

# Une nouvelle méthode de Classification Ascendante Hiérarchique avec contraintes : hclust.mix

L. BELLANGER<sup>a</sup>, A. COULON<sup>a,b</sup> and P. HUSI<sup>b</sup>

<sup>a</sup> *Université de Nantes Laboratoire Jean Leray UMR 6629, 2 rue de la  
Houssinière BP 92208 44322 Nantes Cedex 03, France*

<sup>b</sup> *CNRS/Université de Tours, UMR CITERES, Laboratoire Archéologie et  
Territoires, 40 rue James Watt, ActiCampus 1, 37200 Tours, France*

25 septembre 2020

Workshop virtuel - Nouvelles méthodes pour l'analyse descriptive et  
prédictive de données massives et structurées  
25 septembre 2020

# Plan

- 1 Introduction
- 2 Bref aperçu des méthodes CAH avec contraintes existantes
- 3 La méthode `hclust.mix`
- 4 Applications
- 5 Conclusions
- 6 Références



# Plan

- 1 Introduction
  - Objectifs
  - Motivations
- 2 Bref aperçu des méthodes CAH avec contraintes existantes
- 3 La méthode `hclust.mix`
- 4 Applications
- 5 Conclusions
- 6 Références



# Objectifs

- Il existe de nombreuses méthodes de clustering nécessitant toutes des prétraitements des données spécifiques (eg calcul dissimilarités, ...) (cf [1], [5], [9], [8], ...).
- Les deux algorithmes de clustering les plus répandus sont : (i) partitionnement, (ii) hiérarchique.
- Quand il existe une information *a priori* sur les relations entre les individus (eg relation de voisinage spatial, relation d'ordre (en génomique), ...), l'utilisation de méthodes de clustering avec contraintes permet d'en tenir compte pour construire les groupes.
- Dans ce travail, nous présentons un nouvel algorithme de CAH avec contraintes d'ordre.



# Motivations

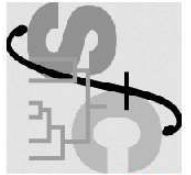


- un projet pluri-disciplinaire : "ModAThom" (projet ANR, 2018-2022) Modèle explicatif de la fabrique d'Angkor Thom : archéologie d'une ville capitale disparue ;
- un problème posé par les archéologues : disposer d'un outil statistique pour les aider à périodiser les ensembles archéologiques étudiés (diagramme des ensembles périodisés) ;
- un nouvel algorithme d'apprentissage semi-supervisé, prenant en compte, deux sources d'information associées aux mêmes observations, potentiellement sujettes à erreur.



# Plan

- 1 Introduction
- 2 **Bref aperçu des méthodes CAH avec contraintes existantes**
- 3 La méthode `hclust.mix`
- 4 Applications
- 5 Conclusions
- 6 Références



# CAH avec contraintes d'ordre : CAHCO (Ferligo et al., 1982 ; Legendre et al., 1985)

La CAH avec contraintes de contiguïté CAHCO ([6], [2]) restreint les regroupements possibles aux objets dits contigus.

- Les méthodes CAH avec stratégies d'agrégation vérifiant la formule de Lance et Williams ([1]) sont facilement modifiables pour intégrer les contraintes de contiguïté.
- La solution la plus courante de regroupement contraint : utiliser des schémas de connexion simples (e.g. Legendre and Legendre, 2012 p774-775).
- **Inconvénient majeur** : seules les dissimilarités entre individus contigus sont prises en compte :
  - trop restrictif dans certains domaines d'application tels que l'archéologie !



# CAH avec contraintes de proximité géographique : hclustgeo (Chavent et al., 2018)

Notations :

- $\mathbf{D}_1$  la matrice  $n \times n$  de dissimilarité normalisée associée aux  $n$  observations dans  $\mathbb{R}^p$ , contenues dans  $\mathbf{X}$  ;
- $\mathbf{D}_2$  la matrice  $n \times n$  de distances géographiques entre les  $n$  observations.

But de hclustgeo [3] : intégrer dans une CAH de "type Ward", les contraintes géographiques contenues dans  $\mathbf{D}_2$ , en utilisant un paramètre de mélange  $\alpha \in [0, 1]$ .





# Principe de la méthode hclustgeo

- Définir  $H(\mathcal{C}_k)$  critère d'hétérogénéité du groupe  $\mathcal{C}_k$  :

$$H(\mathcal{C}_k) = \alpha I(\mathcal{C}_k, \mathbf{D}_1) + (1 - \alpha) I(\mathcal{C}_k, \mathbf{D}_2); \alpha \in [0, 1]$$

où  $I(\mathcal{C}_k, \mathbf{D}_1)$  (resp.  $I(\mathcal{C}_k, \mathbf{D}_2)$ ) est l'inertie des obs. du groupe  $\mathcal{C}_k$  calculée à partir de  $\mathbf{D}_1$  (resp.  $\mathbf{D}_2$ ).

- A chaque itération, minimiser  $\mathcal{H}(\mathcal{P}_K)$  critère d'hétérogénéité de partition, pour une partition  $\mathcal{P}_K$  en  $K$  groupes, avec  $\mathcal{H}(\mathcal{P}_K) = \sum^K H(\mathcal{C}_k)$ .

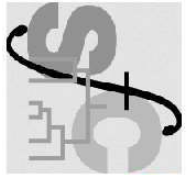
**Inconvénients** : CAH uniquement avec stratégie d'agrégation de type Ward ; choix de  $\alpha$  pas évident ; les groupes obtenus ne sont pas toujours des zones géographiques compactes.

hclustgeo est disponible dans le R package ClustGeo.



# Plan

- 1 Introduction
- 2 Bref aperçu des méthodes CAH avec contraintes existantes
- 3 La méthode `hclust.mix`**
- 4 Applications
- 5 Conclusions
- 6 Références



## hclust.mix : Bellanger, Coulon, Husi (2020) [2]

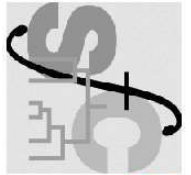
Une approche basée sur la distance est adoptée pour modifier la mesure de dissimilarité dans l'algorithme CAH classique.

