

# Modélisations génériques en vue de simulations sylvicoles au niveau "forêt"

Mise au point d'un **simulateur** (Capsis, Sydy2) d'évolution forestière,  
de **stratégies** sylvicoles et d'évaluation carbone

**ONF Pôle R&D d'Avignon** = réseau RDI x DT Midi-Méd.

+ appui sur réseau ONF RDI :

- pôle de **Dole** : dont traduction Carbone
- pôle de Nancy : types de données  
(inventaire multi-sources, TD)

+ **IGN** : utilisation experte des données de l'Ifn

+ **UMR AMAP** (plate-forme de simulation Capsis, Montpellier) – F. de Coligny

Contacts :

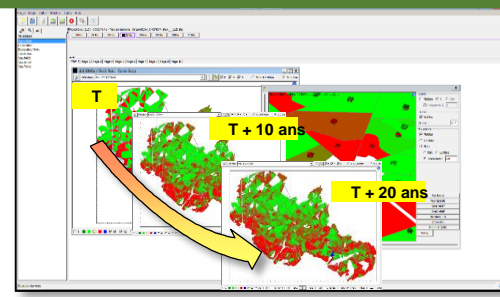
**Timothée Audinot**, Philippe Dreyfus,  
Thomas Cordonnier, Salomé Fournier  
Jean-Pierre Renaud, Christine Deleuze

# Cahier des charges de l'outil

... une idée « ancienne »  
... réactivée par les besoins **d'évaluations 'Carbone'**

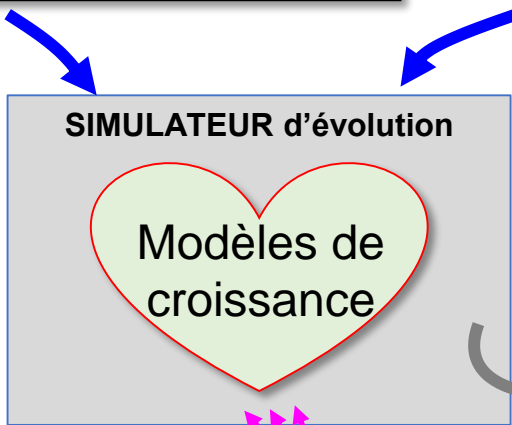
- **Simuler la croissance de toutes forêts** gérées par l'ONF dans l'hexagone
    - Types de forêts (régulières, mélangées, inéquienne, hétérogène)
    - Essences
  - **Simuler des options de gestion diverses (y compris sans intervention)**
  - Données de sortie → **évaluation des flux, du carbone** (en plus des critères classiques dendrométriques, et de composition)
- **Comparer des choix de gestion pour l'ensemble d'une forêt**
- Simuler à partir des **données disponibles en gestion** → détermine en partie l'approche de modélisation
    - Diamètre de précomptage (17,5 cm)
    - Absence / présence donnée « âge »
    - Information ± précise

Schéma de principe d'un simulateur « forêt » :

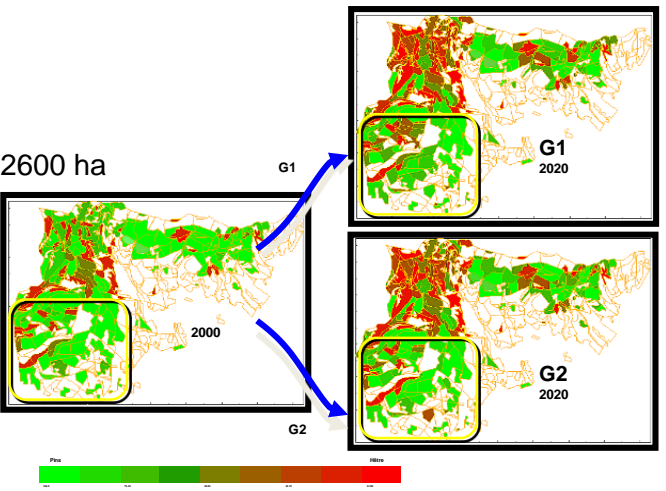


PEUPELEMENTS : composition et structure, dendrométrie, âge (?)

STATIONS : ⇒ potentialités de croissance, survie



évolutions de chacun des peuplements (≈) et de la forêt dans son ensemble



Options de Gestion à COMPARER ► CHOISIR !  
 états simulés à  $t_0 + 5, 10, 15, 20$  ans ... 100 ans (+ incertain !)  
 / ≠ essences & options Itinéraires Sylvicoles (des Guides, ou alternatifs) :

- choix de nouvelles structures, compositions,
- ... d'une « mosaïque » à l'échelle de la forêt
- infos ressources, **stocks, récoltes** (masse et dimensions produits)

entrées & sorties compatibles avec autres outils de gestion

# Des défis et des choix

- Modèles génériques adaptés aux données d'aménagement
  - Croissance
  - Mortalité
  - *Recrutement*

## Focus 1

- Comblent certains manques des données de gestion

## Focus 2

- « Traduire » les sorties du simulateur en carbone
- *Climat → tenir compte du changement climatique, sans approche lourde qui ne pourrait être appliquée à tous types de peuplements*
- *Incertitudes → incertitude d'échantillonnage, d'estimation des paramètres, de la structure des modèles...*
- *Scénarios sylvicoles → décisions multicritères (avec des objectifs variés)*

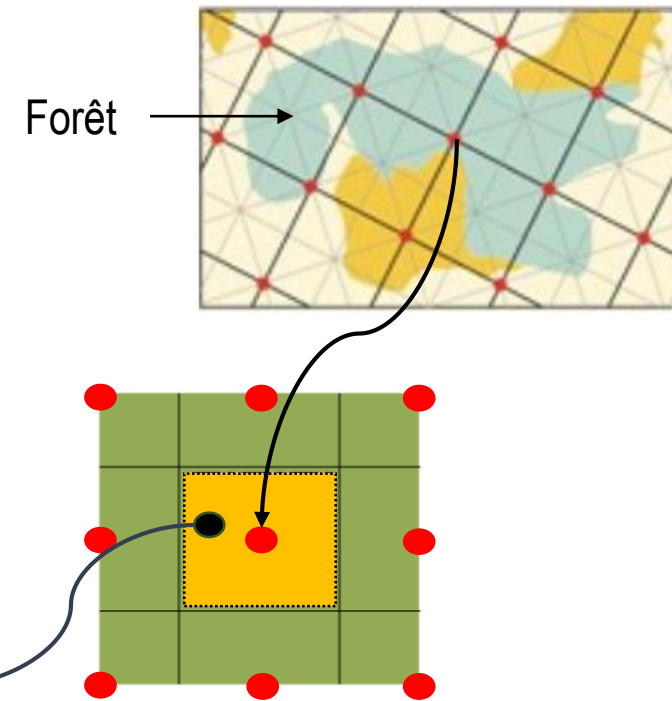
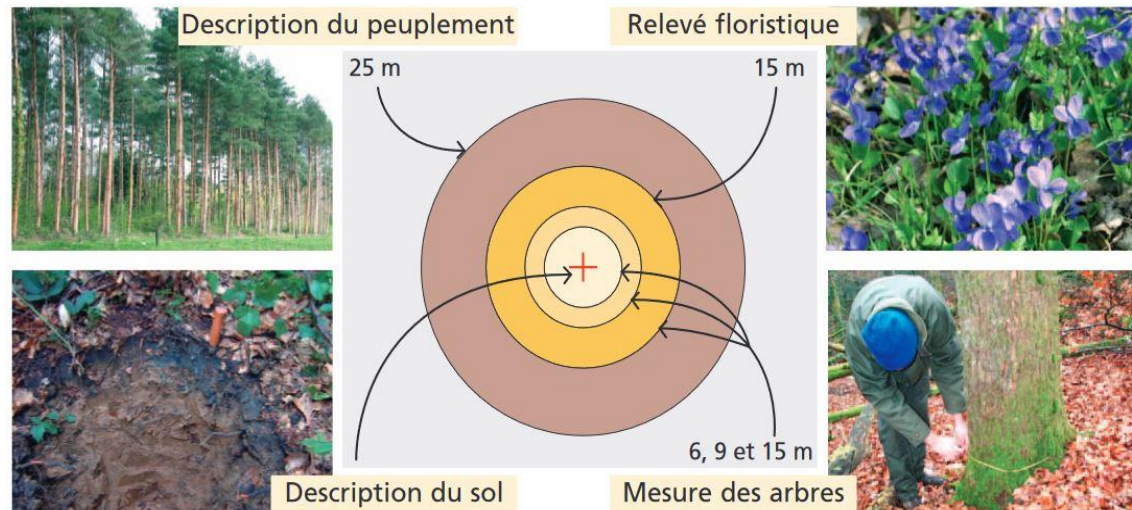
# Focus 1 : Modélisation de la croissance

Utilisation d'une **approche paramétrique (régression)** et exploration d'une **approche non-paramétrique (de type « apprentissage »)**

# Données d'inventaire forestier national (IFN, IGN)

→ **Données représentatives de l'ensemble des forêts françaises**

- Echantillonnage systématique



# Utilisation des données d'IFN pour le développement des modèles

- **Diversité des forêts**

- Grandes régions écologiques (GRECO)
- Essence

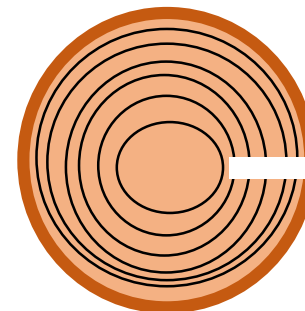
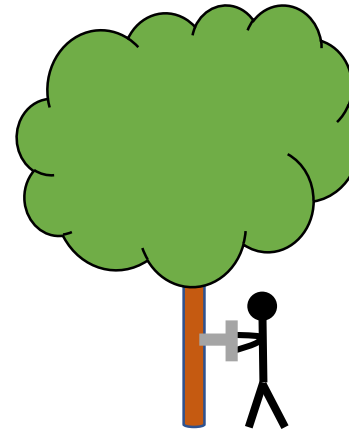
- **Variables à expliquer**

- Etat de l'arbre (mortalité)
- Accroissement radial sur 5 ans (croissance)

- **Variables explicatives**

- Diamètre, Âge
- Surface terrière
- Hauteur

Schéma de la mesure de l'accroissement radial sur 5 ans



# 1<sup>ère</sup> approche → Régressions non-linéaires (RNL)

- Développement de modèles « arbres » non-spatialisés

**Croissance individuelle d'un arbre =**  
*Potentiel de croissance* \* *Réducteur de croissance arbre* \* *Réducteur de croissance peuplement*

