



**HAL**  
open science

## Quelques réflexions sur les modèles : stratégies de conception, formes mathématiques, hypothèses théoriques, généralisation

Jean-Francois Dhote, Patrick Vallet, Christine Deleuze, Laurent Saint-André,  
Jean-Christophe Hervé

### ► To cite this version:

Jean-Francois Dhote, Patrick Vallet, Christine Deleuze, Laurent Saint-André, Jean-Christophe Hervé. Quelques réflexions sur les modèles : stratégies de conception, formes mathématiques, hypothèses théoriques, généralisation. 10. Journées CAQ Croissance Amélioration Qualité des bois, Mar 2007, VILLENAVE D'ORNON, France. 24 p. hal-01194981

**HAL Id: hal-01194981**

**<https://hal.science/hal-01194981v1>**

Submitted on 6 Jun 2020

**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.



# **Quelques réflexions sur les modèles :**

## **Stratégies de conception, formes mathématiques, hypothèses théoriques, généralisation**

Jean-François Dhôte (INRA), Patrick Vallet (Cemagref)

Christine Deleuze (AFoCel), Laurent St-André (CIRAD),

Jean-Christophe Hervé (IFN)

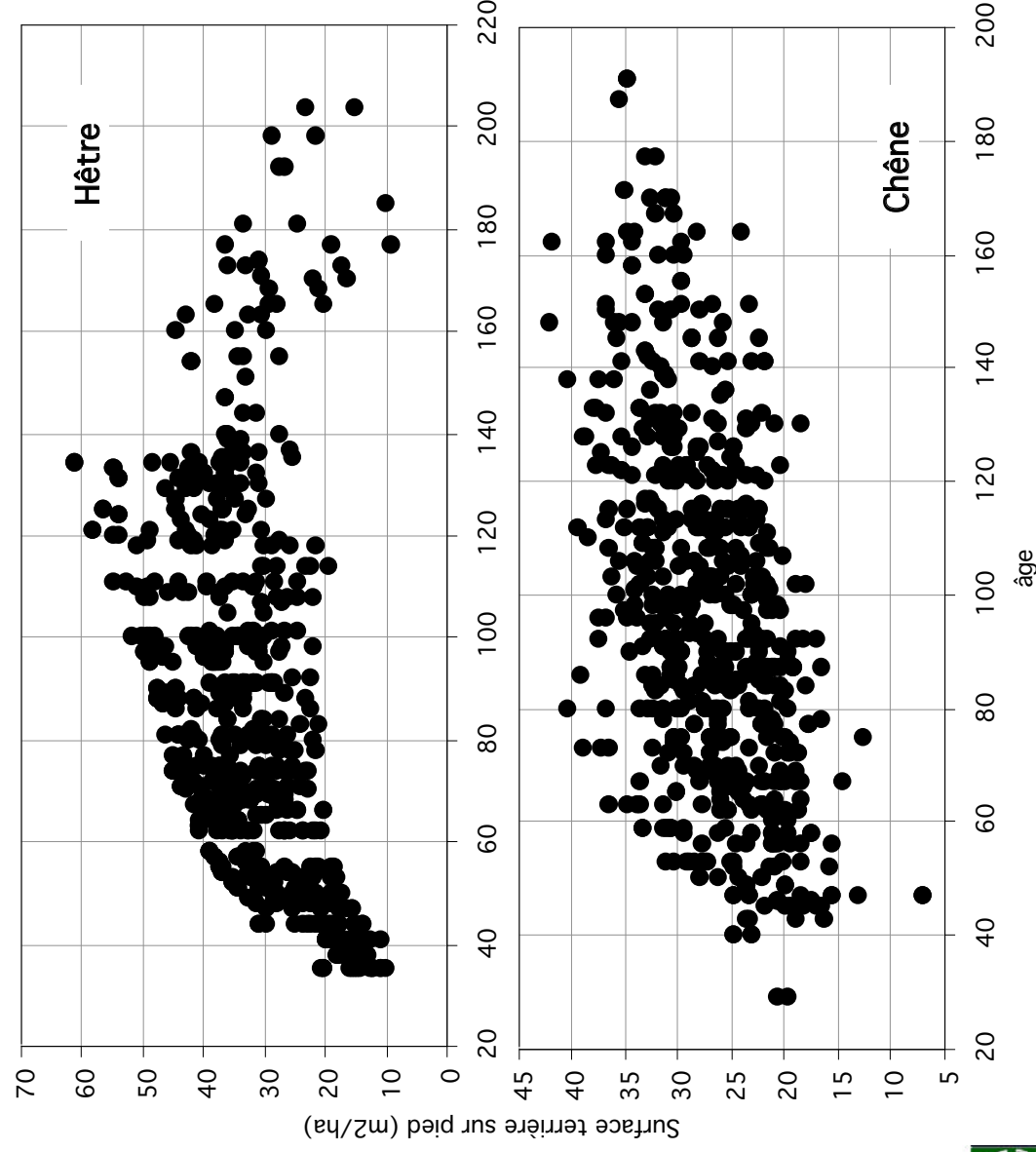
