



**HAL**  
open science

## Clustering de séries chronologiques de patients souffrant de douleurs chroniques

Armel Soubeiga, Jessem Ettaghouti, Violaine Antoine, Alice Corteval, Nicolas Kerckhove, Sylvain Moreno

► **To cite this version:**

Armel Soubeiga, Jessem Ettaghouti, Violaine Antoine, Alice Corteval, Nicolas Kerckhove, et al.. Clustering de séries chronologiques de patients souffrant de douleurs chroniques. Extraction et Gestion des Connaissances (EGC) - 2023, Jan 2023, Lyon, France. . hal-04063254

**HAL Id: hal-04063254**

**<https://hal.science/hal-04063254v1>**

Submitted on 8 Apr 2023

**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.



## Introduction

La douleur chronique touche des millions de patients en France, soit environ 30% de la population générale. Les traitements disponibles sont anciens, ont une efficacité limitée et peuvent entraîner des effets indésirables importants [2]. En plus, le parcours de santé des patients souffrant de douleur chronique est multiple ce qui entraîne des résultats médiocres en terme d'amélioration de leur santé.

### Objectifs

L'identification des typologies des parcours et la détermination des profils des patients permettrait aux corps médical d'améliorer les résultats de soins et de mieux soutenir ces patients.

## Présentation des données

Les données ont été collectées à l'aide de l'application mHealth eDOL [2].



FIGURE 1 – Système et circuit d'acquisition des données eDOL.

## Matériels et méthodes

Pour développer un cadre de clustering basé sur les caractéristiques des baromètres de la douleur, nous avons proposé l'ensemble du processus suivant, résumé à la figure 2.

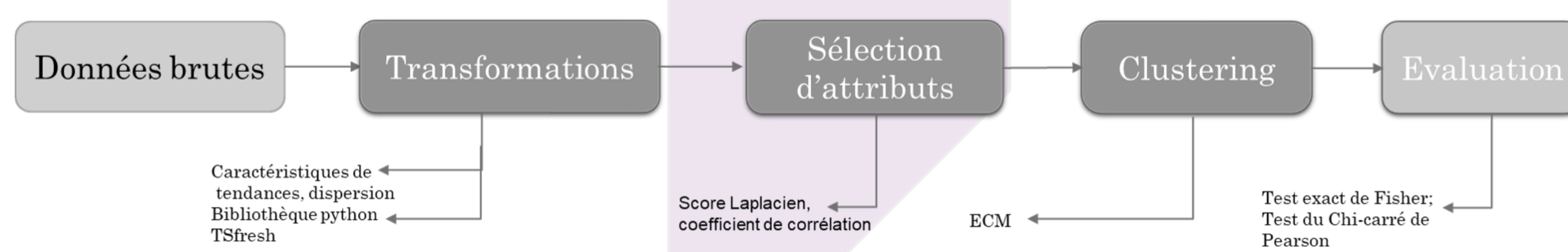


FIGURE 2 – Approche méthodologique proposée

### Transformation et extraction des caractéristiques

Un grand nombre de caractéristiques de tendance, de dispersion, d'homogénéité, d'aplatissement et d'asymétrie ont été calculées.

