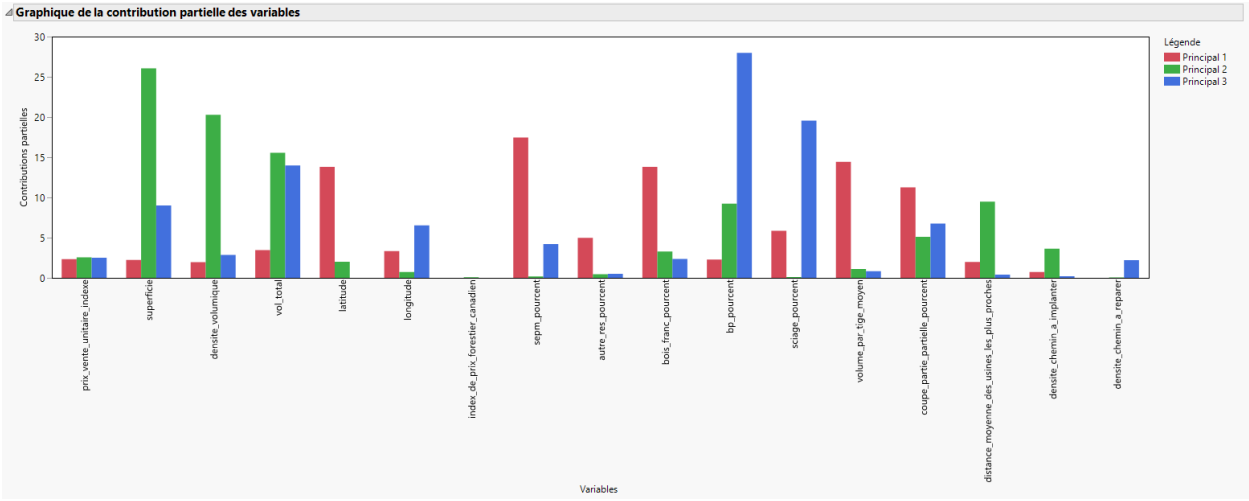


FIGURE 5 – Contribution partielle de la base de données *Minimaliste*



Malgré de légères variations dans les résultats, (Figure 5), nous dégageons les mêmes tendances pour les trois premiers axes de l'analyse sur la base de données *minimaliste*. Ainsi, même avec un nombre de variables réduit, les mêmes relations émergent de l'analyse.

5 Résultats et performance

5.1 Modèles prédictifs

La librairie Scikit-Learn en langage Python [9] sert à la création de nos modèles d'apprentissage. Nous choisissons d'entraîner les modèles à partir de deux bases de données distinctes : la base de données *complète* et celle *minimaliste*. La première contient l'ensemble des 65 variables explicatives alors que la seconde intègre les 18 qui ont étés jugées appropriées selon nos expérimentations (ACP) et notre

jugement (ingénieur forestier dans l'équipe). Pour faciliter l'entraînement, la variable catégorielle du territoire est séparée en deux variables numériques (Nord Sud et Est Ouest). Pour chaque ensemble de données, nous conservons 85% pour l'entraînement et mettons de côté la partie restante pour tester la performance du modèle final. Nous standardisons également les observations pour accélérer la convergence des algorithmes (ex. réseau de neurones).

Le choix des hyperparamètres est réalisé à partir d'une recherche en grille pour la régression linéaire, la forêt aléatoire et le réseau de neurones tandis que la recherche aléatoire est utilisée pour sélectionner les hyperparamètres du SVR. L'espace de recherche est déterminé selon notre connaissance des algorithmes et de la littérature [2]. Les espace de recherche des hyperparamètres sont présentés dans le Tableau 4.

Tableau 4 – Hyperparamètres des modèles pour la recherche (aléatoire ou en grille)

Modèle	**Méthode utilisée**	**Hyperparamètre**	**Valeurs**
Régression linéaire	Recherche en grille	fit_intercept	[True, False]
Forêt aléatoire	Recherche en grille	n_estimators	[10, 50, 100, 500]
		min_samples_leaf	[1, 2, 3]
		max_depth	[50, 500, 5000, None]
Réseau de neurones	Recherche en grille	hidden_layer_sizes	[(100,100,100,100), (110,110,110,110), (120,120,120,120)]
		regularisation L2	[0.005, 0.01, 0.02]
SVR	Recherche aléatoire	kernel	['linear', 'rbf']
		gamma	uniform(0.05, 0.35)
		epsilon	uniform(0.05, 0.35)
		tolérance	uniform(0.001, 0.01)
		regularisation L2	uniform(0.1, 10)

À partir des données d'entraînement, nous réalisons une validation croisée à 5-plis, avec maximisation sur l'indicateur de performance R^2. Cette technique est particulièrement préconisée lorsque le nombre de données disponibles est limité [2], comme dans notre cas. À noter que 50 itérations sont réalisées pour la recherche aléatoire des hyperparamètres du SVR. Les meilleurs hyperparamètres retournés sont illustrés dans le Tableau 5.

Tableau 5 – Hyperparamètres sélectionnés pour chacun des modèles des bases de données

Modèle	**Hyperparamètre**	**Valeur choisie - BD minimaliste**	**Valeur choisie - BD complète**
Régression linéaire	fit_intercept	False	True
Forêt aléatoire	n_estimators	500	500
	min_samples_leaf	2	2
	max_depth	50	50
Réseau de neurones	hidden_layer_sizes	(120, 120, 120, 120)	(110, 110, 110, 110)
	regularisation L2	0.01	0.01
SVR	kernel	rbf	rbf
	gamma	0.065	0.053
	epsilon	0.31	0.066
	tolérance	0.0086	0.0096
	regularisation L2	9.54	8.18

Selon le Tableau 6, la régression linéaire rapporte la pire performance lors de la validation croisée. Cela est cohérent puisque les régressions linéaires multiples sont des modèles généralement plus simples [2]. Enfin, la forêt aléatoire semble prédire le mieux la variable expliquée, à la fois sur la base de données *complète* (R^2 : 0.641 ; R^2 ajusté : 0.514) que sur la base de données *minimaliste* (R^2 : 0.630 ; R^2 ajusté : 0.600).

