

Classification automatique

EL MAZZOUJI Wahel

GILLET Louison

2024/2025



Table des matières

1	Partie 1	3
2	Classification hiérarchique de parcelles forestières tropicales	3
2.1	Données	3
2.2	CAH des parcelles sur les densités de peuplement	3
2.2.1	Indice de Ward	3
2.2.2	Indice du saut maximum	5
2.2.3	Indice du saut moyen	6
2.2.4	Indice du saut minimum	7
2.3	R^2 des différentes partitions	7
2.4	Optimisation d'une partition avec les K-means	7

1 Partie 1

2 Classification hiérarchique de parcelles forestières tropicales

2.1 Données

Nous voulons proposer plusieurs classifications des parcelles décrites par leur peuplement arboré. Pour cela nous disposons des données etc Prépa des données : standardisation

2.2 CAH des parcelles sur les densités de peuplement

Dans cette partie, nous allons traiter différentes méthodes de classification hiérarchique ascendante. Ces classifications se feront à partir de plusieurs indices. Il existe essentiellement deux familles d'indice à savoir les indices d'inertie et les indices de saut.

2.2.1 Indice de Ward

Nous nous intéressons tout d'abord à la méthode de classification par l'indice d'inertie externe, autrement appelé indice de Ward. Rappelons que pour deux centres de gravité A et B , l'indice de Ward est donné par :

$$\mu(A, B) = \ln(\{(\bar{x}_A, w_A), (\bar{x}_B, w_B)\}) = \frac{w_A w_B}{w_A + w_B} \|\bar{x}_A - \bar{x}_B\|_2^2$$

Pour mettre en place cette classification, on crée une matrice des distances euclidiennes par la fonction *dist* de R. Ensuite, nous utilisons la fonction *hclust* :

```
dp <- dist(datapeuple, method = "euclidean")
CAHWard <- hclust(d = dp, method = "ward.D")
```

Ces deux fonctions nous permettent de tracer le dendrogramme de la hiérarchie indiquée par la méthode de Ward représenté par la figure ci-dessous.

