



**HAL**  
open science

## **Les bases de données : outil indispensable de l'expérimentateur**

Alain Bailly, François Guinaudeau, Mohamed Najar

### ► **To cite this version:**

Alain Bailly, François Guinaudeau, Mohamed Najar. Les bases de données : outil indispensable de l'expérimentateur. *Revue forestière française*, 1993, 45 (S), pp.54-62. <10.4267/2042/26501>. <hal-03444057>

**HAL Id: hal-03444057**

**<https://hal.science/hal-03444057v1>**

Submitted on 23 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# **LES BASES DE DONNÉES :**

## **OUTIL INDISPENSABLE**

### **DE L'EXPÉRIMENTATEUR**

**A. BAILLY - F. GUINAUDEAU - M. NAJAR**

Tout expérimentateur a besoin de données entachées du moins d'erreurs possible et facilement utilisables, donc ordonnées et archivées dans des systèmes de Bases de Données.

Les erreurs viennent soit d'erreurs de mesure (lecture sur l'instrument, compas, mètre, perche, ...), soit d'une mauvaise saisie (recopie) de cette mesure avant son traitement mathématique.

Il faut donc des équipes formées, habituées à la lecture des instruments et conscientes des risques de biais de telle ou telle méthode de mesure. Il convient, également, de réduire le plus possible les saisies manuelles de données, opération qui est, de plus, fastidieuse et coûteuse.

C'est pourquoi, l'Association Forêt-Cellulose (AFOCEL) développe, depuis maintenant une vingtaine d'années, un système de traitement informatisé de ses données expérimentales. Ce système a évolué avec les matériels informatiques et la plus grande habitude des personnels de terrain. Il permet actuellement de passer, sans nouvelle saisie, de la mesure sur l'essai à son interprétation statistique et son archivage.

Nous avons ainsi constitué plusieurs bases de données de nos résultats expérimentaux, qui reflètent l'essentiel de nos préoccupations en matière de recherche :

- sélection d'un matériel végétal adapté au milieu et productif,
- sylviculture intensive.

Nous les présenterons dans la suite de cet article, en précisant leur place dans la gestion des essais, leur description et enfin quelques exemples d'utilisation.

#### **LA PLACE DES BASES DE DONNÉES DANS LA GESTION INFORMATISÉE DES ESSAIS**

Le système de gestion informatisé des essais de l'AFOCEL est présenté dans la figure 1 (p. 55), et détaillé ci-après.

##### **La saisie**

Elle se fait directement sur le terrain par l'emploi d'un appareil portatif Husky-Hunter. Décrit par Guinaudeau et Tonarelli en 1987, le logiciel permet la création d'un fichier des mesures individuelles et la mise à jour d'un fichier « plan ».

Toutes les informations font référence à la situation géographique des arbres sur le terrain : les arbres sont repérés dans un système de lignes et de rangées. Chaque ligne de plantation porte un numéro qui correspond à son abscisse et, sur cette ligne, chaque arbre porte un numéro de rangée (son ordonnée).