Selon le principe du rasoir d'Ockham où un modèle plus simple est préférable à un modèle plus complexe à performance égale [2] et tenant compte du R^2 ajusté, notre choix final s'arrête sur la

Tableau 6 – Résultats obtenus en validation croisée pour les quatre modèles d'apprentissage utilisés

Modèle	**Base de données utilisée**	**R^2**	**R^2 ajusté**
Régression linéaire	Complète	0.521	0.352
Forêt aléatoire		**0.641**	**0.514**
Réseau de neurones		0.639	0.512
SVR		0.598	0.456
Régression linéaire	Minimaliste	0.424	0.377
Forêt aléatoire		**0.630**	**0.600**
Réseau de neurones		0.479	0.437
SVR		0.585	0.551

forêt aléatoire entraînée sur la base de données *minimaliste*. C'est donc le modèle que nous évaluons sur notre ensemble de données de test.

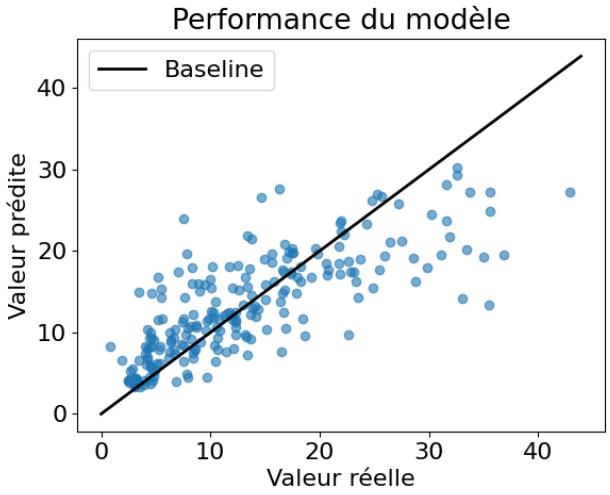
Les résultats obtenus au Tableau 7 montrent que la performance du modèle final en test (R^2 : 0.610 ; R^2 ajusté : 0.570) se rapproche de celle en entraînement (R^2 : 0.630 ; R^2 ajusté : 0.600). Cela suggère qu'il n'y a pas eu de sur-apprentissage. L'erreur absolue moyenne, quant à elle, suggère que notre modèle fait une erreur de prévision de 3.91\$/m^3 en moyenne.

Tableau 7 – Tableau synthèse du modèle final et de sa performance

Modèle	Technique	Forêt aléatoire
	Base de données	Minimaliste
	Nombre de variables explicatives	18
Hyperparamètres	n_estimators	500
	min_samples_leaf	2
	max_depth	50
Performance en test	R^2	0.61
	R^2 ajusté	0.57
	Erreur absolue moyenne	3.91

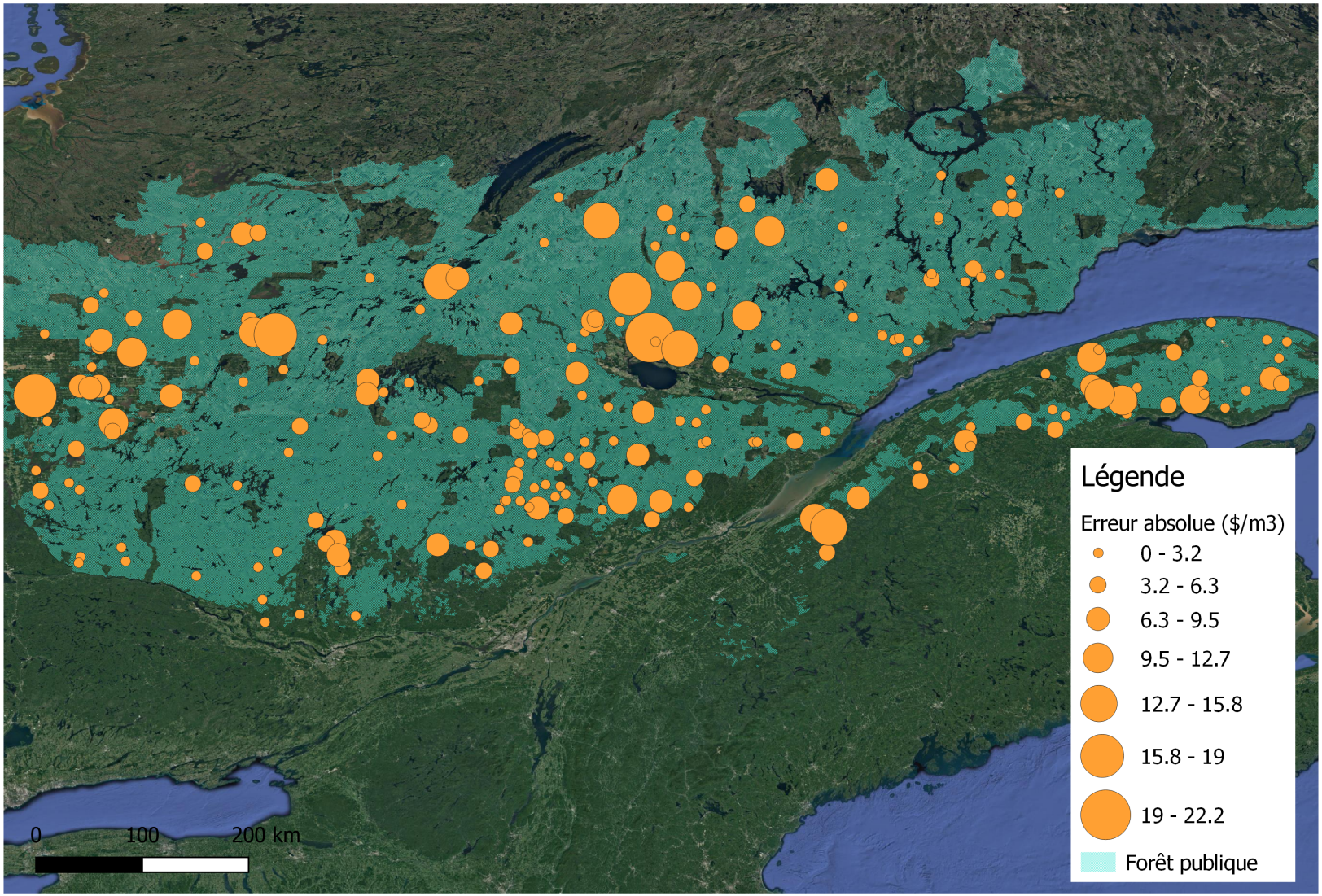
La Figure 6 permet de visualiser les erreurs entre les valeurs prédites et les valeurs réelles. On remarque grâce à ce graphique que le modèle semble systématiquement sous-estimer les valeurs réelles supérieures à 25 \$/m^3.