## Point de vue de la présentation

- L'évaluation de modèles :
  - *On ne parle pas ici d'évaluation a posteriori*
  - *Plutôt de conception de modèles*
- Une situation : fabriquer des modèles d'ingénierie
  - *Objectif : prévision, simulation, optimisation*
  - *Modèles calibrés sur des données*
  - *Recherche nécessaire → faire de bons outils d'ingénieur*
- Une conviction :
  - *La qualité d'un modèle se construit en amont (contexte, question posée, objectifs, données et **connaissances disponibles**)*



# Domaine-cible et domaine de calibration

- Construction d'un modèle de croissance de croissance / feuillus



Les données disponibles sont toutes comprises entre 0.5 et 1 fois la densité maximale

Comment faire un modèle :  
 - juste dans ce domaine,  
 - sain en extrapolation ?

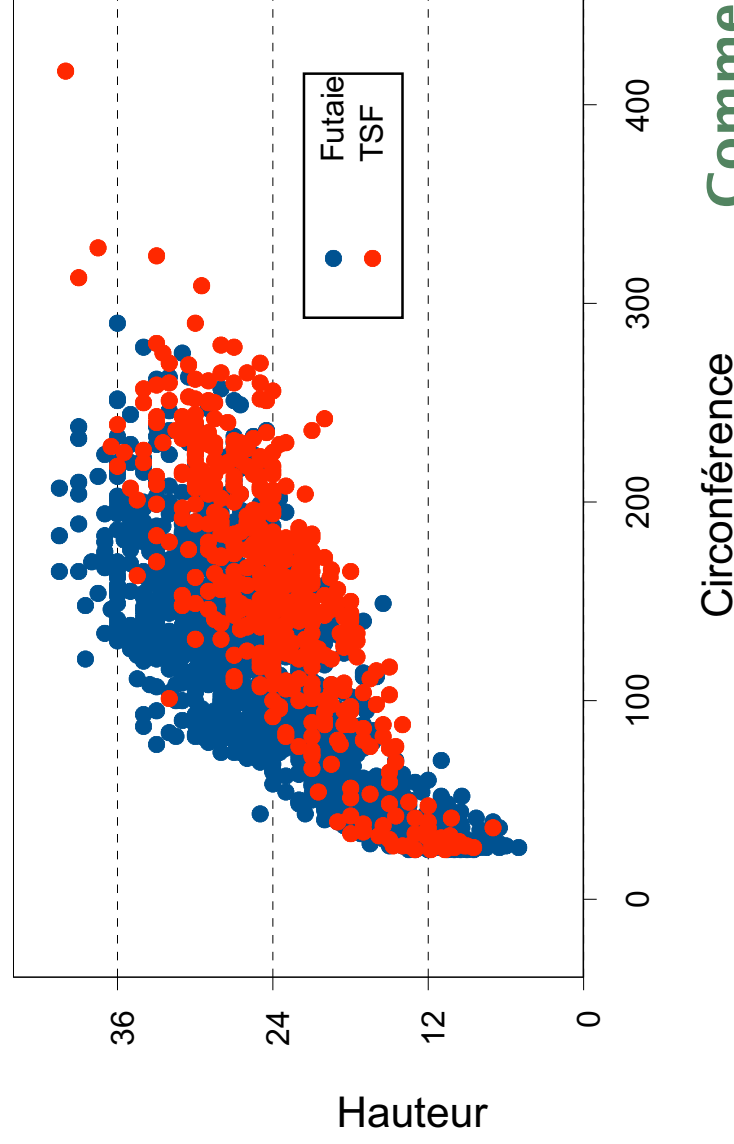


# Domaines : cible & calibration (2)

- Construction d'un tarif de cubage national

Les données disponibles sont toutes en futaie régulière

Source : P. Vallet (2002)



Circonférence

Comment faire un tarif de cubage :  
 -de portée nationale,  
 -juste  $\forall$  la structure ?