=

$a * \text{Accroissement en hauteur dominante}^b *$

$(1 - c * \exp^{-d * \left(\frac{\text{diamètre individuel}}{\text{Hauteur dominante peuplement}}\right)^2}) *$

$(1 + e * \text{Surface terrière} + f * \text{Surface terrière}^2)$



# 1<sup>ère</sup> approche → Régressions non-linéaires (RNL)

- Développement de modèles « arbres » non-spatialisés

*Croissance individuelle d'un arbre =*  
*Potentiel de croissance* \* *Réducteur de croissance arbre* \* *Réducteur de croissance peuplement*

- ✓ Simplicité d'interprétation et de calcul
- ✓ Approche très documentée (littérature abondante, avec de multiples exemples dans Capsis)
- RNL avec et sans l'information « âge » → adaptation aux données d'aménagements
- RNL calibrées pour chaque essence (20) et pour tout contexte → modèles génériques

## 2<sup>ème</sup> approche → par apprentissage (random forest, RF)

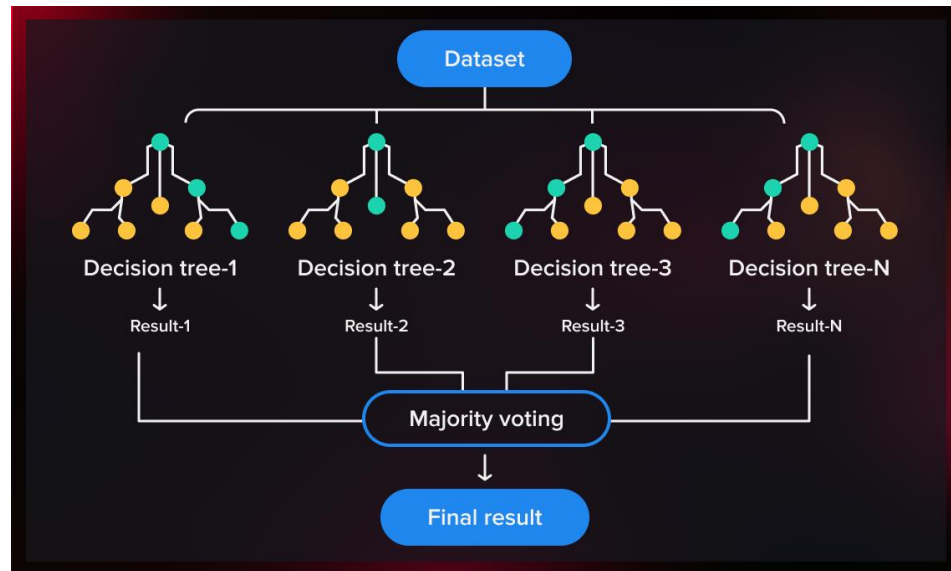
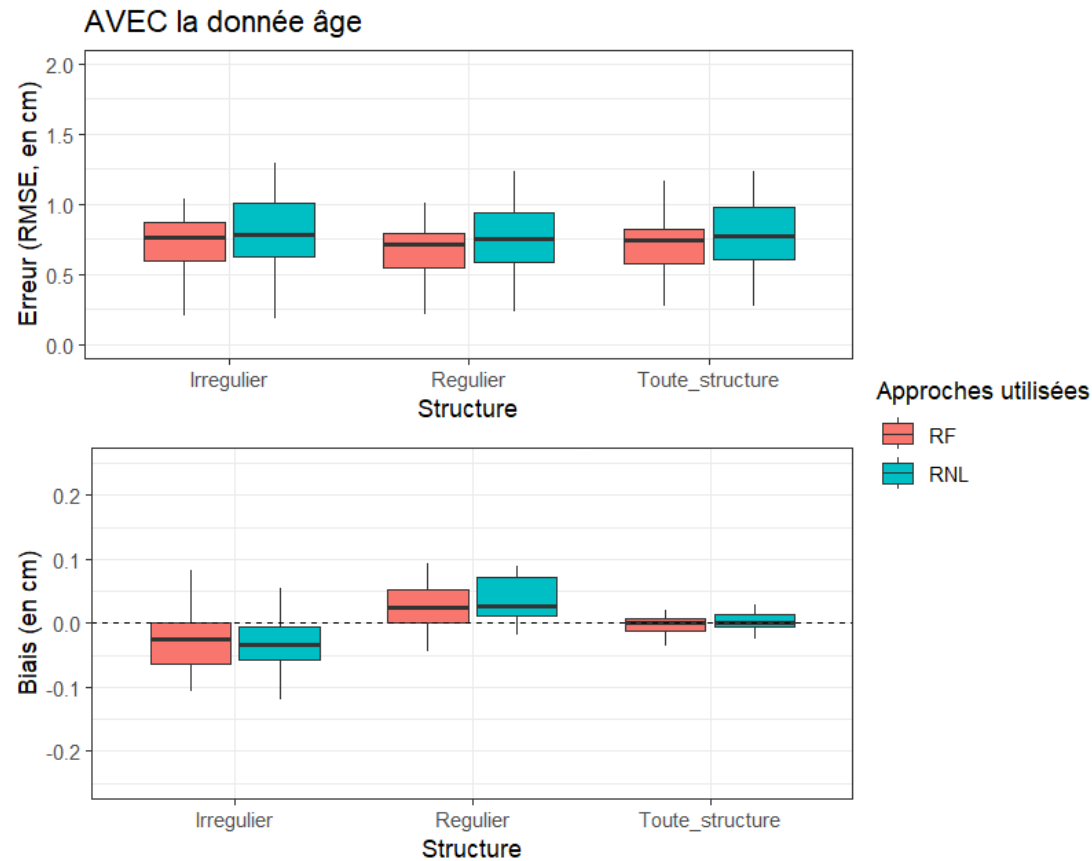


Schéma de principe d'une « forêt aléatoire »

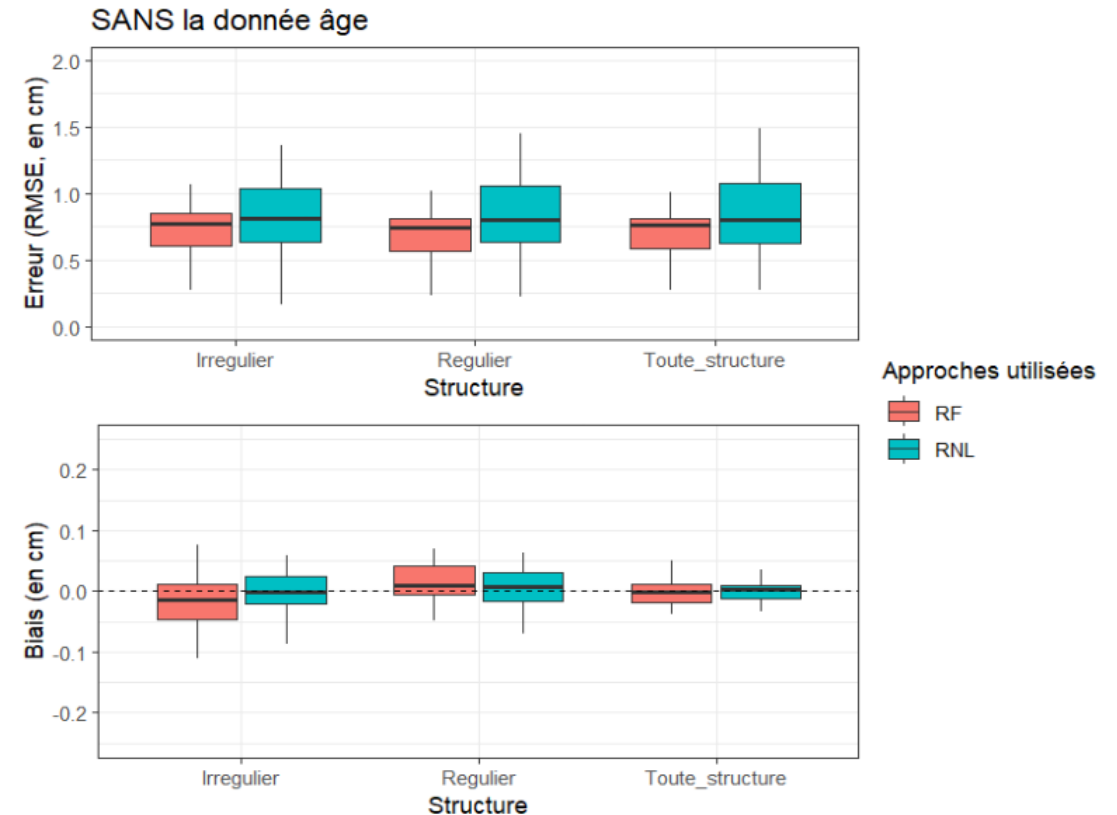
- RF avec et sans l'information « âge »
- RF calibrées pour chaque essence (20) et pour tout contexte

- ✓ Aucun *a priori* sur la forme des relations
- ✓ Meilleures performances que les approches paramétriques
- ✗ Sensible à l'extrapolation
- ✗ Difficile d'interpréter leurs résultats (réputées « boîte noire »)

# Performances régression vs apprentissage



Performances des modèles utilisant la donnée âge sur les 20 essences testées  
1 boxplot = 20 essences



