

La forêt est une ressource économique dont l'exploitation contribue à l'enrichissement collectif. L'enjeu de l'utilisation de toute ressource naturelle renouvelable est évidemment d'en faire une gestion qui ne compromet pas l'intégrité du milieu. Comment y parvient-on ?

Aspects statistiques de

Louis-Paul Rivest
Université Laval
Pierre Bernier
Centre de foresterie
des Laurentides



Copyright © Ministère des Ressources naturelles et de la Faune du Québec

La forêt est aménagée pour la production de bois, selon les principes de rendement soutenu et de développement durable qui nécessitent une prise en compte de l'impact de l'aménagement sur un large ensemble de valeurs. Un tel exercice implique un regard à des échelles spatiales allant du peuplement à celui du territoire dans lequel se situe ce peuplement. Les secteurs impliqués sont aussi multiples, allant de l'infrastructure industrielle dépendante des marchés par le jeu de l'offre et de la demande, aux enjeux sociaux-économiques liés au maintien d'habitat et de paysages en fonction du récréo-tourisme et de l'utilisation traditionnelle du territoire.

Pour planifier de telles opérations, il faut donc bien connaître la forêt dans le territoire sous aménagement. Une partie essentielle de

cette information comprend la composition, l'âge, la densité et la répartition spatiale des peuplements forestiers. L'inventaire forestier permet de recueillir ces informations.

L'inventaire forestier

L'objectif d'un inventaire forestier est d'estimer les caractéristiques de la forêt dans un territoire donné que nous appellerons unité d'aménagement forestier (UAF). Au Québec, les UAF ont des tailles allant de moins de 150 km² à plus de 20 000 km². Nous nous intéresserons ici pour fins d'exemple à deux variables, le volume marchand d'essences feuillues et le volume marchand d'essences résineuses exprimés en m³ par hectare. Le volume marchand est le volume de bois dans les tiges dont le diamètre à une hauteur de

1.3m dépasse une valeur minimale en bas de laquelle les tiges sont considérées comme trop petites pour une utilisation industrielle. Pour déterminer exactement ces deux volumes dans une UAF, il faudrait, en théorie, mesurer précisément le volume de tous les arbres de l'UAF, une tâche évidemment irréaliste. C'est là qu'interviennent les techniques d'échan-

m³ par hectare. Ensuite vient le problème de l'interpolation : il faut prédire les deux volumes d'intérêt dans les parties de l'UAF qui n'ont pas été visitées par les forestiers. Plusieurs méthodes statistiques sont disponibles pour faire ce travail. Nous nous intéresserons ici plus particulièrement à des méthodes de type « plus proche voisin » qui sont

souvent présentées en utilisant l'acronyme anglais « kNN » pour « k nearest neighbors ».

L'inventaire forestier au Québec

Au Québec, la forêt exploitée à des fins commerciales couvre près de 600 000 km² au sud du 52^e parallèle comme illustré dans la figure ci-dessous. On y récolte annuellement environ 40 millions de m³ de bois. Le Québec utilise un inventaire forestier décennal qui couvre annuellement environ 10% de la forêt exploitée. L'inventaire québécois utilise la photo-