## Problème

- On est souvent obligés de construire un modèle :
  - *calibré sur un sous-domaine de la cible visée...*
  - *... mais qui se comporte sainement en dehors*
- Pourquoi est-on souvent dans cette situation ?
  - *Les données forestières sont longues à acquérir*
    - Problèmes ≠ entre t(création) et t(exploitation)
    - Les méthodes d'analyse ont évolué
  - *Les données coûtent cher*
    - Ré-utilisation de données, dispositifs
    - Combinaison de données anciennes et nouvelles
    - Nouvelles technologies : vers une relecture des données



## Comment faire ?

- Bien sûr, on va faire des statistiques, des ajustements, etc...
- Mais essayons de réfléchir à la stratégie de conception :
  - *Approche générale de modélisation*
  - *Connaissances : qualitative, empirique, théorique*
  - *Choix des variables indépendantes*
  - *Choix de la forme mathématique du modèle*
  - *Progresser vers une théorie des phénomènes*



# Exemple : tarifs de cubage découpe 0

## DEA Patrick Vallet (2002) Contrat CARBOFOR

- Un exemple très simple
- pour illustrer des idées
- simple ≠ simpliste
- « *tous les modèles sont faux, quelques-uns sont utiles* »
- c'est difficile que de faire simple et utile





## Contexte, objectifs

- **Besoin d'estimer le stock de C forestier national**
  - *(très) nombreuses essences*
  - *diversité des contextes écologiques (bioclimat, station)*
  - *diversité des conditions sylvicoles (TSF, FR, transition)*
  - *coût très élevé des mesures de biomasse*
  - *intérêt scientifique pour les méthodes génériques*
- **Objectif : construire une méthode de cubage**
  - *objet : volume aérien découpe 0*
  - *tarif de forme homogène pour ≠ essences*
  - *paramétrage minimal (principe de parcimonie)*



## Stratégie d'ensemble (1)

- Qu'est-ce qui se fait ailleurs ?
  - Très nombreux tarifs allométriques :  $Biomasse = a * d_{130}^b$
  - une seule variable indépendante
  - prédiction directe de la biomasse
- Pourquoi cela pose-t-il un problème ici ?
  - Exposant  $b$  : connu pour varier selon structure (pop)
  - Notre jeu de données ne reflète pas la population d'intérêt
  - Faire qq chose de : plus général, robuste vis-à-vis des différences de morphologies entre structures (futaie, TSF)
- Que fait-on ?
  - Au minimum, avoir 2 variables indépendantes ( $d_{130}, h$ )
  - Rendre compte des  $\neq$  de forme à travers ( $d_{130}, h$ )



## Stratégie d'ensemble (2)

- Méthode directe vs indirecte
  - *directe* :  $Biomasse = F(d_{130}, h)$
  - *indirecte* :  $Biomasse = Vol_0(d_{130}, h) * InfraDens$
- Inconvénient de la méthode indirecte :
  - *Composition des erreurs*
- Avantages :
  - *Nombreuses données disponibles (croissance, QB)*
  - *Savoir-faire sous-jacents*
- Anticiper sur les problèmes à venir : changements environnementaux et déterminismes différenciés



## Stratégie d'ensemble (3)

- Réalisations précédentes en France :
  - *Second choc pétrolier (contrat DGRST, 1981)*
  - *Travail de J. Bouchon, J. Pardé & J.M. Ottorini*
  - *Feuillus : 2 procédures distinctes (futaie, TSF)*
- Pourquoi faire autrement ?
  - *Système des procédures distinctes → discontinuité*
  - *Or beaucoup de peuplements sont en transition entre la futaie et le TSF ... et la codification reste subjective*
  - *Hypothèse (théorique) de continuité fondamentale des variations de forme*
  - *Profiter du chantier Carbone pour illustrer l'approche générique*