Notations :

- $\mathbf{D}_1$  la matrice  $n \times n$  de dissimilarité normalisée associée aux  $n$  observations dans  $\mathbb{R}^p$ , contenues dans  $\mathbf{X}$  ;
- $\mathbf{D}_2$  la matrice  $n \times n$  de dissimilarité associée aux  $n$  observations (eg distances géographiques). Pour l'application archéologique  $\mathbf{D}_2 = \mathbf{1}_{(n \times n)} - \mathbf{S}_2$ , où  $\mathbf{S}_2$  la matrice binaire de connectivité entre les observations.

$$\mathbf{D}_\alpha = \alpha \mathbf{D}_1 + (1 - \alpha) \mathbf{D}_2 \quad (1)$$

$\alpha \in [0, 1]$  paramètre fixé donnant l'importance de chaque matrice de dissimilarité.



hclust.mix : critère de choix de  $\alpha$ 

Critère pour choisir  $\alpha \in [0; 1]$

$$\hat{\alpha} = \min_{\alpha} \text{CorCrit}_{\alpha} \quad (2)$$

où

$$\text{CorCrit}_{\alpha} = |\text{Cor}(\mathbf{D}_{\alpha}^{\text{coph}}, \mathbf{D}_1) - \text{Cor}(\mathbf{D}_{\alpha}^{\text{coph}}, \mathbf{D}_2)| \quad (3)$$

$\mathbf{D}_{\alpha}^{\text{coph}}$  matrice cophénétique (ou ultramétrique) (voir [5])  
obtenue à partir du dendrogramme CAH avec  $\alpha$  fixé dans  $\mathbf{D}_{\alpha}$ .

- Le critère (3) "équilibre" le poids de  $\mathbf{D}_1$  et  $\mathbf{D}_2$  dans la classification finale ; mais
- seule une estimation *ponctuelle* est obtenue avec (2) !
- obtention d'un *intervalle de confiance* via un processus de rééchantillonnage.



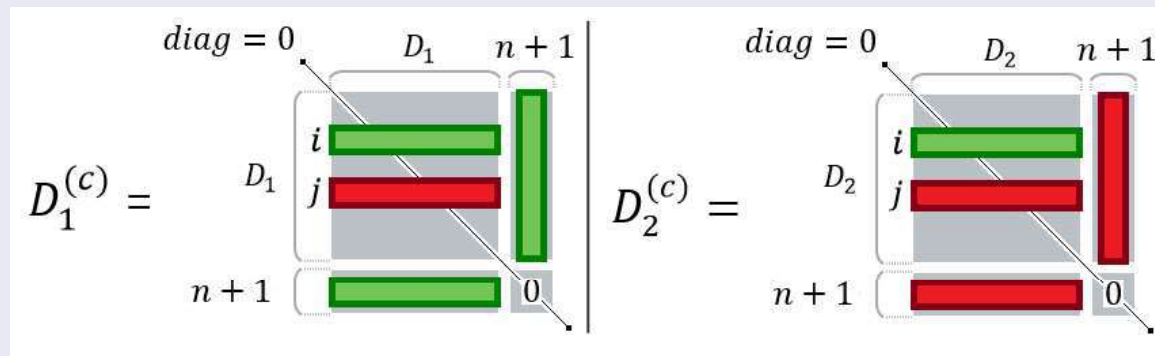
# hclust.mix : $IC_{95\%}(\alpha)$ par procédure de rééchantillonnage

Même esprit que LOO mais "Add" et non "Leave" !

## Construction des clones $D_1^{(c)}$ et $D_2^{(c)}$

Soit  $D_1^{(c)}$  (resp.  $D_2^{(c)}$ ) matrice carrée  $(n+1) \times (n+1)$  définie à partir de  $D_1$  (resp.  $D_2$ )  $c \in \{1, \dots, n(n-1)\}$ . Pour  $c$  fixé, copier à la position "clone"  $(n+1)$  et créer :

- $D_1^{(c)}$  : observation  $i \in \{1, \dots, n\}$  ;  $D_2^{(c)}$  : observation  $j \in \{1, \dots, n\} \setminus \{i\}$



Puis, calcul moyenne, écart-type et IC basé sur les percentiles ([4]).

# hclust.mix : $IC_{95\%}(\alpha)$ par procédure de rééchantillonnage (1)

**for**  $c \in 1, \dots, n(n-1)$  **do**

Generate a clone and create  $\mathbf{D}_1^{(c)}$  and  $\mathbf{D}_2^{(c)}$

Define  $\mathbf{D}_\alpha^{(c)}$  and  $CorCrit_\alpha^{(c)}$  where

$\mathbf{D}_\alpha^{(c)} = \alpha \mathbf{D}_1^{(c)} + (1 - \alpha) \mathbf{D}_2^{(c)}$  and

$CorCrit_\alpha^{(c)} = |Cor(\mathbf{D}_\alpha^{coph(c)}, \mathbf{D}_1^{(c)} - \mathbf{D}_2^{(c)})|$

Evaluate  $\hat{\alpha}^{(c)} = \min_{\alpha \in [0;1]} CorCrit_\alpha^{(c)}$  replication of  $\hat{\alpha}$  for each clone  $c$

**end**

Obtain :

- $\hat{\alpha}^* = \frac{1}{n(n-1)} \sum_{c=1}^{n(n-1)} \hat{\alpha}^{(c)}$ , the replication estimation of  $\alpha$  ;
- $\hat{se}^* = \sqrt{\frac{\sum_{c=1}^{n(n-1)} (\hat{\alpha}^{(c)} - \hat{\alpha}^*)^2}{n(n-1) - 1}}$ , the estimated standard error of  $se(\hat{\alpha})$  using the sample standard deviation of the  $n(n-1)$  replications ;
- A percentile confidence interval  $IC_{95\%}(\alpha)$  based on replications ;

**Algorithm 1:** Estimation of variability of  $\hat{\alpha}$  parameter



hclust.mix : CAH sur  $\mathbf{D}_{\hat{\alpha}}$  ou  $\mathbf{D}_{\tilde{\alpha}}$ 

Un algorithme CAH est alors mis en oeuvre en utilisant

- $\mathbf{D}_{\hat{\alpha}}$  ou  $\mathbf{D}_{\tilde{\alpha}}$  où  $\tilde{\alpha}$  est une valeur interprétable (arrondie) de  $\alpha$  contenue dans  $IC_{95\%}(\alpha)$ .

