

## GENERAL INFORMATION ON RAW DATA

### WHAT HAS CHANGED SINCE THE 2005 INVENTORY SURVEY?

<b>FOREWORD</b>	<b>1</b>
WHAT IS RAW DATA?	1
USING RAW DATA	2
PUTTING RAW DATA ONLINE	2
INVENTORY PLOTS PROVIDED IN DATAIFN	2
'SECOND VISIT' INVENTORY PLOTS	3
<b>WHAT YEARS AND VARIABLES ARE AVAILABLE ONLINE?</b>	<b>3</b>
EXPORT FILES	3
ACCURACY OF COLLECTED DATA (NEW IN 2025)	4
TEMPORAL VALIDITY OF VARIABLES AND UNITS	5
<b>APPENDIX 1: STATISTICAL METHODOLOGY</b>	
<b>SCOPE AND SURVEY UNITS</b>	<b>6</b>
SAMPLING UNITS	6
EXCLUSION AREAS	6
<b>SAMPLING DESCRIPTION</b>	<b>7</b>
SAMPLING FRAME	7
SAMPLING	7

## Foreword

### What is raw data?

Forest inventory "raw data" supplied via DataIFN consists of data collected directly by field teams on plots located in forests (including poplar groves) and copses as part of the "National Forest Inventory" statistical survey (see Appendix 1), labelled by the National Committee for Statistical Information (CNIS) since 2017.

The data provided in the DataIFN tool may, however, sometimes be the result of a calculation.

In fact, it has been decided to provide some non-raw data, particularly in situations where the collected data made available online does not allow these calculated data to be obtained.

This is notably the case for "Volume (V)" data, which is calculated from raw data and scaling rates, or "Tree weighting coefficient (W)" data.

In all other situations, i.e. when synthetic data can be calculated from the raw data made available online, or when the data do not come from the forest inventory (e.g. forest ownership data), the corresponding calculated data is not available through the DataIFN tool.

Some calculated data may be provided in other webpages. This is particularly the case for ecological indices, which are available on the following page: <https://inventaire-forestier.ign.fr/?rubrique262>

## Using raw data

Raw data posted online **can be used to:**

- **visualise** the distribution of one (or more) modality of an attribute (e.g. the distribution of a tree species, a soil type, etc.),
- **carry out data analysis** and/or cartographic analysis, by cross-referencing the coordinates of points with spatialised layers.

As such, raw data may be of interest to the scientific community as well as to a wider public.

A set of documents is attached to the raw data, in particular the metadata.csv file, which provides information on data definition (code, description and definition) and unit(s) (validity period, terms and conditions, etc.).

However, **in no way can raw data be used to recalculate inventory results**, as the user does not have access to either the weight of the points or the post-stratification rules used by the forest inventory calculation service.

To obtain inventory results, IGN provides users with the OCRE tool<sup>1</sup> :

<https://inventaire-forestier.ign.fr/?rubrique226>

## Putting raw data online

Since 2007, the forest inventory has made available the raw data collected using the 'new method' protocol<sup>2</sup>.

The DatalFN tool, launched on 8 December 2021, now provides:

- a single tool for viewing and downloading raw data;
- more variables to explore;
- customised downloading of data by geographical area or time period.

## Inventory plots provided in DatalFN

Only fully inventoried points which fall within field-confirmed woodland cover are included in the DatalFN tool.

As a result, all the following inventory point categories are excluded from the raw data:

- 1- All points with non-wooded ground cover are excluded, including 'heathland' type ground cover points that are otherwise inventoried, or 'non-forest' points close to a linear tree formation;
- 2- All points that are not accessible or are difficult to access, and where it has not been possible to carry out all the survey operations, are excluded;
- 3- All points with wooded ground cover with a land use that doesn't allow thorough surveying are excluded. They include, for example, wooded areas with predominantly agricultural or urban use.

Basically, the inventory points provided only correspond to forest (including poplar groves) and copse locations that have been surveyed, i.e. points on which all the standard inventory operations have been carried out, i.e. around 6,000 points per year. Their characteristics are accessible via the DatalFN tool, whether or not recordable trees ( $D13 \geq 7.5$  cm) have actually been measured on the inventory point.

