Semi-supervised clustering of quaternion time series

Application to gait disorder analysis in Multiple scleross

Pierre Drouin

PhD student (CIFRE) in applied statistics
UmanIT/Mathematics department Jean Leray

June 29, 2022

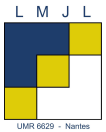


Table of Contents

Actors of the project

Actor	Speciality
UmanIT L. Chevreuil, V. Graillot, Fanny Doistau	Software development Internet of Things (IoT)
PhD in applied statistics P. Drouin (2018 - 2022)	
Mathematics dept. Jean Leray Nantes university L. Bellanger, A. Stamm	Data analysis

In collaboration with reasearch teams of Pr D. Laplaud, P.A. Gourraud and P. Damier, University Hospital Center of Nantes

Context

Two figures on neuro-degenerative diseases in France[1]

- Multiple Sclerosis (MS): $> 100,000$ patients
- Parkinson's disease: $> 200,000$ patients

Walking disabilities analysis:

- One of the most frequently observed symptoms [1].
- Impact on the daily life of patients [2].
- Essential element in neurology diagnosis [3].
- Development perspectives for other pathologies (*e.g.* the elders).

Our Approach: Determination of a biomarker called Individual Gait Pattern (IGP) from data measured by a wearable Inertial Measurment Unit (IMU)

[1] `https://solidarites-sante.gouv.fr/soins-et-maladies/maladies/maladies-neurodegeneratives`

Table of Contents

Description of the device

Gait Measurement Device:

- An Inertial Measurement Unit (*MetaMotionR*[2]) placed on the belt at the right hip.
- Recording of the wearer's hip rotations in the form of *unit quaternions time series*.

[2] https://mbientlab.com/metamotionr/

MetaMotionR orientation

The orientation of the *MetaMotionR* is the rotation between:

- A fixed frame (*i.e.* the Earth frame) $\mathfrak{R}_f = (\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3)$
- Its own frame $\mathfrak{R}_s = (\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3)$

Natural representation of this rotation:

- its **angle of rotation** $\theta \in [0, 2\pi]$
- its **axis of rotation** $\mathbf{u} = (u_1, u_2, u_3) \in \mathcal{S}^2$, where $\mathcal{S}^2$ is the 2-sphere.

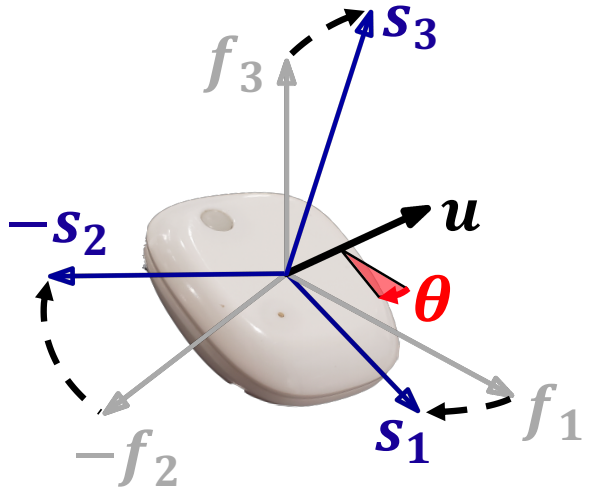


Figure 1: Orientation of the MetaMotionR

The unit quaternion representing this rotation is:

$$\mathbf{q} = w + xi + yj + zk = \cos\frac{\theta}{2} + (u_1 i + u_2 j + u_3 k)\sin\frac{\theta}{2}, \tag{1}$$

with $i^2 = j^2 = k^2 = ijk = -1$ and $||\mathbf{q}|| = \sqrt{w^2 + x^2 + y^2 + z^2} = 1$.
The *unit quaternion time series* represents the hip rotation during time.

Individual Gait Pattern

The biomarker *Individual Gait Pattern* (IGP) represents the orientation of the hip at each percent of duration of a typical walking cycle.
It is a *unit quaternion time series* $Q = (\mathbf{q}_0, \ldots, \mathbf{q}_{100})$.

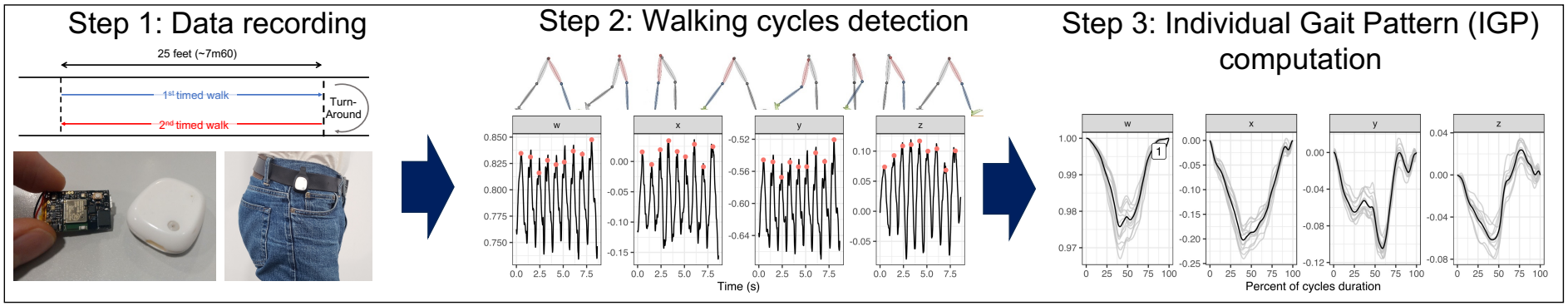


Figure 2: IGP determination method

Hypothesis: The shape of the IGP is dependent on the individual's gait, and thus in part on the presence of gait disorders.

Table of Contents

Presentation of the study

A study was prepared in collaboration with the research teams of Pr D-A Laplaud and Pr P-A Gourraud (University Hospital Center of Nantes):

- Sample: 27 MS patients
- Data measured:
  - Expended Disability Status Scale (EDSS) score [4]
  - IGP durng the *Timed 25 Foot Walk* [3]

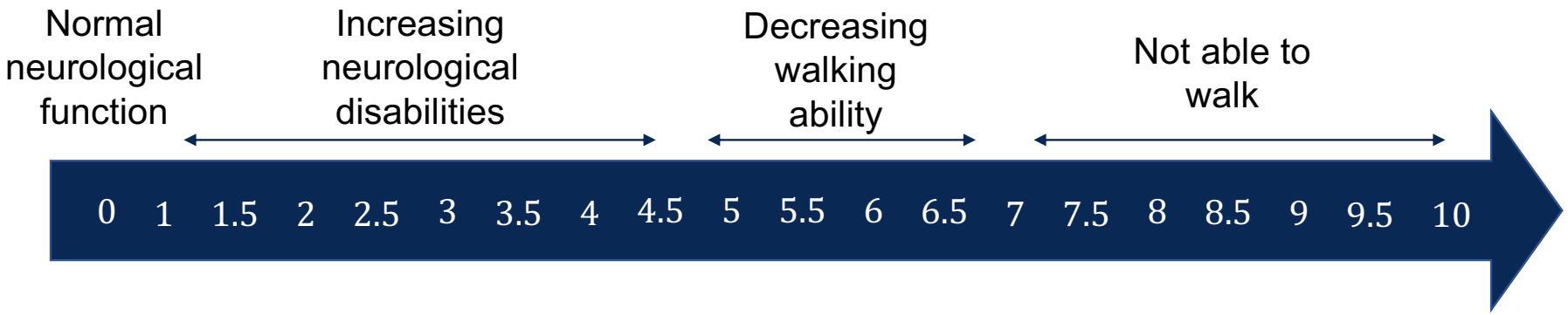


Figure 3: EDSS overview

Presentation of the study

A study was prepared in collaboration with the research teams of Pr D-A Laplaud and Pr P-A Gourraud (University Hospital Center of Nantes):