La fonction `hclust.mix` est implémentée dans le R package SPARTAAS en cours de développement.



# Plan

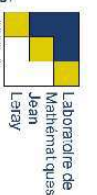
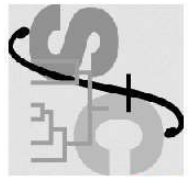
- 1 Introduction
- 2 Bref aperçu des méthodes CAH avec contraintes existantes
- 3 La méthode `hclust.mix`
- 4 Applications**
  - Données archéologiques
  - Données Estuaire
- 5 Conclusions
- 6 Références



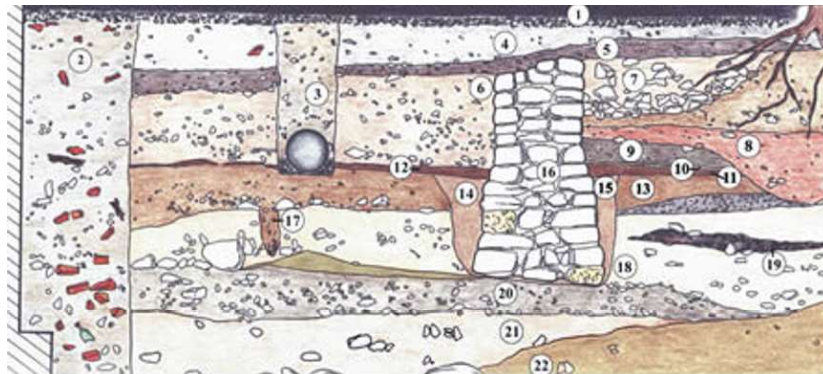


# Données archéologiques disponibles

- L'ensemble des données archéologiques est constitué de la céramique mise à jour sur 3 sites d'Angkor Thom ([7])
- La contrainte utilisée correspond à l'ensemble des relations stratigraphiques entre ensembles archéologiques (cf Fig 1)



# Données archéologiques disponibles



|        | AI01 | AI02 | AI03 | AI04 | AO03 | AI05 | AO01 | AI07 | AI08 | AO02 | AI06 | AO04 | APQR01 | APQR02 | AO05 | APQR03 | AI09 |
|--------|------|------|------|------|------|------|------|------|------|------|------|------|--------|--------|------|--------|------|
| AI01   | 1    | 0    | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 0      | 0      | 0    | 0      | 0    |
| AI02   | 0    | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0      | 0      | 0    | 0      | 0    |
| AI03   | 0    | 1    | 1    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0      | 0      | 0    | 0      | 0    |
| AI04   | 0    | 0    | 1    | 1    | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 0      | 0      | 0    | 0      | 0    |
| AO03   | 0    | 0    | 0    | 0    | 1    | 0    | 1    | 0    | 0    | 1    | 0    | 1    | 0      | 0      | 0    | 0      | 0    |
| AI05   | 1    | 0    | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 1    | 0    | 0      | 0      | 0    | 0      | 0    |
| AO01   | 0    | 0    | 0    | 0    | 1    | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 0      | 0      | 0    | 0      | 0    |
| AI07   | 0    | 0    | 0    | 1    | 0    | 0    | 1    | 1    | 0    | 0    | 0    | 0    | 0      | 0      | 0    | 0      | 0    |
| AI08   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 1    | 0    | 0    | 0    | 0      | 0      | 0    | 0      | 1    |
| AO02   | 0    | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 1    | 0    | 0    | 0      | 0      | 0    | 0      | 0    |
| AI06   | 0    | 0    | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 1    | 0    | 0      | 0      | 0    | 0      | 1    |
| AO04   | 0    | 0    | 0    | 0    | 1    | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 0      | 0      | 0    | 0      | 0    |
| APQR01 | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1      | 1      | 0    | 0      | 0    |
| APQR02 | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1      | 1      | 0    | 1      | 0    |
| AO05   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 0      | 1      | 0    | 0      | 0    |
| APQR03 | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1      | 0      | 1    | 0      | 0    |
| AI09   | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 0    | 1    | 0    | 0      | 0      | 0    | 0      | 1    |

|        | Cat10 | Cat20 | Cat30 | Cat40.50 | Cat60 | Cat70 | Cat80 | Cat100.101 | Cat102 | Cat110.111.113 | Cat120.121 | Cat122 |
|--------|-------|-------|-------|----------|-------|-------|-------|------------|--------|----------------|------------|--------|
| AI01   | 1     | 4     | 18    | 17       | 0     | 3     | 4     | 23         | 0      | 0              | 0          | 0      |
| AI02   | 0     | 8     | 24    | 121      | 0     | 1     | 0     | 4          | 1      | 0              | 0          | 0      |
| AI03   | 0     | 0     | 986   | 874      | 1     | 69    | 10    | 26         | 2      | 22             | 0          | 0      |
| AI04   | 0     | 0     | 254   | 248      | 0     | 54    | 0     | 51         | 2      | 1              | 0          | 0      |
| AO03   | 0     | 0     | 55    | 88       | 0     | 10    | 12    | 31         | 4      | 17             | 0          | 0      |
| AI05   | 0     | 0     | 181   | 413      | 4     | 72    | 38    | 111        | 4      | 21             | 0          | 0      |
| AO01   | 0     | 0     | 43    | 91       | 4     | 7     | 2     | 36         | 13     | 12             | 0          | 0      |
| AI07   | 0     | 0     | 140   | 212      | 3     | 33    | 11    | 47         | 14     | 20             | 0          | 0      |
| AI08   | 0     | 0     | 154   | 272      | 0     | 74    | 38    | 123        | 6      | 30             | 0          | 0      |
| AO02   | 0     | 6     | 177   | 507      | 3     | 36    | 26    | 231        | 6      | 82             | 0          | 0      |
| AI06   | 0     | 0     | 66    | 187      | 0     | 16    | 25    | 106        | 0      | 15             | 66         | 14     |
| AO04   | 0     | 0     | 1     | 40       | 0     | 4     | 1     | 21         | 0      | 22             | 0          | 0      |
| APQR01 | 0     | 0     | 24    | 332      | 0     | 40    | 18    | 128        | 12     | 94             | 58         | 34     |
| APQR02 | 0     | 0     | 15    | 174      | 0     | 5     | 4     | 77         | 5      | 78             | 9          | 5      |
| AO05   | 0     | 0     | 0     | 17       | 0     | 0     | 0     | 10         | 1      | 18             | 0          | 0      |
| APQR03 | 0     | 0     | 31    | 288      | 0     | 17    | 25    | 151        | 17     | 108            | 116        | 134    |
| AI09   | 0     | 0     | 37    | 224      | 0     | 13    | 7     | 114        | 64     | 8              | 184        | 281    |