Figure 6 – Performance du modèle de la forêt aléatoire sur les données de test



La Figure 7, réalisée grâce au logiciel de géomatique *QGIS*, illustre pour sa part les erreurs absolues réparties sur la carte du Québec. Il est possible de noter que les erreurs semblent être aléatoirement réparties sur la carte du Québec. Notons que malgré l'exploration de ces erreurs, nous ne sommes pas en mesure d'identifier précisément les causes derrières celles-ci.

Figure 7 – Erreurs absolues de prédiction sur l'ensemble test



5.2 Importance des variables

Dans les problèmes de prédiction, les variables explicatives sont une part importante de la performance des modèles. L'utilisation de techniques mesurant l'importance des variables permet de mettre en lumière leur contribution individuelle. Par exemple, la technique de permutation permet d'évaluer l'impact de l'absence des variables sur la performance globale du modèle [10][11]. Spécifiquement, cette technique remplace les données d'une variable par des nombres pseudo-aléatoires et évalue la variation de performance selon une métrique désirée. L'intégration de plusieurs réplications permet aussi d'ajouter de la robustesse aux résultats et de réduire l'impact du hasard. Il faut cependant considérer que, dans le cas où deux variables sont corrélées, la permutation peut donner une faible importance à l'une des deux. On ne peut donc pas conclure qu'une variable n'est pas importante pour un modèle. Dans notre analyse, nous utilisons la technique de permutation avec 5 répétitions chacune. Les résultats sont illustrés au Tableau 8. L'importance est exprimée en diminution du R^2 lors de la permutation ($-\Delta R^2$).

Tableau 8 – Importance des variables par permutation de nombres pseudo-aléatoires

Variable	**Importance ($-\Delta R^2$)**	**Écart type**
SEPM pourcent	0.3174	0.021
UA	0.2713	0.0179
Distance moyenne des usines les plus proches	0.1947	0.046
Longitude	0.0737	0.0122
Index de prix forestier canadien	0.0737	0.0253
Latitude	0.0726	0.0111
Sciage pourcent	0.0462	0.0158
Densite volumique	0.0215	0.0209
Densite chemin a implanter	0.019	0.0065
Vol par tige moyen	0.0161	0.0052
Bois franc pourcent	0.0119	0.0019
Densite chemin a reparer	0.0108	0.0039
Coupe partielle pourcent	0.0082	0.0027
Superficie	0.0053	0.0018
Autre res pourcent	0.0039	0.0005
BP pourcent	0.0032	0.0031
Vol total	0.0023	0.0033
Nord Sud	0.0015	0.0009
Est Ouest	-0.0003	0.0005

Selon le tableau, les deux variables avec le plus d'importance sont le pourcentage de bois SEPM dans le secteur de la vente (31.7%) et l'unité d'aménagement où il est situé (27.1%). Autrement dit, les essences destinées aux scieries spécialisées dans le bois d'oeuvre ainsi que la localisation du secteur ont le plus d'importance sur la performance du modèle. Par la suite, la distance moyenne de route entre le point de récolte et l'usine de transformation obtient la troisième place en ordre d'importance (19.4%). Le coût de transport du bois augmentant avec la distance, il est raisonnable de penser que cette variable ait un effet sur la performance du modèle vu son impact sur les coûts de transport de bois. La longitude, la latitude, l'index des prix forestier canadien semblent aussi contribuer positivement à la performance du modèle.

Il est pertinent de constater que plusieurs variables ne sont pas très utiles au modèle telles que les

variables catégorielles du territoire «Nord-sud» et «Est-Ouest». Les autres variables de localisation géographiques (i.e. UA, latitude et longitude) contiennent probablement des informations similaires. Ce résultat montre en partie que des variations du prix s'opèrent à une échelle plus locale que globale. Par ailleurs, la densité volumique (2.2%), la superficie de vente (0.5%), le volume total de bois en vente (0.2%) ont un impact plus faible malgré leur usage courant en terme d'indicateur dans le monde forestier.

À titre indicatif, d'autres techniques d'importance visant le même objectif auraient pu être utilisées (ex. *forward regression* et *backward regression*).

