



FACULTÉ DE
DROIT
SCIENCES
ÉCONOMIQUES
& GESTION
DE NANCY



**UNIVERSITÉ
DE LORRAINE**



BETA
Bureau
d'économie
théorique
et appliquée
UMR 7522

UMR 7522 CNRS/UL/INRA/AgroParisTech

Analyse spatiale des clusters industriels dans le secteur forêt- bois français et leur impact sur la probabilité de coupe

Rapport de stage Master 2 analyse et développement de l'entreprise

Auteur :
Guillaume Salzet

Encadrant :
Sylvain Caurla

2017-2018

Introduction :

Dans la perspective du développement durable national et territorial basé sur la bio ressource bois, l'étude de la filière bois et de son impact est nécessaire afin d'en garantir le faible impact et la neutralité carbone. Cette transition écologique serait bénéfique pour l'environnement et l'économie nationale. En effet, en apportant un cadre réglementaire orientant l'économie nationale vers l'usage du bois comme ressource primaire majeure, un développement d'innovation s'intégrerait à la filière forêt-bois existante tout en apportant capital et main-d'œuvre. Cependant cette politique publique implique une connaissance fine de la filière actuelle et des voies de développement à privilégier. Or la filière forêt-bois regroupe la chaîne de valeur basée sur la ressource bois. Elle présente comme particularité une hétérogénéité importante dans la taille, la rentabilité et la localisation à l'image de celle de la ressource d'origine. Ainsi une étude empirique des agrégats industriels spatialisés se justifie en tant qu'état des lieux préliminaire.

Classiquement en économie géographique, l'étude du lien entre les entreprises et plus généralement un secteur d'activité sur un territoire se fonde sur des effets d'agrégation ou de dispersion selon l'hétérogénéité des facteurs de production comme les salariés (ex : dynamisme des métropoles), le capital, la technologie et enfin la ressource primaire. Cependant les études antérieures se focalisent majoritairement à l'échelle internationale ou nationale mais rarement à une échelle infrarégionale. L'absence de compréhension des dynamiques régionales s'explique par la pauvreté de l'information sur les entreprises à cette échelle. Or le lien entre le secteur forestier et sa ressource est principalement effectué par des entreprises de petite taille présentes en grand nombre pour s'adapter aux besoins de l'exploitation et la transformation.

Pour cela il a été effectué une analyse selon les méthodes d'économétrie spatiale pour la mesure du degré d'agrégation des entreprises de l'amont de la filière (de l'exploitation forestière à la première transformation). Cette première analyse a été ensuite utilisée pour l'étude de l'effet de la filière sur le niveau d'exploitation de la ressource à l'aide d'un modèle statistique de survie à hasard proportionnel. L'utilité de ce modèle est de distinguer finement l'effet dû à hétérogénéité spatiale de la ressource (de manière quantitative et qualitative) de l'effet propre à des pratiques industrielle régionales.

Matériel et méthodes :

Données d'inventaires :

Depuis 2005, l'Inventaire Forestier National (IFN) a effectué un échantillonnage systématique de la France métropolitaine et de la Corse selon une maille de 1 point tous les 10 km². Les campagnes successives assurent une mesure pour l'ensemble des points à un intervalle de 5 ans. A chaque nœud du maillage, un point de mesure est effectué dans une zone de 900 m × 900 m de tolérance. Le protocole de mesure est décrit de façon exhaustive dans IGN (2016).

Chaque point de mesure consiste en trois cercles concentriques de rayons fixés en un même centre où les arbres sont mesurés selon leur diamètre à hauteur de poitrine (DBH, 1.3m de haut). Les arbres sont mesurés selon des surfaces différentes en fonction de leur DBH comme suit :

Diamètre à hauteur de poitrine	Rayon de mesure
7.5 cm – 22.5 cm	6 m
22.5cm – 37.5 cm	9 m
> 37.5cm	15 m

Tableau 1 : Organisation hiérarchique du plan d'échantillonnage des mesures en sous surface de rayon variable selon le diamètre.

La mesure effectuée sur chaque arbre consiste en prise de la circonférence à hauteur de poitrine arrondie au centimètre le plus proche. Sous l'hypothèse d'une section transversale parfaitement circulaire, le diamètre est déduit par une division par Pi. En outre, des observations supplémentaires sont ajoutées pour chaque point de mesure : présence d'un traitement sylvicole sur une tranche de 5 ans et l'inclinaison de la pente.

En conséquence de l'attribution de la mission d'évaluation de la ressource ligneuse à l'échelle nationale, l'IFN a effectué à partir de 2010 une seconde mesure de l'ensemble des points du maillage issu de la campagne de 2005. Sur une période de 5 ans, l'ensemble des 41 380 points de mesures ont été réévalué avec toujours un intervalle de 5 ans entre la mesure initiale. Lors de cette seconde campagne, la présence de coupe a été rajoutée dans les observations.

