

DOCUMENTATION GENERALE SUR LES DONNEES BRUTES DEPUIS LA CAMPAGNE D'INVENTAIRE 2005

Préambule

De l'ancienne à la nouvelle méthode

L'ancienne méthode était basée sur des inventaires forestiers départementaux (un échantillon systématique de points assez dense à l'échelle départementale) renouvelés tous les dix à quinze ans, alors que la nouvelle méthode est basée sur un inventaire national annuel (un échantillon systématique de points moins dense à l'échelle nationale). L'ancienne méthode permettait de générer des résultats nationaux, par agrégation de résultats départementaux asynchrones. La nouvelle méthode fournit des résultats nationaux annuellement, avec une augmentation de la précision par agrégation de plusieurs campagnes annuelles successives : l'agrégation de cinq campagnes annuelles permet d'obtenir des résultats avec une précision suffisante à l'échelle régionale.

La centralisation des opérations d'inventaire a permis la mise en adéquation rapide de certaines définitions nationales avec des définitions internationales (définition de l'arbre et de la forêt) et a facilité l'harmonisation des protocoles au niveau national, ce qui contribue à garantir l'homogénéité des données collectées à l'échelle nationale. Avec la nouvelle méthode, lors de chaque campagne annuelle d'inventaire (une campagne annuelle n se déroule approximativement de novembre de l'année $n-1$ à octobre de l'année n), un ensemble de données est collecté au niveau national sur un échantillon d'environ 7000 « points forêt », c'est-à-dire avec une « couverture du sol boisée » et une utilisation non agricole et non urbaine.

La mise en ligne des données brutes

Cette nouvelle méthode d'échantillonnage permet de rassembler annuellement de nouvelles données sur la forêt française, que ce soit pour fournir des résultats de plus en plus détaillés et précis, ou que ce soit pour mettre à disposition du public des données brutes récentes sur la forêt. Cette volonté de mettre à disposition du public ses données brutes est encouragée par l'État (Etalab, data.gouv.fr, etc.). La première mise en ligne des données brutes a été réalisée en 2007, avec les 2 premières campagnes annuelles 2005 et 2006. En 2008 et 2009, les campagnes respectives 2007 et 2008 ont été publiées.

En 2010, les précédents jeux de données brutes ont été complétés par des jeux de données jusqu'à présent non communiqués, à savoir les données brutes disponibles sur les « **points peupleraie** » (couverture du sol peupleraie) en plus des « points forêt », que ce soit sur les placettes ou sur les arbres vivants. Les protocoles d'inventaire utilisés en peupleraie, bien que très proches des protocoles standard utilisés dans les autres forêts, diffèrent sur un certain nombre de données essentielles (en particulier l'accroissement et les données éco-floristiques), ce qui justifie la séparation des tables de données « forêt » et « peupleraie ».

En plus de ces deux nouvelles tables spécifiques, une table de données sur la description du **couvert des arbres recensables** sur les « points forêt » a complété utilement les données disponibles, avec un détail plus fin que les précédentes années où seules les données synthétiques de composition calculées sur le point d'inventaire étaient transmises.

Enfin, des **données phytosanitaires relevées pour le compte du Département Santé des Forêts** dans le cadre du contrat d'objectifs 2007 – 2012, sur les arbres vivants, ont également été ajoutées aux données brutes.

Chaque année (n), l'ensemble des informations est fourni pour une campagne supplémentaire (année $n-1$).

En 2011, l'ensemble des données relatives aux arbres « morts » (mort et chablis ordinaire) sont ajoutés pour l'ensemble des campagnes.

En 2012, la donnée « âge » est ajoutée dans les campagnes 2008 et 2009 des arbres vivants, en forêt et peupleraie.

En 2013, les données floristiques sont mises à jour pour toutes les campagnes selon la classification d'un unique référentiel (TAXREF v.5).

En 2014, une information sur les sylvoécorégions a été ajoutée systématiquement dans les données de type placette.

En 2015, les données brutes sur les points « deuxième visite » sont mises en ligne pour la première fois.

L'utilisation des données brutes

Les « données brutes », comme leur nom l'indique, sont des données pour la plupart directement « collectées » par les équipes d'inventaire sur les points « forêt » et « peupleraie », à partir desquelles des données synthétiques sont « calculées », pour permettre au service de l'inventaire forestier statistique de produire des résultats d'inventaire aux niveaux national et régional. En plus de « données brutes » issues de la « collecte », un certain nombre de données issues de « calculs » sont également mises à disposition du public, dans toutes les situations où les « données collectées » mises en ligne ne permettent pas d'obtenir ces « données calculées », à savoir la classe d'âge du peuplement dominant, l'essence principale du peuplement, et le volume et le poids des arbres vivants inventoriés. Dans toutes les autres situations, c'est-à-dire lorsque des données synthétiques peuvent directement être calculées à partir des données brutes mises en ligne, les données calculées correspondantes ne sont pas mises à disposition.