6 Discussion et limites

L'objectif du projet consiste à évaluer le potentiel des méthodes d'apprentissage automatique dans le cadre des prédictions des prix de ventes des secteurs de récoltes forestières. Étant donné que la performance en validation des forêts aléatoires explique 22.3% plus de variance que la régression linéaire multiple, nous concluons que les modèles basés sur l'apprentissage automatique pourraient offrir des gains importants dans la prédictions des prix. Même si la base de données utilisée par le BMMB diffère de la nôtre, nous sommes d'avis que l'adoption de telles méthodes amélioreraient leur performance de prédiction.

Malgré l'intégration de plusieurs variables explicatives, près de 40% de la variance n'est pas expliquée par notre modèle final (Tableau 9). Il semble que d'autres variables, absentes de nos bases données, influencent le prix de vente. Certains facteurs affectant le prix misé peuvent être intangibles comme les aspects humains (ex. personnalité, stress) ou encore la situation de l'usine au moment de l'enchère : Essaie-t-elle de combler un déficit de bois ou doit-t-elle gérer des imprévus ? Comme illustré à la Figure 6, le modèle a tendance à sous-estimer la valeur des secteurs qui sont vendus à des prix très élevés. Cela vient appuyer les points précédents à savoir que d'autres informations, contribuant à expliquer le prix de vente, sont absentes.

Selon nos résultats sur l'importance des variables, le volume moyen par tige rapporte une valeur d'importance inférieure à 2%. Compte tenu de la supposée importance de cet indicateur dans le domaine forestier, nous soulevons la question suivante : Est-ce que les industriels valorisent réellement les bois de plus grandes dimensions pour créer des produits à plus grande valeur, ou sont-ils simplement à la recherche de volume de bois pour répondre aux besoins de l'usine ? Dans la même lignée que le paragraphe précédent, d'autres phénomènes expliquent probablement le prix de vente.

Modèle	**Méthode utilisée**	**R^2 ajusté**	
BMMB	Régression linéaire	Modèle SEP	Modèle autres essences
		0.64	0.67
Équipe verte	Forêt aléatoire	Modèle final	
		0.57	

Tableau 9 – Comparaison entre les modèles des différentes équipes

Un enjeu soulevé par le BMMB réside dans l'explicabilité du modèle aux différents industriels. La confiance envers le modèle dépend en partie de la compréhension de la méthode utilisée pour le créer. La régression linéaire multiple, qui est présentement utilisée, est relativement facile à comprendre. Elle est toutefois critiquée lorsque les prix des secteurs sont trop élevés, d'où le défi de mettre en

confiance les industriels. Les méthodes d'apprentissage automatique, ayant tout récemment fait leur apparition dans le monde forestier, sont méconnues des ingénieurs forestiers. L'aspect de la « boîte noire » peut aussi contribuer à cette insécurité et incompréhension. Cependant, puisque que les industriels bénéficieraient de l'incorporation de meilleurs modèles de prévisions des prix, les gains possibles provenant de l'apprentissage automatique pourraient les motiver à apprivoiser ces techniques. Il est également important de noter que certaines techniques d'apprentissage automatique, telles que la forêt aléatoire, sont plus facilement explicables que d'autres, comme les réseaux de neurones. Nous croyons également que l'utilisation de la mesure de performance d'erreur absolue moyenne (MAE) permettrait probablement d'ajouter plus de transparence et d'acceptabilité que la mesure du R^2.

De plus, le nombre restreint d'observations limite les performances des modèles d'apprentissage automatique qui profitent généralement de l'abondance de données pour capter des interactions plus complexes. Ainsi, avec un nombre croissant d'enchères à chaque année, la performance des modèles devrait aussi croître, toutes choses étant égales par ailleurs.

7 Conclusion

Les techniques d'apprentissage automatique plus complexes, comparativement à la régression linéaire, permettent d'estimer plus précisément les prix de ventes des secteurs forestiers. Selon les résultats obtenus, nous recommandons au BMMB d'utiliser la forêt aléatoire pour résoudre ce problème de régression. L'utilisation de la mesure de performance MAE pourrait également contribuer à l'acceptabilité de ce changement. Il faut également souligner qu'une bonne partie de la variance est non-expliquée. Les processus de décisions reposant sur l'humain, comme les enchères, peuvent apporter leur lot de biais et de défis. Malgré tout, l'apprentissage automatique reste un domaine d'avenir pour l'industrie forestière. Le gouvernement du Québec, tout comme les industriels, auront avantage à s'y intéresser.

8 Remerciements

Nous tenons à sincèrement remercier M. François Bélanger ing.f., M. Gabriel Lévesque-Lessard et M. Hugo Therrien ing.f., du BMMB, pour leur temps, leur soutien et leur collaboration dans le cadre de ce projet.

Références

[1] MFFP. *Ressources et industries forestières du Québec : Portrait statistique 2019*. 2019. URL https://mffp.gouv.qc.ca/wp-content/uploads/PortraitStatistique_2019.pdf.

[2] Ethem Alpaydin. *Introduction to Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, 4th edition, 2020.

[3] Banque du Canada. *Commodity Price Index*. 2021. URL https://www.bankofcanada.ca/rates/price-indexes/bcpi/.

[4] MarketWatch. *Random Length Lumber (CME $/bft) Front Month*. 2021. URL https://www.marketwatch.com/investing/future/lb1/download-data?startDate=1/3/2013&endDate=1/3/2014.

[5] Données Québec. *Usine de transformation primaire du bois - Jeux de données*. 2021. URL https://www.donneesquebec.ca/recherche/dataset/usines-de-transformation-primaire-du-bois.

[6] Google. *Google Maps Plateform*. 2021. URL https://developers.google.com/maps.

[7] Banque du Canada. *Inflation calculator*. 2021. URL https://www.bankofcanada.ca/rates/related/inflation-calculator/.

[8] Raymond B. Cattell. The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2) :245–276, 1966. doi : 10.1207/s15327906mbr0102_10. URL https://doi.org/10.1207/s15327906mbr0102_10. PMID : 26828106.