- Sample: 27 MS patients
- Data measured:
  - Expended Disability Status Scale (EDSS) score [4]
  - IGP durng the *Timed 25 Foot Walk* [3]

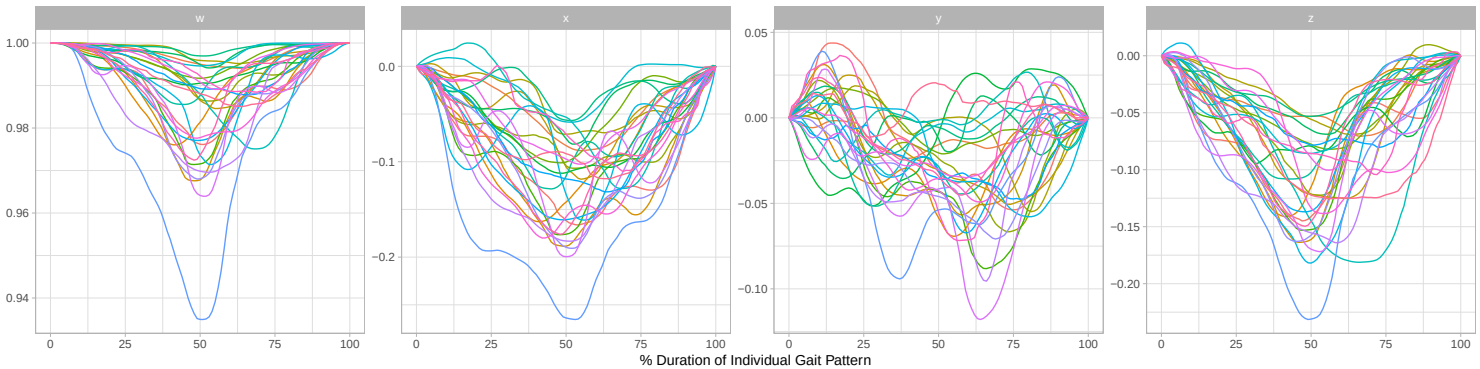


Figure 3: Patients IGPs

Problematic: Build groups of patients with similar gait disorders with a clustering method allowing to consider both sources of information.

Semi supervised clustering

Clustering methods accounting for multiple data sources: *semi-supervised clustering*.
Three categories:

- Constraint based clustering [5]
- Ensemble clustering [6]
- Compromise based clustering [7]

In this study, we compare unsupervised hierarchical ascendant clustering (HAC) and two semi-supervised hierarchical based clustering methods:

- En ensemble clustering method: *mergetrees* [8].
- A compromise based clustering method: *hclustcompro* [7].

Mergetrees [8]

Problem setting:

- $X_1, \ldots, X_m$: m data sets observed on the same n individuals.
- $\boldsymbol{\mathcal{T}} = \{\mathcal{T}_1, \ldots, \mathcal{T}_m\}$: m dendrograms obtained from these data sets with any hierarchical clustering method.

Mergetrees principle:
Build a consensus tree $C(\boldsymbol{\mathcal{T}})$ according to the following rule: *For any individuals i and j in $\{1, \ldots, n\}$, $i \neq j$, if i and j are not in the same cluster in at least one of the trees of $\boldsymbol{\mathcal{T}}$ at a given height h, then they are not in the same cluster in $C(\boldsymbol{\mathcal{T}})$ at height h.*

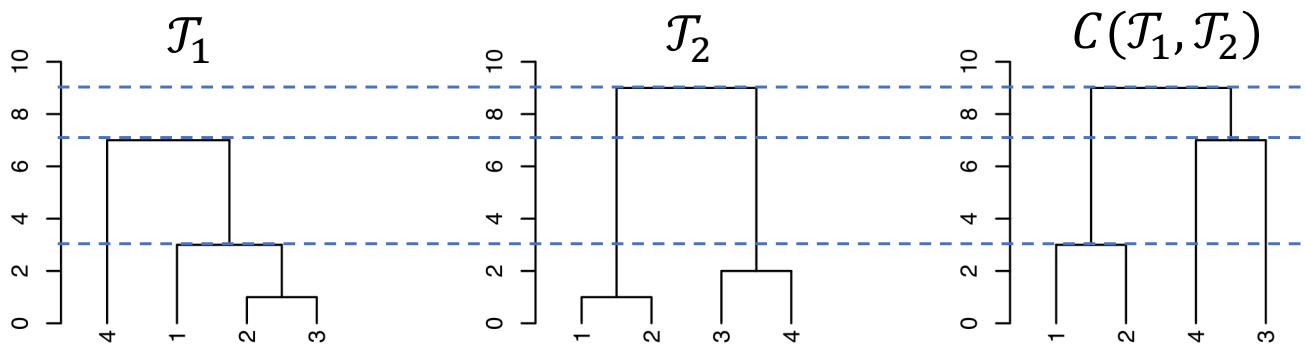
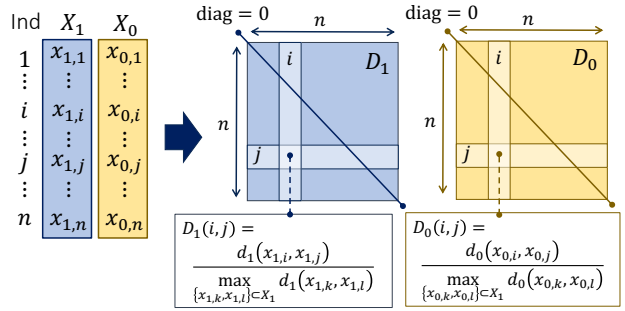


Figure 4: Application of mergetrees on two dendrograms

Hclustcompro [7]

Problem setting:

- $X_1 = \{x_{1,1}, \ldots, x_{1,n}\}$.
- $X_0 = \{x_{0,1}, \ldots, x_{0,n}\}$.
- d_1 and d_0: dissimilarity measures.
- D_1 and D_0: normalized dissimilarity matrices.



(a) Normalized dissimilarity matrices

Hclustcompro [7]

Problem setting:

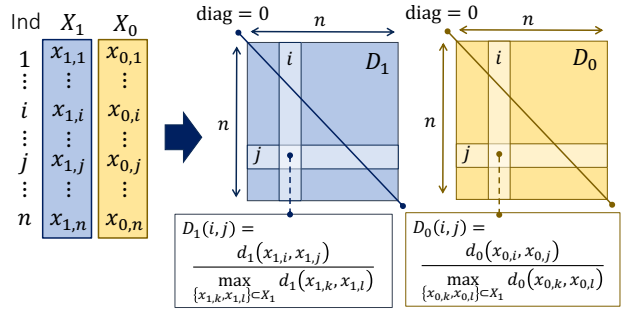
- $X_1 = \{x_{1,1}, \ldots, x_{1,n}\}$.
- $X_0 = \{x_{0,1}, \ldots, x_{0,n}\}$.
- d_1 and d_0: dissimilarity measures.
- D_1 and D_0: normalized dissimilarity matrices.