Les données brutes mises en ligne peuvent permettre, à un public averti, de faire des analyses de données et/ou des analyses cartographiques, par croisement des coordonnées du nœud d'inventaire auquel est rattaché le point d'inventaire avec des couches spatialisées, et à ce titre les données brutes peuvent notamment intéresser la communauté scientifique. En revanche, les données brutes ne peuvent absolument pas permettre à un utilisateur de recalculer des résultats d'inventaire, car l'utilisateur ne dispose ni du poids des points, ni des règles de post-stratification utilisées par le service de l'inventaire forestier. La documentation des données brutes mise à disposition avec les tables de données précise la définition des données, liste les modalités des données qualitatives et précise certaines conditions d'application (données corrélées).

Quels sont les points d'inventaire mis en ligne ?

La constitution de l'échantillon des points à lever repose sur deux étapes successives :

- 1- la création de l'échantillon des points à photo-interpréter ;
- 2- la sélection des points à lever à partir des données issues de la photo-interprétation.

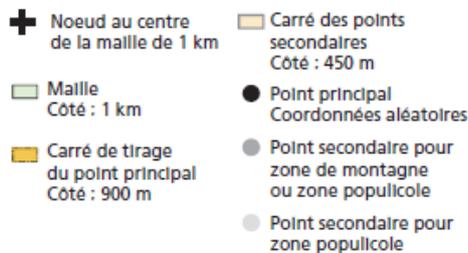
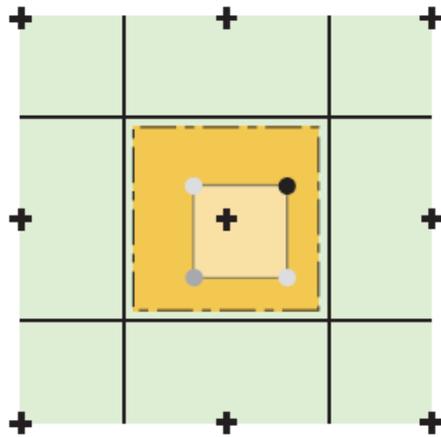
Suite à ces deux étapes, des points de couverture du sol boisée sont effectivement inventoriés par des équipes IGN sur le terrain et ces points fournissent la totalité des données brutes mises en ligne sur Internet.

Création de l'échantillon des points à analyser par photo-interprétation

Jusqu'en 2014

À partir d'une grille systématique qui couvre le territoire de mailles carrées de 1 km de côté, et dont 1/10^e est parcouru chaque année, 1 point d'inventaire (dit « point principal ») est tiré aléatoirement dans chaque maille. Ce dispositif standard est localement densifié, en fonction des objectifs de l'inventaire, en tirant 3 points supplémentaires par maille (les quatre points formant un carré de 450 m de côté) en « zone populicole » (zone cartographiée où la populiculture est plus développée) et en tirant 1 point supplémentaire par maille en « zone de montagne » (altitude supérieure à 1200 m dans les Alpes, les Pyrénées et en Corse).

Ces deux modifications ont entraîné une augmentation sensible de la taille de l'échantillon de points photo-interprétés (près de 60 000 points photo-interprétés pour la campagne 2006, et environ 100 000 points photo-interprétés pour les campagnes suivantes).



■ Zone populiicole ■ Zone montagne
Points d'inventaire de phase 1
(photo-interprétation ponctuelle)

A partir de 2015

À l'issue des dix premières années d'inventaire selon la nouvelle méthode (campagnes 2005 à 2014), l'ensemble de la grille a été parcourue : toutes les mailles de 1 km² ont été visitées au moins par photo-interprétation ponctuelle. L'échantillonnage des cinq années suivantes (campagnes 2015 à 2019) a été alors assis sur une grille correspondant aux mailles formées par les cinq dernières années de la grille décennale initiale (second cycle quinquennal).

Depuis 2010, une partie des points échantillonnés étaient revisités pour l'évaluation des prélèvements. Cependant, les évolutions de la méthode d'inventaire amènent à la mise en place d'un plan d'échantillonnage avec des points revisités systématiquement 5 ans après le passage initial (et pas uniquement les points d'inventaire concernés par les prélèvements à compter de la campagne 2016).