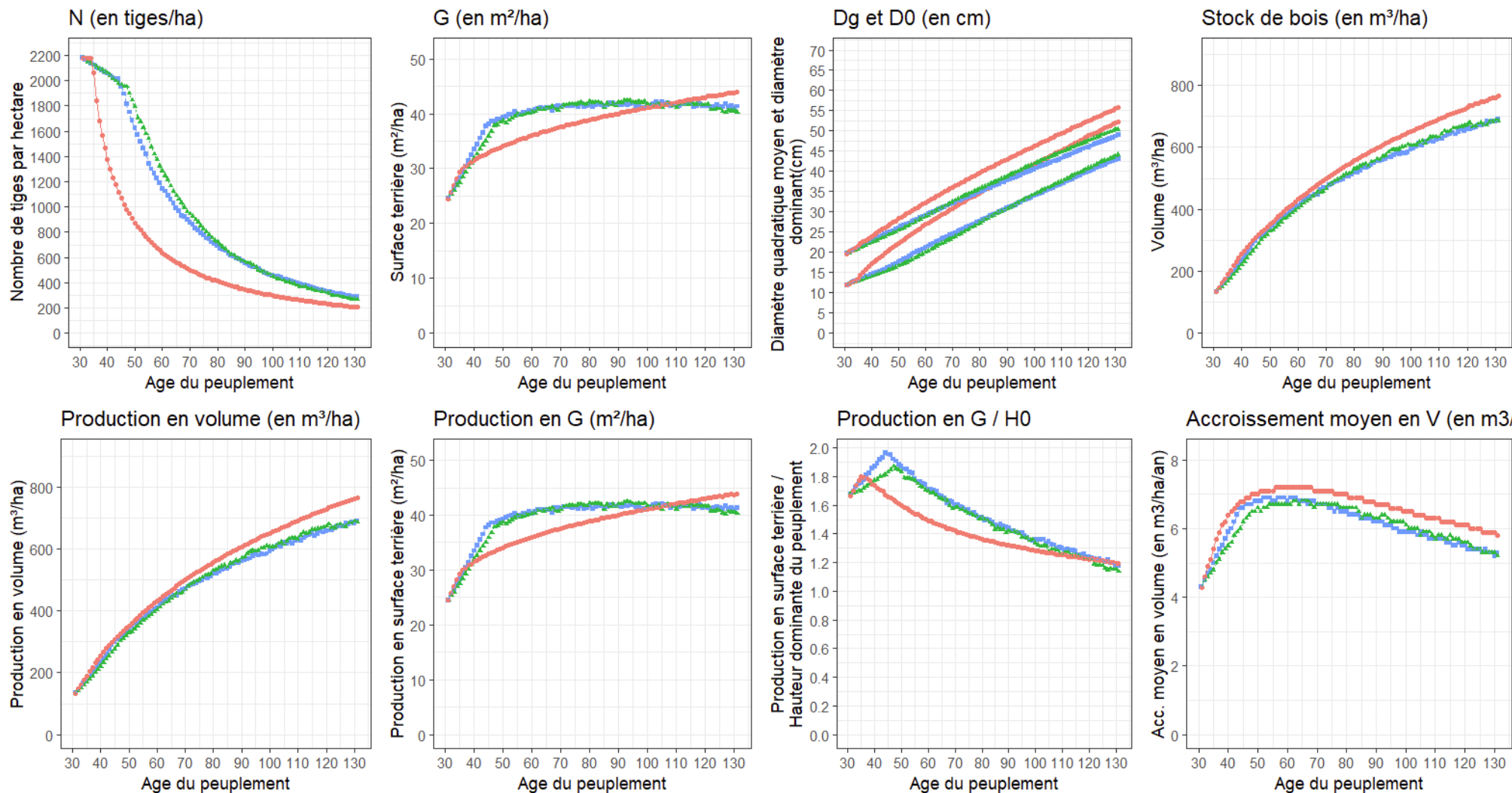
Performances des modèles n'utilisant pas la donnée âge sur les 20 essences testées  
1 boxplot = 20 essences

# Résultats

- **Avec la donnée âge** : diminution moyenne par rapport aux RNL de 6% de l'erreur (non significative)
  - **performances apprentissage  $\approx$  régression**
  - conservation de l'approche par régression (+ simple à interpréter, simulation + rapide)
- **Sans la donnée âge** : diminution moyenne par rapport aux RNL de 13% de l'erreur ( $p < 0.1$ )
  - **performances apprentissage  $>$  régression**
  - conservation de l'approche par apprentissage

# Comparaisons de modèles en simulation (1)

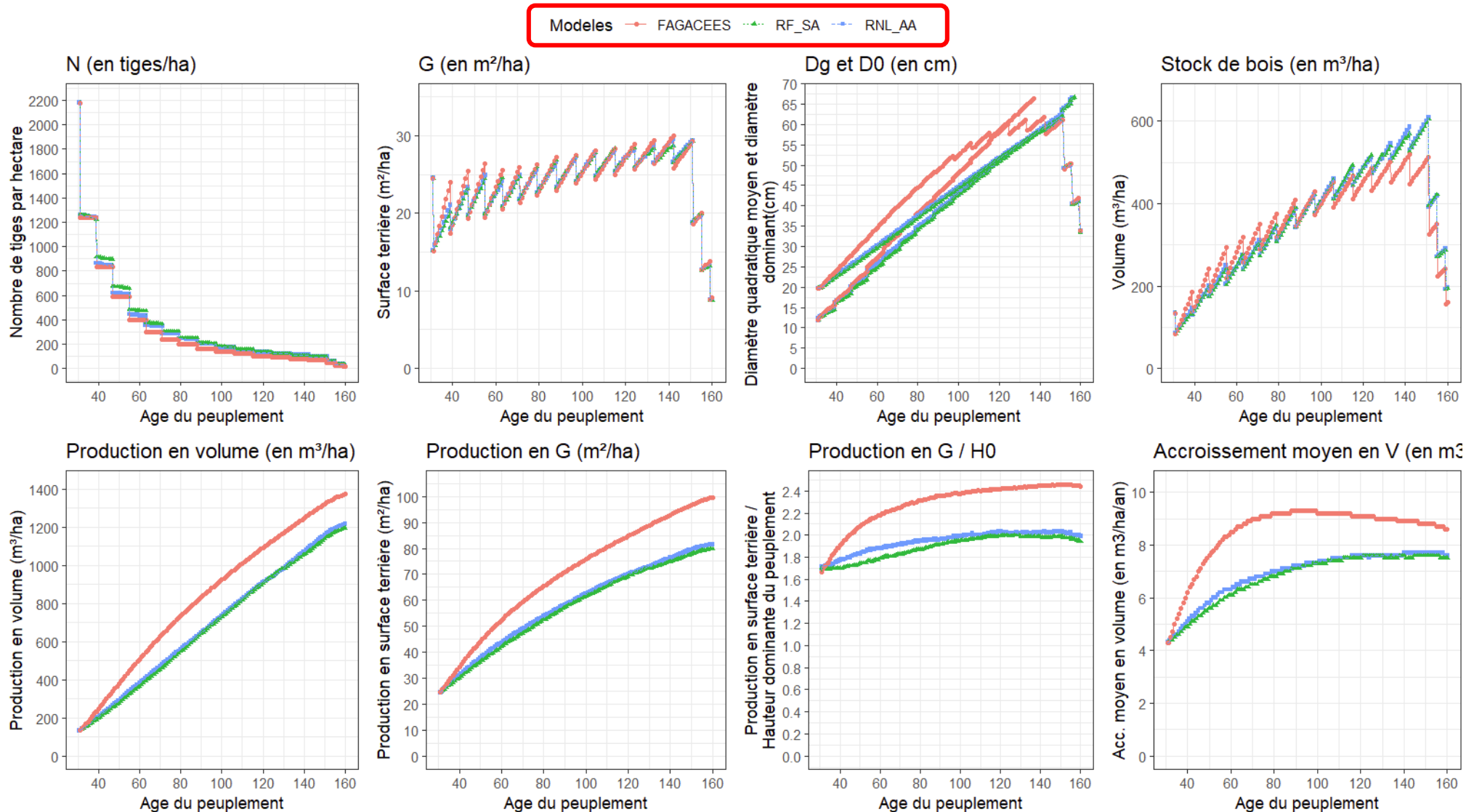
Modeles —●— FAGACEES —▲— RF\_SA —◆— RNL\_AA



Simulation de 3 modèles sur un peuplement régulier de chêne sessile sans sylviculture

→ Les simulations sont relativement proches

# Comparaisons de modèles en simulation (2)



Simulation de 3 modèles sur un peuplement régulier de chêne sessile avec sylviculture

→ Les simulations sont relativement proches

# Focus 1 : Bilan

→ **Modélisations établies pour 20 essences (représentant  $\approx$  92% de la superficie des forêts publiques)**

- Approche par **apprentissage**  $\approx$  ou  $>$  approche par **régression**
- Approche par **apprentissage opérationnelle en simulation** (avec ou sans sylviculture) → intégrée à Capsis dans Sydy2 (lien entre Java et R)

→ **mais durée de simulation + importante** que les RNL

- Reste à faire :
  - Validation des modèles sur données indépendantes (ex : GIS-Coop)
  - Simulations sur forêts « tests » (2024 – 2025)



## Focus 2 : Formation d'agrégats de points d'inventaire (notion de peuplement virtuel)

Application au passage d'une dendrométrie « 17,5 cm » à une dendrométrie « 7,5 cm »



# Objectif

- **Problématique: diamètre précomptage IFN < diamètre précomptage classique en gestion forestière**

- Données des aménagistes : diamètre précomptage = 17,5 cm
- Modèles Sydy2 : diamètre précomptage = 7,5 cm

## → Biais lors des simulations !

- Recalibrer les modèles sans les arbres de diamètre < 17,5cm dans les données IFN
  - Risque de ↘ des performances des modèles dans Sydy2
  - Gestion des stages « jeunes » déterminante → simulation utile !
- **Faire le lien entre une dendrométrie « 17,5 cm » et une dendrométrie « 7,5 cm »**

# Méthode

- Estimation variables dendrométriques → avec/sans les arbres de diamètre < 17,5 cm
  - Relations faisant le lien entre les 2 dendrométries
  - ...à quelle échelle estimer ces variables dendrométriques ?
    - 1 unique point d'inventaire → ne représente pas un peuplement forestier de façon assez fine  
→ risque de biais (et/ou d'erreurs très élevées)
  - **Formation d'agrégats de points d'inventaire → *constitution de peuplements forestiers plus consistants***
- **Les agrégats ainsi constitués seront dits « peuplements virtuels », car**
- **disjoints dans le temps (points d'inventaire issus de campagnes d'inventaire différentes)**
  - **disjoints dans l'espace (surfaces non continues)**