FIGURE 1 – Exemple de données disponibles : cas d'un unique site



# Problème archéologique

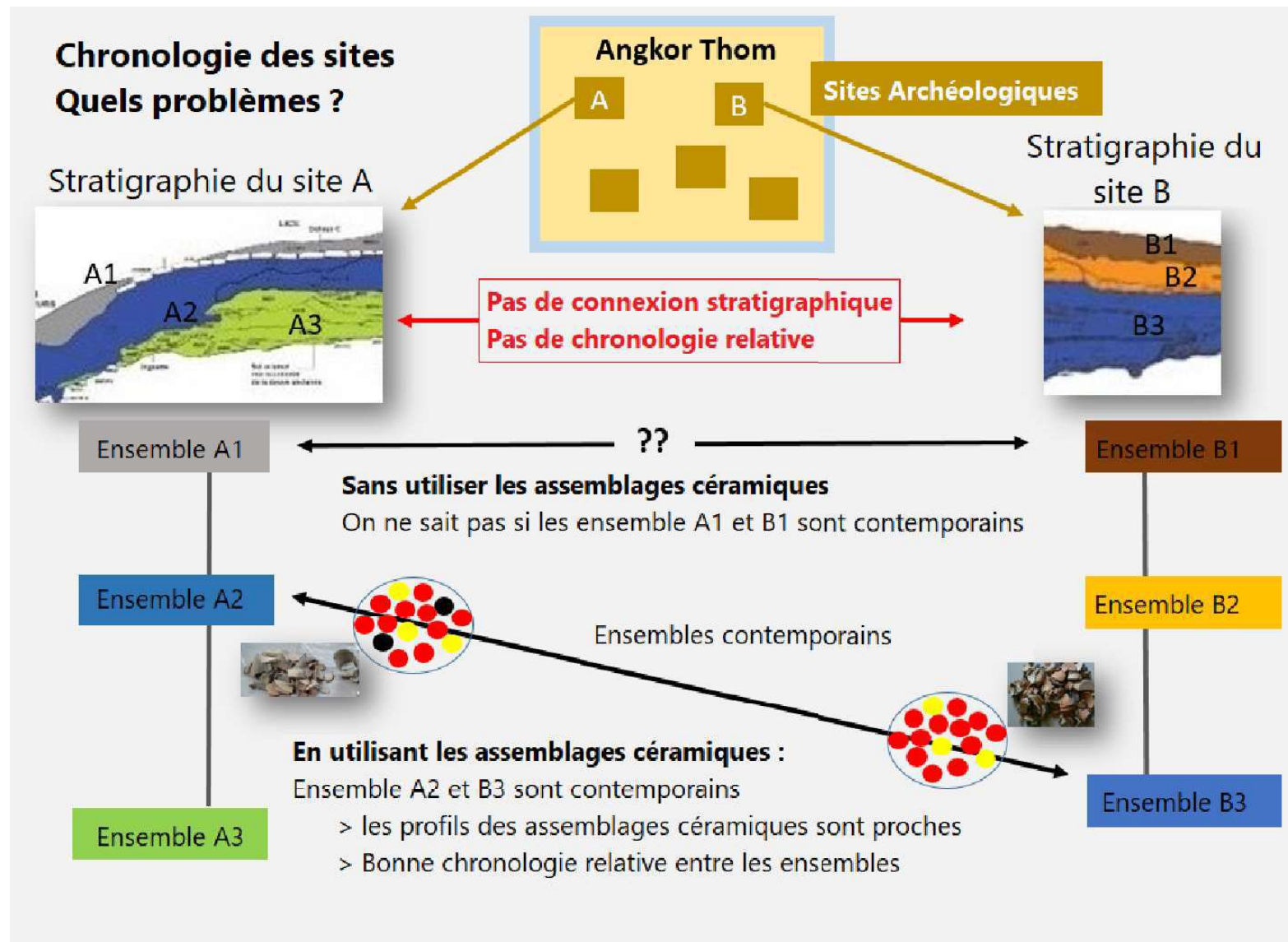
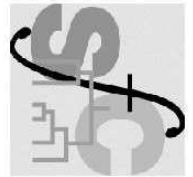


FIGURE 2 – La chronologie générale des sites



# Description des données d'Angkor Thom

Etude de la céramique et de la stratigraphie pour

- 3 sites d'Angkor Thom, en relation avec le système d'enceinte d'Angkor Thom ([7]);
- 17 ensembles stratigraphiques :
  - présentant une grande amplitude chronologique,
  - caractérisés par un corpus de données céramiques qualitativement et quantitativement intéressant.



# De la céramique à $D_1$

$X$  : matrice  $17 \times 12$ , quantification en NMI de la céramique pour les 3 sites d'Angkor Thom.

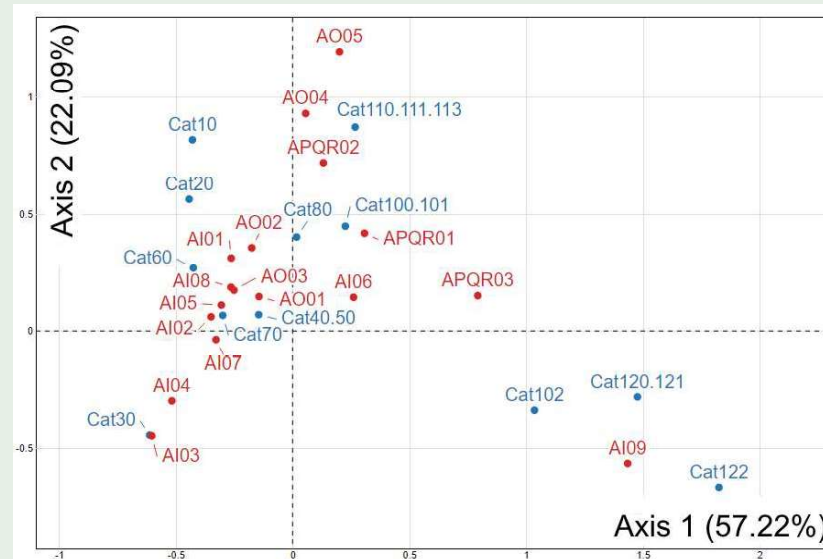


FIGURE 3 – Plan 1-2 AFC sur  $X$

- Effet Guttman : sériation chronologique des ensembles par la céramique uniquement.
- $D_1$  : matrice  $17 \times 17$  des distances euclidiennes calculées à partir de l'ensemble des composantes de l'AFC sur  $X$ . (quantification céramique).