# Hypothèses et ressources utilisées

- Hypothèses :
  - *Le couple (d130, h) contient l'information nécessaire et suffisante pour rendre compte des variations de forme*
  - *Pas de covariation avec région biogéographique, station, structure*
  - *Deux espèces « botaniquement proches » ← même tarif*
- **Ressource en données : le fonds d'archives INRA**
  - *10aines de milliers de cubage bois-fort + menus bois*
  - *Période de relevé : 1920-1950*
  - *Principales essences forestières*
  - *Futaies régulières, pureté ≥ 80% dans 90% des cas*
  - *... mais une gamme suffisante de morphologies individuelles*



# Forme mathématique

1. Tarif J. Bouchon, bois-fort total Chêne :
  - $V = \alpha + \beta.d + \gamma.d^2 + \delta.d^3 + \varepsilon.h^3 + \eta.d^2.h$
2. Idem pour le Hêtre :
  - $V = \pi.d^2.h . [\alpha + \beta.d^2 + \gamma.d^3 + \delta/h + \varepsilon^*h] . [1 + \theta.d + \kappa.h]$
- (Combinaison de) polynômes :
  - *Forme (1) : adaptée pour une publication tabulée, mais comportement en extrapolation non maîtrisable*
  - *Forme (2) : mieux adaptée pour*
    - Décomposer les phénomènes influents
    - Homogénéiser la variance (hétéroscédasticité des données de volume)
    - Faire du test d'hypothèses
- Substitution : volume → facteur de forme



## Transformation des variables indépendantes

- Le diamètre et la hauteur : corrélés !
- Pourtant, information non-redondante
  - Transformation de  $(d130, h)$  en  $(d130, \text{robustesse})$ 
    - où  $\text{robustesse} = \sqrt{d130 / h}$
- En outre, cette nouvelle variable :
  - = principale source des variations de forme (*feuillus*)
  - discrimine bien futaie et TSF
  - possède une gamme suffisante dans nos données
  - a un effet linéaire raisonnablement extrapolable



## Modèle

- On substitue le facteur de forme au volume :
  - $Forme = Volume_0 / [ \pi \cdot d^2 \cdot h / 40000 ]$  (sans unité)
- Modèle :
  - $Forme = [\alpha + \beta \cdot d + \gamma \cdot robustesse] \cdot [1 + \delta / d^2]$





## Principaux résultats

- Groupes de comportements spécifiques :
  - Feuillus : légère différence entre Hêtre et Chêne sessile
  - Résineux « sombres » (Sapin, Épicéa, Douglas) : comportements similaires
  - Pins : ressemblent aux feuillus, mais forme <
  - Feuillus et Pins : la robustesse est prépondérante
  - **Groupes cohérents cf études / autres continents**
- Fusion d'espèces :
  - Toutes sont ≠, sauf les Pins (voir zoom)
- Variations
  - Pas de tendance temporelle (vérification partielle)
  - Légères variations spatiales, sans structuration géographique
  - Effet structure (TSF, taillis) : vérifié par l'IFN