<sup>1</sup> The 'General public' version accessible on the website provides inventory results at national and regional level, for the last 5 inventory surveys, and for a selection of data.

'Pro' versions can be configured to broaden access (department, all surveys, more data, etc.). Access to the 'Pro' version requires specific training, the signing of an agreement to make the tool available and the configuration of users' accounts.

<sup>2</sup> The 'new inventory method' was introduced at the end of 2004 (for the 2005 survey). The main difference between this method and the previous one is that all the departments in mainland France are surveyed every year, during each annual survey (continuous national survey).

An annual survey runs from approximately November of year n-1 to October of year n. A set of data is collected nationwide on a sample of around 6,000 to 7,000 'forest sites', i.e. with 'wooded ground cover' and non-agricultural, non-urban use. With the "previous method", used between the 1960s and 2004, only a small number of departments were surveyed each year and a cycle of 10 to 15 years was needed to cover the entire country). It did not allow new data to be collected annually on the whole of the French forest (either to provide results or to make recent raw data on the forest available to the public).

---

## 'Second visit' inventory plots

Raw data is also provided for 'second visit' plots.

The trees and plots concerned are included in the raw data files along with the data collected on the first visit.

The 'VISITE' variable in the plot.csv file indicates whether the plot is covered by the first or second visit protocol.

From **2010 to 2014**, the forest inventory carried out a repeat visit of the plots visited 5 years before. The sole purpose of this revisit was to quantify forest removals. It was only carried out on the inventory plots that had at least one tree that could be recorded during the first visit. On these plots, the felling (or not) of trees present during the first visit was the only information recorded.

With the **2015** survey, information is enhanced as more attributes are recorded during the second visit to the plot. All the trees measured during the first visit are measured again (or noted as felled or dead) and the newly recordable trees present on a 5-m plot are counted.

**As of 2016**, all the plots visited once will be visited a second time. In addition to the observations and measurements made on the trees, the field officers record new observation, such as land cover and land use, and collect data on changes in the stand attributes (planting, felling, etc.).

The data collected during the 2nd visit is therefore generally more limited (especially before 2016) and some data is not collected (ecological variables in particular).

---

## What years and variables are available online?

**All inventory surveys since 2005 are accessible.**

By default, the tool filters through the last 5 surveys.

It is possible to select more or fewer surveys.

**Over 160 variables are available.** The list is displayed on the filter criteria selection screen (as well as on the export data selection screen).

However, each variable is valid for a specific period, therefore not all variables are available simultaneously. The variable list adjusts to the selected inventory surveys. **Greyed-out variables are not available for the entire selected period.**

Some variables are not offered as a filter criterion and are automatically supplied for export. This applies to the plot identifier and geographical coordinates, as well as the "tree weighting coefficient (W)".

Each year, a new inventory survey is uploaded into the tool, adding a new survey and, if applicable, new data.

The tool's home page allows you to download two files:

- One containing all the raw data made available, for all surveys since 2005.
- The other containing all the raw data made available, for the latest survey available.

These files are updated annually, when a new inventory survey is released.

The database (and the 2 files) may be updated during the course of the year. The date of the last update is indicated on the site's home page.

---

## Export files

Variables are grouped into 7 files (data table):

- ARBRE.csv
- BOIS\_MORT.csv
- COUVERT.csv
- ECOLOGIE.csv
- FLORE.csv
- HABITAT.csv
- PLACETTE.csv

These files are supplied with an associated PDF documentation file.

The number of files supplied for export depends solely on the variables requested for filtering and export.

## Raw forest inventory data available online on DataIFN

Other documentation files may be supplied:

- espar-cdref13.csv
- metadata.csv
- IGN\_DB\_general\_documentation.pdf
- precision-donnees\_2020-2024.csv

The "espar-cdref13.csv" correspondence file is only supplied if the "Reference taxon (CD\_REF)" variable is requested.

Metadata and general documentation files are systematically supplied.