# Du diagramme stratigraphique à $D_2$

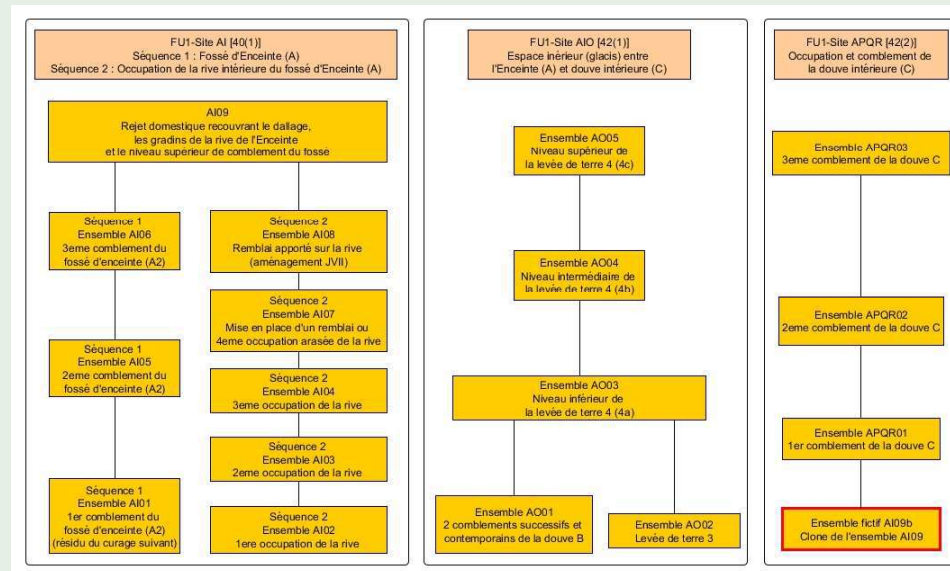


FIGURE 4 – Diagramme stratigraphique des 3 sites d'Angkor Thom étudiés

- Les ensembles reliés par une ligne sont en connexion stratigraphique. Les 3 sites géographiquement distincts ne sont pas connectés entre eux.
- $D_2$  : matrice  $17 \times 17$  des dissimilarités calculées à partir du diagramme stratigraphique

$$D_2 = \begin{cases} 0 & : \text{s'il existe une relation stratigraphique directe avant/après} \\ 1 & : \text{sinon} \end{cases}$$

# Comparaison des dendrogrammes : CAH sur $D_1$ et $D_2$

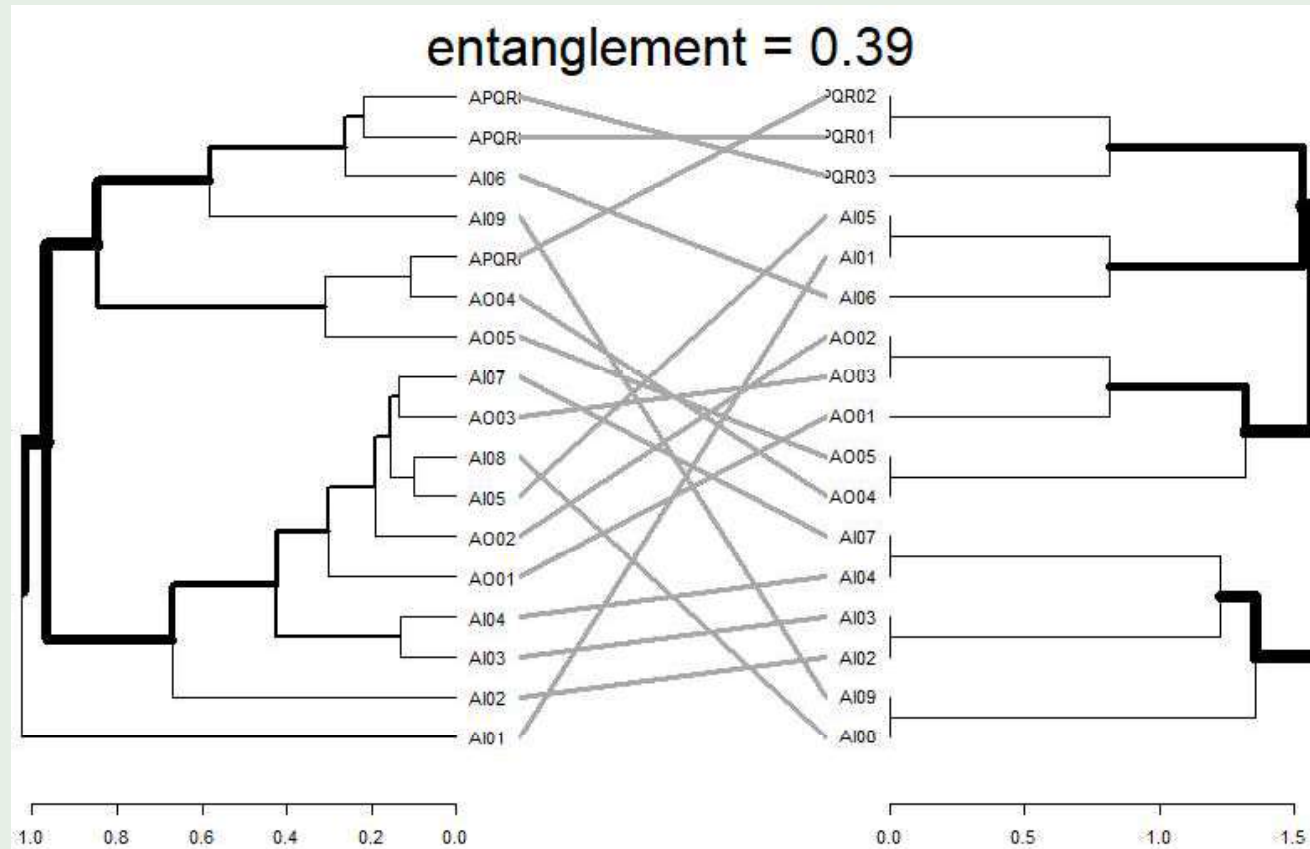


FIGURE 5 – Tanglegram : Comparaison des 2 dendrogrammes obtenus à partir de la CAH Ward sur  $D_1$  céramique (gauche) et  $D_2$  stratigraphie (droite)

Valeur d'entanglement 0.39 : dendrogrammes similaires mais pas identiques (0 si identiques) !