Les points de mesures localisés dans les zones protégées et ceux où l'usage du sol n'est pas la production de bois ont été ignorés. De plus les points n'ayant aucun bois commercialement utilisables (point vide ou ayant uniquement des arbres ayant un DBH < 7.5 cm) ont été enlevés. On obtient finalement 40 077 points de mesures. A partir des mesures individuelles, la surface terrière, qui correspond à la somme des surfaces transversales à 1.30 m et la densité de tige ont été calculés.

Données spatialisées de la filière forêt-bois :

La définition de la filière forêt-bois se fonde sur la production, la transformation et la vente de biens et services en lien avec la ressource bois. Cette définition agrège des industries de différentes sections de la nomenclature des activités françaises (NAF) (INSEE 2008) (e.g 02 Sylviculture et exploitation forestière ; 16 Travail du bois et fabrication d'artisanat en bois et liège sauf des meubles – fabrications artisanales de vannerie et

sparterie ; 17 Industries du papier et carton). Aucune divisions ou emboitements ne rassemblent la totalité du secteur forestier comme un *forest cluster* au sens de Lammi (1996). En conséquence, la reconstruction partielle de la partie amont de la filière (en limitant aux entreprises d'exploitation forestières et la première transformation du bois par sciage) s'effectue par l'agrégation de sous sections précises comme en [Annexe I](#).

L'ensemble des entreprises françaises sont enregistrées par un identifiant unique, le code SIREN. La base de données regroupant les informations basiques sur l'activité économique des entreprises est la base SIRENE. La présence de l'adresse normalisée permet la spatialisation précise des structures référencées. Cette base de données spatialisées est fourni par l'open Data du gouvernement français et est géocodé par les scripts de Christian Quest (Open Database License 1.0). La base a été extraite sous la version de Juillet 2018 (N = 70023).

Econométrie spatiale :

La concentration industrielle consiste en la mesure de la déviation locale à la moyenne globale en termes de densité d'entreprises par unité de surface. Les unités de productions peuvent être soit dispersées autour de la ressource, ici la forêt, soit rassemblées. On parle respectivement de *dispersion* et de *clustering* local. Afin de distinguer les patrons de distributions, différentes statistiques peuvent être utilisées avec comme hypothèse nulle une distribution uniforme sur le territoire.

Notre étude se fonde sur l'analyse dite *Hotspot & Coldspot analysis*. Elle est réalisée pour délimiter spatialement les agrégats d'industries en se basant sur la statistique G_i^* de Getis-Ord à rayon fixe à l'aide du logiciel ArcGIS (version 10.6, ESRI). La statistique *Z score* résultante permet d'identifier l'état d'agrégation. Pour des valeurs significativement positives (négative) de Z score, plus la valeur de cette métrique est importante (faible), plus l'état d'agrégation est fort (faible). Cette mesure se fonde sur l'analyse de l'environnement local en comparaison à l'environnement global.

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}x_j - \bar{X} \sum_{j=1}^n w_{ij}}{S \sqrt{\frac{n \sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2}{n-1}}}$$

Avec x_j la valeur prise dans l'observation j , w_{ij} est le poids spatial entre les observations i et j , n est le nombre total d'observation. On pose les valeurs suivantes : $\bar{X} = \frac{1}{n} \sum_{j=1}^n x_j$ et $S = \sqrt{\frac{1}{n} \sum_{j=1}^n x_j^2 - (\bar{X})^2}$.

Modélisation par régression de Cox :

Les données de récolte de bois sont sous la forme d'intervalles temporels censurés. La date de la coupe n'est pas observée avec précision, seule son occurrence entre les deux mesures pour chaque point est disponible. L'utilisation d'un modèle de survie permet de tenir compte de cette structure temporelle.

On postule que la date de récolte T suit une fonction de répartition de probabilité $F(T \leq t)$. De façon usuelle, la probabilité que l'événement de coupe n'ait pas eue lieu au moment t suit une fonction de répartition de

longévité $S(t) = F(T > t) = 1 - F(T \leq t)$. L'utilisation de données avec censure par intervalle oblige une définition de la probabilité d'occurrence entre t_1 et t_2 sous la forme : $\mathbb{P}(t_1 < T < t_2) = S(t_1) - S(t_2)$.