Ainsi, la grille a été reformatée pour que les points nouveaux et les points à revisiter d'une même campagne soient dans la même maille (attachés au même nœud, même niveau de tirage). Ainsi, une grille formée de cinq ensembles de mailles rectangulaires de 2 km se dessine.

Le choix de ce schéma est un compromis entre deux conditions contraires :

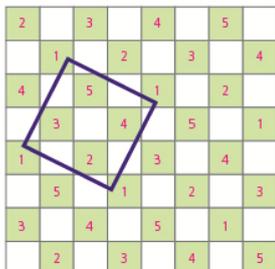
- proximité des points afin d'optimiser les coûts de déplacement dès lors que point « première visite » et point « deuxième visite » font partie de l'échantillon terrain (tend à minimiser l'écart de distance) ;
- écartement des points afin de minimiser la perte d'information statistique due à une trop grande corrélation spatiale entre point « première visite » et point « deuxième visite » (tend à maximiser l'écart de distance), les deux points faisant partie du même échantillon statistique pour produire des résultats.

Le tirage reste aléatoire, puisque :

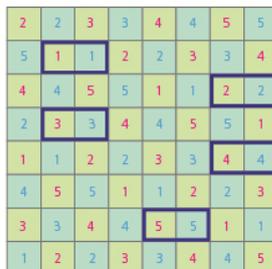
- la grille initiale a été positionnée aléatoirement
- les points ont été tirés aléatoirement dans chaque maille 5 ans auparavant.

Cela se traduit de la manière suivante :

Fractions annuelles à l'intérieur
du premier cycle quinquennal
(2005 - 2009)



Toutes les mailles sont parcourues
en 2 cycles appariés de 5 fractions
annuelles chacun

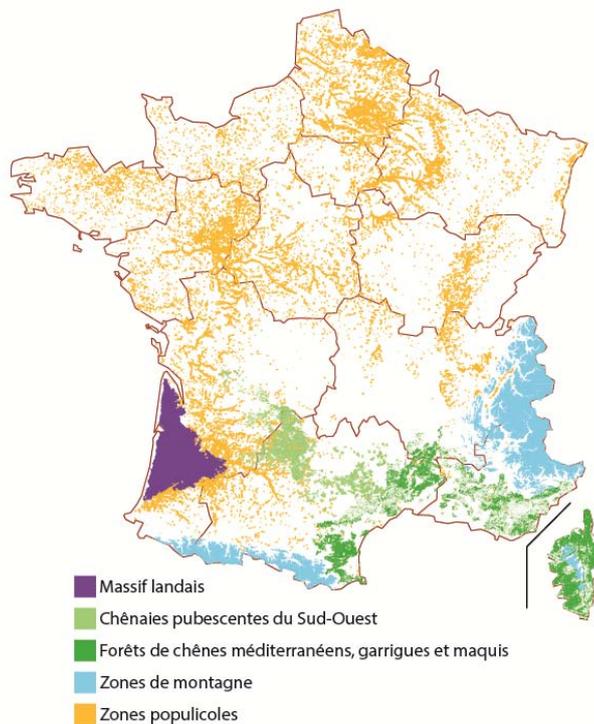


Premier cycle quinquennal
(2005 - 2009)

Deuxième cycle quinquennal
(2010 - 2014)

Afin de limiter les coûts de collecte, le territoire a de plus été partagé en différentes zones dans lesquelles la densité d'échantillonnage des points à visiter peut être divisée par deux ou des optimisations du tirage mises en œuvre. On trouve :

- des étendues homogènes de forêts, comme le massif landais ou les chênaies pubescentes du Sud-Ouest ;
- des forêts de type garrigue ou maquis, présentant un intérêt limité pour la production de bois ;
- les forêts de montagne (altitude > 1200 m et pente > 30 % ou altitude > 1700 m) ;
- les zones populières avec une forte probabilité de présence de peupleraies.



Sélection du sous-échantillon des points à inventorier sur le terrain

En moyenne, un point « photo-interprété forêt » sur deux est sélectionné pour constituer l'échantillon des points à inventorier sur le terrain, sauf dans les zones dites allégées, ou seulement un point « photo-interprété forêt » sur quatre est sélectionné. Un point « photo-interprété forêt » est un point dont la couverture du sol semble boisée à partir de l'analyse de la photo-aérienne : la couverture du sol est ensuite confirmée ou infirmée par les équipes de terrain. Les zones dites « allégées », sont des zones dans lesquelles les zones boisées sont relativement homogènes (massif landais, chênaies pubescentes du sud-ouest, maquis du sud-est), pour lesquelles un échantillonnage plus dense n'apporterait pas plus de précisions sur les estimateurs.