# hclust.mix : Sélection du paramètre $\alpha$

$\mathbf{D}_\alpha = \alpha \mathbf{D}_1 + (1 - \alpha) \mathbf{D}_2$ ;  $\alpha \in [0; 1]$  déterminé par

- une valeur obtenue en minimisant le critère (2)
- un intervalle de confiance.

```
perio.selectAlpha(D1_cera, D2_strati, acc=2,
resampling=TRUE, method="ward.D2", iter=16)
```

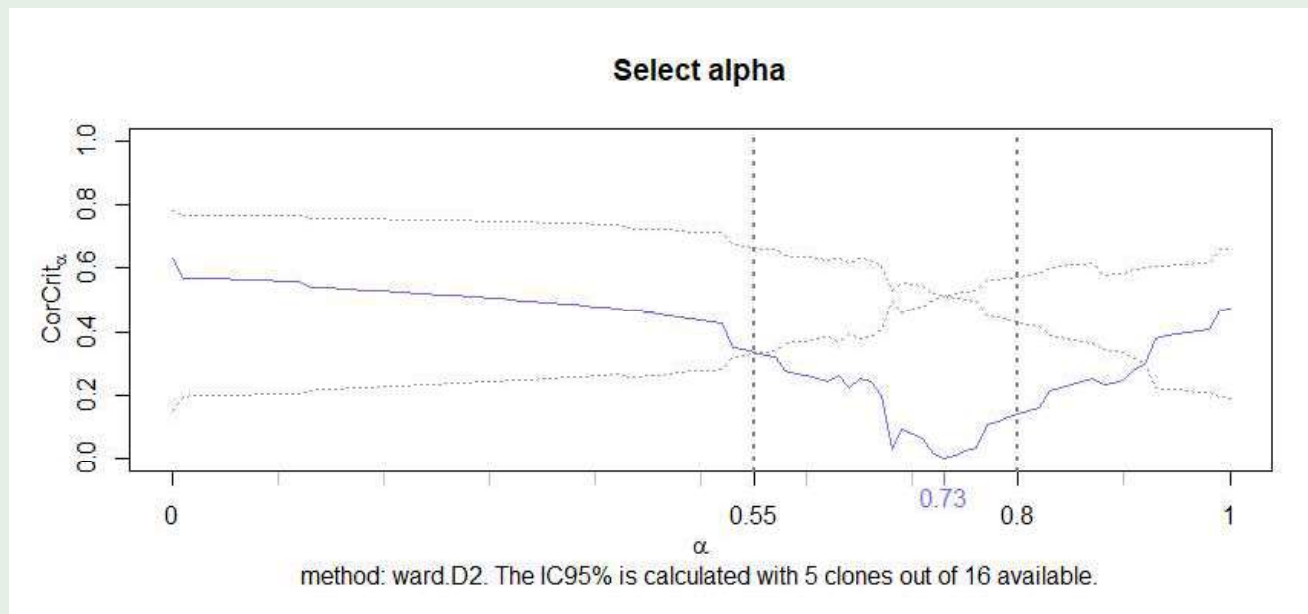


FIGURE 6 –  $CorCrit_\alpha$  en fonction de  $\alpha$ .  $\hat{\alpha} = 0.73 \in [0.55; 0.8]$ ;  $\tilde{\alpha} = 0.7$

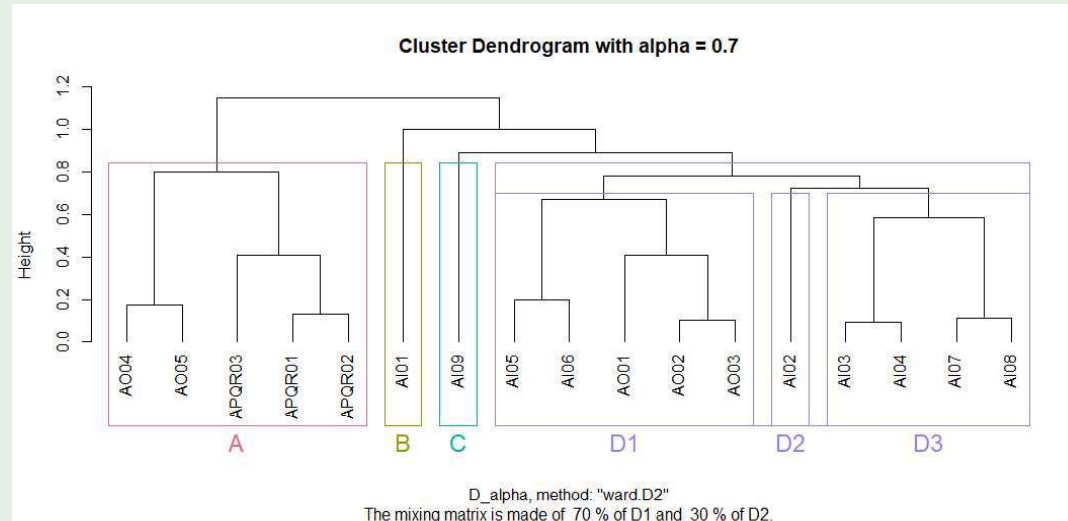




# HAC et nombre de groupes

```
clustering <- perioclust(D1_cera, D2_strati, alpha =
0.7, method = "ward.D2")
```

```
clustering <- perio.subdivide(clustering, cluster = 4,
nb_class = 3)
```



**FIGURE 7** – Dendrogramme : 4 groupes avec un groupe subdivisé en 3 :  
Justification : (i) Examen de l'échelle des indices d'agrégation associé au dendrogramme ; (ii) Connaissance archéologique des sites.