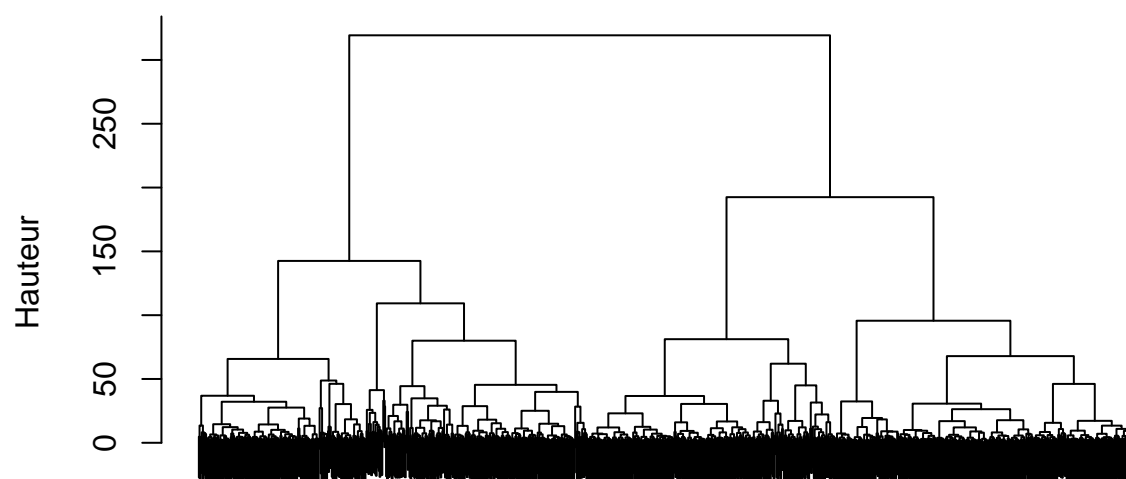


Figure 1: Dendrogramme - Méthode de Ward

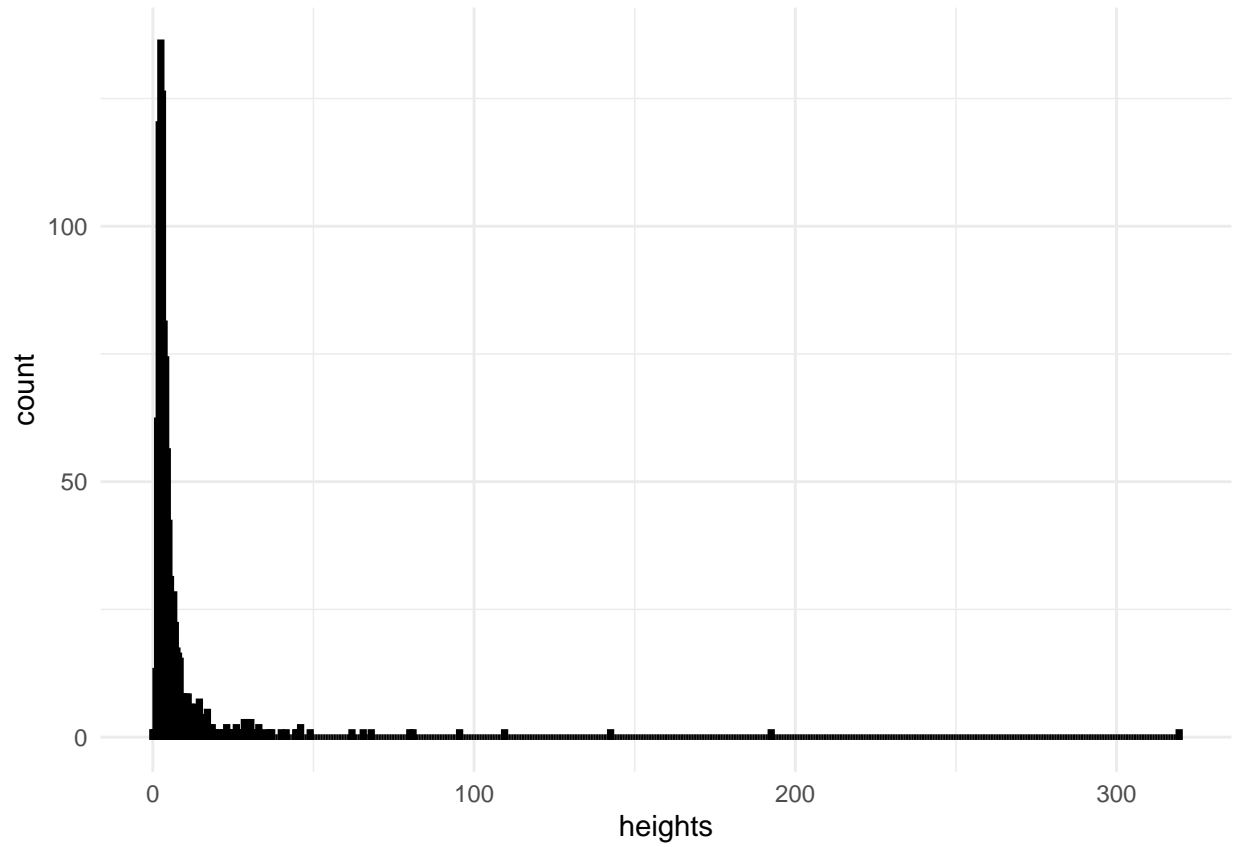


Figure 2: Histogramme - Ward

2.2.2 Indice du saut maximum

Pour deux classes A et B , l'indice de saut maximum est donné par

$$\mu(A, B) = \max_{a \in A, b \in B} (d(a, b))$$

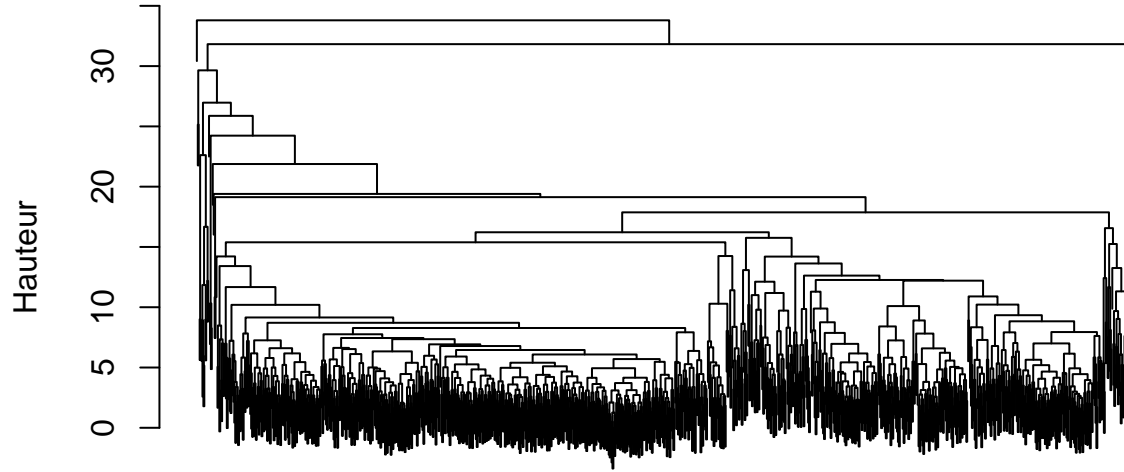


Figure 3: Dendrogramme - Méthode du saut maximum

2.2.3 Indice du saut moyen

Pour deux classes A et B , l'indice de saut moyen est donné par