Dans le cas où on constate que le point de mesure n'a pas encore subi de récolte à t_1 , alors la probabilité conditionnelle de récolte entre t_1 et t_2 est de la forme :

$$\mathbb{P}(t_1 < T < t_2 | T > t_1) = \frac{S(t_1) - S(t_2)}{S(t_1)} = 1 - \frac{S(t_2)}{S(t_1)}$$

La fonction S peut suivre des distributions paramétriques classiques comme les lois exponentielles ou de Weibull. Cependant, les modèles de type régression de cox à hasard proportionnel permettent de tenir compte d'effets hiérarchiques spatialement (*i.e.* valable soit pour tous les points, l'échelle nationale, où pour certains groupes de points, l'échelle régionale et enfin local). On suppose dans ce cadre que la fonction de longévité est de la forme :

$$S(t_i) = e^{h(x_i, \beta) \int_0^{t_i} h_o(t, g_i, \gamma) dt}$$

Avec x_i et g_i deux vecteurs de variables explicatives du point de mesure i , β et γ sont les vecteurs de paramètres pour l'ensemble de la population. La fonction $h(x_i, \beta)$ correspond au risque proportionnel alors que la fonction $h_o(t, g_i, \gamma)$ correspond au niveau de risque basal accumulé au cours du temps. Le modèle étudié peut être réécrit sous forme discrète à l'aide des probabilités conditionnelles :

$$\mathbb{P}(t_1 < T < t_2 | T > t_1) = 1 - e^{h(x_i, \beta) \sum_{t_1}^{t_2} h_o(t, g_i, \gamma)} = \pi_i$$

On peut finalement effectuer l'estimation des paramètres par maximisation de la vraisemblance à l'aide de l'expression de la log-vraisemblance suivante :

$$\mathcal{L}(y|\mathbf{X}, \mathbf{G}, \beta, \gamma) = \sum_{i=0}^n y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)$$

Avec y_i la variable binaire d'occurrence de coupe (0 : absence de coupe, 1 : coupe observée), \mathbf{X} et \mathbf{G} sont les matrices associés aux variables explicatives respectivement à x_i et g_i pour toutes les observations.

L'évaluation de l'ajustement du modèle est effectuée selon le critère d'information d'Akaike (AIC) afin de garder ou non une variable. Après ajout d'une ou plusieurs variables, il est calculé l'ensemble des approximations des résidus de Pearson *versus* les covariables. Dans le cas de cette étude, l'approximation des résidus de Pearson peut être estimé comme suivant :

$$r_{\text{Pearson}, k} = \frac{p_k - \widehat{\pi}_k}{\sqrt{\frac{\widehat{\pi}_k(1 - \widehat{\pi}_k)}{n_k}}}$$

Avec k , l'index de la covariable, p_k la proportion de point effectivement récolté, $\widehat{\pi}_k$ est la moyenne prédite par les probabilités de coupe ($\widehat{\pi}_k = \sum_{i \in s_k} \frac{\widehat{\pi}_i}{n_k}$) où s_k est le sous-échantillon de la k -ième classe de la covariable

et n_k est le nombre d'observation de ce sous-échantillon. Les classes pour les variables continues ont été discrétisées selon un pas de $5\text{m}^2.\text{ha}^{-1}$. On suppose que l'approximation normale de la variable de Bernoulli est valable pour n_k suffisamment large. Par conséquent, les résidus sont distribués normalement.