Each raw data file (with the exception of the metadata file) begins with an initial header line containing the name of the variables supplied. Each of the following lines corresponds, depending on the file, to:

- an inventory plot ("PLACETTE", "ÉCOLOGIE" and "HABITAT" files)
- a tree ("ARBRE" and "BOIS\_MORT" files)
- a species ("FLORE" and "COUVERT" files).

An missing value (NULL in the database) can be identified by the absence of characters between two semicolons.

The espar\_cdref13 metadata and match files are used to join the modality code of a variable to its label or definition.

Providing the modality code of a factorial variable in the file, and the matches between codes and labels in a separate file, has two advantages: on the one hand, raw data files are less voluminous, and therefore quicker to download, and, on the other hand, they are easier to import into statistical processing software or databases.

**Geographical coordinates** are entered in the "PLACETTE" file. For other data categories, the unique plot identifier (IDP variable) is used as a link with the inventory plot location.

**Please note** that automatic opening of csv files in certain spreadsheet programs (e.g. Microsoft Excel) may unwittingly convert the format of certain columns (10/01 → October 1).

It is therefore recommended that you open your software and then open (import) the file, specifying a ";" delimiter and a text format for the columns (especially the first two columns).

---

### Accuracy of collected data (new in 2025)

The precision-donnees\_2020-2024.csv file provides information on the accuracy with which data is collected by field teams. To assess this, their data are compared to those of the "field quality monitoring" teams.

The data collected can be of two types: qualitative or quantitative.

- In the case of quantitative data (e.g., circumference measurement), the *precision* is provided in a unit of measurement (indicated in the *Unité* column), simply by comparing the two numerical values.
- In the case of qualitative data:
  - When the data modalities are ordered (e.g., the abundance of a species), the comparison of the *correspondance* is done by searching for an exact match of the given modality, or a close match (with one or two modalities).
  - When the data modalities are not ordered (e.g., a type of accident), the comparison of the correspondence is done according to a broader criterion (*critere* : presence/absence, identical gender, etc.) and then a correspondence calculation.

For some data, accuracy or correspondence is provided for both the first and second *visite* collection.

The "2020-2024" file combines the five campaigns from 2020 to 2024 and provides the average accuracy of these five campaigns for over a hundred data.

The "field quality monitoring" is carried out on 4% of the inventory plots, over a part of the territory. These plots are reviewed on average 2 months after the teams' survey. This may lead to some discrepancies in the data between the two teams, without this being considered abnormal and being taken into account in the correspondence and precision calculations.

## Temporal validity of variables and units

The forest inventory protocol may evolve with each new survey in various ways if:

- a given variable gets discontinued;
- variables are added;
- there is change in the modalities of the variable's unit – which stops the use of a unit and starts a new unit associated with a given variable.

The “metadata.csv” file shows the surveys for which a variable and its unit are valid.

For example, it shows:

- 3 successive units for the ESPAR variable (ESPAR from 2005 to 2006, ESPAR0 from 2007 to 2008 and ESPAR1 since 2009). The variable has therefore existed since 2005, but its units (and modalities) have changed.

```
//Unit definition
//Data / Unit / Survey / Type / Label / Definition
ESPAR / ESPAR / 2005-2006 / Qualitative / Tree species / Botanical tree species or detailed species group
ESPAR / ESPAR0 / 2007-2008 / Qualitative / Tree species / Botanical tree species or detailed species group
ESPAR / ESPAR1 / 2005-2006 / Qualitative / Tree species / Botanical tree species or detailed species group
```

- 2 successive units for the SFO variable (SF from 2005 to 2006 and SF0 from 2007 to 2013). The variable existed from 2005 to 2013, and SFO no longer exists as of 2014.

```
Data / Unit / Survey / Type / Label / Definition
```

```
SFO / SF / 2005-2006 / Qualitative / Forest structure / Complex characteristic of a stand combining regime and vertical
```

```
SFO / SF0 / 2007-2013 / Qualitative / Forest structure / Complex characteristic of a stand combining regime and vertical
```

These modifications may also depend on the study area.

For many surveys, the protocol for plots in poplar groves was slightly different from that for other “forest” plots.