# Interprétation archéologique

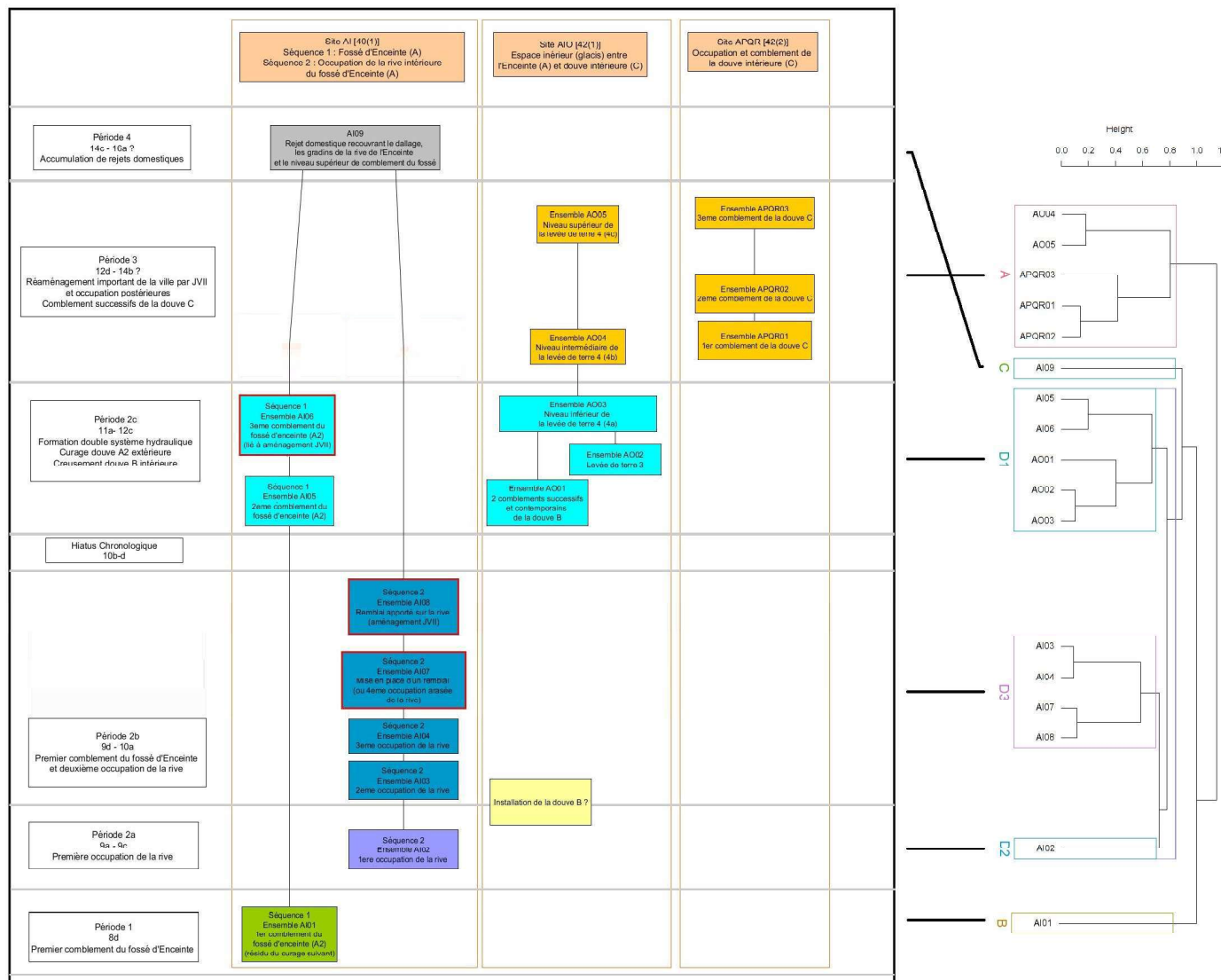


FIGURE 8 – Diagramme des ensembles périodisés



# Comparaison avec hclustgeo (Chavent et al., 2018)

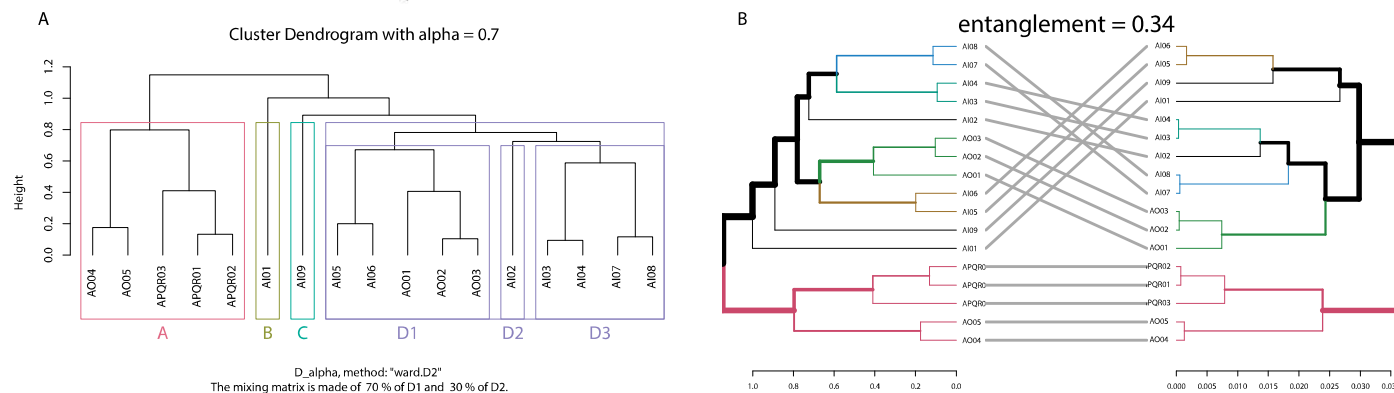


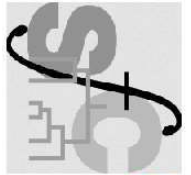
FIGURE 9 – A Dendrogramme obtenu avec hclust.mix. B Tanglegram hclust.mix (gauche) et hclustgeo (droite),  $\alpha = 0.7$

- En appliquant hclustgeo avec  $\alpha = 0.7$  : résultats différents de hclust.mix (cf Fig 9).
- Interprétation archéologique en faveur de hclust.mix : hclustgeo regroupe très vite les ensembles A109, A105 et A106. Or A109 est beaucoup plus récent que les autres ensembles et devrait donc rester isolé comme c'est le cas avec hclust.mix.