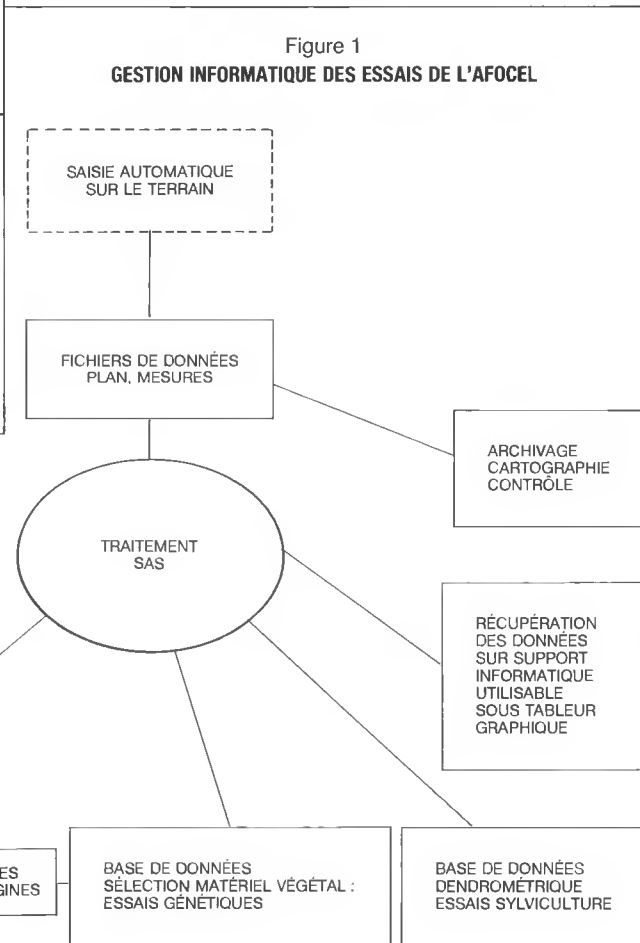
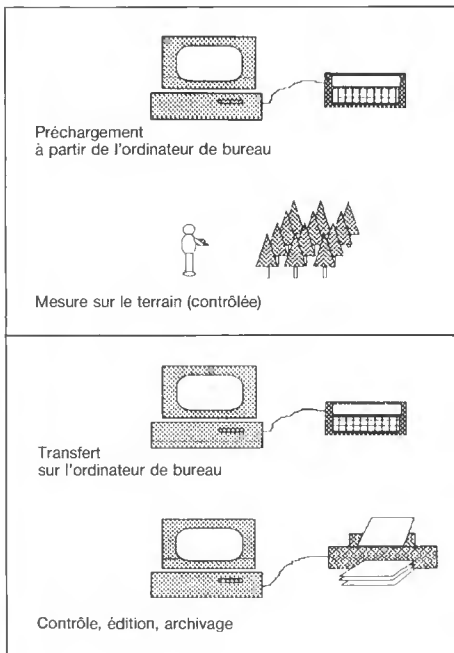
## Le stockage et la communication des données

Le « plan » est un fichier unique établi à l'installation de l'essai et mis à jour à chaque passage, afin de consigner, pour chaque arbre, les événements exceptionnels survenus depuis le dernier passage (mortalité, accident, coupe d'éclaircie, élagage, ...).

Au niveau du terrain, la saisie des données individuelles avec contrôle de la position géographique de l'arbre permet de vérifier la mesure, soit dans l'absolu, en fixant pour chaque essai des bornes inférieures et supérieures à la valeur à saisir, soit sur l'accroissement par rapport à des mesures antérieures. En cas de problème, un signal sonore est émis par le micro-ordinateur portable, ce qui rend la saisie des données beaucoup plus fiable.

### Les traitements

Dès son retour au bureau, l'équipe de mensuration transfère les fichiers « PLAN » et « MESURE » sur un micro-ordinateur équipé du système d'exploitation DOS. Un logiciel spécifique permet de fusionner ces deux fichiers, le dépouillement statistique étant réalisé à l'aide du logiciel SAS.



À ce stade, le système permet :

— de cartographier la parcelle expérimentale, en générant une image où chaque arbre est représenté par un symbole. Le critère de choix de la cartographie peut être une identification (clone, provenance, dose d'engrais, d'herbicide, ...), un état (mort, regarnis, éclairci, ...), une mesure (hauteur, circonférence, ...), une note (de froid, de résistance à un pathogène, ...).

— de choisir, en fonction des objectifs de l'expérimentateur et de sa connaissance du terrain, les individus sur lesquels peut porter l'analyse statistique. Ainsi, nous pouvons éliminer du dépouillement statistique les regarnis, les arbres accidentés pour des causes étrangères à l'expérimentation (gibier, façons culturales, ...), les traitements non comparables, les répétitions non homogènes, ...