# Variables utilisées pour constituer et caractériser les peuplements virtuels

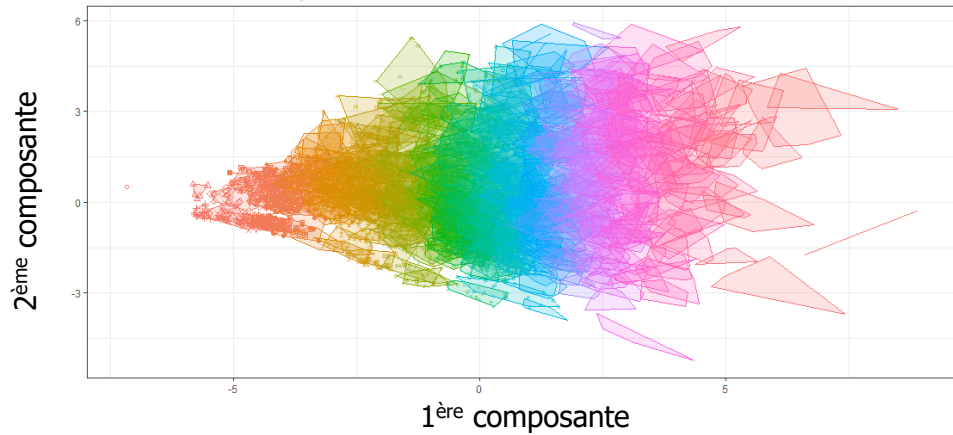
- Les points d'inventaires sont en premier lieu regroupés selon leur :
  - appartenance à une **GRECO**
  - **essence dominante** en termes de surface terrière
- Puis les points d'inventaire sont décrits par 11 variables:
  - **Dg** (diamètre quadratique moyen)
  - **DO** (diamètre dominant)
  - **H0** (hauteur dominante)
  - **G** (surface terrière)
  - **N/ha** (nombre de tiges par hectare)
  - **RDI** (relative density index)
  - **Age maximal** observé sur le point d'inventaire
  - **Indice de Shannon** (composition spécifique du point d'inventaire)
  - **Proportion de l'essence dominante en surface terrière**
  - **Indice de Gini** (structure forestière verticale)
  - **H0 / Age maximal** (indicateur du potentiel de croissance sur le point d'inventaire)

# Méthode HCPC (*Hierarchical Clustering on Principal Components*, Husson et al., 2010)

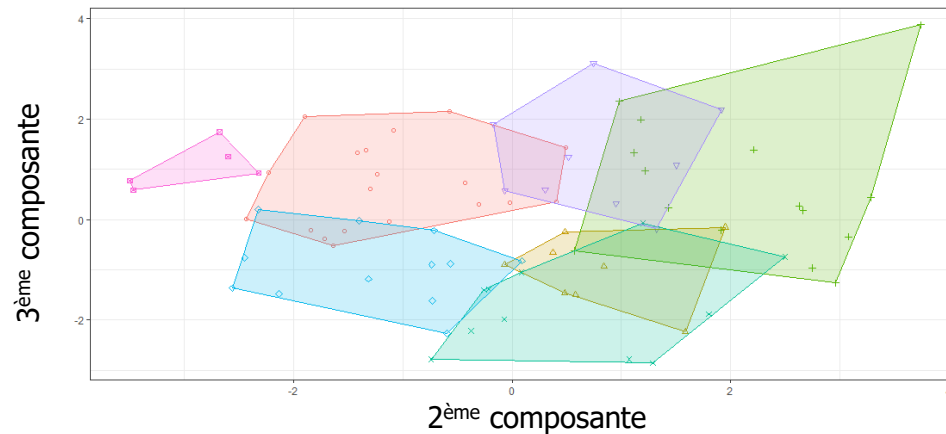
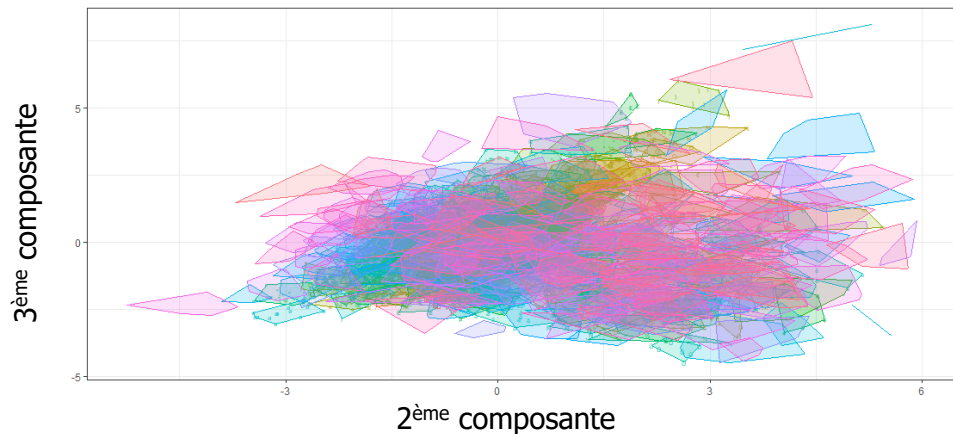
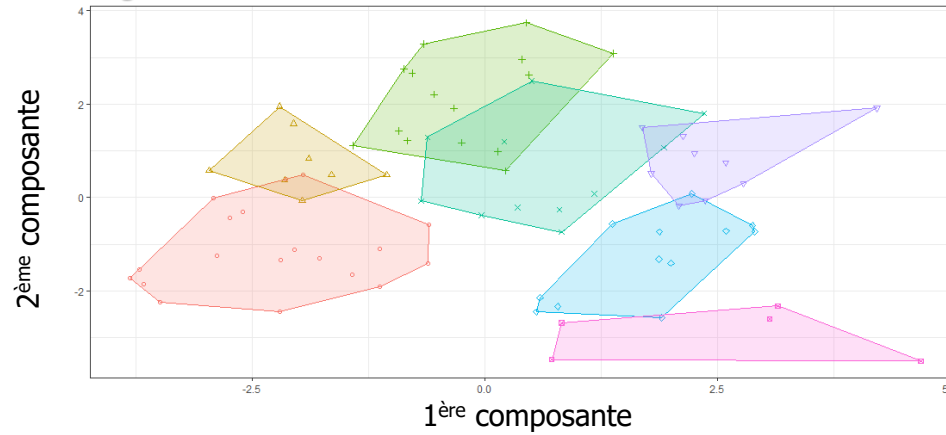
- Méthode de partitionnement comportant 3 étapes :
  - **Analyse des composantes principales (ACP)** → Réduction de la dimensionnalité de l'ensemble de variables
    - Elimine les variables corrélées et peu explicatives (retire le « bruit »)
    - Optimise les temps de calcul de la méthode
  - **Classification ascendante hiérarchique (CAH)** → Choix du nombre de peuplements virtuels
    - Fournie une mesure de l'homogénéité intra-agrégat (*within-class variance*) et une mesure de l'hétérogénéité inter-agrégat (*between-class variance*) selon le nombre de clusters utilisés
  - **Méthode des k-moyennes (*k-means*)** → Consolidation du dendrogramme issu de la CAH
    - Nécessite d'avoir prédéterminer au préalable le nombre d'agrégats à former

# Visualisation des peuplements virtuels

Peuplements virtuels du pin maritime dans le Sud-Ouest océanique



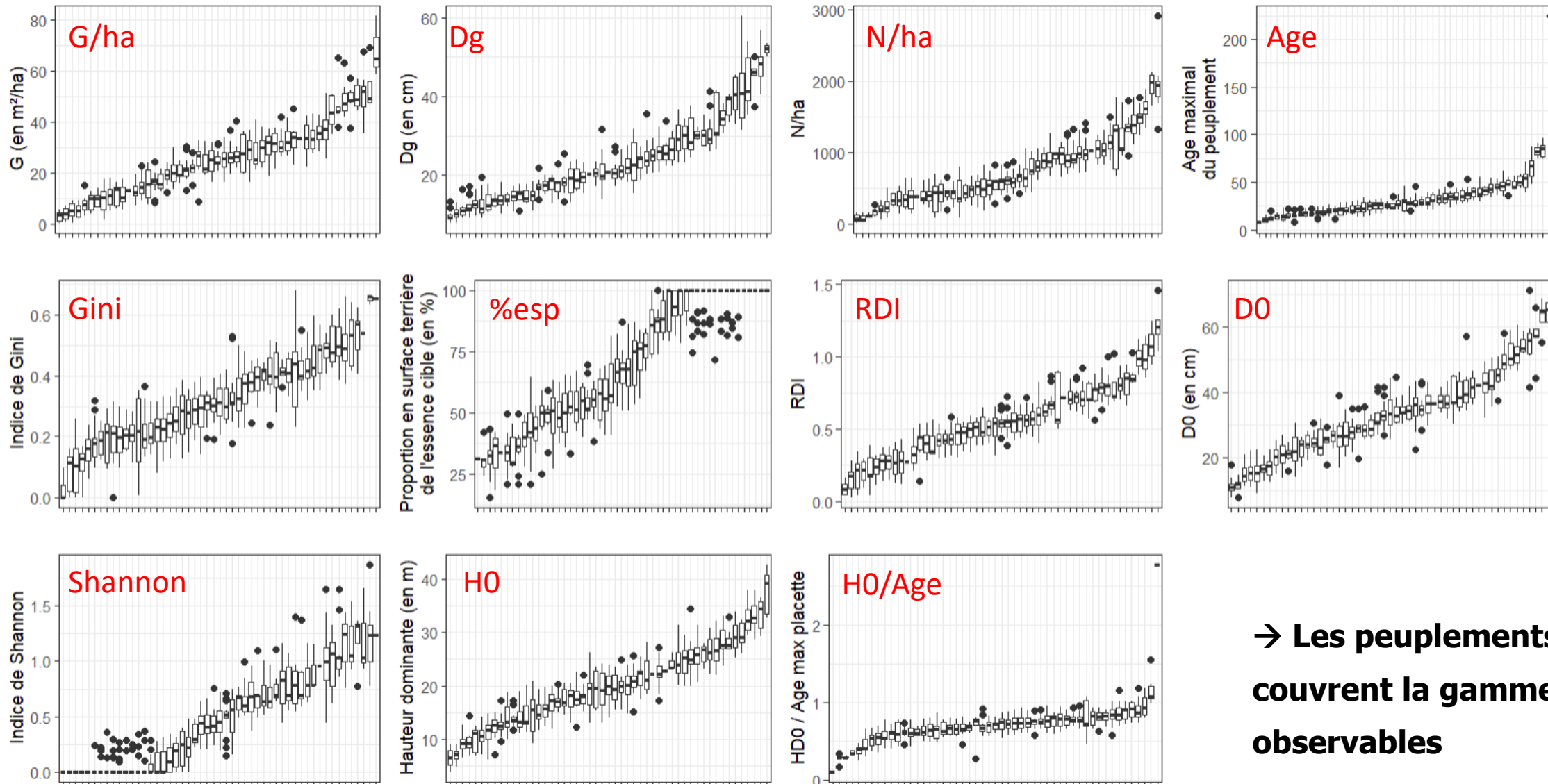
Peuplements virtuels du cèdre de l'Atlas dans la région méditerranéenne



→ Les peuplements virtuels semblent être distincts les uns des autres

Représentation des peuplements virtuels du pin maritime et du cèdre de l'Atlas sur les trois premières composantes (ou dimensions)

Diagramme de dispersion des différentes variables caractérisant les points d'inventaire par peuplements virtuels dans le Grand Ouest cristallin et océanique pour le Douglas



Peuplements virtuels

→ Les peuplements virtuels couvrent la gamme des observables

# Focus 2 : Bilan

- Partitionnement → **méthode générique et souple**
  - Peuplements virtuels distincts, couvrant la gamme des données observées
- A partir des peuplements virtuels, on a :
  - Relations faisant le lien entre les 2 dendrométries (7,5 et 17,5 cm)
  - Distribution diamétrique entre 7,5 et 17,5 cm reconstituée pour chaque peuplement dont on veut simuler l'évolution
- Les peuplements virtuels peuvent être utilisés à plusieurs fins → modélisation de la croissance, de la mortalité...du recrutement?