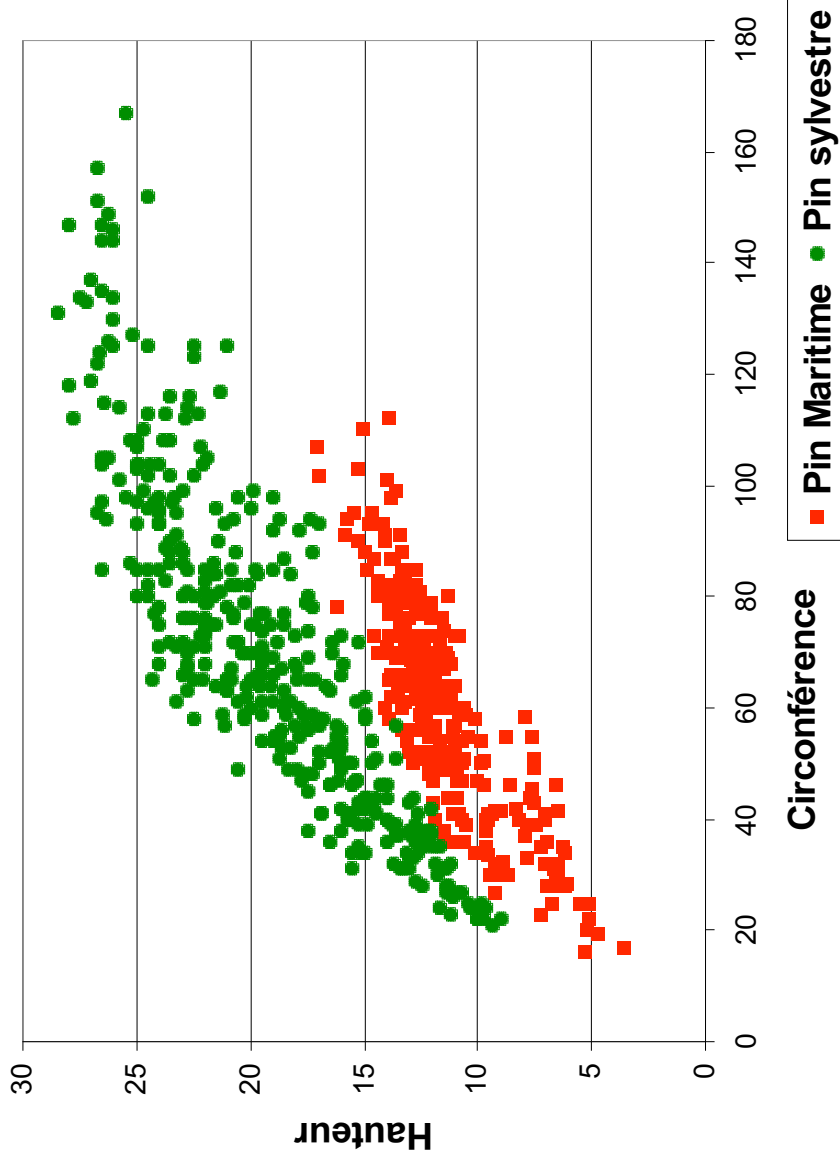


# Le cas des pins

- Pins sylvestre et maritime : fortes  $\neq$  des caractéristiques dendrométriques, sylvicoles et écologiques