For example, the floristic survey and ecological variable have only been carried out in poplar groves since 2016 (and were released in autumn 2022 when the 2021 survey was released), and some data is specific to poplar groves (e.g. “Poplar grove maintenance (ENTP)”).

This difference does not show up directly in the metadata, as all points (poplar grove or “forest”) are grouped together in the files. The metadata shows the broadest period (for example, branch mortality data collection (MORTB) started in 2006 on forest plots, whereas it started in 2007 on poplar grove plots. Therefore the variable is noted as available as of 2006). For the 2006 survey, the variable takes a ‘NULL’ for poplar grove plots.

## APPENDIX 1: STATISTICAL METHODOLOGY

### Scope and survey units

The scope of the survey covers two populations:

1. The whole of mainland France and Corsica, described in terms of (projected) surface areas according to land cover and land use, in which **forest**, and more particularly the forest known as “available for production”, is the main purpose. This population is made up of all the points (infinite population) belonging to the territory, with known area A.
2. All recordable trees (with a minimum circumference of 23.5 cm at a height of 1.30 m) in forests “available for wood supply”. These trees are a finite population.

According to the international definition applied in France, a “forest” is defined as an area covered by more than 10% trees, over a surface area of at least 0.5 ha, with a width of 20 m. A tree is a woody specie capable of growing to more than 5 m at maturity *in situ*. Areas that meet these definitions but are used primarily for agricultural or leisure activities are not considered as part of the forest. Thus, a wooded park is not a forest.

This is a continuous survey, based on the annual drawing of a sample covering the whole country.

### Sampling units

Formally, there are two types of sampling unit:

1. locations on the territory (infinite population)
2. trees (finite population)

These two types of unit are closely linked, with trees selected around inventory points via circular observation plots (of constant area) centred on the points. These plots are used to collect a range of data and to select the sample of trees to be measured *in the field*. Tree sampling is therefore indirect, based on the selection made possible by the concentric plots (6, 9 and 15 m radius) around the point.

### Exclusion areas

The only exclusion areas are those parts of the country that are strictly prohibited to aerial photography for security reasons (military bases, airfields, etc.). These areas are determined by the military authorities and change over time. The selected inventory points in these areas are directly classified in a distinct land use category (“black-out areas”), and they do not get visited, even though some of these areas undoubtedly fall within the definition of forest.

## Sampling description

### Sampling frame

The **study area** is the territory, i.e. the country (or an administrative division: e.g. department). Within this territory, the forest constitutes an unknown area of interest prior to the survey. This territory, with its known outline and surface area (the boundaries of which correspond to those of the BD Carto IGN of 2002 in the Lambert II extended projection) is the "sampling frame" for all surface quantities, particularly forest area.

All the trees located in the forest, which form the tree population targeted by the survey, are unlocated, of finite (but very large) size and unknown prior to the survey. There is therefore no formal tree sampling frame.

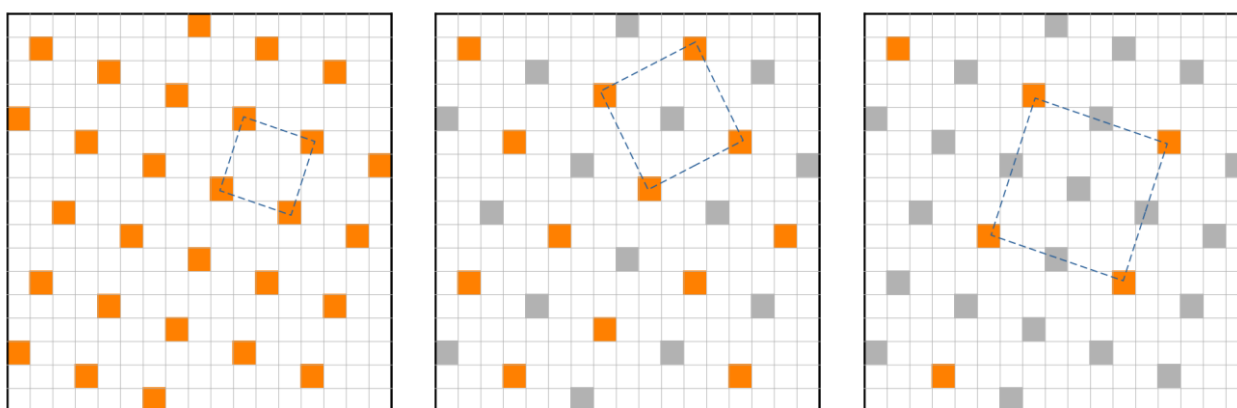
### Sampling

The sampling plan consists of two statistical phases.