À l'issue de ces traitements, l'expérimentateur dispose des résultats complets sur supports papier et informatique. Ces derniers sont directement utilisables par les principaux outils logiciels (tableurs graphiques) et permettent une personnalisation de la présentation des résultats.

### **L'archivage**

La saisie des données individuelles permet de stocker, sur support informatique, les différentes mesures individuelles par essai et par campagne de mesures. Ainsi, la trajectoire de croissance de chaque arbre mis sur le terrain, dans les expérimentations, est conservée. Cela représente, en France, environ 1 000 essais régulièrement mesurés, pour un total de 1 750 hectares et 2 000 000 d'arbres. Actuellement, le flux annuel de mesures traitées par la station de la Biométrie de l'AFOCEL est d'environ 400 000 arbres.

Après agrégation des mesures individuelles, les résultats alimentent les différentes bases de données qui assurent l'archivage des résultats sous une forme facilement accessible.

L'AFOCEL en distingue trois types, en fonction du niveau de recueil de l'information, à savoir l'essai, le matériel végétal (clone, provenance, ...) et le peuplement.

## **LE CONTENU DES DIFFÉRENTES BASES DE DONNÉES**

### **La base documentaire des essais**

C'est la base de données la plus ancienne.

Elle permet de consigner toutes les informations relatives aux essais installés, qu'ils fassent encore l'objet de mesures régulières ou non. Elle concerne les essais de terrain, mais aussi la plupart des essais réalisés en pépinière.

Elle est indexée par un numéro d'essai unique.

Sont consignées, en premier lieu, les informations relatives aux caractéristiques de l'essai, sa localisation, son objet, défini en clair mais aussi grâce à des mots-clés en rapport avec nos axes de recherches (amélioration génétique, fertilisation, installation de peuplements, éclaircies, ...), le programme général dont il dépend, l'espèce concernée, la surface, le dispositif statistique, la date d'installation, ... Ces informations sont fixes, pour un essai donné.

Suivent des informations qui font l'objet de mises à jour régulières. C'est le cas des mesures réalisées, avec l'année, l'âge des arbres, les hauteurs et circonférences moyennes pour l'essai, l'existence d'un compte-rendu, les observations faites...

## **Le stockage et la communication des données**

Cette base, saisie régionalement, est mise à jour au niveau national une fois par an. En 1992, elle comprenait 2 566 essais.

Cette base est régulièrement transmise au réseau national d'essais forestiers, géré par le CEMAGREF.

### **Les bases de sélection du matériel végétal**

Tout matériel végétal, qu'il soit sous forme de graines ou de boutures, issu de peuplements naturels ou de créations variétales, est répertorié dans un fichier nommé « Livre des Origines », dès son entrée en culture dans une pépinière de l'AFOCEL.

Dans ce fichier sont consignés, pour chaque clone et chaque descendance, repérés par un numéro unique à 6 chiffres (2 pour l'année d'entrée dans le fichier, 4 pour le numéro d'ordre annuel), tous les renseignements concernant l'espèce, le type de matériel mis en culture (graines, boutures, ...), l'origine génétique, le lieu, la date de récolte, l'expéditeur, les quantités, ...

Parallèlement, pour les sélectionneurs, nous alimentons des bases réunissant les performances, sur le terrain, du matériel végétal.

Les informations invariables sont identiques à celles décrites précédemment (espèce, origine génétique, date, lieu de récolte, ...). Les données, mises à jour régulièrement, reprennent les résultats de chaque clone, chaque provenance sur tous les essais où il est présent et pour chaque année de mesure.

Cela permet de connaître rapidement, pour chaque clone ou provenance, le nombre d'essais où il a été installé, l'effectif, le taux de survie, la croissance moyenne à différents âges. Y est consignée également une notion de performance, par l'indication du nombre d'essais où le clone, la descendance, la famille, sont classés dans le premier tiers ou les premiers 15 %.