[9] Scikit-Learn. *Scikit-Learn : Machine Learning in Python*. 2021. URL https://scikit-learn.org/stable/modules/permutation_importance.html.

[10] Leo Breiman. Random forests. *Mach Learn.*, 45 :5–33, 2001.

[11] Scikit-Learn. *Permutation feature importance*. 2021. URL https://scikit-learn.org/stable/modules/permutation_importance.html.

[12] MFFP. *Guide explicatif du modèle de transposition des prix de marché*. 2019. URL https://bmmb.gouv.qc.ca/media/61604/transposition_sep_2020-2021.pdf.

[13] MFFP. *Cartes des unités d'aménagement – Période 2008-2013*. 2012. URL https://mffp.gouv.qc.ca/forets/amenagement/CAAF-pdf/carte/uaf_pgaf_PQ.pdf.

A Annexe

FIGURE 8 – Analyses descriptives

```
      territoire  prix_vente_unitaire_indexe       ua        superficie     densite_volumique   vol_total      vol_obligatoire_total vol_optionnel_total     annee        latitude       longitude      dist_vol_oiseau_1
Centre-Nord:270   Min.   : 0.82             9351   :  84   Min.   :  17.0   Min.   : 28.98   Min.   :   840   Min.   :   840       Min.   :    0       Min.   :2011   Min.   :45.28   Min.   :-79.51   Min.   :0.0200
Centre-Sud :331   1st Qu.: 5.52             4151   :  82   1st Qu.: 154.0   1st Qu.: 84.29   1st Qu.: 16150   1st Qu.: 15550       1st Qu.:    0       1st Qu.:2015   1st Qu.:47.34   1st Qu.:-75.51   1st Qu.:0.3000
Nord-Est  :195    Median :10.67             9751   :  73   Median : 266.0   Median :102.51   Median : 27700   Median : 26800       Median :    0       Median :2016   Median :48.29   Median :-72.97   Median :0.4800
Nord-Ouest :238   Mean   :12.86             4352   :  72   Mean   : 361.4   Mean   :107.26   Mean   : 36063   Mean   : 34381       Mean   : 1683       Mean   :2016   Mean   :48.31   Mean   :-72.68   Mean   :0.5698
Sud-Est   :201   3rd Qu.:18.22             4251   :  60   3rd Qu.: 458.0   3rd Qu.:126.03   3rd Qu.: 47250   3rd Qu.: 44710       3rd Qu.: 1010       3rd Qu.:2018   3rd Qu.:49.20   3rd Qu.:-69.78   3rd Qu.:0.7300
Sud-Ouest :258    Max.   :49.91             2571   :  50   Max.   :3265.0   Max.   :233.96   Max.   :261000   Max.   :261000       Max.   :47460       Max.   :2021   Max.   :51.20   Max.   :-64.55   Max.   :2.3300
                                            (Other):1072

dist_vol_oiseau_2 randomlenghts   index_de_prix_forestier_canadien sepm_pourcent    autre_res_pourcent bois_franc_pourcent  bp_pourcent       sepm_sciage     sepm_trituration autres_res_sciage autres_res_trituration
Min.   :0.0700    Min.   :226.8   Min.   :302.3                   Min.   :0.0000   Min.   :0.00000    Min.   :0.000000    Min.   :0.00000   Min.   :     0   Min.   :    0   Min.   :    0.0   Min.   :   0.0
1st Qu.:0.4500    1st Qu.:299.8   1st Qu.:349.6                   1st Qu.:0.5322   1st Qu.:0.00000    1st Qu.:0.000000    1st Qu.:0.05081   1st Qu.:  6700   1st Qu.:  2450   1st Qu.:    0.0   1st Qu.:   0.0
Median :0.6300    Median :346.8   Median :375.9                   Median :0.7976   Median :0.00000    Median :0.003215    Median :0.12476   Median : 13150   Median :  5800   Median :    0.0   Median :   0.0
Mean   :0.7095    Mean   :376.6   Mean   :399.8                   Mean   :0.7113   Mean   :0.02073    Mean   :0.089663    Mean   :0.17827   Mean   : 17780   Mean   :  9678   Mean   :  421.6   Mean   : 100.9
3rd Qu.:0.9100    3rd Qu.:390.5   3rd Qu.:423.6                   3rd Qu.:0.9306   3rd Qu.:0.01091    3rd Qu.:0.083082    3rd Qu.:0.26034   3rd Qu.: 23350   3rd Qu.: 12300   3rd Qu.:  150.0   3rd Qu.:  50.0
Max.   :2.4900    Max.   :959.9   Max.   :675.4                   Max.   :1.0000   Max.   :0.92344    Max.   :0.990741    Max.   :0.82812   Max.   :156500   Max.   :129800   Max.   :15150.0   Max.   :3200.0

bois_franc_sciage bois_franc_trituration   bp_sciage       bp_trituration  sciage_pourcent   trituration_pourcent sepm_sciage_pourcent sepm_trituration_pourcent autres_res_sciage_pourcent autres_res_trituration_pourcent
Min.   :    0.0   Min.   :    0         Min.   :    0.0   Min.   :    0   Min.   :0.06481   Min.   :0.02425      Min.   :0.0000       Min.   :0.0000            Min.   :0.000000           Min.   :0.000000
1st Qu.:    0.0   1st Qu.:    0         1st Qu.:   50.0   1st Qu.: 1050   1st Qu.:0.45863   1st Qu.:0.38049      1st Qu.:0.3792       1st Qu.:0.1267            1st Qu.:0.000000           1st Qu.:0.000000
Median :   10.0   Median :   60         Median :  250.0   Median : 3050   Median :0.53586   Median :0.46414      Median :0.4926       Median :0.2292            Median :0.000000           Median :0.000000
Mean   :  573.3   Mean   : 1462         Mean   :  632.4   Mean   : 5415   Mean   :0.53249   Mean   :0.46751      Mean   :0.4733       Mean   :0.2381            Mean   :0.016111           Mean   :0.004617
3rd Qu.:  300.0   3rd Qu.: 1550         3rd Qu.:  750.0   3rd Qu.: 7250   3rd Qu.:0.61951   3rd Qu.:0.54137      3rd Qu.:0.5990       3rd Qu.:0.3384            3rd Qu.:0.008028           3rd Qu.:0.002003
Max.   :16050.0   Max.   :33300         Max.   :13200.0   Max.   :88750   Max.   :0.97575   Max.   :0.93519      Max.   :0.9664       Max.   :0.6260            Max.   :0.622010           Max.   :0.301435