# Données Estuaire

Données disponibles dans le R-package `ClustGeo` :

- 303 communes situées sur la côte atlantique entre Royan et Mimizan (Sud-Ouest de la France) décrites par 4 variables quantitatives.
- Les 2 matrices de dissimilarité considérées sont :
  - (i)  $D_1$  : distances euclidiennes entre les municipalités calculées à partir des 4 variables socio-économiques
  - (ii)  $D_2$  : distances géographiques.
- Nous conservons 5 groupes comme dans [3] et comparons les résultats obtenus avec `hclust.mix` et `hclustgeo`.



- Pour `hclustgeo`,  $\alpha = 0.8$  a été retenu par [3]. En fixant  $\alpha = 0.8$  dans `hclust.mix`, nous obtenons une valeur d'entanglement de 0.84, ce qui indique que les dendrogrammes obtenus sont très différents.
- Pour `hclust.mix`, la procédure de sélection de  $\alpha$  renvoie  $\alpha = 0.45$  ( $IC95\%(\alpha) = [0.33; 0.50]$ ). Une valeur d'entanglement de 0.8, indique que les dendrogrammes obtenus avec les valeurs  $\alpha$  adaptées à chaque méthode restent différents !

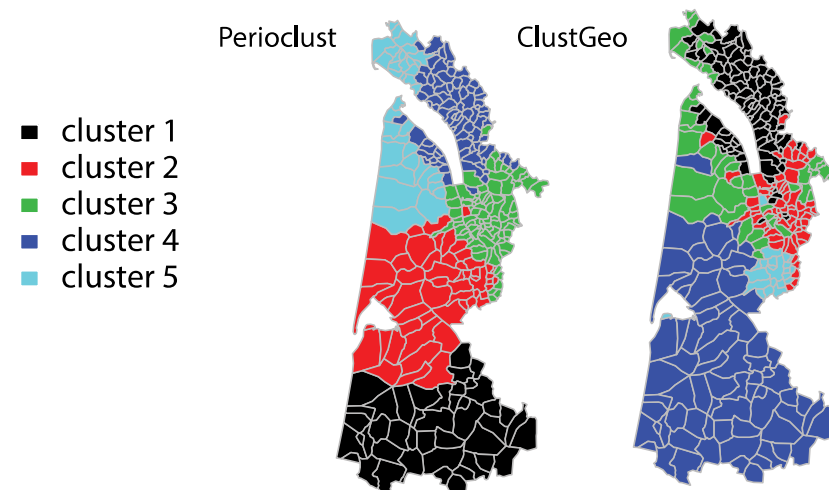


FIGURE 10 – Estuaire : partition en 5 groupes :  $\alpha = 0.45$  pour `hclust.mix` (à gauche),  $\alpha = 0.8$  `hclustgeo` (à droite).

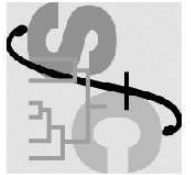
# Plan

- 1 Introduction
- 2 Bref aperçu des méthodes CAH avec contraintes existantes
- 3 La méthode `hclust.mix`
- 4 Applications
- 5 Conclusions
- 6 Références



# Conclusions

- Présentation d'une nouvelle méthode de CAH avec contraintes (`hclust.mix`), inspirée par un problème archéologique.
- Notre méthode fournit un outil statistique qui
  - intègre 2 sources de données ;
  - permet de trouver un compromis entre les sources d'information contenant potentiellement des erreurs.
- `hclust.mix` sera bientôt disponible dans le R package SPARTAAS R-package sur Framagit et le CRAN.



# Plan

- 1 Introduction
- 2 Bref aperçu des méthodes CAH avec contraintes existantes
- 3 La méthode `hclust.mix`
- 4 Applications
- 5 Conclusions
- 6 Références





# Références



Aggarwal, C. and Reddy, C., 2014. *Data Clustering : Algorithms and Applications*. Boca Raton : Chapman and Hall/CRC.



Bellanger L., Coulon A. Husi P. (accepté avril 2020) PerioClust : a simple hierarchical agglomerative clustering approach including constraints. Springer Series “Studies in Classification, Data Analysis, and Knowledge Organization.



Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J., 2018. ClustGeo : an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 33(4), pp. 1799-1822.



Efron, B. and Tibshirani, R., 1993. *An Introduction to the Bootstrap*. New York : Chapman & Hall/CRC.



Everitt, B., Landau, S. and Morven, L., 2001. *Cluster Analysis*. 4th ed. ed. Oxford : Oxford University Press Inc.



Ferligoj, A. and Batagelj, V., 1982. Clustering with relational constraint. *Psychometrika*, 47, 4, 413-426.



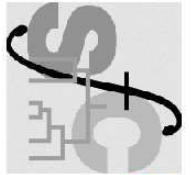
Gaucher, J. and Husi, P., 2014. L'archéologie urbaine appliquée à un site archéologique : l'exemple d'Angkor Thom (Cambodge), capitale du royaume khmer angkorien. In : R. X. Lorans E., ed. *Archéologie de l'espace urbain. Tours : Col. Perspectives Villes et Territoires*, coed PUF/CTHS, 121-131.



Henning, C., Meila, M., Murtagh, F., Rocci, R., 2016. *Handbook of cluster analysis*. Boca Raton : Chapman and Hall/CRC.



Kaufman, L. and Rousseeuw, P., 2005. *Finding Groups in Data : An Introduction to Cluster Analysis*. New York : Wiley-Interscience.



# Références suite



Lance, G. and William, W., 1967. A general theory of classificatory sorting strategies : 1. Hierarchical systems. *Computer Journal*, Volume 9, 373-380.



Legendre, P., Dallot, S. and Legendre, L., 1985. Succession of Species within a Community : Chronological Clustering, with Applications to Marine and Freshwater Zooplankton. *American Naturalist*, 125(2), 257-288.