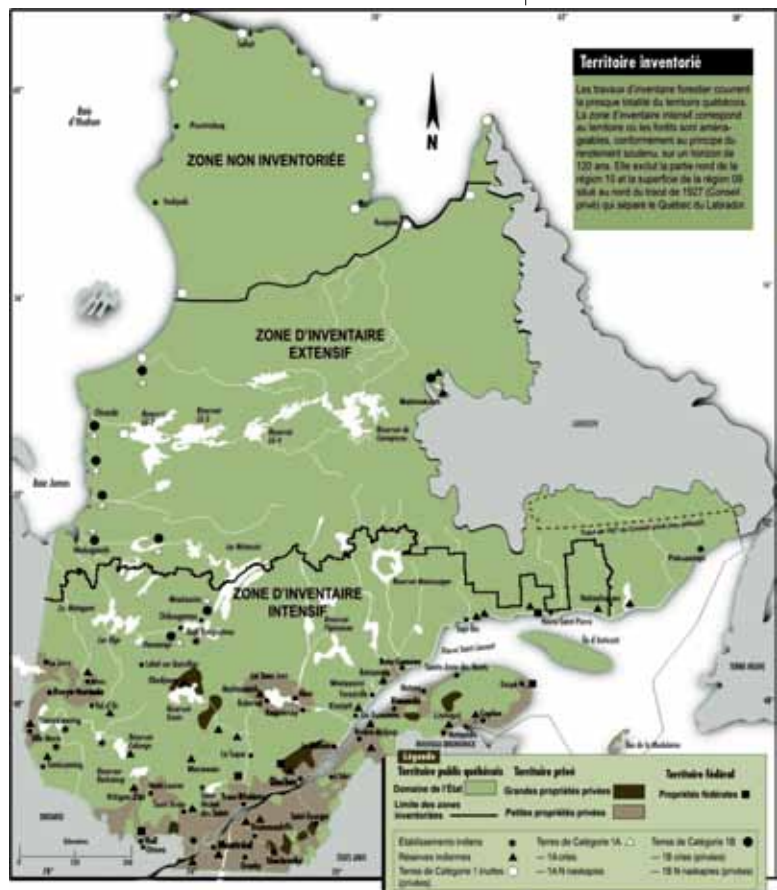
La zone d'inventaire intensif de la forêt québécoise se trouve au sud du 52^{ème} parallèle.

la gestion forestière

tillonnage statistique : on détermine dans un premier temps les deux volumes dans une portion de l'UAF judicieusement sélectionnée. On extrapole ensuite les résultats à l'UAF toute entière.

En pratique un inventaire forestier est réalisé en deux étapes. On mesure d'abord des variables associées aux volumes d'intérêt dans l'ensemble du territoire à l'étude. Pour ce faire on peut par exemple utiliser des photographies aériennes à haute résolution. En étudiant attentivement les photos, des analystes expérimentés découpent une UAF en polygones forestiers homogènes. Sans aller sur le terrain, ils peuvent estimer la hauteur des arbres du polygone, le pourcentage de la surface occupé par les feuillus et le pourcentage occupé par des résineux. L'interprétation de photo aérienne est la méthode classique de réalisation de la première étape d'un inventaire. Dans un avenir rapproché l'UAF pourrait plutôt être découpée en « pixels » forestiers correspondant à l'aire couverte par un des « points » de l'image numérique. Des variables telles que l'intensité de la réflectance des bandes spectrales dans chaque pixel, ou la variation spatiale de la réflectance entre les pixels voisins extraites de ces images, pourraient alors être associées aux variables forestières désirées telles que les volumes de bois présents dans le pixel ou dans le groupe de pixels forestiers.

Pour la deuxième étape il faut sélectionner au hasard des placettes dans le territoire et envoyer dans chacune une équipe d'inventaire qui, en mesurant tous les arbres qui s'y trouvent, évaluent les volumes des essences résineuses et de feuillues, en



La carte et les deux figures de l'article ont été gracieusement fournies par le Ministère des Ressources naturelles et de la Faune du Québec

Une photo interprétation réalisée pour un inventaire forestier québécois.

Les lignes blanches délimitent les polygones.

Chaque polygone a un nom qui résume l'information extraite de la photo.



Copyright © Ministère des Ressources naturelles et de la Faune du Québec

interprétation faite à partir de photos prises à l'échelle 1/15 000 pour délimiter des polygones forestiers homogènes d'une superficie de quelques hectares. La figure accompagnant le titre de cet article présente, de façon très simplifiée, la construction des polygones forestiers à partir d'une photo où le cartographe a déterminé les frontières entre des peuplements homogènes. En pratique la photo-interprétation est une opération beaucoup plus complexe impliquant l'identification et la codification d'un nombre important de variables à partir de paires de photos vues en stéréoscopie. La figure ci-haut donne une vraie photo-interprétation réalisée lors d'un inventaire. Chaque polygone reçoit un nom qui résume l'information recueillie par le photo-interprète.