Principle of hclustcompro:

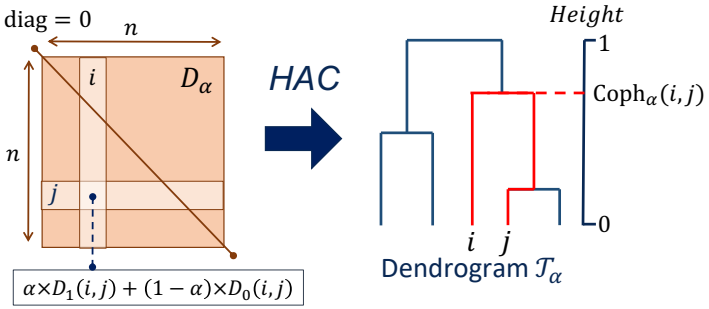
- D_α, convex combination of D_1 and D_0:

$$D_\alpha = \alpha D_1 + (1-\alpha) D_0, \qquad (2)$$

  with $\alpha \in [0, 1]$
- Apply HAC on D_α to obtain $\mathcal{T}_\alpha$.



(a) Normalized dissimilarity matrices



(b) Hierarchical clustering on D_α

Hclustcompro [7]

Problem setting:

- $X_1 = \{x_{1,1}, \ldots, x_{1,n}\}$.
- $X_0 = \{x_{0,1}, \ldots, x_{0,n}\}$.
- d_1 and d_0: dissimilarity measures.
- D_1 and D_0: normalized dissimilarity matrices.

Principle of hclustcompro:

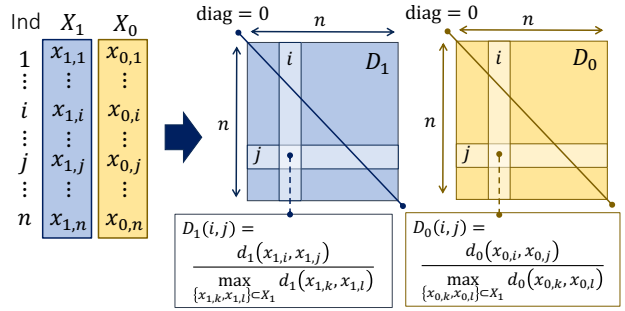
- D_α, convex combination of D_1 and D_0:

$$D_\alpha = \alpha D_1 + (1-\alpha) D_0, \qquad (2)$$

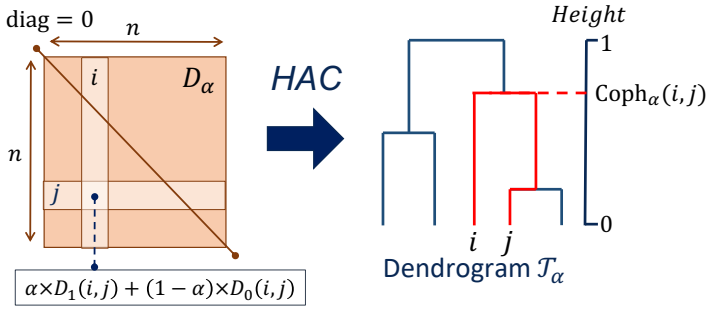
  with $\alpha \in [0, 1]$
- Apply HAC on D_α to obtain $\mathcal{T}_\alpha$.

Determination of the weighting parameter α:

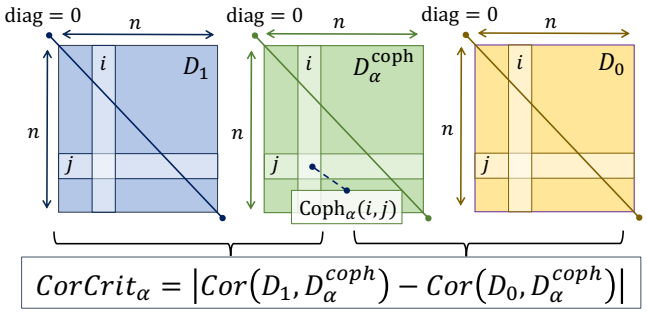
- D_α^{coph}: Cophenetic matrix of $\mathcal{T}_\alpha$ [9].
- $CorCrit_\alpha = |Cor(D_\alpha^{Coph}, D_1) - Cor(D_\alpha^{Coph}, D_0)|$.
- $\hat{\alpha} = \arg\min_\alpha CorCrit_\alpha$.