Cette base est particulièrement utile pour faciliter et accélérer la sélection génétique.

En 1992, l'AFOCEL dispose d'une base de ce type pour douze espèces.

### **Les bases de sylviculture et de production**

Ces bases de données sont destinées à rassembler les résultats des parcelles expérimentales pour lesquelles un calcul de production est effectué (mesures de hauteur et circonférence nécessaires). Ces parcelles sont des dispositifs avec répétitions et plusieurs modalités ou de simples parcelles de référence (placette permanente de production), avec une seule modalité et aucune répétition.

Elles permettent de prendre en compte tous les descripteurs usuels, mesurés dans de telles parcelles : circonférences moyennes, surface terrière, production en surface terrière depuis l'origine, hauteurs, volumes, caractéristiques du peuplement dominant, classes de diamètre, ... Les différentes régressions (tarif de cubage, hauteur  $\times$  diamètre, ...) utilisées dans les calculs y sont intégrées.

Dans chaque base, on distingue les caractéristiques du peuplement avant et après éclaircie, ainsi que celles du peuplement enlevé en éclaircie.

L'archivage, pour un essai donné, se fait par parcelle élémentaire, aussi appelée unité expérimentale. Il se fait également par agrégation de différentes unités expérimentales. Un code est alors adopté pour distinguer l'agrégat de tous les niveaux d'un facteur, par exemple tous les niveaux d'apport d'une même fertilisation au moment d'une éclaircie, ou de toutes les répéti-

tions d'une parcelle expérimentale, par exemple toutes les répétitions dans un essai des parcelles éclaircies une ligne sur trois, ...

En 1992, ces bases sont opérationnelles pour l'Épicéa, le Douglas, le Peuplier, en cours de constitution pour le Pin maritime, et permettent d'envisager divers traitements de données portant sur :

- la modélisation de la production totale, en fonction de la sylviculture ;
- la représentation graphique des données de peuplement relatives à une même espèce ;
- la recherche d'unités expérimentales ayant des caractéristiques proches de celles étudiées.

Tableau I  
Exemple des informations contenues dans la base de données Peuplement,  
pour l'essai AFOCEL n° 19015, situé à Tarnac

	1 <sup>re</sup> éclaircie (18 ans)			Mesures 22 ans	2 <sup>e</sup> éclaircie (25 ans)		
	Avant éclaircie	Éclaircie	Après éclaircie		Avant éclaircie	Éclaircie	Après éclaircie
<b>TÉMOIN SANS 1<sup>re</sup> ÉCLAIRCIE</b>							
Effectif	191		191	191	191	5	186
N/ha	1 531		1 531	1 531	1 531	40	1 491
D moyen (cm)	15		15	17,2	18,6	10,6	18,8
D quadratique (cm)	15,3		15,3	17,7	19,2	11	19,4
G/ha (m <sup>2</sup> /ha)	28,3		28,3	37,6	44,5	0,4	44,1
Hauteur (m)	10,7		10,7		14,8	11,2	14,9
V unitaire (m <sup>3</sup> )	0,09		0,09		0,2	0,05	0,2
V/ha (m <sup>2</sup> /ha)	142,1		142,1		322,3	2	320,3
ID (cm) depuis dernière mesure				0,5	0,5		
IG (m <sup>2</sup> /ha) depuis dernière mesure				2,3	2,3		
ID (cm) depuis dernière éclaircie					0,5		
IG (m <sup>2</sup> /ha) depuis dernière éclaircie					2,3		
IV (m <sup>3</sup> /ha) depuis dernière éclaircie					25,7		
Prod. volume (m <sup>3</sup> /ha)	142,1		142,1		322,3		
Prod. G/ha (m <sup>2</sup> /ha)	28,3		28,3	37,6	44,5		
<b>ÉCLAIRCIE 1 LIGNE SUR 5</b>							
Effectif	152	64	88	88	88	26	62
N/ha	1 456	613	843	843	843	249	594
D moyen (cm)	14,2	12,9	15,1	18,2	20,3	18,7	21
D quadratique (cm)	14,6	13,5	15,3	18,5	20,7	19	21,4
G/ha (m <sup>2</sup> /ha)	24,4	8,8	15,6	22,8	28,5	7	21,5
Hauteur (m)	10,4	9,9	10,8		15,5	15	15,8
V unitaire (m <sup>3</sup> )	0,08	0,07	0,09		0,2	0,2	0,3
V/ha (m <sup>2</sup> /ha)	119,2	40,8	78,3		210,4	50,7	159,7
ID (cm) depuis dernière mesure				0,8	0,7		
IG (m <sup>2</sup> /ha) depuis dernière mesure				1,8	1,9		
ID (cm) depuis dernière éclaircie					0,8		
IG (m <sup>2</sup> /ha) depuis dernière éclaircie							
IV (m <sup>3</sup> /ha) depuis dernière éclaircie							
Prod. volume (m <sup>3</sup> /ha)							
Prod. G/ha (m <sup>2</sup> /ha)							