bois_franc_sciage_pourcent bois_franc_trituration_pourcent bp_sciage_pourcent bp_trituration_pourcent vol_par_tige_moyen vol_tige_moyen_sepm vol_tige_moyen_autres_resineux vol_tige_moyen_bois_franc
Min.   :0.0000000          Min.   :0.000000                Min.   :0.000000   Min.   :0.0000          Min.   :  16.52   Min.   :   0.0     Min.   :   0.0                Min.   :   0.0
1st Qu.:0.0000000          1st Qu.:0.000000                1st Qu.:0.002457   1st Qu.:0.0453          1st Qu.: 157.90   1st Qu.: 154.6     1st Qu.:   0.0                1st Qu.:   0.0
Median :0.0000445          Median :0.002682                Median :0.009155   Median :0.1105          Median : 216.56   Median : 182.6     Median :   0.0                Median : 113.5
Mean   :0.0251784          Mean   :0.064485                Mean   :0.017935   Mean   :0.1603          Mean   : 291.34   Mean   : 225.3     Mean   :  387.8               Mean   : 219.6
3rd Qu.:0.0126659          3rd Qu.:0.065872                3rd Qu.:0.023837   3rd Qu.:0.2310          3rd Qu.: 347.39   3rd Qu.: 267.0     3rd Qu.: 531.0                3rd Qu.: 340.0
Max.   :0.5373728          Max.   :0.935185                Max.   :0.261818   Max.   :0.8047          Max.   :1515.15   Max.   :1143.5     Max.   :6610.0                Max.   :1863.5

vol_tige_moyen_bouleau_peuplier superficie_cp    superficie_cr     cp_pourcent       cr_pourcent     secteur_contrainte   dist_sepm     disp_peuplier      dist_ft        dist_fs        dist_pins     vol_desire_global
Min.   :   0.0                  Min.   :   0.00   Min.   :   0.00   Min.   :0.00000   Min.   :0.0000   0:1461               Min.   : 16.80   Min.   : 26.21   Min.   : 11.0   Min.   : 14.22   Min.   : 16.63   Min.   :   840
1st Qu.: 169.2                  1st Qu.:   2.99   1st Qu.:  97.88   1st Qu.:0.01752   1st Qu.:0.7054   1:  32               1st Qu.: 85.75   1st Qu.:107.28   1st Qu.:111.0   1st Qu.: 99.89   1st Qu.:274.56   1st Qu.: 15500
Median : 229.0                  Median :  25.91   Median : 205.36   Median :0.07701   Median :0.9230                        Median :124.99   Median :158.51   Median :168.0   Median :167.71   Median :429.41   Median : 26360
Mean   : 304.1                  Mean   :  70.90   Mean   : 290.49   Mean   :0.22193   Mean   :0.7781                        Mean   :134.89   Mean   :180.32   Mean   :206.3   Mean   :225.76   Mean   :416.38   Mean   : 33959
3rd Qu.: 382.5                  3rd Qu.:  87.89   3rd Qu.: 368.56   3rd Qu.:0.29458   3rd Qu.:0.9825                        3rd Qu.:161.55   3rd Qu.:238.01   3rd Qu.:272.0   3rd Qu.:331.75   3rd Qu.:551.46   3rd Qu.: 44060
Max.   :3681.0                  Max.   :1005.14   Max.   :2748.06   Max.   :1.00000   Max.   :1.0000                        Max.   :371.36   Max.   :539.99   Max.   :690.0   Max.   :713.52   Max.   :863.97   Max.   :256800

vol_desire_sepm  vol_desire_autre_res vol_desire_peup vol_desire_autre_feuillus vol_desire_sepm_pourcent vol_desire_autre_res_pourcent vol_desire_peup_pourcent vol_desire_autre_feuillus_pourcent
Min.   :     0   Min.   :    0.0      Min.   :    0   Min.   :    0             Min.   :0.0000           Min.   :0.000000             Min.   :0.000000         Min.   :0.00000
1st Qu.:  9500   1st Qu.:    0.0      1st Qu.:    0   1st Qu.:    0             1st Qu.:0.5505           1st Qu.:0.000000             1st Qu.:0.000000         1st Qu.:0.00000
Median : 18890   Median :    0.0      Median :  110   Median : 1770             Median :0.8436           Median :0.000000             Median :0.008093         Median :0.07834
Mean   : 26959   Mean   :  335.4      Mean   : 2356   Mean   : 4309             Mean   :0.7406           Mean   :0.016052             Mean   :0.077843         Mean   :0.16546
3rd Qu.: 34900   3rd Qu.:  100.0      3rd Qu.: 2500   3rd Qu.: 5880             3rd Qu.:0.9941           3rd Qu.:0.004293             3rd Qu.:0.110763         3rd Qu.:0.25293
Max.   :256800   Max.   :15450.0      Max.   :57800   Max.   :56070             Max.   :1.0000           Max.   :0.923445             Max.   :0.739826         Max.   :0.99074

distance_moyenne_des_usines_les_plus_proches chemin_a_implanter chemin_a_reparer  densite_chemin_a_implanter densite_chemin_a_reparer
Min.   :  6.735                              Min.   :  0.000    Min.   :  0.000   Min.   :0.0000000          Min.   :0.0000000
1st Qu.: 56.282                              1st Qu.:  3.837    1st Qu.:  0.000   1st Qu.:0.0001844          1st Qu.:0.0000000
Median : 87.833                              Median :  7.500    Median :  1.382   Median :0.0002725          Median :0.0000491
Mean   : 98.237                              Mean   : 10.339    Mean   :  5.038   Mean   :0.0002998          Mean   :0.0001654
3rd Qu.:126.723                              3rd Qu.: 13.550    3rd Qu.:  6.349   3rd Qu.:0.0003813          3rd Qu.:0.0002336
Max.   :328.259                              Max.   :103.000    Max.   :144.550   Max.   :0.0025984          Max.   :0.0034632
```