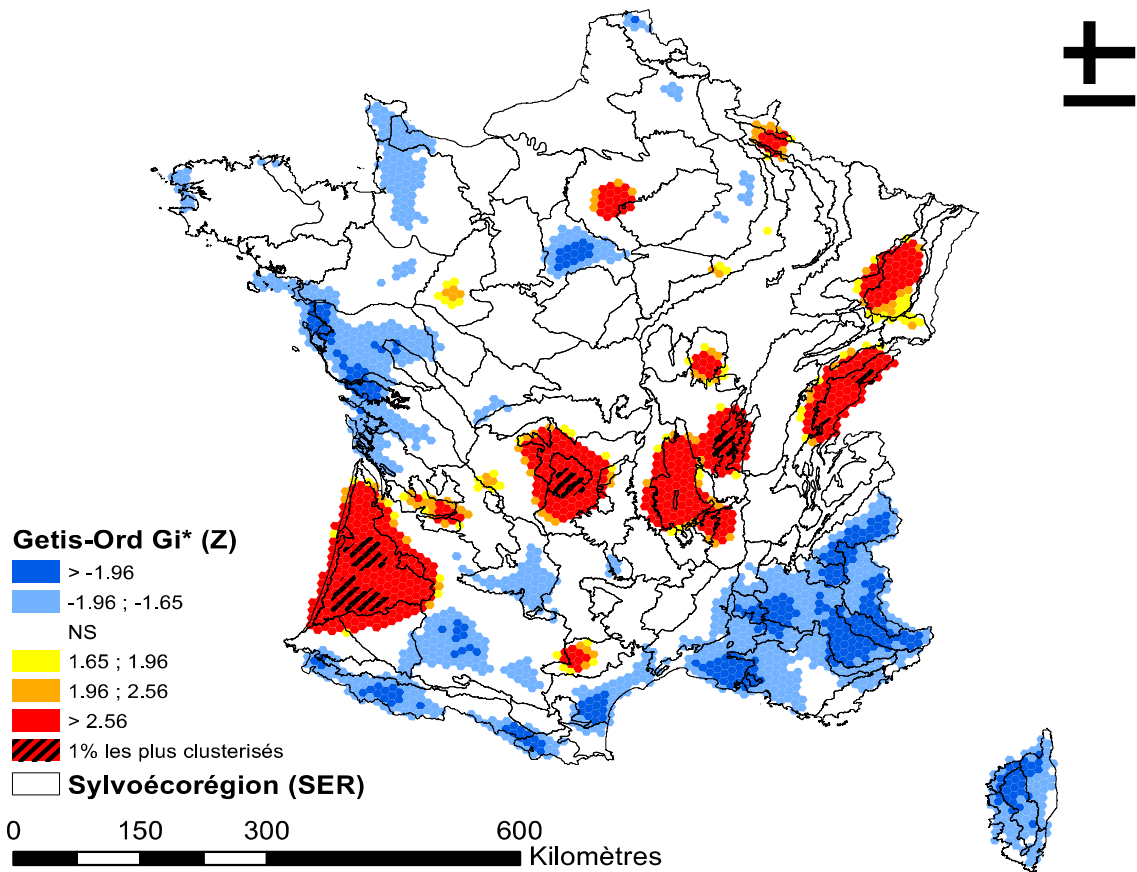


Figure 1 : Agglomération de la filière forêt-bois (NAF 02, 16, 17 – INSEE 2018). Observation d'une congruence forte avec les sylvoécórégions.

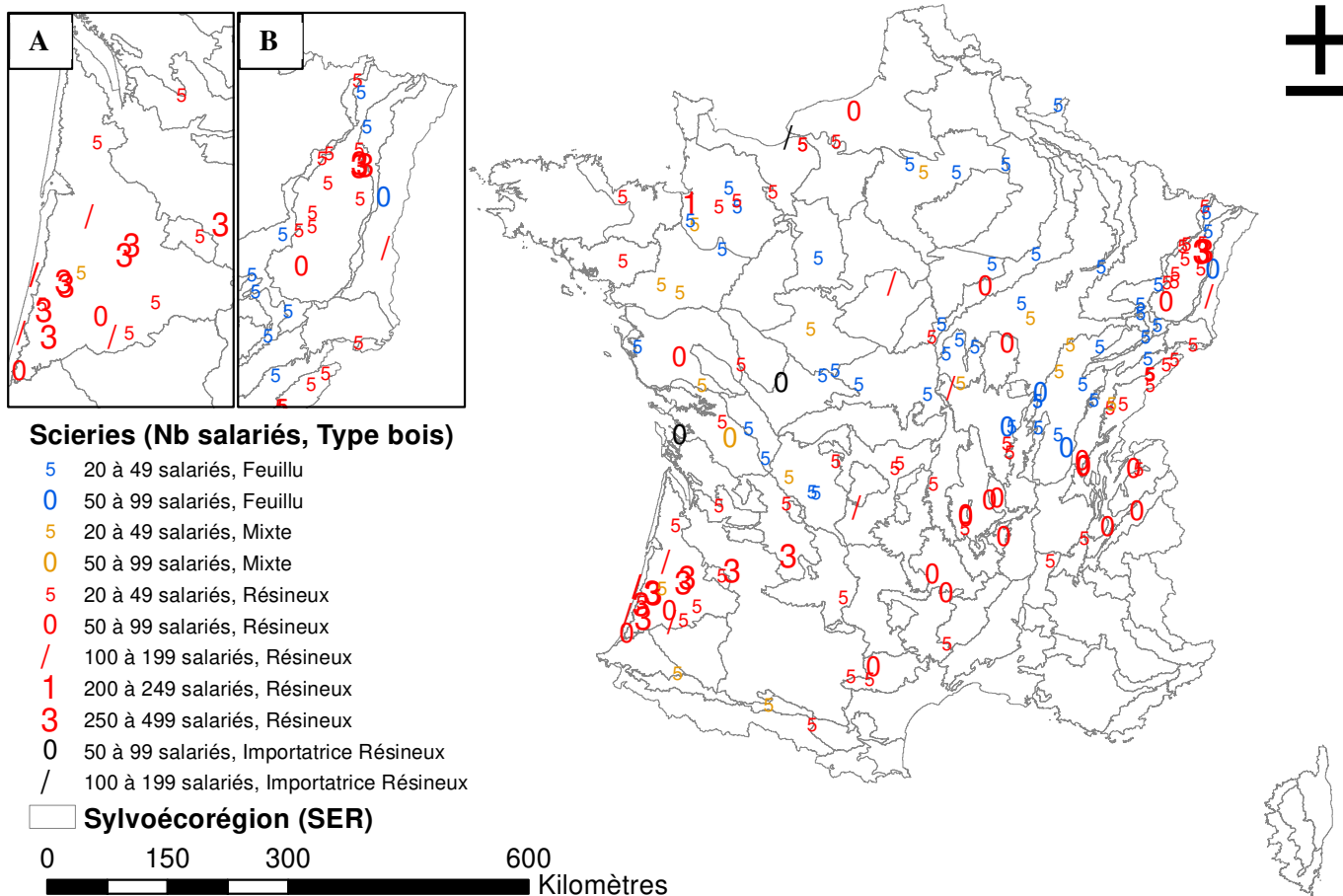


Figure 2 : Distribution des scieries selon leur effectif et le type de bois scié. Encarts sur les agrégats Landais A et Vosgien B.