$$\mu(A, B) = \frac{1}{w_A w_B} \sum_{a \in A, b \in B} w_a w_b d(a, b)$$

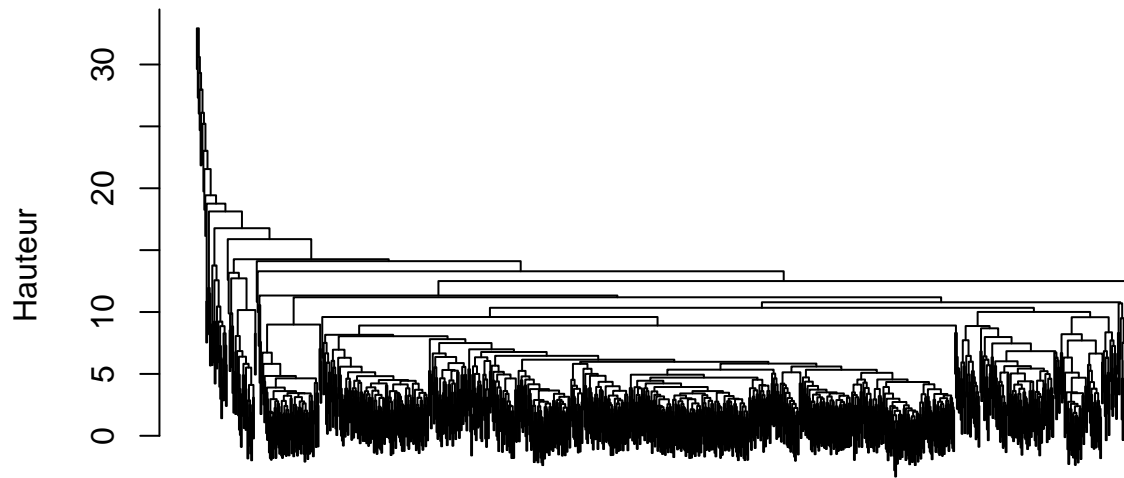


Figure 4: Dendrogramme - Méthode du saut moyen

2.2.4 Indice du saut minimum

Pas traité car : justification

2.3 R^2 des différentes partitions

2.4 Optimisation d'une partition avec les K-means

Nous cherchons désormais à optimiser chacune de nos partitions jugées prometteuses issues de la CAH grâce à la méthodes des K-means.