Estimation par la méthode du plus proche voisin : Un exemple

En pratique de grands ensembles de données sont récoltés aux deux étapes d'un inventaire. Evidemment une UAF normale contient des dizaines de milliers de polygones. À des fins d'illustration nous considérons une UAF fictive, de forme carrée, qui se divise en 25 unités homogènes que nous nommerons ici « pixels » car ils sont décrits par leurs propriétés

spectrales et non pas par photo-interprétation. Cette UAF est représentée dans le tableau ci-dessous. Deux mesures ont été prises sur les 25 pixels de la photo, l'intensité de rouge (x_1) et de bleu (x_2). Les valeurs de ces deux variables pour les 25 pixels de la photo sont données dans le tableau de la page suivante. Les trois premières colonnes de ce tableau représentent l'information à la fin de la première étape de l'inventaire.

Unité d'aménagement forestier (UAF) contenant 25 pixels forestiers

1	2E	3	4	5E
6	7	8	9	11
11	12	13E	14	15
16	17E	18	19E	30
21	22	23	24	25

A l'étape 2, on tire 5 des 25 pixels de la forêt au hasard ; les pixels tirés sont identifiés par la lettre E dans le tableau précédent. Ils portent les numéros 2, 5, 13, 17 et 19. Des techniciens forestiers se rendent sur le terrain et mesurent les volumes en essences résineuses et de feuillues en m³ par ha dans les 5 pixels sélectionnés. On retrouve le plus de feuillues dans le pixel 13 (87 m³ par ha) alors que c'est le pixel 17 qui contient le plus de résineux (41 m³ par ha). Ces données apparaissent aux colonnes 4 et 5 du tableau 1, seulement pour les cinq pixels sélectionnés.

Il faut maintenant prédire les volumes dans les 20 pixels non visités, tâche que nous allons réaliser au moyen de la méthode du plus proche voisin (kNN avec k=1 car un seul voisin intervient dans les prédictions). On va déterminer, pour chacun de ces 20 pixels, le pixel visité le plus proche dans l'espace spectral déterminé par les variables x₁ et x₂. Pour évaluer la proximité de deux pixels, on utilise la distance euclidienne calculée avec les intensités de couleur x₁ et x₂. Considérons le pixel numéro 1 avec x₁ = 2.9 et x₂ = 3.5. La distance euclidienne entre ce pixel et le pixel numéro 2 pour lequel x₁ = 3.3 et x₂ = 3 est

$$d(1,2) = \sqrt{(2,9-3,3)^2 + (3,5-3)^2} = 0,64$$

On calcule de même la distance entre le pixel 1 et les quatre autres pixels visités ; ces quatre distances sont 2,24 ; 3,11 ; 2,13 et 3,51. Parmi les 5 pixels visités, le plus proche voisin du pixel 1 est donc le pixel 2.

Dans le tableau à droite et en haut de la page, la colonne intitulée « voisin » prend donc la valeur 2 pour le pixel 1. Pour ce pixel, les variables volume de feuillu et volume de résineux vont donc être prédites par les valeurs du pixel 2 (Résineux=31, Feuillu=48). Les deux dernières colonnes du tableau 1 donnent les valeurs prédites pour chaque