Résultats :

1. Hétérogénéité spatiale de la filière forêt-bois :

La distribution du secteur forestier sur le territoire métropolitain révèle de fortes hétérogénéités spatiales. L'analyse a porté sur la partie *stricto sensu* en amont de la filière forêt-bois (exploitation forestière, sciage et production de pâte à papier, $N = 70023$). L'organisation de la filière respecte le principe d'atomicité pour les entreprises les plus en lien avec la ressource. En effet, les entreprises d'exploitation forestière sont majoritairement de faible taille (inférieur à 20 employés). Par conséquent, l'analyse peut être effectuée avec une pondération égale pour chaque entreprise. Cette hypothèse n'est pas vérifiée pour les industries situées à un niveau supérieur dans la chaîne de valorisation de la ressource. Pour cela une cartographie exhaustive des scieries de plus de 20 employés a été faite en indiquant le type de bois scié.

On observe l'existence d'agrégats industriels significatifs localisés dans un nombre restreint de sylvoécotérritoires. Cette congruence est plus robuste avec les sylvoécotérritoires à forte concentration de résineux (SER) qu'avec les régions administratives (pour les anciennes et les nouvelles régions). Quinze agrégats significatifs ($G_i^* > 1.96$ à $p = 0.05$) sont identifiables sur le territoire. Les agrégats disposant des valeurs les plus fortes (1% des plus hautes valeurs de G_i^*) sont le massif des Landes (SER : F21) avec deux noyaux d'aggrégations, les massifs des plateaux Ouest (SER : G21) et la bordure Est (SER : G41) du massif central et le massif du second plateau du Jura (SER : E20).

L'agrégat des Vosges du Sud (SER : D11 & D12) présente la particularité de rassembler des ensembles forestiers très différents en termes d'écologie et *in fine* de ressources exploitables. L'ensemble central est principalement composé d'industries valorisant des conifères alors que l'anneau extérieur tourné vers la Champagne valorise les feuillus (distinction visible sur la répartition des types de scieries).

Enfin l'agrégat en région parisienne correspond à un ensemble d'entreprise de service forestier principalement. Cependant, on ne peut écarter pas l'hypothèse d'une zone d'attraction car la zone dispose d'une importante main d'œuvre et est dynamique économiquement.

On observe ensuite l'existence de zones de dispersion de la filière significativement peu présentes. Ces zones sont de deux types : les premières correspondent à des zones faiblement boisées où la ressource est dispersée (Littoral vendéen, SER : A30, et Normand, SER : A13) potentiellement soumis à des imports avec la présence d'infrastructure portuaires (cf. scieries importation de résineux); les secondes sont des zones boisées à faible productivité (Zone méditerranéenne SER : J- et littoral corse SER : K11 & K13) ou difficilement accessible (Alpes SER : H-, Pyrénées, SER : I- et montagnes corses SER : K12).