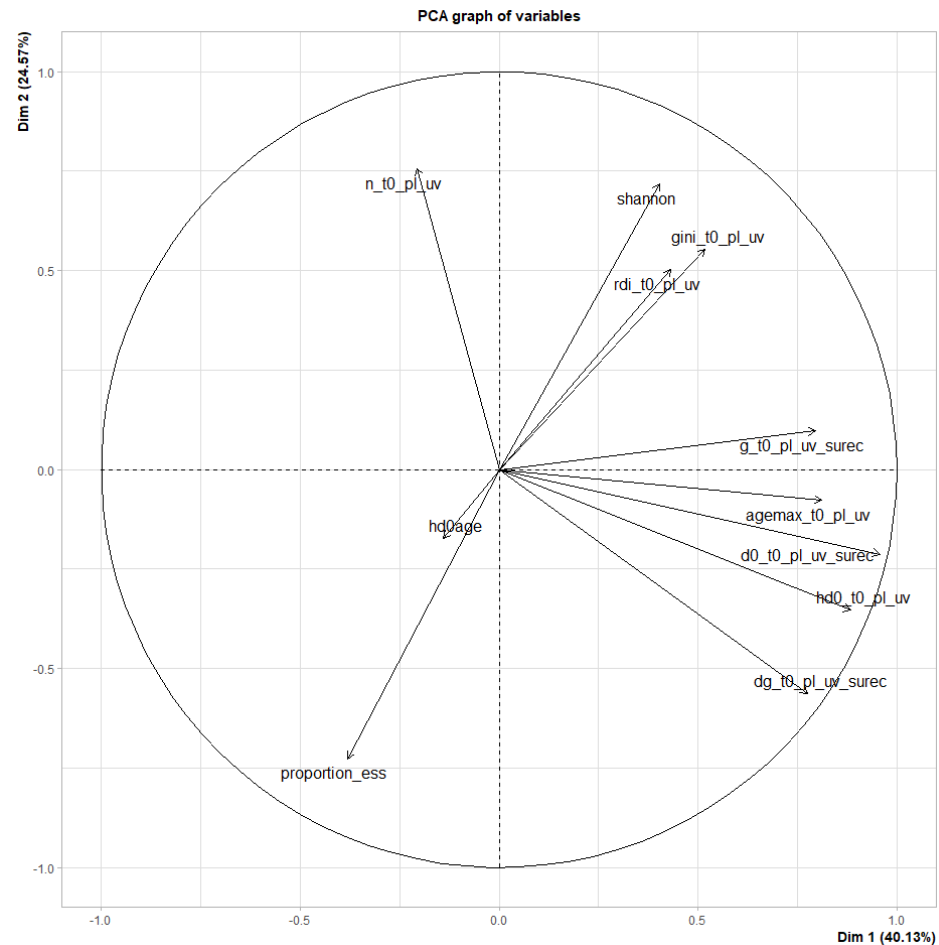
# Conclusion

- **Données IFN (IGN)** → modèles de croissance couvrant une large gamme (**respect du cahier des charges**)... mais ne s'appuie pas sur des expérimentations
    - **Comparaisons nécessaires avec l'expérimentation**
      - **via modèles existants** (modèles dans Capsis qui sous-tendent la plupart des guides sylvicoles (ONF et +))
      - évaluation sur des **évolutions mesurées en expérimentation** (ex : réseaux GIS-Coop de données pour la modélisation de la croissance)
  - Approche par **random forest convaincante**
    - Amélioration des performances relativement peu élevée (6% et 13%) → **résultat dépendant des essences** (ex : ↘ de l'erreur de 34% pour le sapin)
    - Approche limitée car gourmande en données + temps de calcul + réputée plus sensible que les RNL à l'extrapolation
- Comparer les RF avec une autre approche que les RNL → par exemple utiliser des modèles mixtes :
- Kuehne et al. (2022) → ↘ de 11% de l'erreur

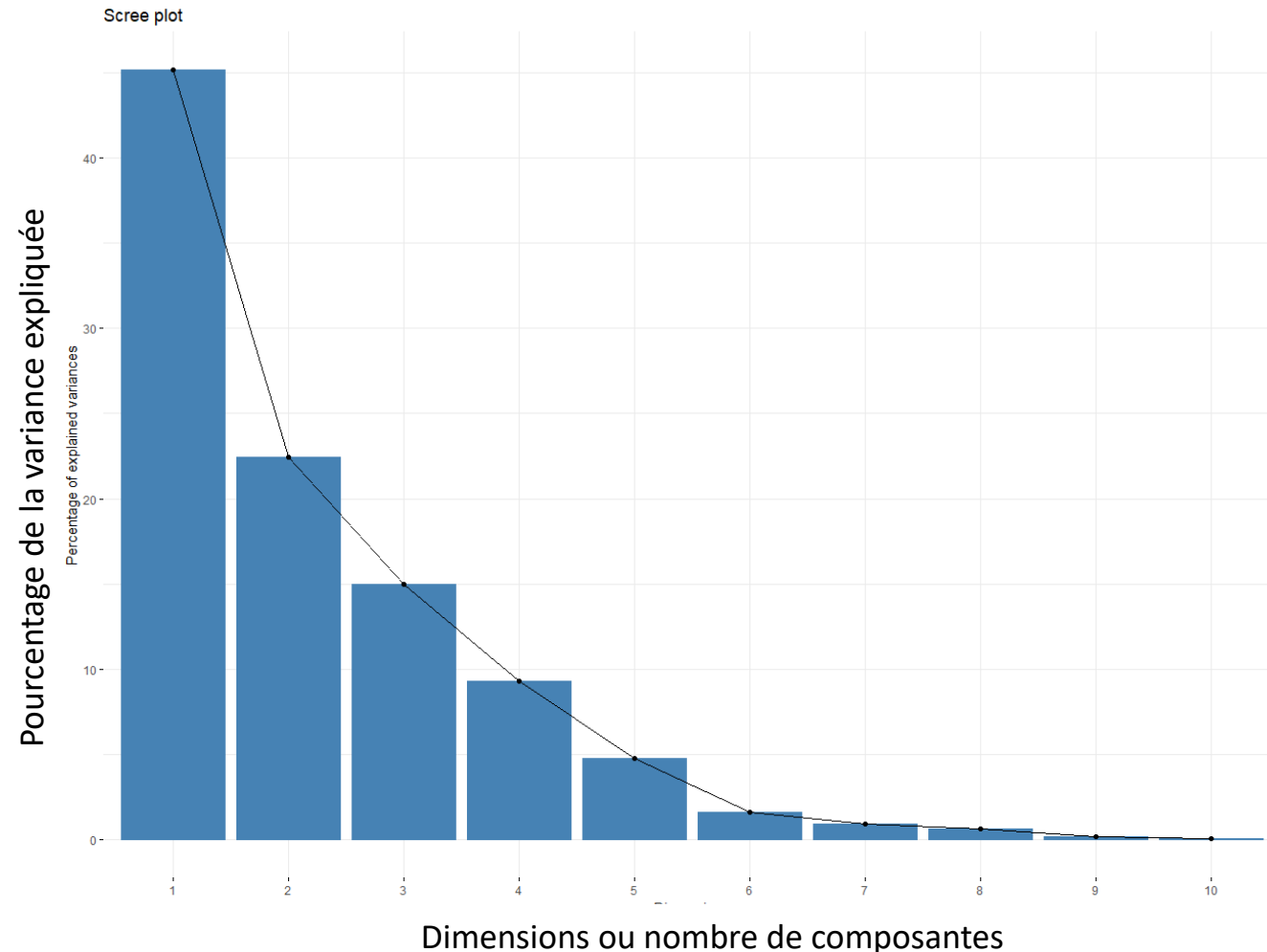


# Etape 1 → Réduction de la dimensionnalité

(illustration pour le Douglas en GRECO du Grand Ouest cristallin et océanique)