Figure 9 – Distribution des variables

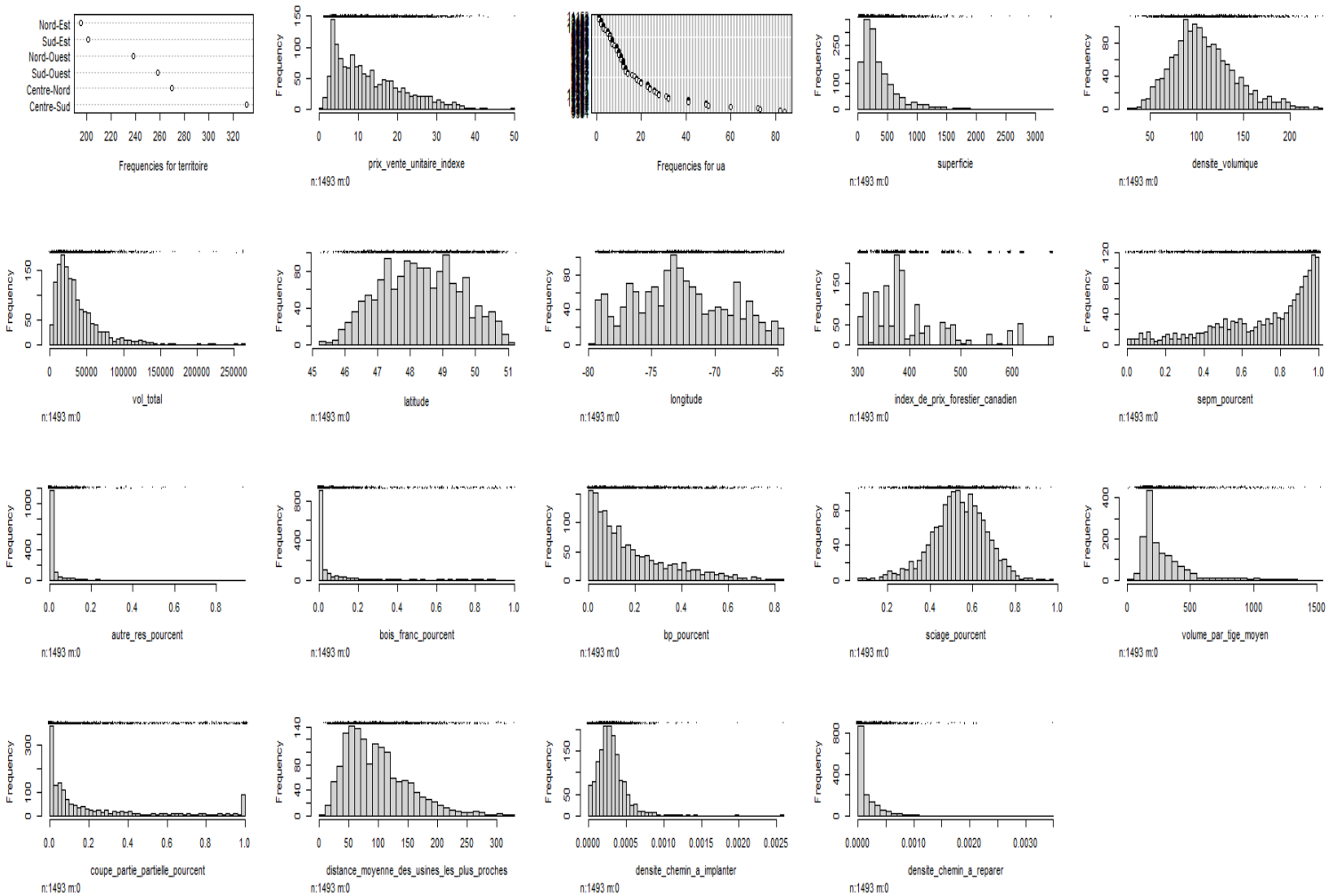


Tableau 10 – Variables explicatives et coefficient du modèle de régressions SEP du BMMB [12]