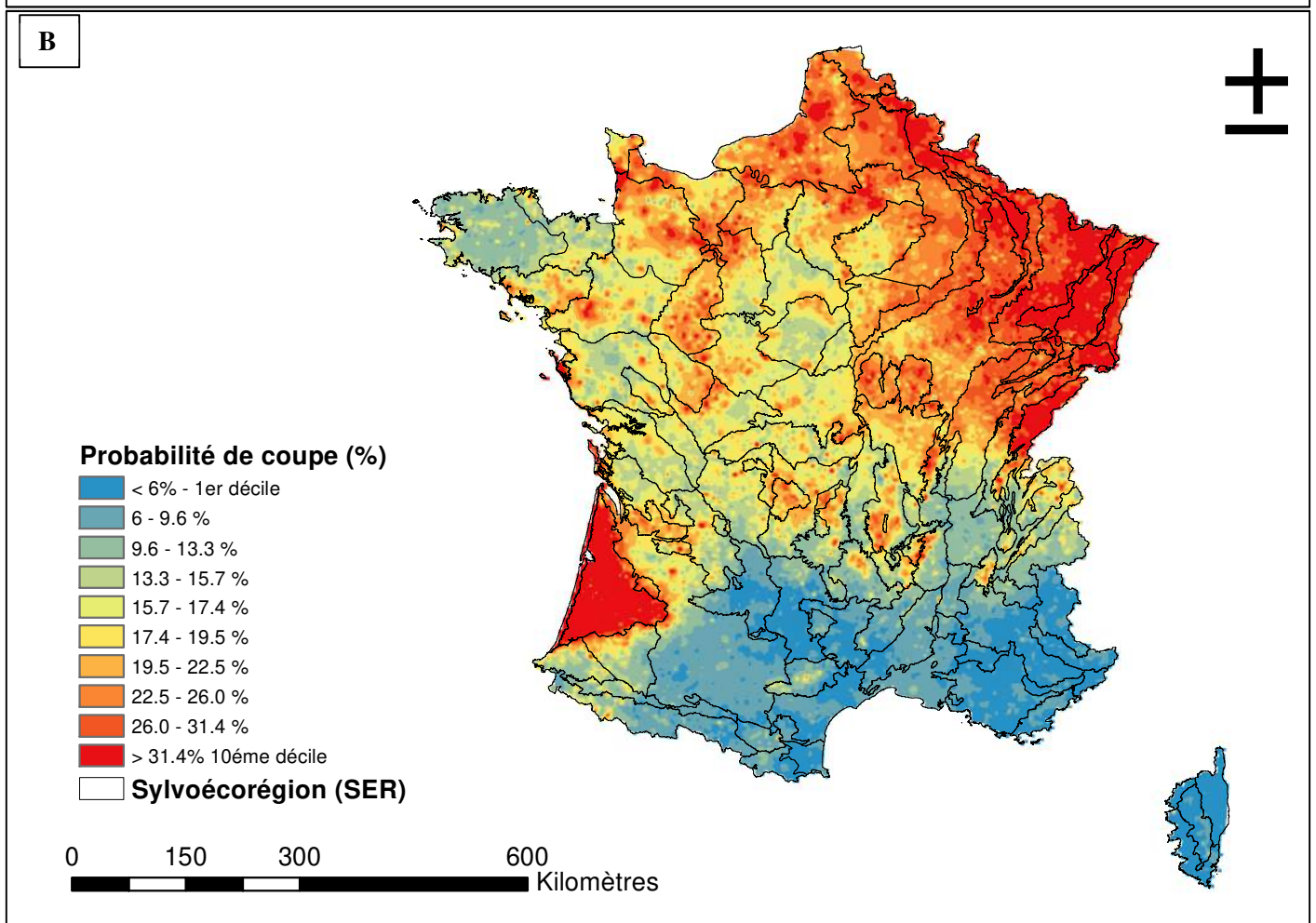
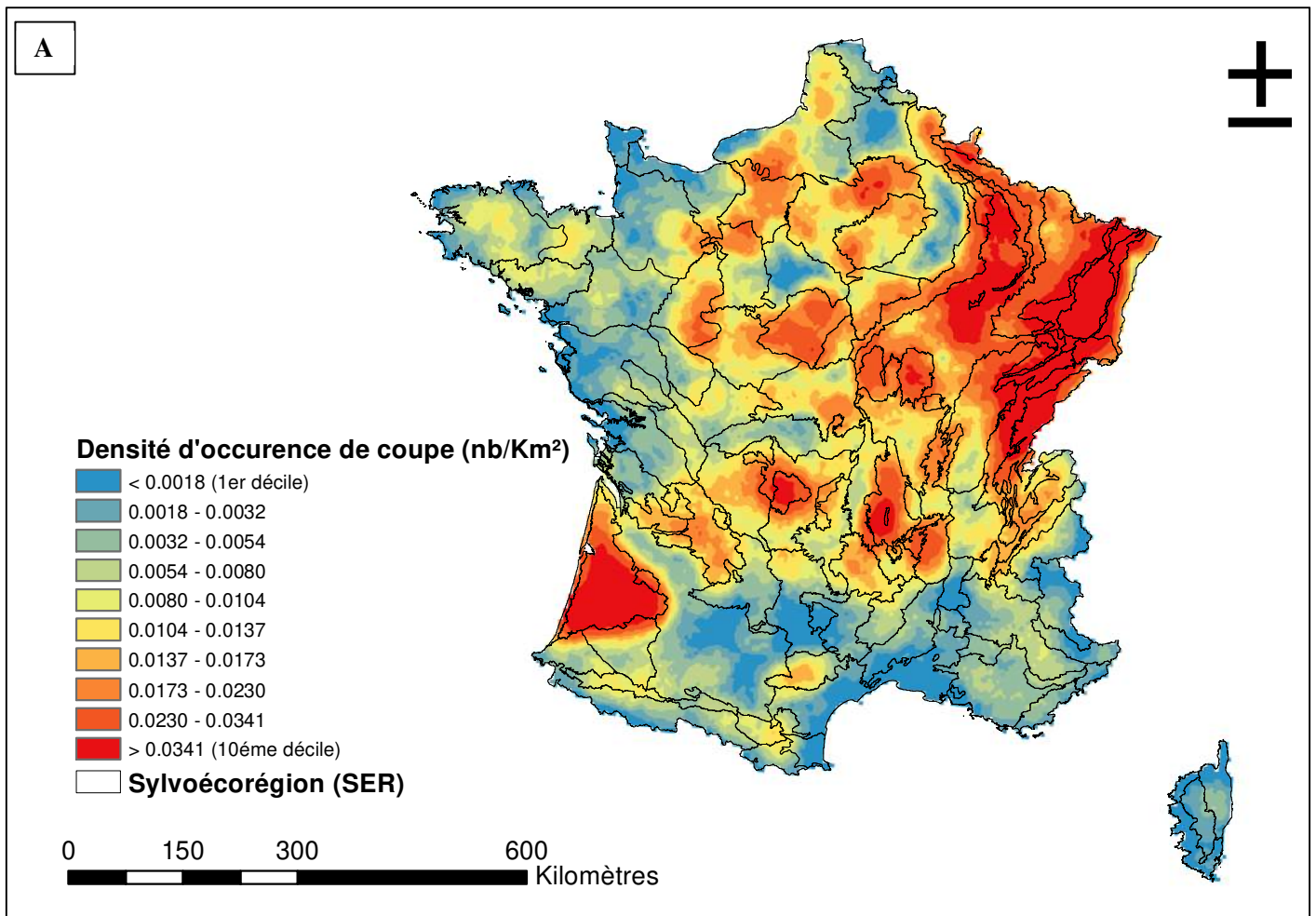