(a) Normalized dissimilarity matrices

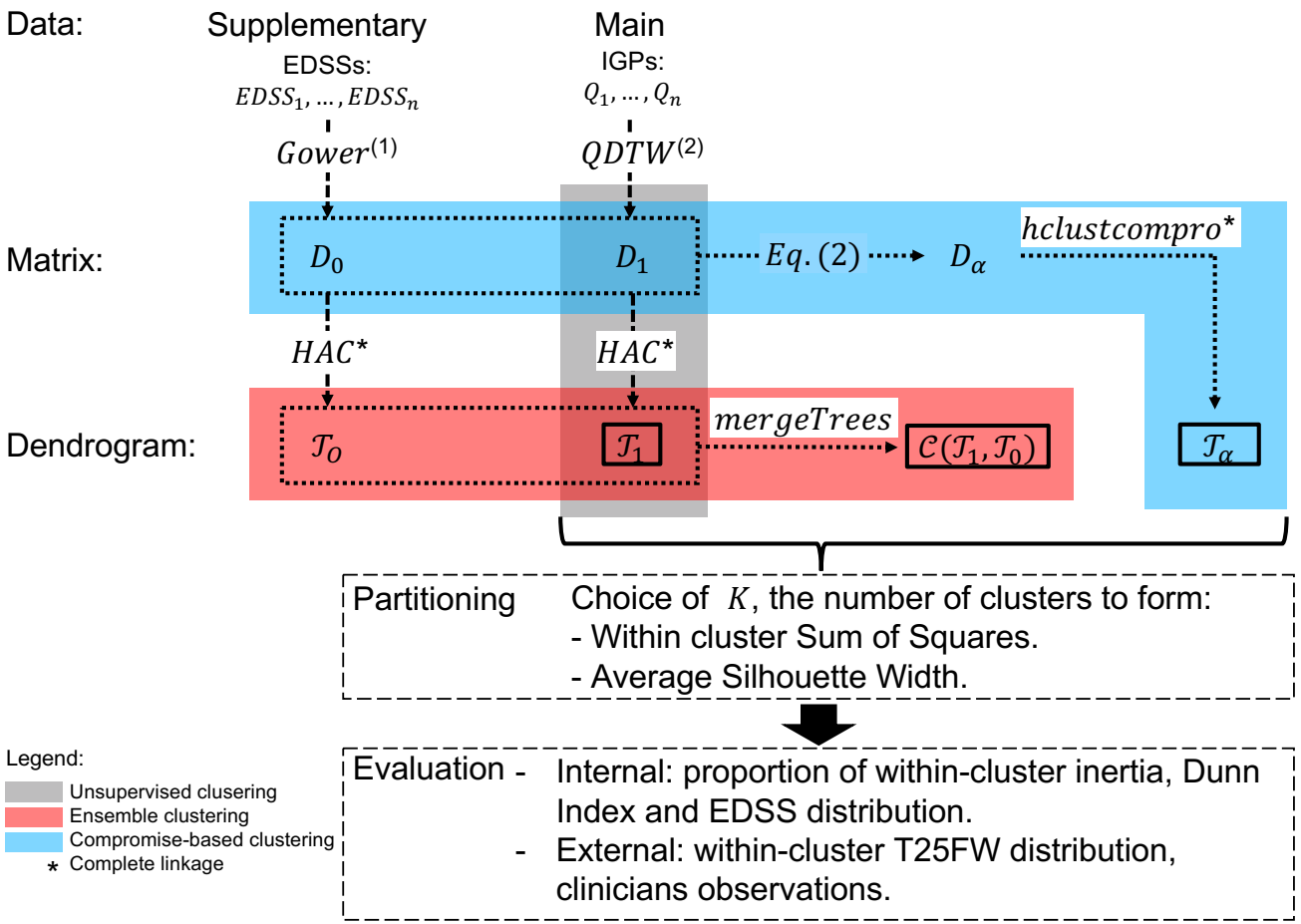


(b) Hierarchical clustering on D_α



(c) Weighting parameter optimization

Experimental design



(1) Gower dissimilarity [10].

(2) $QDTW$: Quaternion Dynamic Time Warping [11], a generalization of the method Dynamic Time Warping to unit quaternion time series.

Dendrograms

- Hclustcompro weighting parameter estimation: $\hat{\alpha} = 0.69$.
- Optimal number of clusters: $K = 5$

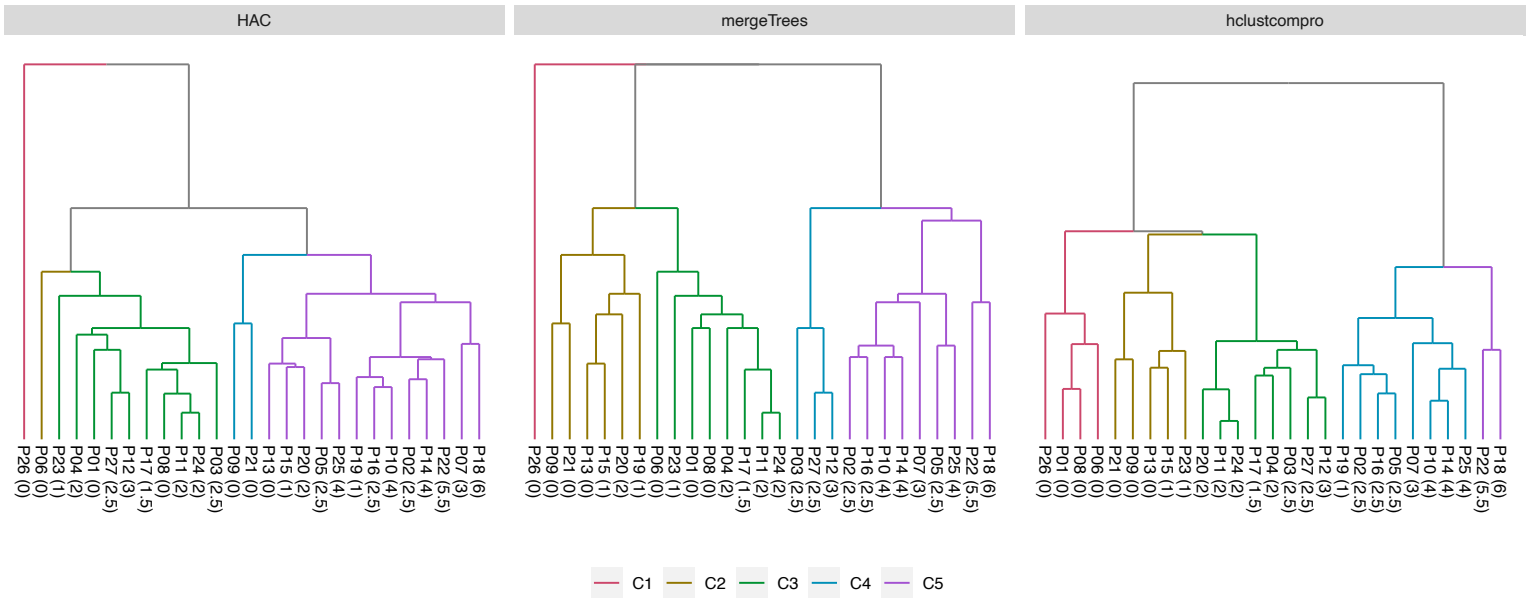


Figure 6: Dendrograms obtained with HAC, mergeTrees and hclustcompro

Interpretation of the results

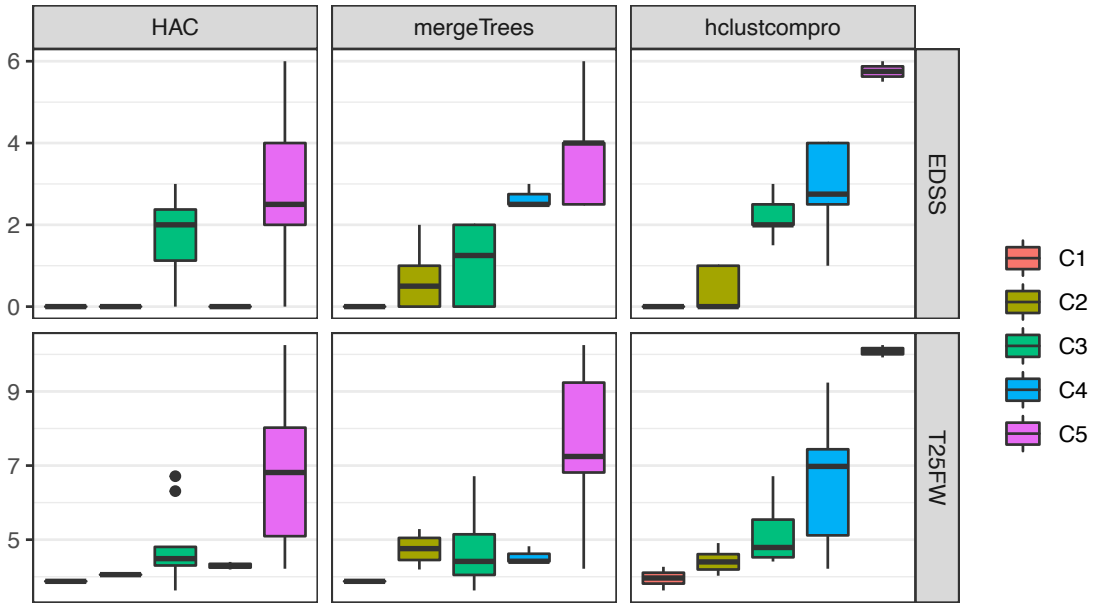


Figure 7: Clusters description

- HAC: Heterogeneous clusters (number of patients and disabilities).