Cercle des corrélations établi par l'ACP

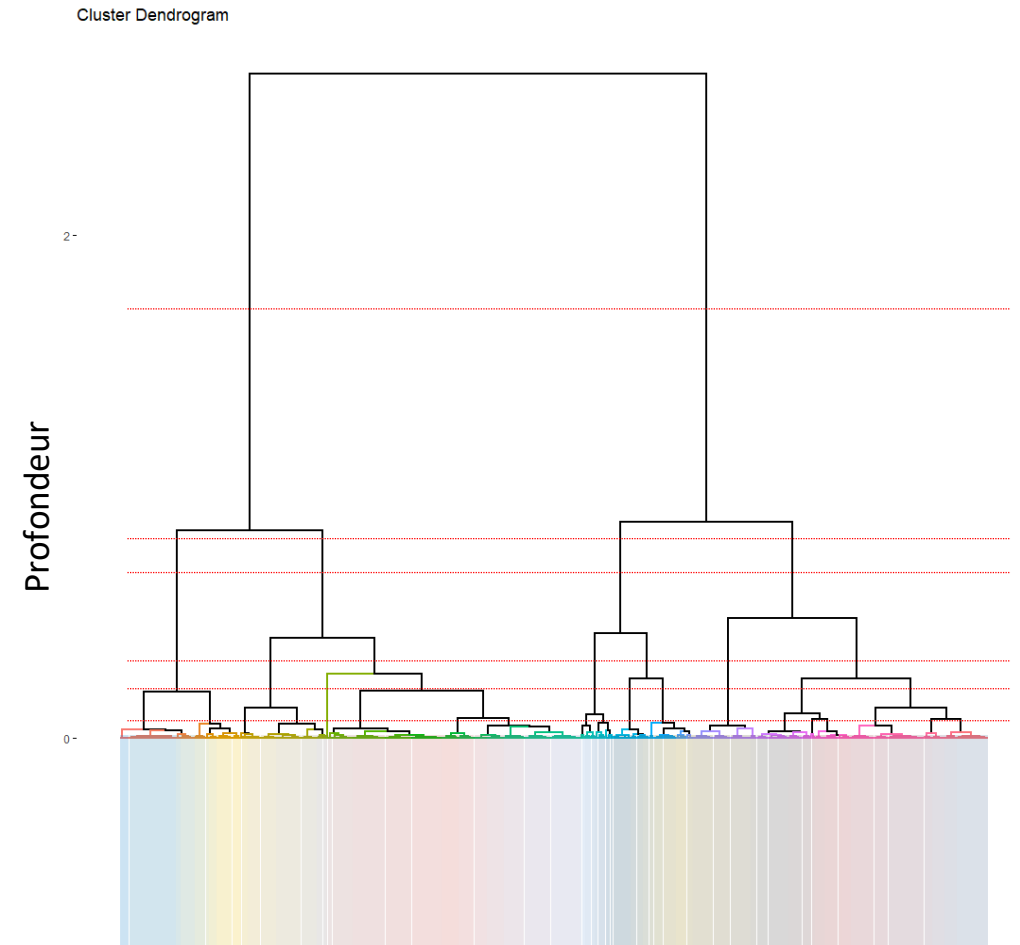


Choix du nombre de dimensions à utiliser selon la variance expliquée

# Etape 2 → Détermination du nombre de peuplements virtuels

- Classification ascendante hiérarchique (CAH)

→ Où couper l'arbre ?



Dendrogramme issu de la CAH pour le Douglas dans la GRECO Grand Ouest cristallin et océanique

# Détermination du nombre de peuplements virtuels → critère de décision

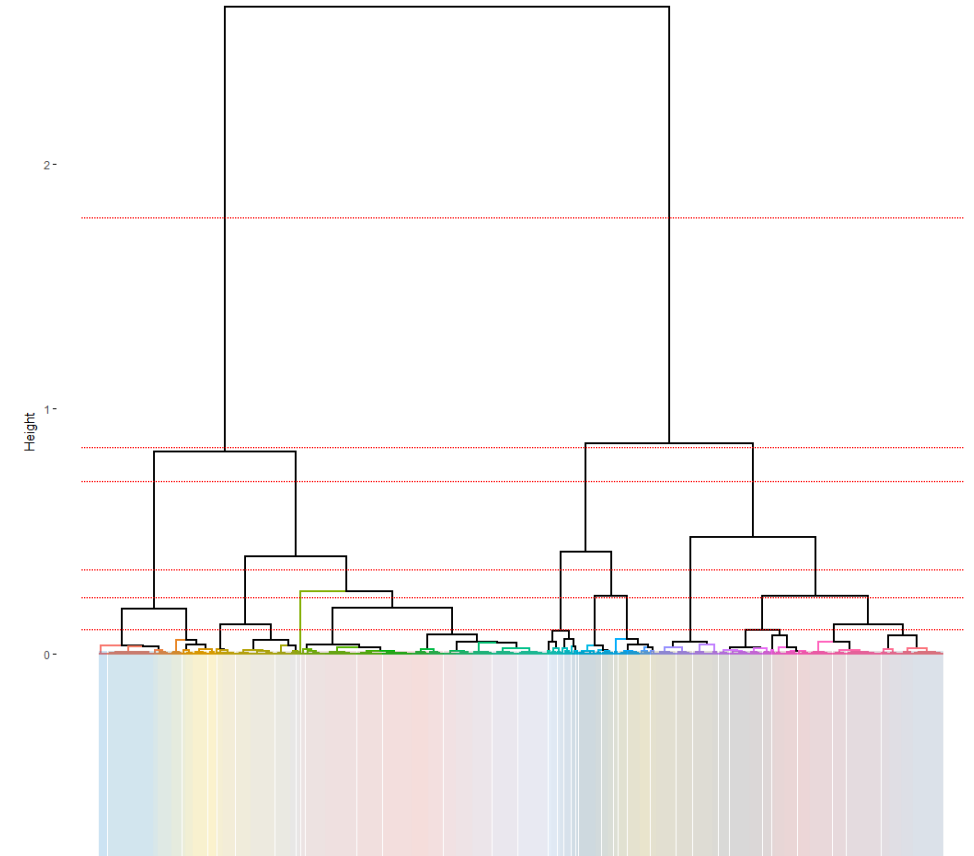
- **Critère quantitatif basé sur le nombre de points d'inventaire constituant les peuplements virtuels**

- Classes de peuplements virtuels :

- 1<sup>ère</sup> classe : 1 point d'inventaire
- 2<sup>nd</sup>e classe : 2 à 4 points d'inventaire
- **3<sup>ème</sup> classe : 5 à 24 points d'inventaire**
- 4<sup>ème</sup> classe : 25 à 49 points d'inventaire
- 5<sup>ème</sup> classe : + de 50 points d'inventaire

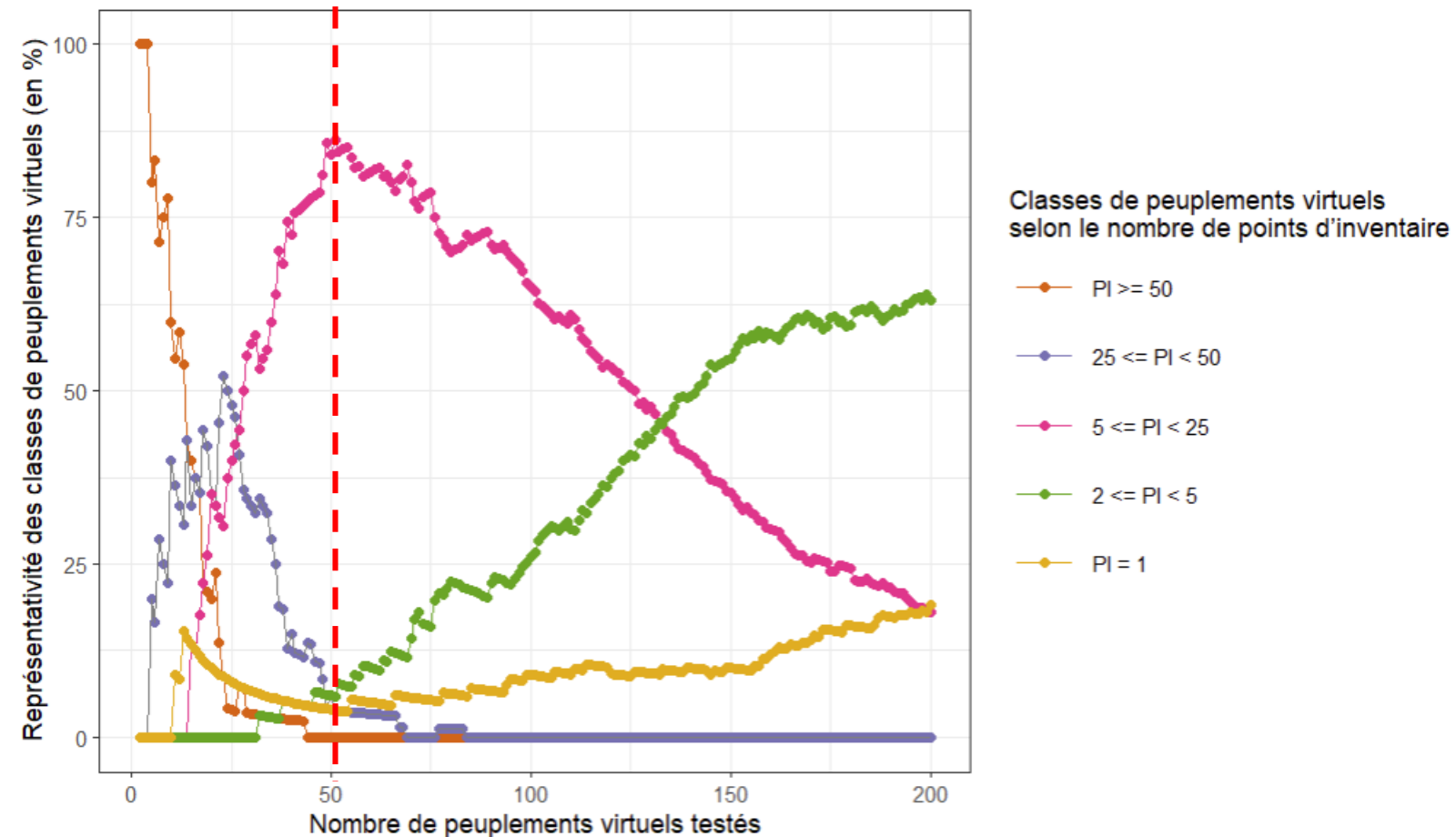
→ Quelle « coupe » permet de maximiser la représentativité des peuplements virtuels composés de 5 à 24 points d'inventaire ?