## QUELQUES EXEMPLES

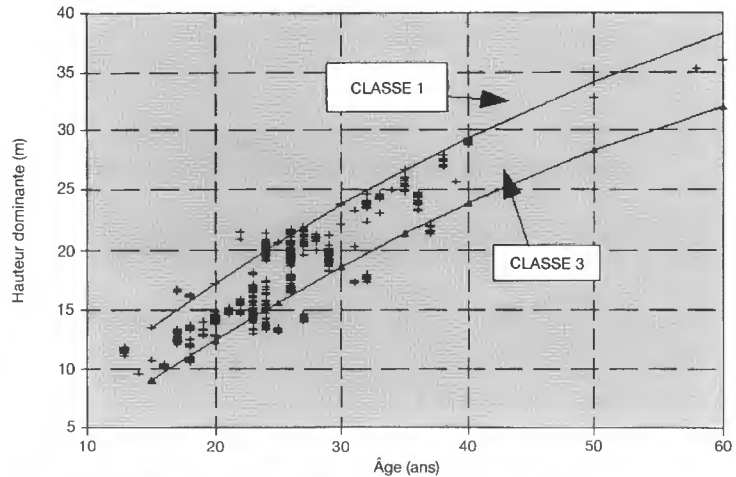
### La base de données Sylviculture Douglas

Créée en 1992, cette base de données regroupe actuellement 26 essais d'éclaircie et 9 parcelles de production, pour un total de 287 parcelles. Elle est gérée par le logiciel PARADOX de BORLAND.

Un exemple des informations contenues dans la base est fourni ci-après, avec l'essai 19015 Tarnac (tableau I, p. 58). Les résultats présentés sont agrégés par traitement sylvicole, alors que, dans la base originale, ils sont renseignés par parcelle. Cette agrégation nécessite un traitement, réalisé très rapidement et facilement par le logiciel de gestion de la base de données.

Un exemple de l'utilisation de la base complète est donné ensuite en positionnant toutes les parcelles des essais d'éclaircie de l'AFOCEL sur Douglas par rapport aux tables de production de Décourt pour l'ouest du Massif Central (figure 2, ci-dessous).

Figure 2  
POSITION DES PARCELLES DE DOUGLAS DE L'AFOCEL PAR RAPPORT À LA TABLE DE PRODUCTION DÉCOURT POUR L'OUEST DU MASSIF CENTRAL



L'interrogation des bases de données, autrefois réservée aux informaticiens professionnels, est devenue une opération facile et largement répandue avec les développements de la technologie.



**La base de données Clones de *Sequoia sempervirens***

Dans le cadre de son programme d'introduction et de sélection de clones de *Sequoia sempervirens*, l'AFOCEL a installé 26 expériences comparatives de clones sur tout le territoire national.

