



*Forest disturbances impose challenges to the monitoring of forest resources*

## **Échantillonnage dynamique et inférence pour un monitoring intelligent des forêts, Application à l'Inventaire Forestier National**

Responsable scientifique : Olivier BOURIAUD, Laboratoire d'Inventaire Forestier, IGN, Nancy

Partenaires Labex : IECL (Elie Cartan)

Collaborations : Dr Guillaume CHAUVET (HDR), ENSAI, UMR 6625 IRMAR, Rennes

Action thématique concernée : WP2

### **Contexte —**

Les effets des changements globaux et la perte de biodiversité sont devenus des problématiques présentes dans les sphères publiques et politiques, tandis que la prise de conscience de l'empreinte des actions humaines accroît les attentes envers les forêts. Ces nouvelles attentes soulignent la nécessité de quantifier et de suivre non seulement l'état des forêts mais aussi leur capacité à soutenir l'approvisionnement de ces services écosystémiques. Cependant, les méthodes d'inventaire à grande échelle ont été développées avant tout pour un suivi de la ressource et à un pas de temps pluriannuel. Les nouvelles attentes, ainsi que les perturbations plus intenses et fréquentes, imposent un examen de ces méthodes, pour les faire évoluer vers une plus grande réactivité.

Les méthodes de sondage et d'estimation de l'inventaire forestier national français ont été repensées en 2004 dans l'objectif de permettre une couverture annuelle du territoire et une adaptabilité de l'effort d'échantillonnage. Les données recueillies depuis 15 ans constituent une base empirique inégalée, qui doit permettre d'évaluer l'efficacité et les limites des méthodes actuelles.

Par ailleurs, les compétences nécessaires à l'analyse des méthodes de sondage et d'estimation sont extrêmement rares, particulièrement dans le domaine forestier. La disparition prématurée des statisticiens de référence (Jean-Christophe Hervé, Adrian Lanz) laisse un manque profond dans une communauté de très petite taille. Le rôle du projet est de contribuer au renouvellement de ces compétences et de développer de nouvelles stratégies d'échantillonnage et d'inférence, en rapprochant dans le cadre du projet des personnes possédant chacune une expérience spécifique et complémentaire.

### **Objectifs —**

Le but du projet est de développer des méthodes de sondage, allant de l'échantillonnage jusqu'à l'estimation, permettant de répondre à des contraintes de modulation rapides de l'intensité d'échantillonnage, tout en conservant des hauts niveaux de précision. Bien que développées dans le contexte de l'inventaire forestier national, le caractère général de ces méthodes sera conservé et renforcé pour permettre d'être employé aux enquêtes des écosystèmes.

Un premier objectif est de décrire et d'évaluer les estimateurs statistiques post-stratifiés actuels, et de les comparer avec les estimateurs existants dans des contextes critiques. Un deuxième objectif est d'évaluer et d'optimiser les schémas de stratification / post-stratification des forêts avec comme perspectives le compromis variance-biais et les problèmes de strates de petite taille. Un troisième objectif est de reconcevoir la stratégie d'échantillonnage de cet inventaire et d'accroître sa flexibilité.

### **Démarche —**

La démarche comporte deux aspects très complémentaires : l'axiomatisation et l'analyse quantitative.

L'axiomatisation constitue une formalisation des méthodes de sondage et d'estimation de l'inventaire forestier national. Cette formalisation représente un effort d'ancrage des méthodes d'inventaire dans la théorie des sondages, et représente donc une activité essentiellement théorique. Cette activité vise à offrir aux travaux des fondations statistiques propres à la discipline ainsi qu'une base de références bibliographiques. La formalisation assure également le développement des propriétés limites et des hypothèses de validité des estimateurs statistiques.

L'analyse quantitative repose sur la mise en valeur de l'exceptionnelle base de données de l'inventaire forestier national, et en premier lieu, d'en réaliser une analyse quantitative critique. Les hypothèses et valeurs limites théoriques seront confrontées aux étendues et variance estimées sur la base des données collectées lors de la quinzaine d'années de fonctionnement de l'inventaire.