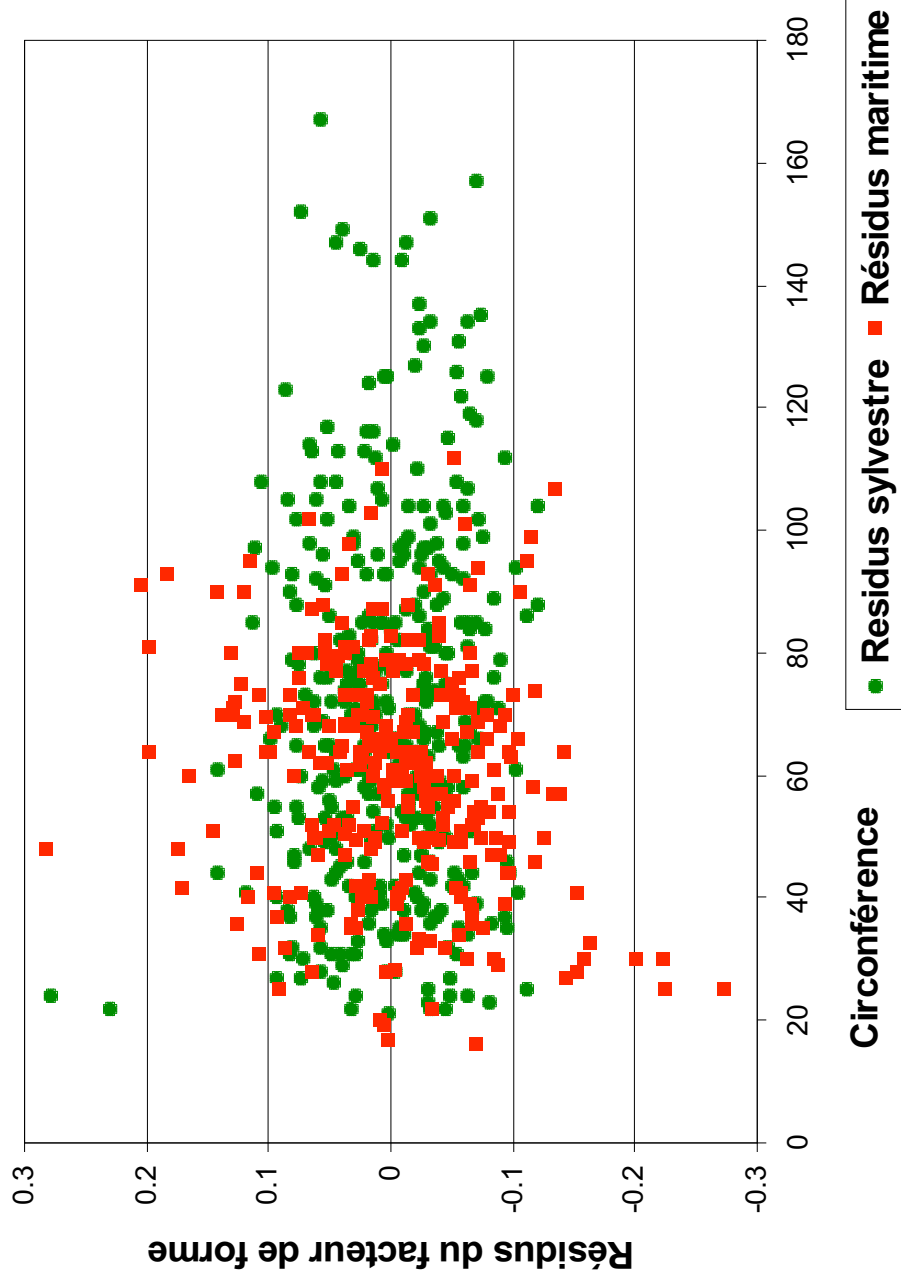
Nuage de points circonférence hauteur pour les Pins sylvestre et maritime



# Le cas des pins

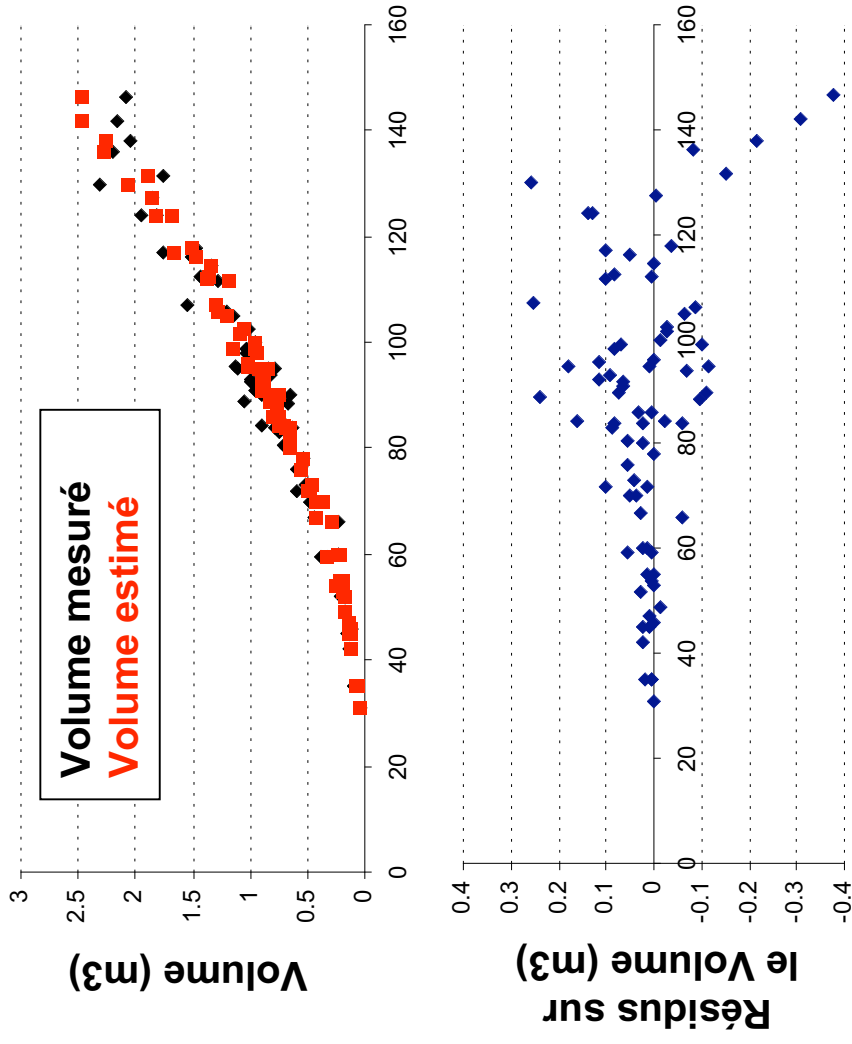
- Ajustement d'un tarif de cubage commun