The aim of the first phase is to identify likely or potential woodland cover points and definite non-woodland cover points on orthorectified aerial photographs (BD Ortho).

In the second phase, a sub-sample is drawn from the potential or likely woodland cover to be checked in the field and measured where necessary. The other categories of points not visited in the field are sub-sampled at 100%, i.e. all the points are retained in the phase-2 sample.

The sampling plan is based on a systematic grid with a 1-kilometre square mesh covering the whole of mainland France and Corsica. The grid is divided into 10 sections, each being used for annual sampling. This is referred to as an annual fraction of the initial grid or, more simply, as an 'annual grid', with each fraction also covering the territory in square meshes with a surface area of 10 km<sup>2</sup> (Figure A).



1 - 10-km<sup>2</sup> grid

2 - 20-km<sup>2</sup> grid

3 - 40-km<sup>2</sup> grid

Figure A: First three levels of an annual fraction (annual grid).

Furthermore, each annual grid is structured into nested levels that allow the sampling level to be adjusted by a factor of 2 between each level (see Figure A). The points selected for the second phase are based directly on the nested levels of an annual fraction.

The statistical weight of a point is then determined by its sampling level. It is unitless. By convention, this weight is 1 for the points in the level 1 mesh of the complete annual grid. The weight is 2 if the sample consists of level 2 points.

### First-phase annual sample

The first-phase sample is systematically distributed spatially. It is based on the sampling of one point (random coordinates) within each kilometre grid cell (1 x 1 km) belonging to level 1 of the annual fraction of the grid. This results in 1 point per 10 km<sup>2</sup> (see figure A).



For meshes bordering the territory, the points are drawn in the portion of the mesh belonging to the territory, only for meshes the centre of which is located in the territory.

The first phase is carried out each year by photo-interpretation (PI), on IGN's BD Ortho®, of a circular plot with a radius of 25 m, centred on the inventory location drawn at random from the grid. This phase covers around 55,000 points, and provides information on the coverage and extent of the forest. It also provides information on land cover and land use. When land cover or land use is heterogeneous over the surface of the plot, only the part where the point fell is described. If there is any doubt between two categories, the photo-interpreter must classify the point in a category that will be the subject of a field check (sub-sampling) and possible measurements in the second phase.

### Second-phase annual sample

The second-phase sample is formed by sub-sampling locations from the first phase.

The selection of points in phase 2 is based on their classification (ground cover, use, etc.) in the first phase, which distinguishes between the 'visited' categories, which will be sub-sampled, and the 'unvisited' categories, for which the information gathered in the first phase (photo-interpretation) is considered definitive. For this second point category, it means that they are not likely to contain wooded cover corresponding to the international definition of forest (e.g. locations on agricultural land, in recreation parks, in urban areas, etc.) and are not directly targeted by the survey. These points are all part of the final (phase 2) sample and retain their phase 1 weight ('sub-sampling' at 100%).

For the categories visited, the second-phase points form a sub-sample of the phase-1 points. Sampling rates are typically as follows:

- one point in two for woodland cover locations (closed, open, copse and poplar grove): level-2 grid;
- one point out of four for moorland cover locations: level-3 grid;
- one point out of eight for linear woodland formations: level-4 grid.

The sampling is therefore of **unequal probability**.

Figure B illustrates the sub-sampling of 50% for forest (visited category) and 100% for the non-visited category, simplified as 'Non-forest' (e.g. farmland, urban area, etc.).