Au total, cela représente 740 clones mis en test et 13 000 arbres régulièrement mesurés.

À chaque nouvelle mesure, la base de données est alimentée avec, pour chaque clone présent, l'effectif mesuré, le taux de survie, la moyenne de la mesure, le rang de classement, la performance, ...

Un traitement automatique permet de renseigner la zone d'en-tête qui récapitule les performances du clone dans ses différentes introductions.

Ainsi, dans l'exemple ci-après (tableau II, ci-dessous), nous savons immédiatement que le clone 770365, sélectionné en 1977 dans l'arboretum national des Barres, a été installé dans 20 essais de l'AFOCEL en France, pour un total de 482 individus. Lors des dernières mesures de ces essais, il en restait 373, soit 78 % de vivants. Il était parmi les 15 % meilleurs clones dans 12 essais sur 20 et parmi les 30 % meilleurs clones dans 17 essais sur 20.

Tableau II  
Exemple d'utilisation de la base de données *Sequoia sempervirens*, pour le clone n° 770365

BASE DE DONNÉES SEQUOIA SEMPERVIRENS								
N° LO : — 770365      N° travail : — 4 SÉLECTION : date — 061077    lieu — LES BARRES (45)    nom sélect. BOULAY dépt. : — 45 alt. :    lat. :    long. :								
COMMENTAIRES : — prél. 4, ligne 4 type de mesure ou note : — H      âge des clones mesurés ou notés : — 99 nombre d'essais : — 20    effectif total : — 482    effectif mesuré : — 373 taux de vivants : — 78    Nb essais perf. > 85 : 12    perf. > 70 : 17								
ORIGINE DU CLONE								
RÉCAPITULATIF DES MESURES DE TOUS LES ESSAIS								
<ul style="list-style-type: none"> <li>• présent sur 20 essais</li> <li>• 482 individus plantés</li> <li>• 373 individus mesurés</li> <li>• 78 % de vivants</li> <li>• sur les dernières mesures de hauteur de chaque essai :                             <ul style="list-style-type: none"> <li>- dans les 15 % meilleurs sur 12 essais</li> <li>- dans les 30 % meilleurs sur 17 essais.</li> </ul> </li> </ul>								
RÉSULTATS DES DIFFÉRENTES MESURES DE TOUS LES ESSAIS SUR LESQUELS CE CLONE EST PRÉSENT								
Essai mesure	âge	Eff. tot.	Eff. mesuré	% viv.	moyenne	rang	performance	type
11129-H86	1	10	10	100	0,33	4	75	H
16014-D90	11	30	24	80	10,8	1	100	D
16014-H83	4	30	30	100	1,91	1	100	H
16014-H86	7	30	24	80	3,48	2	92,9	H
16014-H88	9	30	24	80	5,37	2	85,7	H
16014-H60	11	30	24	80	6,14	1	100	H
16059-H88	3	30	29	97	1,14	1	100	H
16059-H90	5	30	29	100	2,36	1	100	H
16064-H90	5	30	23	77	1,43	4	85	H
26011-H	2	89	66	75	0,39	32		H
26011-H89	5	8	33	7	6	17	90	H
30023-H90		20	17	85	0,96		84,4	H
30029-H89	2	23	23	100	0,41	3	95,8	H
30029-H91	4	23	9	39	1,12	2	97,6	H



Photo AFOCEL-ARMEF

**À gauche :** La mise sur le terrain avec un micro-ordinateur portable des mesures et observations représente à la fois un gain de temps et une plus grande fiabilité des données.

**À droite :** la mesure des hauteurs.

## CONCLUSIONS

L'évolution de la micro-informatique nous permet de faire fonctionner, aujourd'hui, ces systèmes de traitement des données expérimentales en « routine ». Les équipes de mensurations s'y sont très facilement adaptées, si bien qu'un retour en arrière, vers la saisie sur bordereau informati- que, ne leur semble plus envisageable.