Figure 3 : Répartition des densités surfaciques de coupe (# de coupe.Km⁻¹) **CORRECTION NECESSAIRE**, carte A, et leur modélisation par le modèle de hasard proportionnel intégrant les données économiques et écologiques, carte B.

2. Impact de la filière forêt-bois sur l'exploitation de la ressource bois :

a. Spécification du modèle :

En accord avec Melo *et al.* (2017), La probabilité basale inclut les macrovariables définies à l'échelle nationale et qui peuvent varier dans le temps. La partie proportionnelle de la régression, est supposée tenir compte des effets régionaux à infrarégionaux rassemblés sous l'appellation de variables locales.

Les macrovariables rassemblent principalement l'évolution des cours du bois par espèce sous la forme de l'évolution de la valeur de la placette sans l'occurrence de la coupe. Cette évaluation découle des cours nationaux du bois sous la forme des prix moyens fournis par l'office nationale des forêts (ONF) pour les terrains publics sur la période 2006-2016. Des cours additionnels pour les espèces *Pseudotsuga menziesii* et *Populus spp.* en plantation et sur les terrains privés ont été obtenus à partir des associations nationale des propriétaires forestiers. Ces valeurs sont exprimées en euros constants. Cette évaluation se fonde sur un traitement sylvicole ayant une espèce objectif identifiée. L'effet sur la placette correspond à la valeur selon les cours du bois global auquel on ajoute un effet spécifique de cette espèce supposée lors de l'inventaire. Un effet distinct entre résineux et feuillu est ajouté pour prendre en compte de manière intégrée les différences de traitement sylvicole.

Les variables locales correspondent à des spécificités liées à des déterminants écologiques, techniques et économiques.

Pour les déterminants écologiques, nous avons testés : la surface terrière (*i.e.* la somme des surface des sections des arbres à 1.30m de hauteur) car comme cela a été montré précédemment (Antón-Fernández and Astrup, 2012; Melo et al., 2017) l'occurrence de coupe lui est corrélé positivement ; la densité de tige car elle représente l'état de maturité des arbres selon la loi d'auto-éclaircissement or plantation ; la pente puisque cela influence d'une part l'accessibilité et *in fine* l'occurrence de coupe (Antón-Fernández and Astrup, 2012; Melo et al., 2017) et correspond à un proxy de la fertilité des sols en zone montagneuse. Ce dernier effet a été distingué entre résineux et feuillu.

Pour les déterminants techniques, la présence de trace d'aménagement forestier est un facteur important dans le cas français puisque la gestion forestière n'est pas systématique au sein des propriétaires forestiers français. En France, les forêts de plus de 25 ha doivent avoir un plan simple de gestion si elles sont privée afin d'en assurer l'aménagement (l'ensemble des forêts publiques sont sous ce type de réglementation). Ainsi cet effet a été distingué selon le type de propriétaire, publique ou privé. Afin de garder la représentativité des points, l'IGN garde secret la localisation exact des placettes autour du point de mesure dans une surface de 900 m × 900 m. par conséquent le type de propriété ne peut être que estimé à partir des données des forêts publiques gérées par l'ONF disponible gratuitement. L'estimation de la probabilité d'une propriété privée a été estimée en effectuant le ratio de la surface privée dans zone d'incertitude divisé par la surface de cette zone. Synthèse des résultats d'ajustement : Finalement ces traces d'aménagement ont été relevées pour chaque relevé.