Interpretation of the results

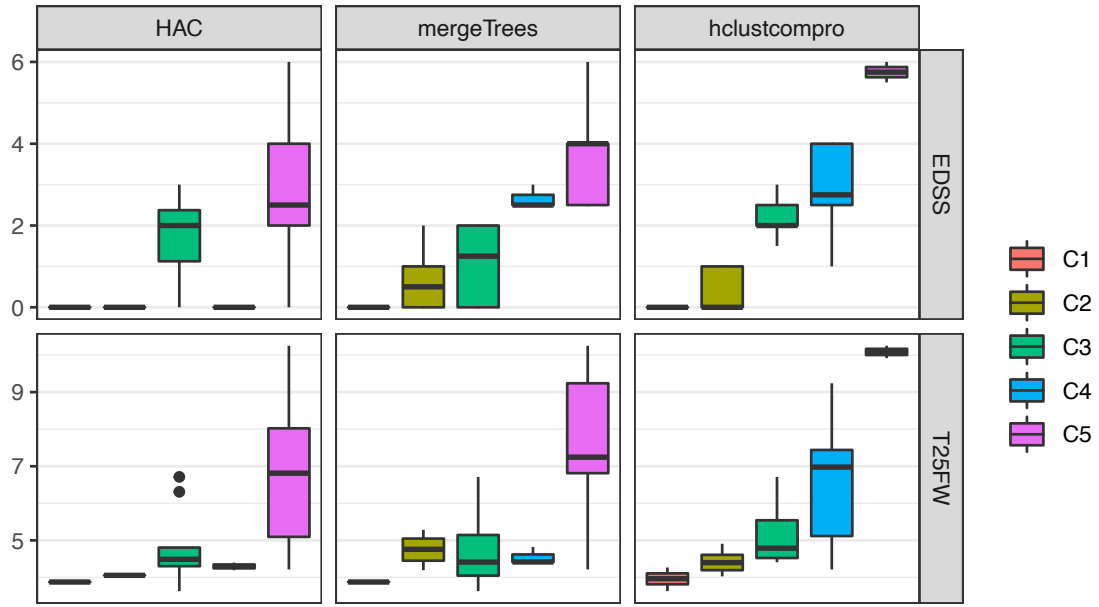


Figure 7: Clusters description

- Mergetrees:
  - Patients with stronger disabilities in C4 and C5.
  - Patients with similar disability split in different groups.
  - No ordered difference in terms of walking speed.

Interpretation of the results

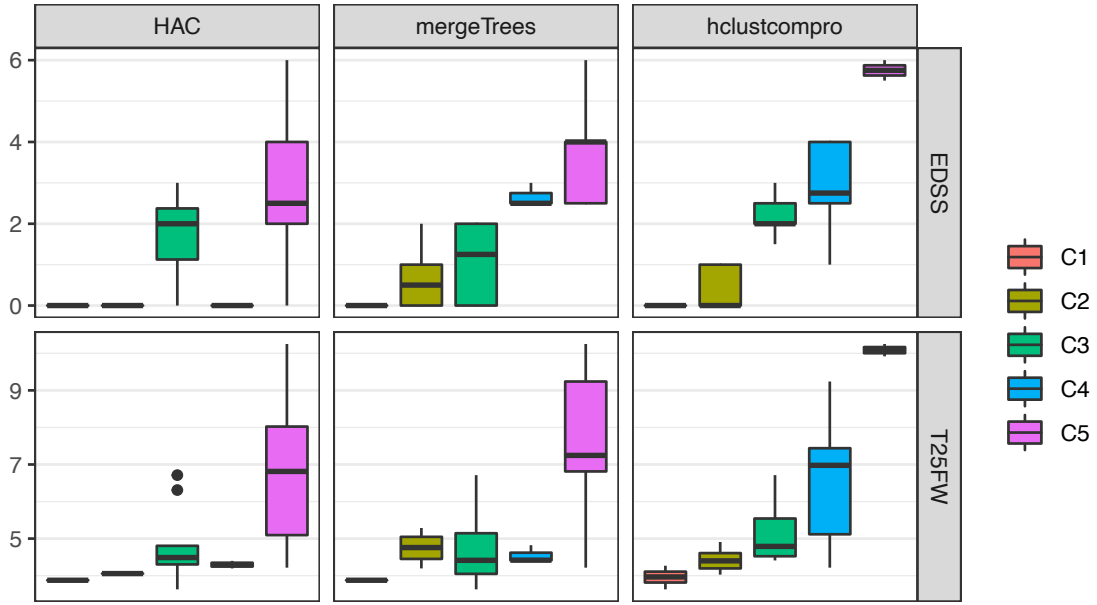


Figure 7: Clusters description

- Hclustcompro:
  - C1 and C2: patients with normal to moderate neurological disabilities.
  - C3 and C4: patients with moderate disabilities.
  - C5: patients with severe disabilities.

Table of Contents

Conclusion

Discussion of the results

- Additional information improve clustering interpretability.
- Compromise based method *hclustcompro* provide better results in this application.
- Results need to be confirmed on a larger cohort.

Perspectives

- Extend *hclustcompro* to the cases of more than two sources of information. (as in factorial method *STATIS* [12, 13]).
- Investigate of how well the IPG correlates with state-of-the-art spatio-temporal parameters (Gaitrite [14]).

References I

N. G. LaRocca, "Impact of Walking Impairment in Multiple Sclerosis," *The Patient: Patient-Centered Outcomes Research*, vol. 4, pp. 189–201, Sept. 2011.

C. Heesen, J. Böhm, C. Reich, J. Kasper, M. Goebel, and S. Gold, "Patient perception of bodily functions in multiple sclerosis: gait and visual function are the most valuable," *Multiple Sclerosis Journal*, vol. 14, no. 7, pp. 988–991, 2008.

R. W. Motl, J. A. Cohen, R. Benedict, G. Phillips, N. LaRocca, L. D. Hudson, R. Rudick, and M. S. O. A. Consortium, "Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis," *Multiple Sclerosis Journal*, vol. 23, no. 5, pp. 704–710, 2017.

J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (edss)," *Neurology*, vol. 33, no. 11, pp. 1444–1444, 1983.

D. Dinler and M. K. Tural, "A survey of constrained clustering," in *Unsupervised learning algorithms*, pp. 207–235, Springer, 2016.

A. Cornuéjols, C. Wemmert, P. Gançarski, and Y. Bennani, "Collaborative clustering: Why, when, what and how," *Information Fusion*, vol. 39, pp. 81–95, 2018.

References II

L. Bellanger, A. Coulon, and P. Husi, "Perioclust: A simple hierarchical agglomerative clustering approach including constraints," in *Data Analysis and Rationality in a Complex World* (T. Chadjipadelis, B. Lausen, A. Markos, T. R. Lee, A. Montanari, and R. Nugent, eds.), (Cham), pp. 1–8, Springer International Publishing, 2021.

A. Hulot, J. Chiquet, F. Jaffrézic, and G. Rigaill, "Fast tree aggregation for consensus hierarchical clustering," *BMC Bioinformatics*, vol. 21, p. 120, Mar. 2020.