### **Résultats marquants —** (présentés sous forme de puces séparées)

- Les méthodes statistiques d'estimation de l'Inventaire Forestier Français (IFN) ont été décrites conformément à la théorie des sondages pour la première fois depuis la création de la 'nouvelle méthode d'inventaire' (2004). Elles confirment les estimateurs de totaux employés par l'IFN.
- L'analyse théorique des estimateurs statistiques (i.e. l'axiomatisation) apporte une contribution originale en proposant une nouvelle génération d'estimateurs statistiques capables de fonctionner avec des taux d'échantillonnage variables dans le temps et dans l'espace. Il n'existe aucun estimateur de ce type à l'heure actuelle dans la littérature, ce qui positionne favorablement les travaux de thèse dans une perspective de publications scientifiques de spécialité.
- L'analyse théorique a permis la création d'estimateurs statistiques reliant les données de niveau arbre aux estimations cibles à l'échelle du territoire. C'est une première : le niveau arbre avait jusqu'à présent été tout à fait absent car simplement absorbé par le calcul de valeurs moyennes de niveau placette. Ce progrès a des répercussions théoriques car il permet de résoudre de vieux problèmes forestiers d'échantillonnage.
- L'analyse théorique des estimateurs a révélé un ensemble de conditions et d'hypothèses nécessaires à l'absence de biais dans les estimations. Ces hypothèses vont jouer un rôle très important dans les réflexions à venir sur l'optimisation des méthodes d'échantillonnage dans l'IFN. Ces résultats auront donc des retombées scientifiques mais également des retombées pratiques très importantes pour l'IFN.

- Un simulateur d'échantillonnage et d'estimation a été développé. Il s'agit d'une suite de fonctions dans R permettant de simuler des échantillonnages complexes et d'implémenter diverses méthodes d'estimation. Ce simulateur va être accouplé à un jumeau numérique forestier de grande taille : pour la première fois, toutes les étapes de l'échantillonnage à l'estimation vont être groupées dans un seul simulateur capable d'implémenter des méthodes très complexes et représentatives des IFNs.

### ***Principales conclusions incluant des points-clés de discussion —***

- Tout d'abord il convient de souligner le fait que la thèse se déroule dans d'excellentes conditions, avec une doctorante très motivée et travailleuse, et les progrès sont conformes au programme déterminé lors du dépôt du projet Labex.
- L'analyse théorique des estimateurs de l'IFN, et plus généralement, des estimateurs complexes des inventaires forestiers nationaux modernes, constituait le premier objectif clé du projet. Il a été réussi et est achevé. Cette étape était la plus difficile de la thèse car elle supposait un très fort investissement de la part du doctorant. Elle a ouvert de très nombreuses perspectives.
- Les résultats obtenus au cours de cette première année sont déjà tout à fait considérables et sont actuellement synthétisés dans le but de soumettre une publication. Ce travail d'écriture vient de commencer.
- L'investissement fort dans des aspects théoriques va être complété par des travaux portant sur des simulations et l'analyse de données réelles. Ici aussi les retombées sont potentiellement très fortes tant dans le champ disciplinaire des statistiques de sondage que dans celui des sciences de l'environnement.

### ***Perspectives —***

- Cette première année marathon de la thèse nous a permis d'ouvrir un grand nombre de pistes. Cette thèse est la première thèse portant sur les méthodes d'inventaire forestier en France depuis 1986 (thèse de F. Houllier). Le déficit est donc substantiel et cette thèse offre un support pour reconstruire des compétences. D'où aussi la diversité des perspectives et des besoins.
- La doctorante a permis l'accrétion de compétences autour des travaux de thèse. Cette masse critique de connaissance devrait être mise en valeur, par exemple par de nouveaux travaux connexes.

### ***Valorisation —***

#### 1. Présentation lors d'une conférence

PhD / Postdocs Labex day 2022, Location: INRAE Champenoux, Date: October 11th, 2022

Type: Oral Presentation

Titre : Dynamic sampling and inference for a smart forest monitoring with application to the French NFI

#### 2. Présentation lors d'une conférence

Journée scientifique du LIF; Location: IGN, LIF, Nancy ; Date: November 28th, 2022

Type: Oral Presentation

Titre : Sampling and estimation methods of the French NFI

#### 3. Présentations lors d'une conférence (à venir, acceptée)

12<sup>th</sup> International Conference on Surveys ; Location : Aubervilliers Campus Condorcet, Date: 22nd to 24th March 2023

Type: Oral Presentation

Titre : Formalization of the sampling design and of the estimators used in the French NFI

## **Effet levier du projet —**

Dans un contexte où les spécialistes de sondage sont de plus en plus rares (en France et globalement dans le monde), tandis que les demandes et attentes de monitoring environnemental sont croissantes, la thèse permet de renforcer la communauté dans ces compétences clés ayant des débouchés bien identifiés.

Le rayonnement international est un élément très important de l'effet levier du projet. Ce rayonnement passe par deux « leviers » complémentaires :

- le levier de la collaboration : il est prévu que la doctorante fasse un séjour de recherche dans le laboratoire de sondages de Neuchâtel pour travailler avec Alina Matei. Ce laboratoire est celui dans lequel travaille aussi Yves Tillé, une référence mondiale en matière de sondages.
- le levier des publications : de nombreux aspects sont nouveaux et permettent de se repositionner parmi les laboratoires de pointe travaillant sur les aspects de monitoring et d'inventaire des ressources naturelles.