Données et prédictions kNN pour l'exemple d'inventaire forestier

Pixel	x ₁	x ₂	Résineux	Feuillu	Voisin	RésineuxP	FeuilluP
1	2,9	3,5			2	31	48
2	3,3	3,0	31	48			
3	4,1	2,3			2	31	48
4	4,5	5,3			5	32	58
5	4,9	4,5	32	58			
6	2,3	4,7			17	41	8
7	2,8	4,5			17	41	8
8	3,6	3,0			2	31	48
9	3,7	1,9			2	31	48
10	4,4	1,0			2	31	48
11	1,7	5,3			17	41	8
12	2,1	4,4			17	41	8
13	6,0	3,3	24	87			
14	7,8	2,9			19	16	81
15	7,4	2,3			19	16	81
16	1,2	5,5			17	41	8
17	1,3	4,9	41	8			
18	5,7	4,4			5	32	58
19	6,4	3,3	16	81			
20	6,6	3,0			19	16	81
21	0,4	6,3			17	41	8
22	0,7	5,7			17	41	8
23	4,8	5,0			5	32	58
24	5,2	4,2			5	32	58
25	5,9	3,7			13	24	87

pixel par la méthode du plus proche voisin. Les valeurs prédites sont les valeurs observées pour les plus proches voisins donnés à la colonne « Voisin ».

Dans ce tableau, on voit que le pixel 2 est le plus proche voisin de 5 des pixels non visités. On dénombre également que les pixels 5, 13, 17, 19 sont les plus proches voisins de respectivement 4, 1, 7, et 3 des pixels non visités. Les estimations, par la méthode du plus proche voisin, des volumes des essences résineuses et feuillues pour l'UAF sont des moyennes de 25 valeurs qui sont soit observées soit prédites. Pour le volume de résineux on a :

Résineux

$$\frac{6 \times 31 + 5 \times 32 + 2 \times 24 + 8 \times 41 + 4 \times 16}{25} = 32,04 \text{ m}^3.$$

Alors que pour les feuillus l'estimation est

$$\begin{aligned} & \text{Feuille} \\ & \frac{6 \times 48 + 5 \times 58 + 2 \times 87 + 8 \times 8 + 4 \times 81}{25} \\ & = 44,28 \text{ m}^3. \end{aligned}$$

Notons que pour réaliser ces estimations il a fallu déterminer les plus proches voisins des 20 pixels non visités, ce qui a nécessité le calcul de $20 \times 5 = 100$ distances euclidiennes. Même pour ce petit exemple, il faut utiliser l'ordinateur pour automatiser le calcul de ces distances. Il n'est pas question de les évaluer à la main. Pour de vrais inventaires, avec des centaines de polygones visités et des dizaines de milliers de plus proches voisins à calculer, un ordinateur avec une grande puissance de calcul est nécessaire pour faire ce travail.

Précision des estimations

Une alternative simple à l'estimation par la méthode du plus proche voisin est de simplement prendre la moyenne des valeurs observées pour estimer la moyenne dans toute l'UAF. Cette façon de faire donne

$$\begin{aligned} & \text{Résineux} \\ & \frac{1 \times 31 + 1 \times 32 + 1 \times 24 + 1 \times 41 + 1 \times 16}{25} \\ & = 28,80 \text{ m}^3. \end{aligned}$$

au lieu du $32,04 \text{ m}^3$ de la section précédente. De même pour les feuillus la moyenne des 5 volumes observés donne $56,4 \text{ m}^3$ par hectare

au lieu de $44,28 \text{ m}^3$. Cette méthode, par la moyenne, est-elle plus (ou moins précise) que la méthode du plus proche voisin ?

Pour comparer les deux méthodes, on utilise une technique appelée « validation croisée ». On retire à tour de rôle un des pixels visités du jeu de données et on utilise les quatre pixels visités restants pour assigner au pixel retiré une valeur pour chaque variable d'intérêt, selon la méthode du proche voisin ou la méthode d'estimation par la moyenne. Les calculs sont résumés au tableau ci-dessous.