R. R. Sokal and F. J. Rohlf, "The comparison of dendrograms by objective methods," *Taxon*, pp. 33–40, 1962.

J. C. Gower, "A general coefficient of similarity and some of its properties," *Biometrics*, pp. 857–871, 1971.

B. Jablonski, "Quaternion dynamic time warping," *IEEE transactions on signal processing*, vol. 60, no. 3, pp. 1174–1183, 2011.

Y. Escoufier, "Operators related to a data matrix," *Recent Developments in Statistics*, pp. 125–131, 1977.

C. Lavit, Y. Escoufier, R. Sabatier, and P. Traissac, "The act (statis method)," *Computational Statistics & Data Analysis*, vol. 18, no. 1, pp. 97–119, 1994.

References III


J. J. Sosnoff, M. Weikert, D. Dlugonski, D. C. Smith, and R. W. Motl, "Quantifying gait impairment in multiple sclerosis using gaitrite™ technology," *Gait & posture*, vol. 34, no. 1, pp. 145–147, 2011.


D. Q. Huynh, "Metrics for 3D Rotations: Comparison and Analysis," *Journal of Mathematical Imaging and Vision*, vol. 35, pp. 155–164, Oct. 2009.


M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On Clustering Validation Techniques," *Journal of Intelligent Information Systems*, vol. 17, pp. 107–145, Dec. 2001.

Unit quaternion algebra

Quaternions are hyper-complex numbers with 4 dimensions

$$\mathbf{q} = w + ix + jy + kz, \tag{3}$$

where $i^2 = j^2 = k^2 = ijk = -1$
$\rightarrow$ the product of two quaternions (*i.e.* Hamilton product) is noncommutative

$$\begin{aligned}\mathbf{q}_1\mathbf{q}_2 =& w_1 w_2 - x_1 x_2 - y_1 y_2 - z_1 z_2 \\ &+ (w_1 x_2 + x_1 w_2 + y_1 z_2 - z_1 y_2)i \\ &+ (w_1 y_2 - x_1 z_2 + y_1 w_2 + z_1 x_2)j \\ &+ (w_1 z_2 + x_1 y_2 - y_1 x_2 + z_1 w_2)k\end{aligned} \tag{4}$$

Unit quaternion algebra (2)

Some definitions:

- $Re(\mathbf{q}) = w$: the *scalar part* of $\mathbf{q}$.
- $Im(\mathbf{q}) = ix + jy + kz$: the *vector part* of $\mathbf{q}$.
- $\mathbf{q}^t = Re(\mathbf{q}) - Im(\mathbf{q})$: the *conjugate* of $\mathbf{q}$.
- $||\mathbf{q}|| = \sqrt{\mathbf{q}\mathbf{q}^t} = \sqrt{w^2 + x^2 + y^2 + z^2}$: the *norm* of $\mathbf{q}$.
- $\mathbf{q}^{-1} = \frac{\mathbf{q}^t}{||\mathbf{q}||}$: the *reciprocal* of $\mathbf{q}$

Unit quaternion algebra (3)

Quaternions of *norm one* form the set of *unit quaternions* $\mathbb{H}_u = \{\mathbf{q}; ||\mathbf{q}|| = 1\}$.
Unit quaternions represent rotation in 3 dimensions space.
Natural representation of a $3D$ rotation:

- its **angle of rotation** $\theta \in [0, 2\pi]$
- its **axis of rotation** $\mathbf{u} = (u_1, u_2, u_3) \in \mathcal{S}^2$, where $\mathcal{S}^2$ is the 2-sphere.

The unit quaternion representing this rotation is:

$$\mathbf{q} = w + xi + yj + zk = \cos\frac{\theta}{2} + (u_1 i + u_2 j + u_3 k)\sin\frac{\theta}{2}, \tag{5}$$

A possible distance between two unit quaternions $\mathbf{q}_1$ and $\mathbf{q}_2$ is the **geodesic distance**[15]:

$$d(\mathbf{q}_1, \mathbf{q}_2) = 2\arccos\left(\mathbf{q}_1^{-1}\mathbf{q}_2\right) \tag{6}$$

Quaternion Dynamic Time Warping [11]

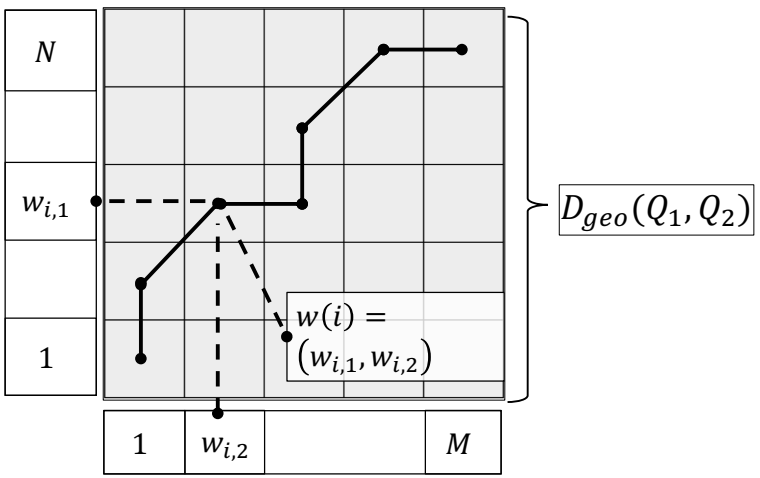


Figure 8: QDTW: Warping path

QDTW purpose: find the *warping path* W such as:

$$\mathrm{QDTW}(Q_1, Q_2) = \min_W \sum_{i=1}^{K} d(\mathbf{q}_{1,w_{i,1}}, \mathbf{q}_{2,w_{i,2}}), \tag{7}$$

with $d(.,.)$ the geodesic distance between two unit quaternion and W following these rules:

(1) $w_{1,1} = w_{1,2} = 1$, $w_{K,1} = N_1$, $w_{K,2} = N_2$.

(2) Each element of Q_1 is aligned with at least one element of Q_2.

(3) $w_{k,1} \leq w_{k+1,1}$ and $w_{k,2} \leq w_{k+1,2}$, $\forall k \in \{1, \ldots, K-1\}$.

(4) $w_{k,1} - w_{k-1,1} \leq 1$ and $w_{k,2} - w_{k-1,2} \leq 1$.