Le schéma de traitement informatique développé permet aux ingénieurs et techniciens de terrain de rester maîtres des résultats de la machine. En effet, en fonction de leur connaissance des conditions de milieu, des accidents, ils choisissent les arbres, ou unités expérimentales, qui peuvent être intégrés au traitement statistique, à la mise à jour des bases de données. Ils peuvent ainsi éliminer toutes les mesures entachées d'un biais indépendant des facteurs expérimentaux testés.

Cette possibilité de contrôle, ainsi qu'un outil de consultation à la fois simple et efficace, ont été des conditions indispensables à la réussite du système.

L'informatisation de la saisie et du traitement des données a conduit à une plus grande rigueur dans la conception, le suivi et l'analyse de nos dispositifs expérimentaux.

Elle permet également, sur la base de protocoles harmonisés, de plus grands échanges de données entre les organismes impliqués dans la recherche forestière, au niveau français, mais aussi étranger. En effet la mise en commun de mesures effectuées par chaque centre de recherches permet d'asseoir les recherches de chacun sur un plus grand nombre de mesures, et d'augmenter ainsi leur fiabilité. C'est l'objet du projet « Cooperatives de données », en cours de

lancement, sur quatre espèces (Douglas, Pin maritime, Chêne, Merisier), avec divers organismes français (INRA, CEMAGREF, IDF, ONF, ...). L'AFOCEL est associée à ce projet pour le Douglas et le Pin maritime. La France ne fait là que suivre (en l'adaptant) l'exemple d'autres pays, comme les USA.

A. BAILLY  
Directeur  
AFOCEL-ARMEF  
Quartier Saillans  
26120 MALISSARD

F. GUINAUDEAU - M. NAJAR  
AFOCEL Biométrie  
30, boulevard Foch  
77000 FONTAINEBLEAU

## BIBLIOGRAPHIE

- BOUCHON (J.). — Protocole d'échantillonnage : base de données, considérations générales. — AIP. Projet n° 6, 1990, pp. 168-194.
- BRUNET (A.M.), MENDIBOURE (P.). — Un fichier central des plans d'expérience sous forme de base de données. — *Annales AFOCEL*, 1975, pp. 83-99.
- GUINAUDEAU (F.). — Direct data recording on portable computer in forest experimentation. — *Bulletin des Recherches agronomiques de Gembloux*, vol. 25, n° 1, 1990, pp. 123-131.
- GUINAUDEAU (F.). — Base de données des expériences forestières de l'AFOCEL. — *Bulletin des Recherches agronomiques de Gembloux*, vol. 25, n° 1, 1990, pp. 19-37.
- GUINAUDEAU (F.), TONARELLI (M.). — Saisie directe de mesures sur ordinateur portatif dans les expérimentations forestières. — *Informations-Forêt*, n° 3, 1987, fasc. 323, pp. 163-174.
- GUINAUDEAU (F.), BOUVET (A.), BAILLY (A.). — Fichier de sélection de clones de *Sequoia sempervirens*. — Poster, AFOCEL-IUFRO « Mass propagation technology for genetically improved fast growing forest tree species », Bordeaux, septembre 1992, p. 463.
- HOULLIER (F.). — Etude de faisabilité d'une coopérative de données en vue de l'analyse et de la modélisation de la croissance et de la production des peuplements forestiers. — Compte-rendu de convention entre la DERF et l'ENGREF, octobre 1991.
- MENDIBOURE (P.), EARITH (G.). — Système automatisé d'analyse des données de plans d'expériences. — *Annales AFOCEL*, 1975, pp. 55-82.
- NAJAR (M.). — Cartographie automatique des parcelles expérimentales forestières : une étape vers de nouveaux usages de l'informatique pour la gestion forestière. — *Informations-Forêt*, n° 2, 1991, fasc. 441, pp. 125-134.