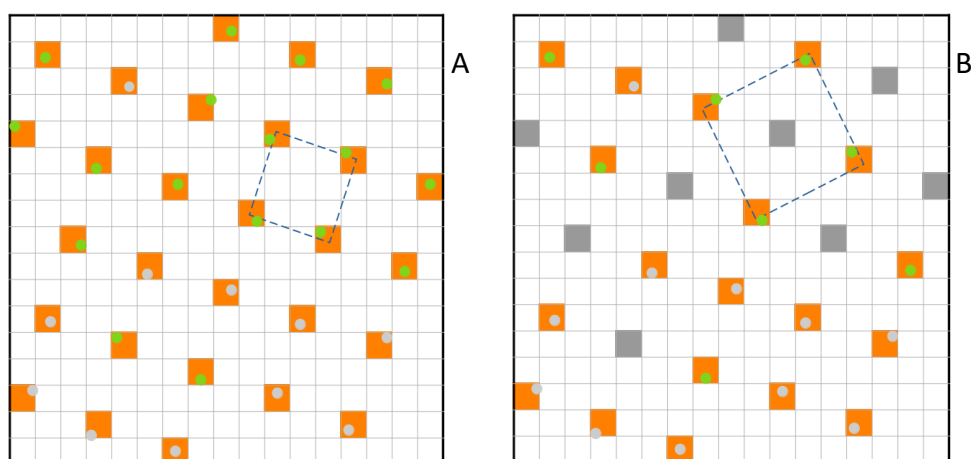


Figure B : Illustration of phase-1 (A) and phase-2 (B) samples.

(A) In the first phase, a point is drawn at random from all the meshes in the level-1 annual grid (orange meshes forming a square grid of 10 km<sup>2</sup> symbolised by dashed lines). The green points are photo-interpreted 'Forest' points, a category to be visited in the field. The light grey points belong to the 'Non-forest' category, not visited in the field.

(B) In the second phase, all the 'Out of Forest' points are retained (100% sub-sampling, level 1 of the grid) and one point out of 2 in the forest (50% sub-sampling) based on the level-2 square grid of 20 km<sup>2</sup>, represented in dashed lines. The grey squares are the grid cells containing the 'Forest' points from phase 1, which were not retained in phase 2. The orange squares are those containing the points from the phase 2 sample.



The selection rates for the second phase may be revised annually by the department in charge of the inventory (SIF) depending on the results of the first phase (number of points in the forest category, in a context of rapidly expanding French forests ~ +80,000 ha per year) and the resources allocated (number of field staff) in order to ensure that field measurements are carried out within the allotted time (12 months). These adjustments are based on the levels of the grid and thus involve deleting high-level points in order to bring into line the number of points visited.

### Selecting and measuring trees in phase 2

The description and measurement system depends on the type of point visited. In forests, circular plots centred on the inventory location are used to collect information on trees (living, dead, etc.) and a large number of variables according to a complex survey protocol, which is updated for each new annual survey.

The plots are used to sample the targeted tree population in the vicinity of the second phase points. In the estimation method used, the variables measured on the trees are all expressed in terms of surface density (per hectare, for example for volume, in m<sup>3</sup>/ha), and thus brought back to the central point by a system of weights assigned to each tree. Since the trees are measured on circular plots with radii varying from 6 to 15 m, the size of the plot on which each individual is selected has to be factored in when extrapolating values to a per-hectare basis.

In the simplest case scenario (absence of forest/non-forest boundary), the weight  $w$  of the trees is directly provided by the inverse of the size of the survey plot. The weights calculated in this way are, in 'number of stems per hectare',  $10,000 / (\pi \times \text{plot\_radius}^2)$ , i.e. :

- 88,4 for trees with a small circumference (6-m plot);
- 39,3 for trees with a moderate circumference (9-m plot);
- 14,1 for trees with a large circumference (15-m plot).

Where a plot overlaps a boundary (forest/non-forest boundary) or two different stands, tree measurements are only taken if the point is in the forest, and only on the stand concerned. A correction system (a function of the relative positions of the trees and the boundaries) is then applied to take account of the lower probability of selection of the edge trees, which are clearly different from those in the centre of the stand.