Choice of the number of clusters: WSS

Let $Q_1, \ldots, Q_n$ be a sample of n QTS grouped into K clusters $C_1, \ldots, C_K$. In addition, let $V_k := \{\ell \in [\![1, n]\!] : Q_\ell \in C_k\}$.
The WSS is given by:

$$\mathrm{WSS} = \sum_{k=1}^{K} \sum_{i \in V_k} \mathrm{QDTW}\left(Q_i, \widetilde{Q}_k\right)^2, \tag{8}$$

where $\widetilde{Q}_k$ is the medoid of the cluster C_k computed as:

$$\widetilde{Q}_k = Q_{i_k}, \quad \text{where} \quad i_k = \underset{i \in V_k}{\arg\min} \sum_{j \in V_k, j \neq i} \mathrm{QDTW}\left(Q_i, Q_j\right)^2.$$

For an observation i assigned to a cluster k:

- *Cohesion coefficient*

$$A(i,k) := \frac{1}{|V_k| - 1} \sum_{j \in V_k, j \neq i} \mathrm{QDTW}(Q_i, Q_j).$$

- *Separation coefficient*

$$B(i,k) := \min_{\ell \neq k} \frac{1}{|V_\ell|} \sum_{j \in V_\ell} \mathrm{QDTW}(Q_i, Q_j).$$

- *Silhouette width*

$$s(i,k) := \frac{B(i,k) - A(i,k)}{\max\left(A(i,k), B(i,k)\right)}.$$

The *average silhouette width* (ASW) of a partition $(V_1, \ldots, V_K)$:

$$\mathrm{ASW}\left(V_1, \ldots, V_K\right) := \frac{1}{n} \sum_{k=1}^{K} \sum_{i \in V_k} s(i,k). \tag{9}$$

Cluster validation: Within-cluster inertia

Within-cluster inertia $I_W^{(k)}$ of each cluster C_k:

$$I_W^{(k)} := \frac{1}{|V_k|} \sum_{i \in V_k} \mathrm{QDTW}^2(Q_i, \widetilde{Q}_k).$$

Between-cluster inertia $I_B^{(k)}$:

$$I_B^{(k)} = \mathrm{QDTW}^2 \left(\widetilde{Q}_k, \widetilde{Q} \right),$$

where $\widetilde{Q}$ is the overall medoid of the data set:

$$\widetilde{Q} = Q_i, \quad \text{where} \quad i = \underset{i \in [\![1,n]\!]}{\arg\min} \sum_{j \in [\![1,n]\!], j \neq i} \mathrm{QDTW}^2 \left(Q_i, Q_j \right).$$

Proportion of within-cluster inertia:

$$p_W^{(k)} := \frac{I_W^{(k)}}{I_W^{(k)} + I_B^{(k)}}. \tag{10}$$

At the level of the entire partition:

$$I_W := \frac{\sum_{k=1}^K |V_k| I_W^{(k)}}{\sum_{k=1}^K |V_k|}, \quad I_B := \frac{\sum_{k=1}^K |V_k| I_B^{(k)}}{\sum_{k=1}^K |V_k|} \quad \text{and} \quad p_W = \frac{I_W}{I_W + I_B}. \tag{11}$$

Dunn index

Dunn Index:

$$\mathrm{DI} := \frac{\min_{k,\ell \in [\![1,K]\!]^2, \ell \neq k} \delta\left(C_k, C_\ell\right)}{\max_{k \in [\![1,K]\!]} \Delta(C_k)}, \tag{12}$$

with

$$\delta\left(C_k, C_\ell\right) = \min_{i \in V_k, j \in V_\ell} \mathrm{QDTW}\left(Q_i, Q_j\right) \text{ (separation) and}$$
$$\Delta\left(C_k\right) = \max_{i,j \in V_k} \mathrm{QDTW}\left(Q_i, Q_j\right) \text{ (cohesion).}$$

This index is expected to be large when clusters are compact and well-separated [16].