Résidus du facteur de forme après ajustement d'un tarif commun



# Le cas des pins

- Utilisation du tarif commun sur 77 Pins laricio



Circonférence (cm)

## Qu'a-t-on obtenu/appris ?

- En vue des applications :
  - *Huit tarifs de même structure (C, H, E, S, P, D, M, divers)*
  - *Une proposition de méthode de portée nationale*
- Cette méthode procède par généralisation (= expertise)
  - *Hypothèse « de proximité botanique » non-contredite*
  - *Groupements d'espèces : cohérents avec autres continents*
  - *Application d'un tarif unique aux essences du même genre*
- **Évaluer un modèle : c'est multi-critères !**
  - *Choix de construction, hypothèses testées, propriétés qualitatives, simplicité, utilité, vérification numérique...*
  - *C'est l'ensemble qui donne confiance dans le modèle*



## Résumé des hypothèses

- Trois hypothèses explicites (= écrites dans l'article)
  - *Le couple (d130, h) contient l'information nécessaire et suffisante pour rendre compte des variations de forme*
  - *Pas de covariation avec région biogéographique, station, structure*
  - *Deux espèces « botaniquement proches » ← même tarif*
- Critique a priori :
  - *Ces hypothèses restent vagues dans leur formulation*
  - *Elles ne contiennent pas leur propre critère de mise à l'épreuve*
- Deux hypothèses implicites :
  - *Continuité fondamentale des variations de forme*
  - *L'âge n'est pas un facteur autonome, il agit à travers les variations dimensionnelles*



## Pourquoi faire des hypothèses ?

- Organiser un problème, clarifier ses idées
- Étape utile pour :
  - *Sélectionner des variables, choisir des formes mathématiques*
  - *Justifier ses choix*
  - *Modèles  $\neq$  données : plus généraux grâce aux hypothèses*
  - *Hypothèses et connaissances : arbitrer en modèles concurrents*
- Dialoguer avec des collègues qui croient à la méthode expérimentale
  - *Hypothèse  $\rightarrow$  expérience  $\rightarrow$  décision*
- Faire accepter ses publications



## Franchir une étape : vers une théorie des phénomènes

- Prendre conscience des « théories implicites » :
  - Exemple : Croissance = Potentiel . Réducteur
  - Lois d'invariance : Eichhorn, Pressler, autoéclaircie
  - Théorie dendrométrique de la production des PR
- Il s'agit bien de théories (phénoménologiques) :
  - Organiser des phénomènes, constituer un champ disciplinaire
  - Associer concepts abstraits (compétition, fertilité) et indicateurs concrets (RDI, IF)
  - Forme mathématique achevée, analyse qualitative
  - Insertion dans une démarche expérimentale : concevoir de nouvelles expériences (cf Coop), faire des prédictions testables
  - On sait comment progresser !





## Pas de complexes avec l'idée de théorie !

- On n'est pas dans un monde où :
  - *Théorie = modèles de l'écophysiologie*
  - *Empirisme, statistiques = modèles croissance-QB*
- Où trouver ses sources d'inspiration ?
  - *Écologie (cycles, populations, communautés)*
  - *Anatomie et physiologie*
  - *Botanique et architecture des plantes*
  - *Biomécanique*
  - *Physique*