Pour les déterminants économiques, l'accessibilité de la placette a été prise en compte globalement par la distance à la route la plus proche à partir des données géographiques de l'IGN sans distinction du type de route. Il a été retenu comme échelle représentative de la distribution des scieries, les entreprises de plus de 20 employés indiqué dans la base INSEE. Chacune de ces entreprise a été vérifiée individuellement afin de déterminer le type de bois (feuillu/ mixte/ résineux) principalement utilisé. Enfin il a été mesuré la distance minimale entre la placette et une scierie du type adéquat (le type mixte est commun aux deux types). Enfin la dernière mesure résulte du comptage des entreprises par SER.

b. Résultat d'ajustement du modèle :

Le modèle final après ajustement a inclus les variables précédemment présentées. En termes de contribution pour le modèle, les variables écologiques ont été prépondérantes dans la baisse de l'AIC (prix du volume sur pied et pente). La valeur de l'AUC (*Area Under the Curve*) est de 0.743 est ce qui considéré comme acceptable par Hosmer *et al.* (2013).

Le modèle ajusté peut être écrit comme suivant :

$$\mathbb{P}(t_1 < T < t_2 | T > t_1) = 1 - e^{-h(x_i, \beta) \int_{t_1}^{t_2} h_o(t, g_i, \gamma)}$$

$$h(x_i, \beta) = -e^{\beta_1 ST_i + \beta_2 \frac{ST_i N_i}{1000} + (\beta_3 + \beta_4 c) \text{Pente}_i + (\beta_5 + \beta_6 Am_i) \mathbb{P}_{\text{priv},i} + \beta_7 \text{Dist}_{\text{route},i} + \beta_8 \text{Type} \text{Dist}_{\text{scierie,type},i} + \beta_9 \text{Dens}_{\text{ind,SER},i} + \beta_{10,r}}$$

$$h_o(t, g_i, \gamma) = e^{\gamma_0 + \gamma_{1,c} + (\gamma_2 + \gamma_{3,e}) \text{Valeur}_{e,t}}$$

Avec ST la surface terrière ($\text{m}^2 \cdot \text{ha}^{-1}$) ; N la densité de tige (nombre de tige. ha^{-1}) ; la pente (%) ; Am est la variable binaire où la valeur 1 indique la présence de traces d'aménagements dans l'intervalle de 5 ans, 0 sinon ; \mathbb{P}_{priv} est la probabilité que le point de mesure se situe sur un terrain privé ; $\text{Dist}_{\text{route}}$ est la distance minimale à vol d'oiseau à partir du point de mesure pour atteindre une route ; $\text{Dist}_{\text{scierie,type}}$ est la distance minimale à vol d'oiseau à partir du point de mesure pour atteindre une scierie correspondant au type de bois de l'essence objectif (ayant une taille supérieure à 20 employés) ; $\text{Dens}_{\text{ind,SER}}$ est la densité d'entreprise par SER par unité de surface (nombre d'entreprise. Km^{-2}) ; $\text{Valeur}_{e,t}$ est la valeur sur pied de l'essence objectif e à l'année t selon les prix ONF ; i est l'index des points de mesure ; c est l'index d'une essence résineuse ; e est l'index des essences objectifs.

Annexe 1 : Liste des codes NAF retenus pour la filière bois

DIVISION	NAF	ACTIVITE
02 Sylviculture et exploitation forestière	0210Z	Sylviculture et autres activités forestières
	0220Z	Exploitation forestière
	0230Z	Récolte de produit forestiers non ligneux poussant à l'état sauvage
	0240Z	Services de soutien à l'exploitation forestière
16 Travail du bois et fab. d'art. en bois et en liège sauf des meubles. Fab. art. vannerie et sparterie	1610A	Sciage et rabotage du bois, hors imprégnation
	1610B	Imprégnation du bois
	1621Z	Fabrication de placage et de panneaux de bois
	1622Z	Fabrication de parquets assemblés
	1623Z	Fabrication de charpentes et d'autres menuiseries
	1624Z	Fabrication d'emballages en bois
	1629Z	Fabrication d'objets divers en bois ; fabrication d'objets en liège, vannerie et sparterie
17 Industrie du papier et carton	1711Z	Fabrication de pâte à papier
	1712Z	Fabrication de papier et de carton
	1721A	Fabrication de carton ondulé
	1721B	Fabrication de cartonnages
	1721C	Fabrication d'emballages en papier
	1722Z	Fabrication d'articles en papier à usage sanitaire ou domestique
	1723Z	Fabrication d'articles de papeterie
	1724Z	Fabrication de papiers peints
	1729Z	Fabrication d'autres articles en papier ou en carton