Pour le pixel 2, les distances euclidiennes, calculées à l'aide des variables x_1 et x_2 , avec les 4 autres pixels visités sont respectivement 2,19; 2,72; 2,76 et 3,11. Le plus proche voisin, à une distance de 2,19, est le pixel 5. Les valeurs prédites par la méthode du plus proche voisin pour les volumes des essences résineuses et de feuillus du pixel 2 sont 32 m^3 et 58 m^3 respectivement, telles que données dans les colonnes « RP » (pour Résineux Prédits) et « FP » (pour Feuillus Prédits). Pour l'estimation par la moyenne les valeurs prédites sont les moyennes pour les quatre pixels restants. Ceci donne

$$\frac{32 + 24 + 41 + 16}{4} = 28,25 \text{ m}^3$$

et $\frac{58 + 87 + 8 + 81}{4} = 58,5 \text{ m}^3.$

pour les essences résineuses et feuillus du pixel 2. Ces valeurs apparaissent dans les colonnes « RM » et « FM » du tableau « Calcul

Calcul des erreurs de prédiction pour l'estimation par la moyenne et par la méthode du plus proche voisin

Pixel	x_1	x_2	Résineux	Feuille	Voisin	RP	FP	RM	FM	ERP	EFP	ERM	EFM
2	2,9	3,0	31	48	5	32	58	28,25	58,50	1	10	-2,75	10,5
5	4,9	4,5	32	58	13	24	87	28,00	56,00	-8	29	-4,00	-2,0
13	6,0	3,3	24	87	19	16	81	30,00	48,75	-8	-6	6,00	-38,3
17	1,3	4,9	41	8	2	31	48	25,75	68,50	-10	40	-15,3	60,5
19	6,4	3,3	16	81	13	24	87	32,00	50,25	8	6	16,00	-30,8

des erreurs de prédiction ». Les erreurs de prédictions pour les deux méthodes sont les différences entre la vraie valeur et la valeur prédite. Elles apparaissent aux colonnes « ERP », « EFP » pour le plus proche voisin et « EFM » et « EPM » pour la moyenne. L'erreur de la prédiction par la méthode du plus proche voisin est, en général, inférieure en valeur absolue à celle de la prédiction par la moyenne. Pour le plus proche voisin la moyenne des valeurs absolues des erreurs pour la prédiction du volume de résineux de nos cinq pixels visités est de

$$\frac{1+8+8+10+8}{5}=7$$

alors que pour la prédiction par la moyenne cette moyenne est

$$\frac{2,75+4+6+15,3+16}{5}=8,8.$$

Les valeurs correspondantes pour les feuillus sont de 18,2 et 28,4. La méthode du plus proche voisin donne donc de meilleures estimations.

Dans le tableau ci-haut, on retrouve les plus grandes erreurs pour les volumes de feuillus. Cette variable semble donc plus difficile à estimer que le volume de résineux. On aimerait bien utiliser les erreurs du tableau « Calcul des erreurs de prédiction » pour mesurer la précision des estimations obtenues par la méthode du plus proche voisin, mais ces estimés d'erreurs ne capturent qu'une partie de l'erreur véritable d'imputation. Le calcul rigoureux de l'erreur est un problème compliqué et il n'existe pas encore une théorie mathématique qui permette d'estimer avec confiance la précision de ces estimations même si elles sont couramment utilisées pour réaliser des inventaires forestiers.

Discussion

L'inventaire forestier est une source privilégiée d'information concernant la forêt. En pratique plusieurs dizaines de variables sont mesurées dans les deux phases de l'inventaire et utilisées à plusieurs fins. On peut ainsi jumeler l'inventaire avec des modèles de croissance des arbres pour prédire l'effet de pratiques sylvicoles, ou utiliser l'information spatiale pour évaluer la qualité de paysages forestiers en tant qu'habitats fauniques. Les données d'inventaire interviennent également dans la comptabilisation du carbone forestier, et dans l'évaluation de l'impact de perturbations naturelles telles que les épidémies d'insectes ou les feux. La vaste majorité des utilisations requièrent le jumelage de l'information ponctuelle provenant des placettes échantillon et de l'information spatiale provenant présentement de la photo-interprétation. La prise en compte des erreurs d'estimation de l'inventaire dans ces calculs complexes fait aujourd'hui l'objet de nombreux travaux de recherche en mathématiques, en statistique et en foresterie.