Variable	Unité	Description
Nombre de mises	Nb discret	Nombre de mises conformes reçues sur le secteur vendu.
Revenus nets usine	$/m³	Revenus nets moyens (qualité B, C et M) du modèle usine calculés pour chaque secteur à l'aide de la distribution de tiges, selon les prix PRIBEC trois mois précédant le mois de publication de l'appel d'offres.
Coûts de transport aux cinq usines et aux marchés	$/m³	Coûts de transport moyens pondérés aux 5 usines de sciage de SEPM les plus proches, plus le coût de transport moyen des 5 usines aux marchés.
Nombre de fois que le secteur a été mis en vente	Nb discret	Tient compte du fait que les secteurs ne trouvent pas toujours preneur lors d'un premier affichage.
Proportion SEPM	%	Proportion de SEPM (bois d'œuvre et trituration) sur le volume total dans le secteur.
Autres coûts	$/m³	Autres coûts pondérés selon les volumes de récolte du secteur : hébergement, camp et administration.
Coûts de chemin	$/m³	Coûts de construction de chemin, de réfection, d'entretien et des infrastructures extraordinaires (ex. : pont) diminués des sommes reçues dans le cadre du Programme de remboursement des coûts de chemins multiressources.
Housing market index (Indice de mise en chantier)	Indice positif	Indice de prévision des mises en chantier aux États-Unis pour les états du Nord-Est fourni par la National Association of Home Builders (NAHB).
Territoire structuré	Indice entre 0 et 7	Indice marquant la présence d'une ou de plusieurs structures d'organisation du territoire forestier.
Vol/ha	m³/ha	Volume désiré à l'hectare du secteur.
Difficulté terrain	%	Moyenne pondérée des pourcentages de correction de chaque machine en fonction de leur poids de coût. Cette correction est en fonction de la pente, du drainage et de la rugosité.
Volume aux permis par superficie de forêt publique	m³/km²	Total du volume aux permis des usines de sciage d'une région d'attribution divisé par la superficie en kilomètre carré de la forêt publique de cette même région.
Distance de transport aux 3ᵉ, 4ᵉ, 5ᵉ usines	km	Distance moyenne pondérée aux usines de sciage SEPM et ce, pour les 3ᵉ, 4ᵉ et 5ᵉ usines les plus proches.
Superficie	ha	Superficie totale du secteur en hectare.
Proportion de pin gris	%	Proportion de pin gris (bois d'œuvre et trituration) sur le volume SEPM dans le secteur.
Proportion de sapin	%	Proportion de sapin baumier (bois d'œuvre et trituration) sur le volume SEPM dans le secteur.

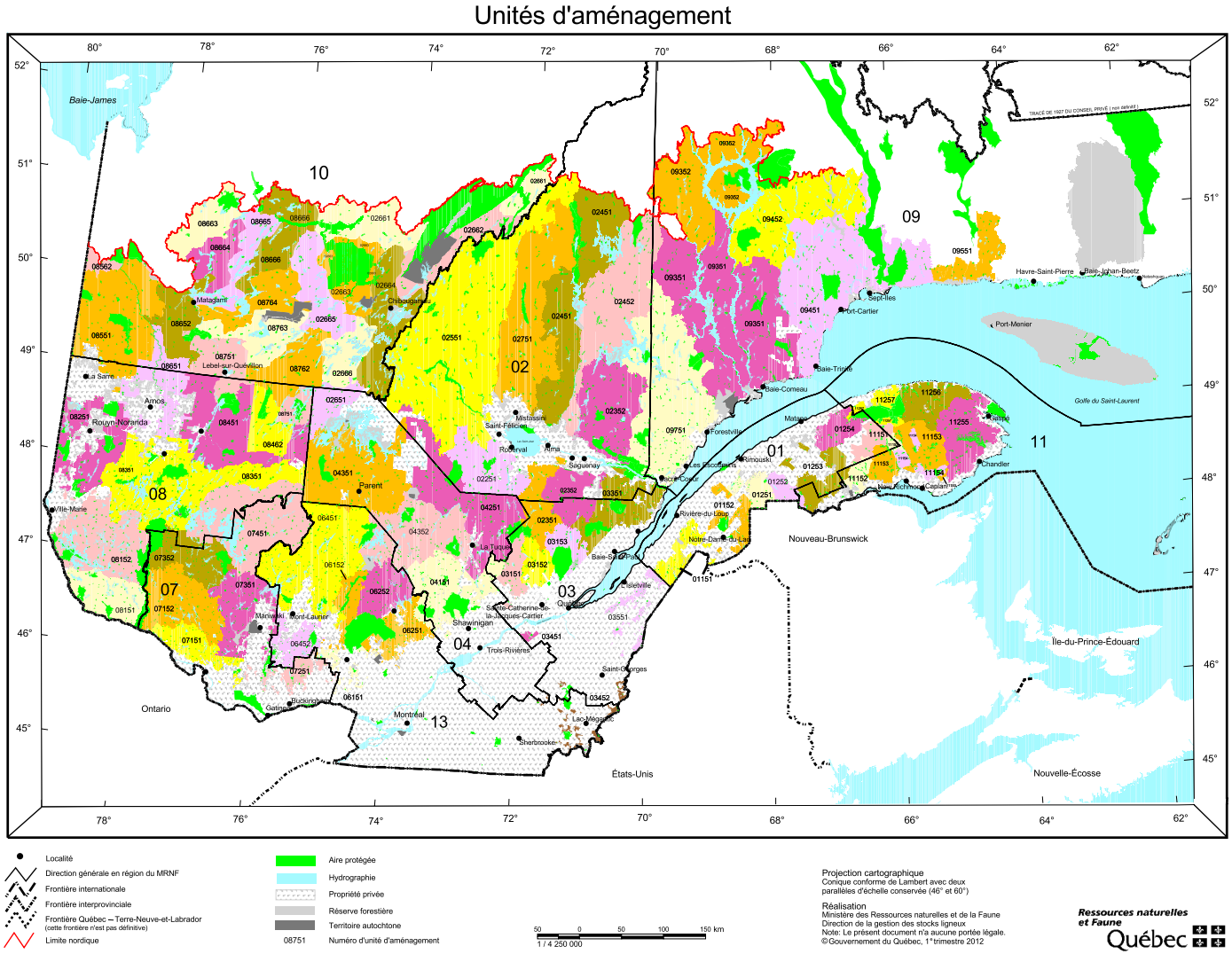



FIGURE 10 – Carte des unités d'aménagements du Québec [13]