Cluster Dendrogram



Dendrogramme issu de la CAH pour le Douglas dans la GRECO Grand Ouest cristallin et océanique

# Détermination du nombre de peuplements virtuels



→ La « coupe » regroupant les points d'inventaire en 51 peuplements virtuels est sélectionnée

→ **Ce critère minimise le nombre de peuplements virtuels avec très peu, et beaucoup de points d'inventaire**

Proportion de classes de peuplements virtuels de Douglas dans la GRECO du Grand Ouest cristallin et océanique en fonction de la « coupe » testée sur la CAH

# Etape 3 → Consolidation du partitionnement ( $k$ -moyennes)

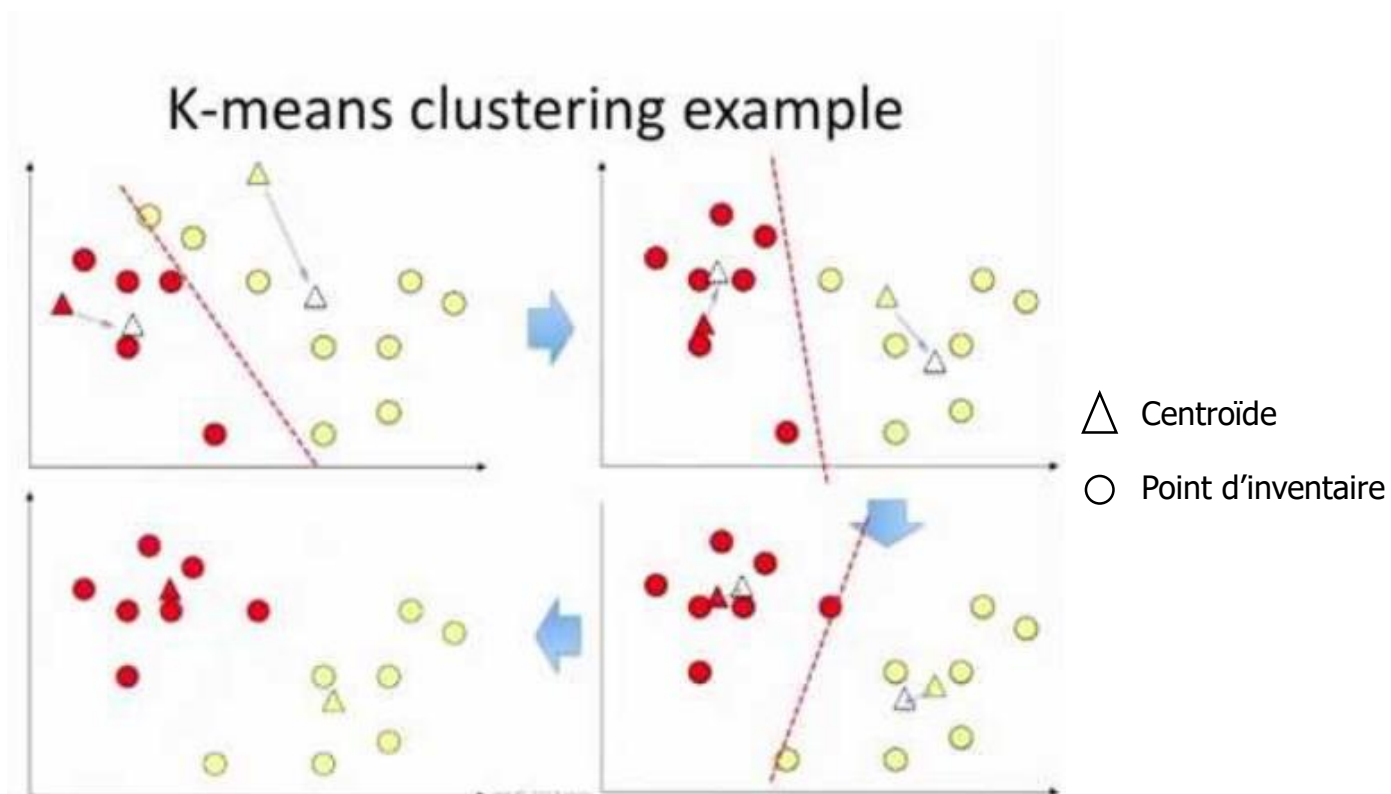


Schéma du principe de fonctionnement de la méthode  $k$ -moyennes

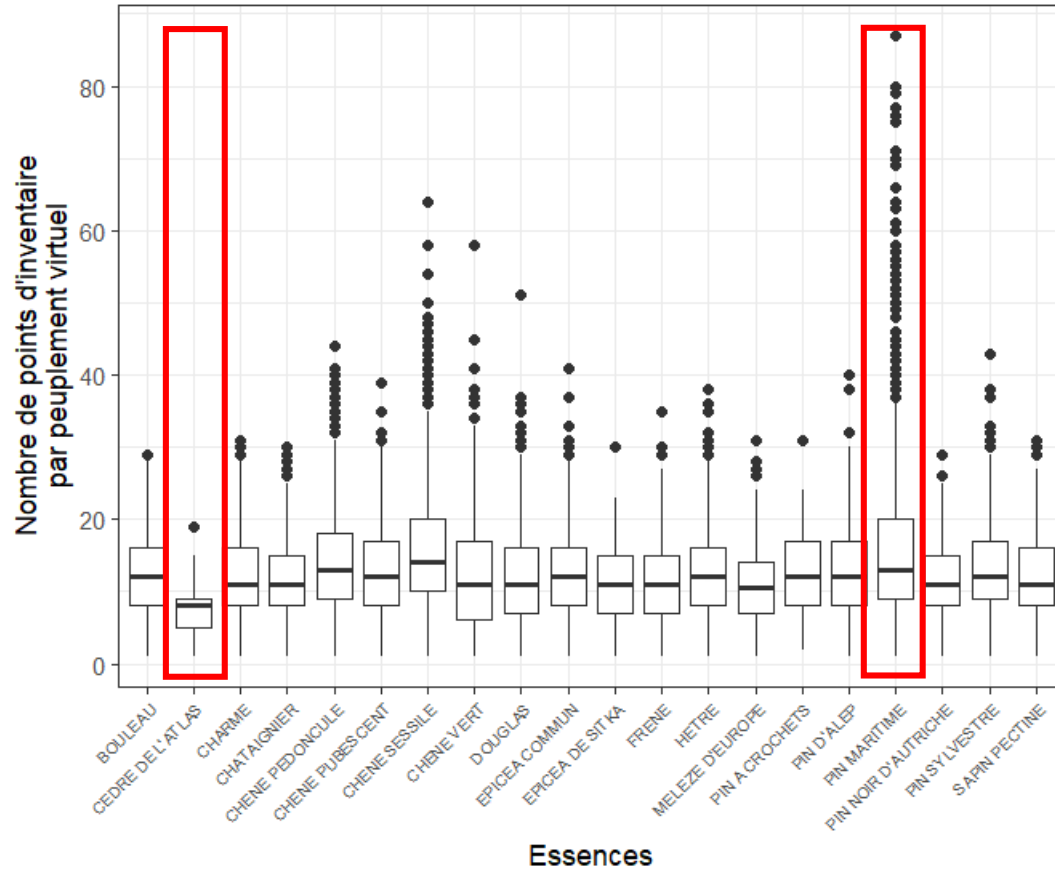
1)  $k$  centroïdes sont tirés aléatoirement dans un plan à  $n$  dimensions (correspondant aux nombres de variables dans le jeu de données) → il y a autant de peuplements virtuels que de centroïdes

2) La distance euclidienne entre chaque centroïde et chaque observation est calculée

3) L'observation est ensuite associée au centroïde le plus proche

4) Les 3 étapes sont réitérées jusqu'à la stabilisation des  $k$  centroïdes

# Résultats pour les 20 essences testées



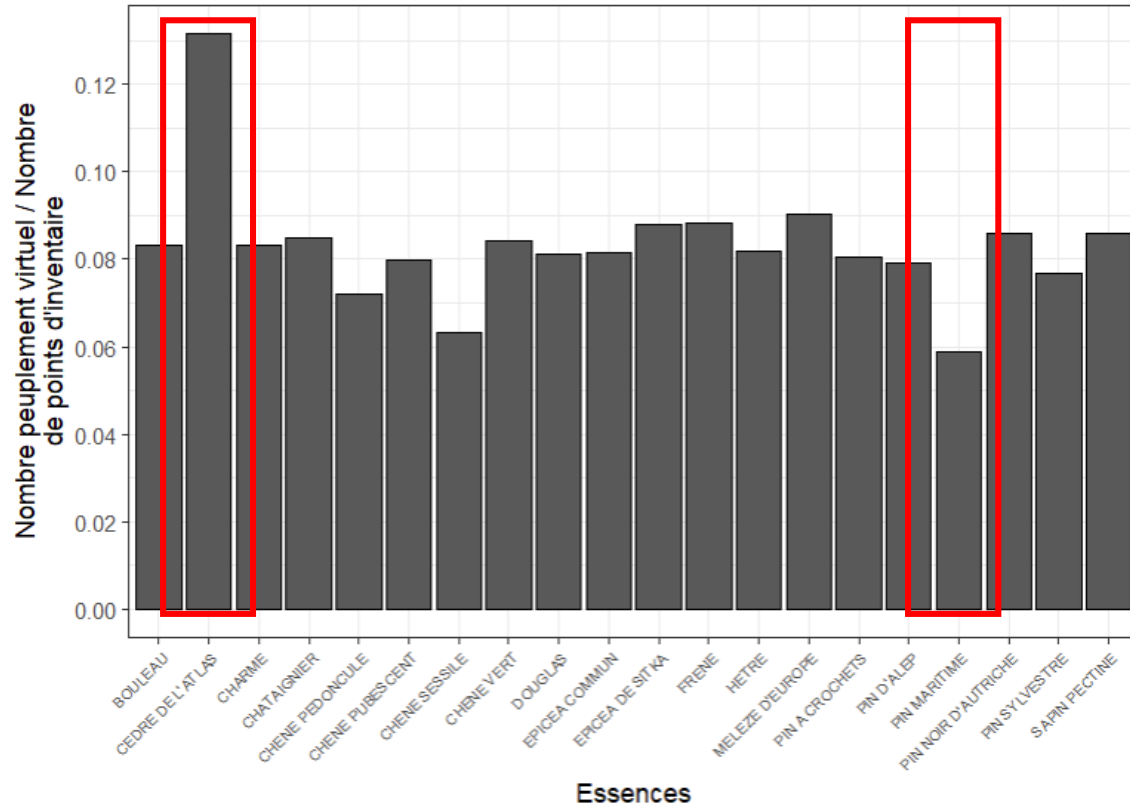
→ En moyenne : 826 peuplements virtuels par essence ( $\pm 605$ )

→ En moyenne : 13,0 points d'inventaire par peuplement virtuel ( $\pm 7,4$ )