Sur 100 000 points photo-interprétés, environ 7000 points sont inventoriés sur le terrain, dont 6500 points environ correspondent à des points d'inventaire de couverture du sol boisée (fermée, ouverte ou peupleraie).

Points d'inventaire retenus dans les données brutes

Seuls les points d'inventaire de couverture boisée confirmée sur le terrain et complètement inventoriés sont retenus dans les données brutes.

Par conséquent, toutes les catégories de point d'inventaire suivantes sont exclues des données brutes :

- 1- exclusion de tous les points avec une couverture du sol non boisée, y compris les points de couverture du sol de type « lande » par ailleurs inventoriés, ou les points « non forêt » à proximité d'une formation linéaire arborée ;
- 2- exclusion de tous les points d'accès impossible ou difficile, sur lesquels toutes les opérations d'inventaire n'ont pu être menées à terme ;
- 3- exclusion de tous les points de couverture du sol boisée, dont l'utilisation du sol ne conduit pas à réaliser un levé complet, en raison d'un intérêt limité.

Au bilan, les points d'inventaires mis en ligne correspondent aux seuls points forêt et peupleraie levés, c'est-à-dire à des points sur lesquels toutes les opérations d'inventaire standard ont été mises en œuvre, soit donc environ 6 500 points par an. Tous les points forêt et peupleraie effectivement levés sur le terrain, sont mis en ligne, que des arbres recensables ($D_{13} \geq 7,5$ cm) aient été effectivement mesurés ou non sur le point.

NB : Depuis 2015, des données brutes sont également fournies sur les points « deuxième visite » qui sont un sous-échantillon des points levés cinq ans auparavant.

Quelles sont les années et données mises en lignes ?

Un ensemble de table de données est mis en ligne pour chaque campagne annuelle d'inventaire.

Une campagne d'inventaire annuelle n de déroule de novembre de l'année $n-1$ à octobre/novembre de l'année n .

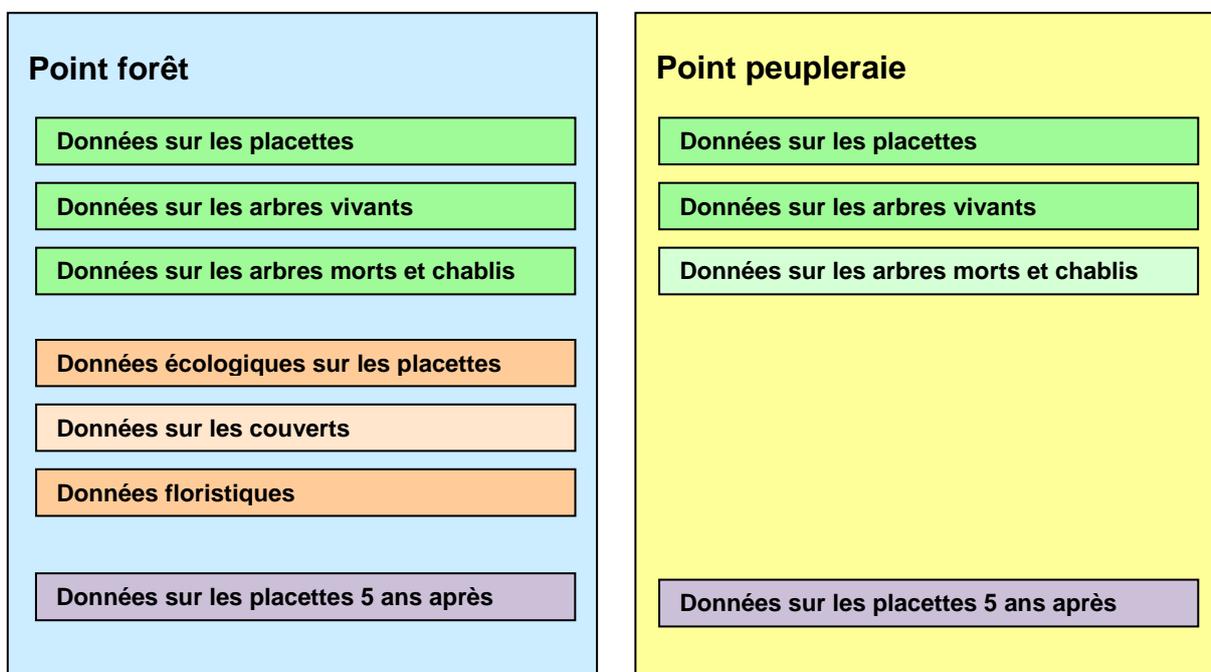
La première campagne annuelle d'inventaire (campagne 2005) selon la nouvelle méthode s'est déroulée de novembre 2004 à octobre 2005.