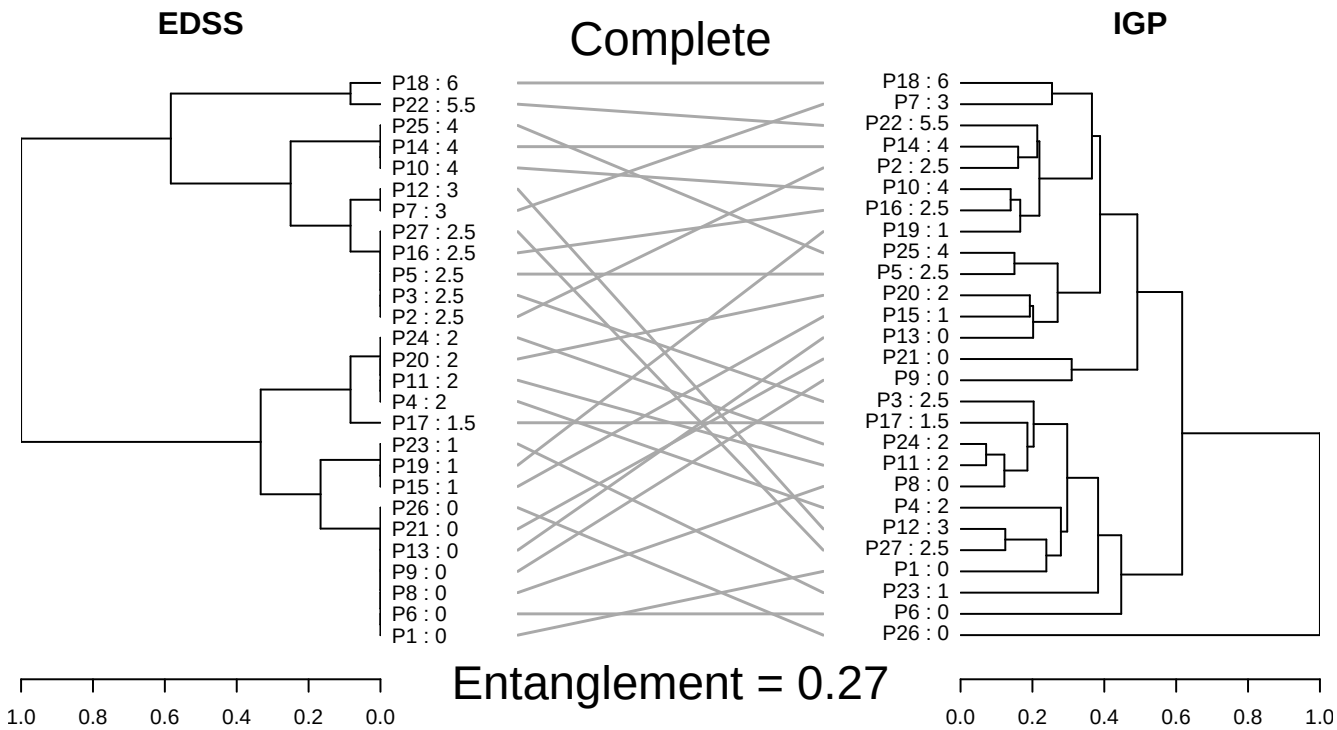


Figure 9: Tanglegrams

Number of clusters

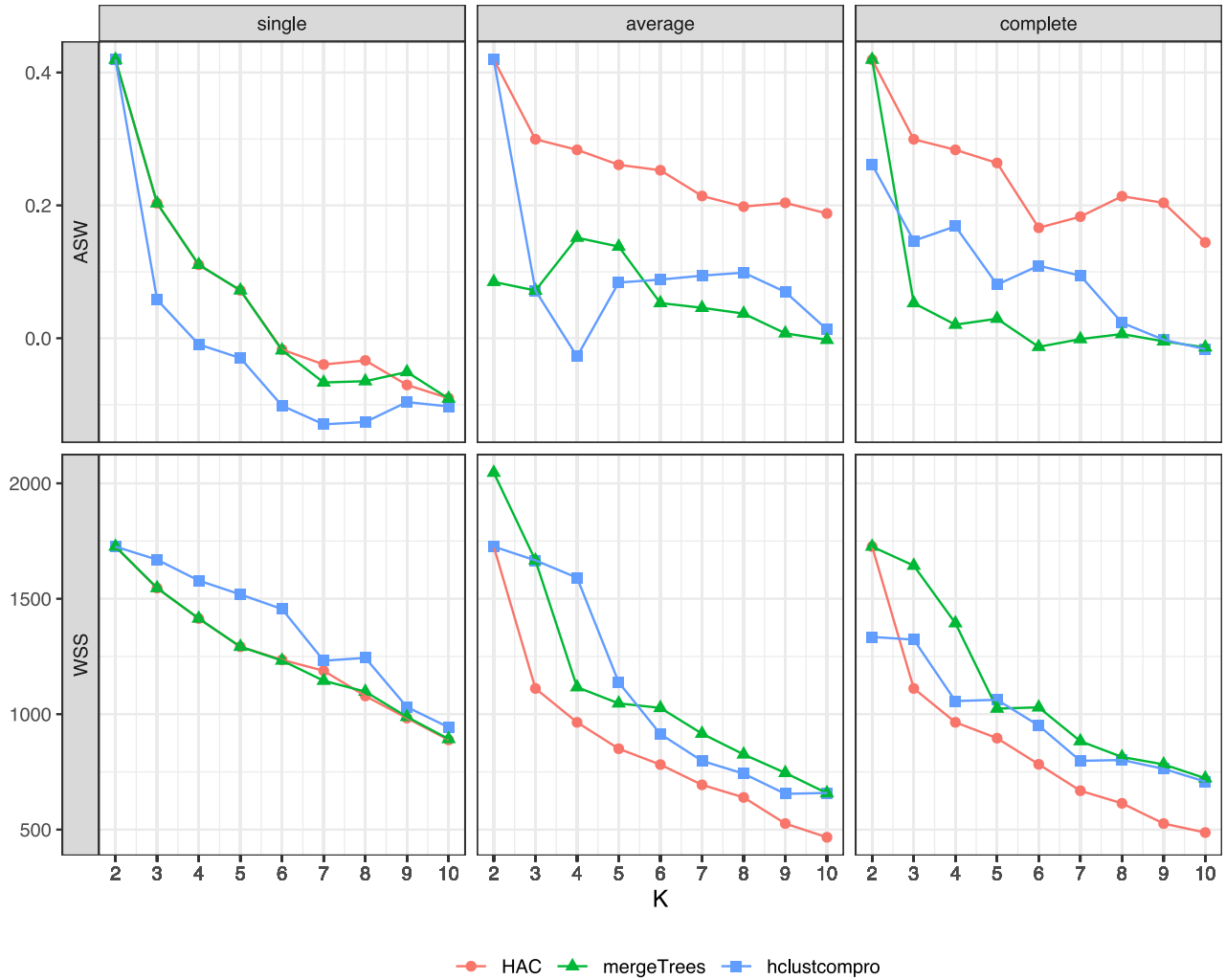


Figure 10: Within-cluster sum of squares (WSS) and average silhouette width (ASW).

Internal criteria

Table 1: **Summary of the clusters in terms of size (n), proportion of within-cluster IGP inertia $\left(p_W^{(k)}\right)$ and median EDSS (EDSS).** For a given cluster, its proportion of within-cluster inertia is computed on the IGP data according to eq. (10).

Classification	Critère de liaison	C1			C2			C3			C4			C5		
		n	$p_W^{(1)}$	EDSS	n	$p_W^{(2)}$	EDSS	n	$p_W^{(3)}$	EDSS	n	$p_W^{(4)}$	EDSS	n	$p_W^{(5)}$	EDSS
`CAH`	average	1	0.0	0	1	0.0	1	10	47.0	2	2	28.6	0	13	43.8	2.5
`mergeTrees`	average	1	0.0	0	11	52.6	2	5	65.5	0	8	37.0	3	2	43.5	5.5
`hclustcompro`	average	1	0.0	0	8	69.8	0	8	52.0	2	8	33.9	3	2	43.5	5.5
`CAH`	complete	1	0.0	0	1	0.0	0	10	50.5	2	2	28.6	0	13	43.8	2.5
`mergeTrees`	complete	1	0.0	0	6	63.6	1	8	55.2	1.5	3	22.3	2.5	9	38.0	4
`hclustcompro`	complete	4	64.7	0	5	63.6	0	8	52.0	2	8	33.9	3	2	43.5	5.5

Internal criteria (2)

Table 2: Performance metrics of IGP data at the partition level.

Clustering	Linkage	Using all clusters		Using only clusters that are never singletons	
		p_W	DI	p_W	DI
HAC	average	34.0	0.377	42.5	0.377
mergeTrees	average	42.7	0.288	49.2	0.288
hclustcompro	average	45.9	0.192	42.1	0.467
HAC	complete	35.0	0.428	43.8	0.428
mergeTrees	complete	40.2	0.288	41.4	0.329
hclustcompro	complete	51.6	0.214	42.1	0.467

Clusters medoids

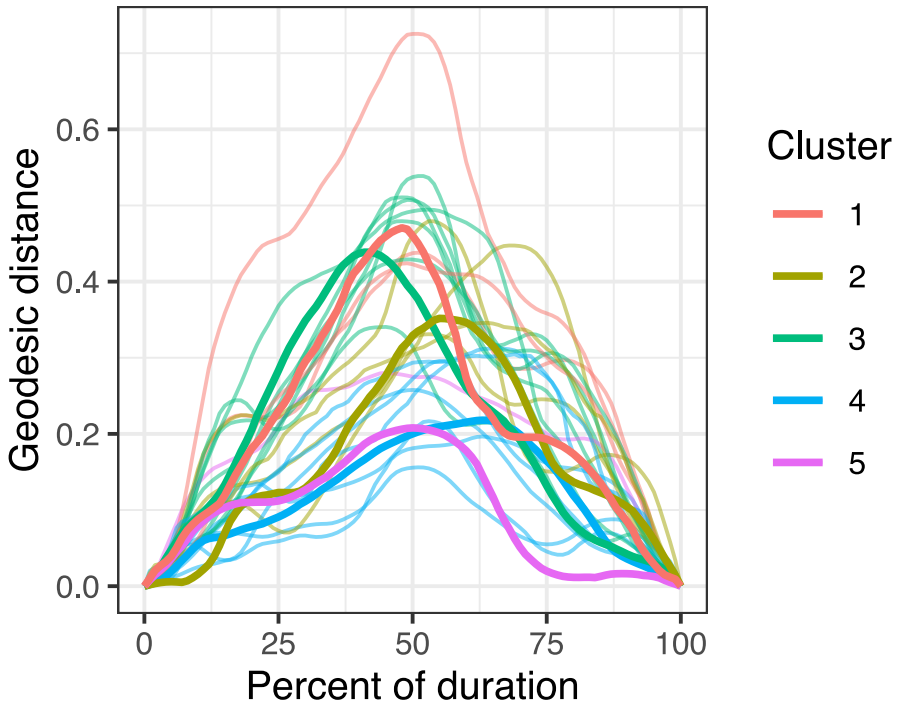


Figure 11: Geodesic distance of the clusters medoids

An intuitive easy-to-read representation of the IGP is the time series of the angles of its rotations:

$$\theta_i = \{2 \arccos \mathrm{Re}\left(\mathbf{q}_{i,0}\right), \ldots, 2 \arccos \mathrm{Re}\left(\mathbf{q}_{i,100}\right)\} \tag{13}$$