Nombre de points d'inventaire au sein des peuplements virtuels selon l'essence

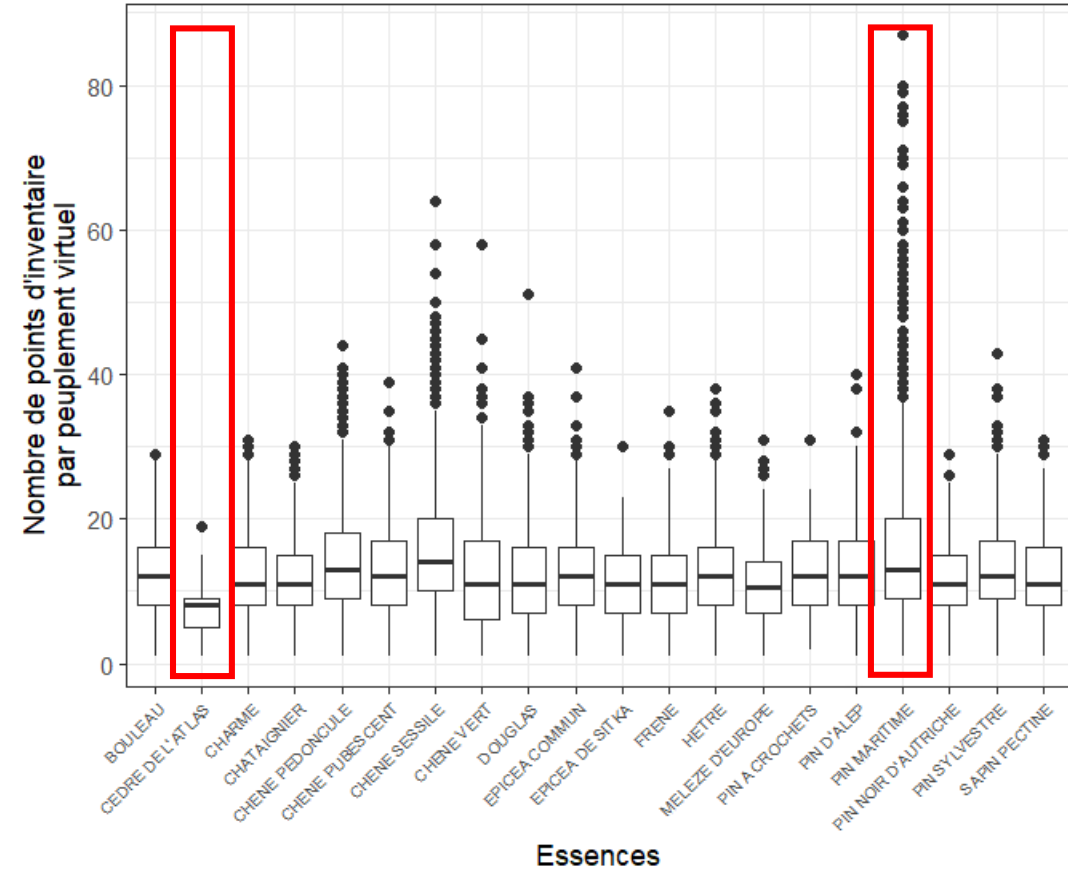
# Résultats pour les 20 essences testées

Nombre de peuplements virtuels par rapport au nombre de points d'inventaire par essence



Nombre de peuplements virtuels par essences

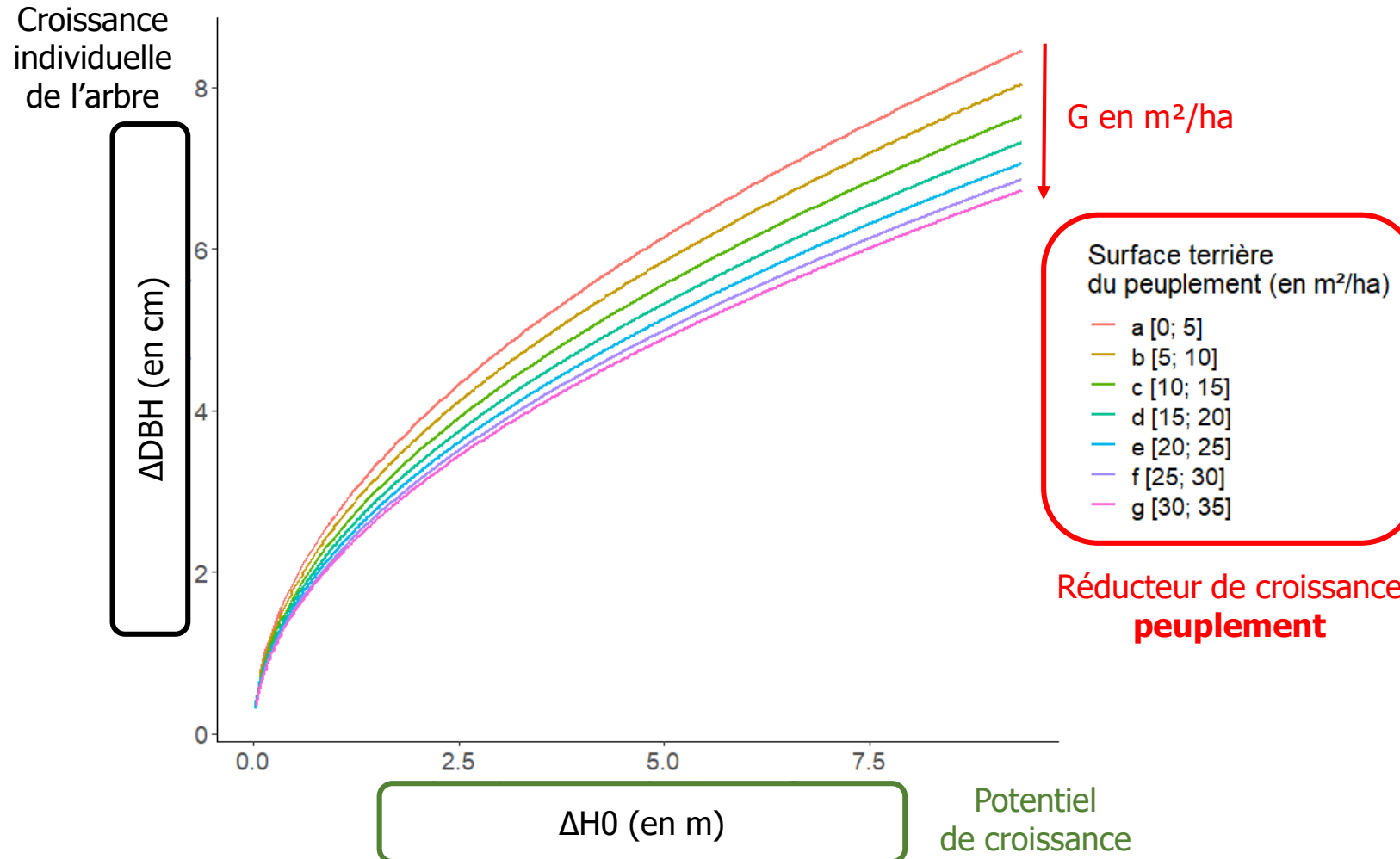
→ En moyenne : 826 peuplements virtuels par essence (± 605)



Nombre de points d'inventaire au sein des peuplements virtuels selon l'essence

→ En moyenne : 13,0 points d'inventaire par peuplement virtuel (± 7,4)

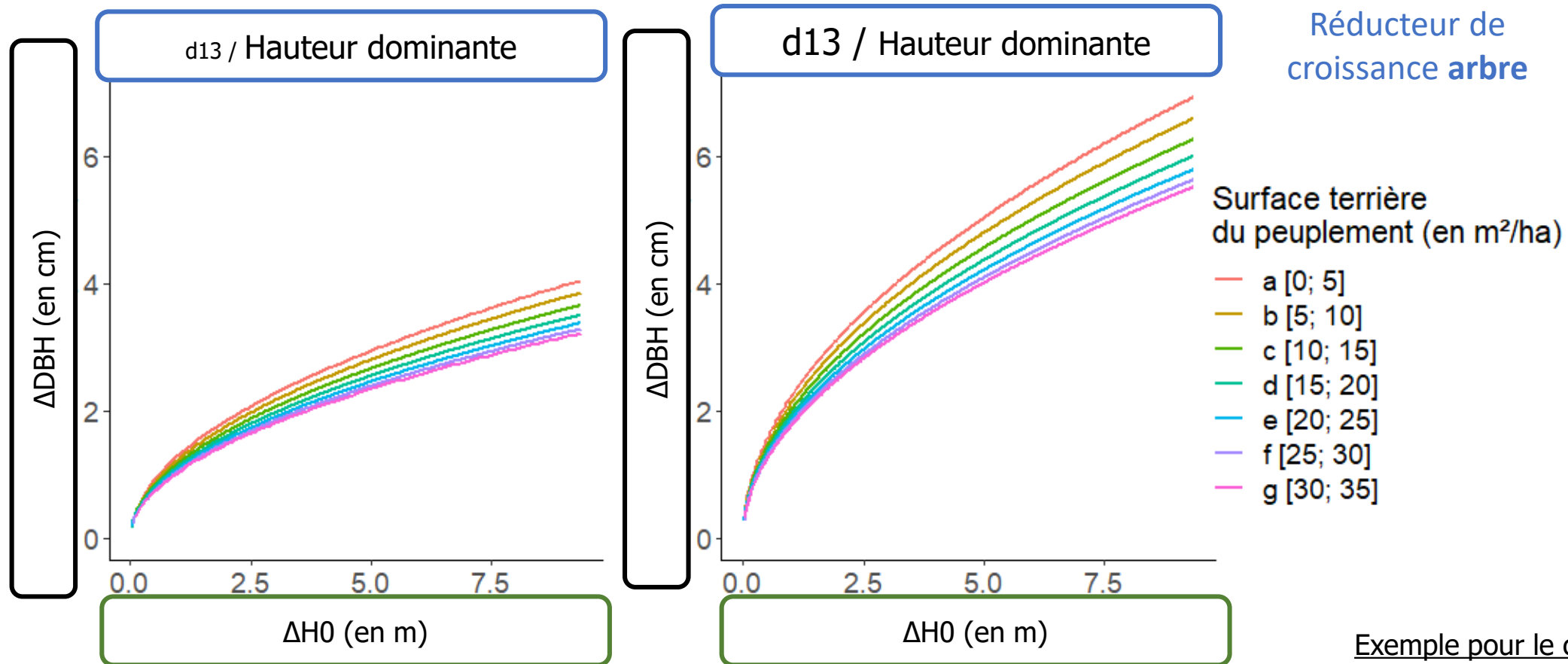
# Réponse de la croissance aux variables explicatives sélectionnées (1)



Exemple du chêne pédonculé



# Réponse de la croissance aux variables explicatives sélectionnées (2)



Exemple pour le chêne  
pédonculé