Pour chaque table de données, et pour chaque année d'inventaire si nécessaire, sont mis en ligne :

- 1- une documentation technique des variables mises en ligne ;
- 2- un fichier de correspondance codes-libellés-définitions pour toutes variables qualitatives.

Pour les informations sur les points « deuxième visite », celles-ci sont disponibles en vis-à-vis des données de l'année $n-5$.

Tables de données mises en ligne



Les 7 tables de données (en vert et orange vif) :

- « données sur les placettes » (forêt et peupleraie) ;
- « données sur les arbres vivants » (forêt et peupleraie) ;
- « données sur les arbres morts et chablis » (forêt) ;
- « données écologiques sur les placettes » (forêt) ;
- et « données floristiques » (forêt).

sont disponibles pour les toutes les campagnes d'inventaire.

La table de données « données sur les arbres morts et chablis », en point peupleraie, est disponible depuis la campagne 2008.

La table de données « données sur les couverts », en point forêt, est disponible depuis la campagne 2006.

Les deux tables de données sur les placettes revisitées 5 ans après portent l'année de la campagne d'inventaire $n-5$ (ex. : les points de la campagne 2005 revisités en 2010 sont nommés « _5_2005 »).

Nom des fichiers

Les fichiers de données brutes sont des fichiers CSV. Leur nom est composé de la thématique, du domaine d'étude et de la campagne.

Les fichiers de documentation sont des fichiers PDF ou des fichiers CSV. Ces derniers servent à faire la jointure entre le code d'une modalité d'une donnée et son libellé ou sa définition.

Par exemple, pour la campagne d'inventaire initiale 2008 :

Thématique des données	Domaine d'étude	Nom du fichier de données	Nom des fichiers de documentation
Arbre vivant	Forêt	arbres_foret_2008.csv	Doc-DB_arbres_PF_20xx.pdf Doc-DB_arbres_PP_20xx.pdf + documentation_2008.csv
	Peupleraie	arbres_peupleraie_2008.csv	
Arbre mort et chablis	Forêt	arbres_morts_foret_2008.csv	Doc-DB_morts-chablis_PP_20xx.pdf Doc-DB_morts-chablis_PP_20xx.pdf + documentation_2008.csv
	Peupleraie	arbre_morts_peupleraie_2008.csv	
Couverts	Forêt	couverts_foret_2008.csv	Doc-DB_couverts_PF_20xx.pdf + documentation_2008.csv
Ecologie	Forêt	ecologie_2008.csv	Doc-DB_ecologie_PF_20xx.pdf + documentation_2008.csv
Flore	Forêt	flore_2008.csv	Doc-DB_flore_PF_20xx.pdf + documentation_flore.csv
Placettes première visite (=initiale)	Forêt	placettes_foret_2008.csv	Doc-DB_placettes_PF_20xx.pdf Doc-DB_placettes_PP_20xx.pdf
	Peupleraie	placettes_peupleraie_2008.csv	
Placettes deuxième visite (5 ans après la visite initiale)	Forêt	placettes_foret_5_2008.csv	Doc-DB_retour-placette_PF-PP_20xx.pdf + documentation_5_20xx.csv
	Peupleraie	placettes_peupleraie_5_2008.csv	

NB1 : Pour les données des points deuxième visite (relevés cinq ans après la première visite), le fichier porte la même année que celle de son observation initiale. Les points seconde visite sont à rattacher aux points première visite par l'identifiant de placette.

NB2 : Le fait de fournir le code d'une donnée modale dans le fichier, et la correspondance entre codes et libellés dans un fichier à part entière, présente deux avantages : les fichiers de données brutes sont moins volumineux, d'une part, et donc plus rapides à télécharger, et, d'autre part, ils sont plus simples à importer dans des logiciels de traitements statistiques ou des bases de données.

Format des fichiers de données

Les fichiers de données brutes (CSV) ont un séparateur point-virgule (;) pour éviter toute confusion avec la virgule en tant que séparateur décimal.

Chaque fichier de données brutes commence par une première ligne d'en-tête contenant le nom des données fournies. Chacune des lignes suivantes correspond, selon le fichier, à un point inventorié (données « point »), à un arbre (données « arbre ») ou à une espèce (données « flore »). Une valeur non renseignée (NULL en base de données) se repère par l'absence de caractères entre deux points-virgules.

Les **coordonnées géographiques** sont renseignées dans les fichiers "placettes" (point forêt ou point peupleraie). Pour les autres thématiques, l'identifiant placette unique permet de faire le lien avec la placette concernée.