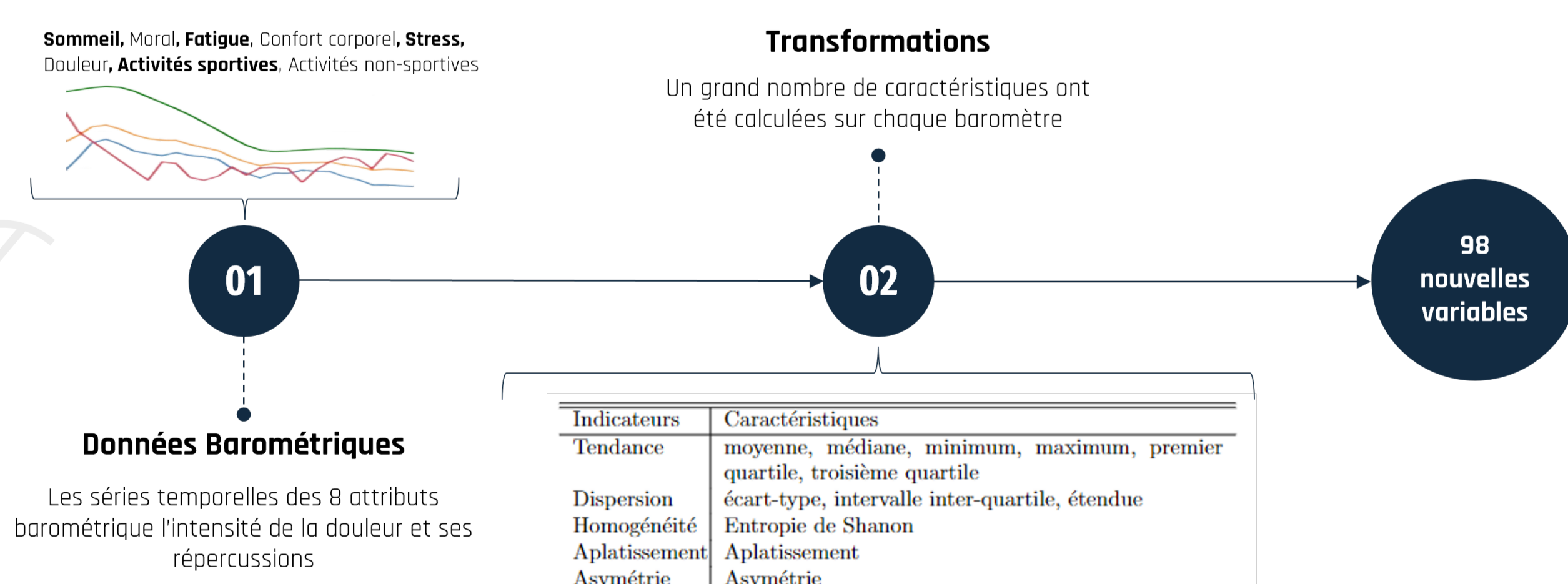


FIGURE 3 – Processus d'extraction des caractéristiques

### Sélection des caractéristiques

Pour récupérer les caractéristiques relativement plus concises, des techniques de sélection de variables ont été adoptées. La méthode de sélection proposée prend en compte, la similarité entre les observations, par le calcul du score Laplacien (1) et la corrélation de Pearson (2).

#### (1) Similarité de Laplacien

— Le score Laplacien de la variable  $x$  est défini :

$$L_x = \frac{\sum_{ij}(x_i - x_j)^2 S_{ij}}{\sum_i (x_i - \mu)^2 D_{ii}} \quad [1]$$

#### (2) Corrélation de Pearson

— Le coefficient de corrélation entre les variables  $x_1$  et  $x_2$  est défini :

$$C_{x_1 x_2} = \frac{\sum_i (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_i (x_{1i} - \bar{x}_1)^2 \sum_i (x_{2i} - \bar{x}_2)^2}}$$

### Algorithme de clustering ECM

L'algorithme Evidential c-means (ECM) permet de représenter un large éventail de situations concernant l'appartenance d'un objet à une classe. Elle recherche une partition appelée crédal  $M$  et l'ensemble de prototypes  $V$  qui minimisent la fonction objective suivante [3] :

$$\begin{cases} J_{ECM}(\mathbf{M}, \mathbf{V}) = \sum_{i=1}^n \sum_{A_j \neq \emptyset} |A_j|^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta \\ \text{s.c. } \sum_{j/A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij} + m_{i\emptyset} = 1 \quad \forall i = 1, \dots, n \end{cases} \quad (1)$$

Avec  $\alpha$  l'exposant de pondération pour la cardinalité,  $\beta$  l'exposant de pondération pour le caractère flou,  $\delta$  la distance à l'ensemble vide.

## Résultats

Au total, les données de 246 patients ont été analysées. Nous avons extrait 98 attributs des séries brutes des baromètres dans la phase de transformation et 8 attributs pertinents ont été choisis selon la similarité et la corrélation durant la phase de sélection.

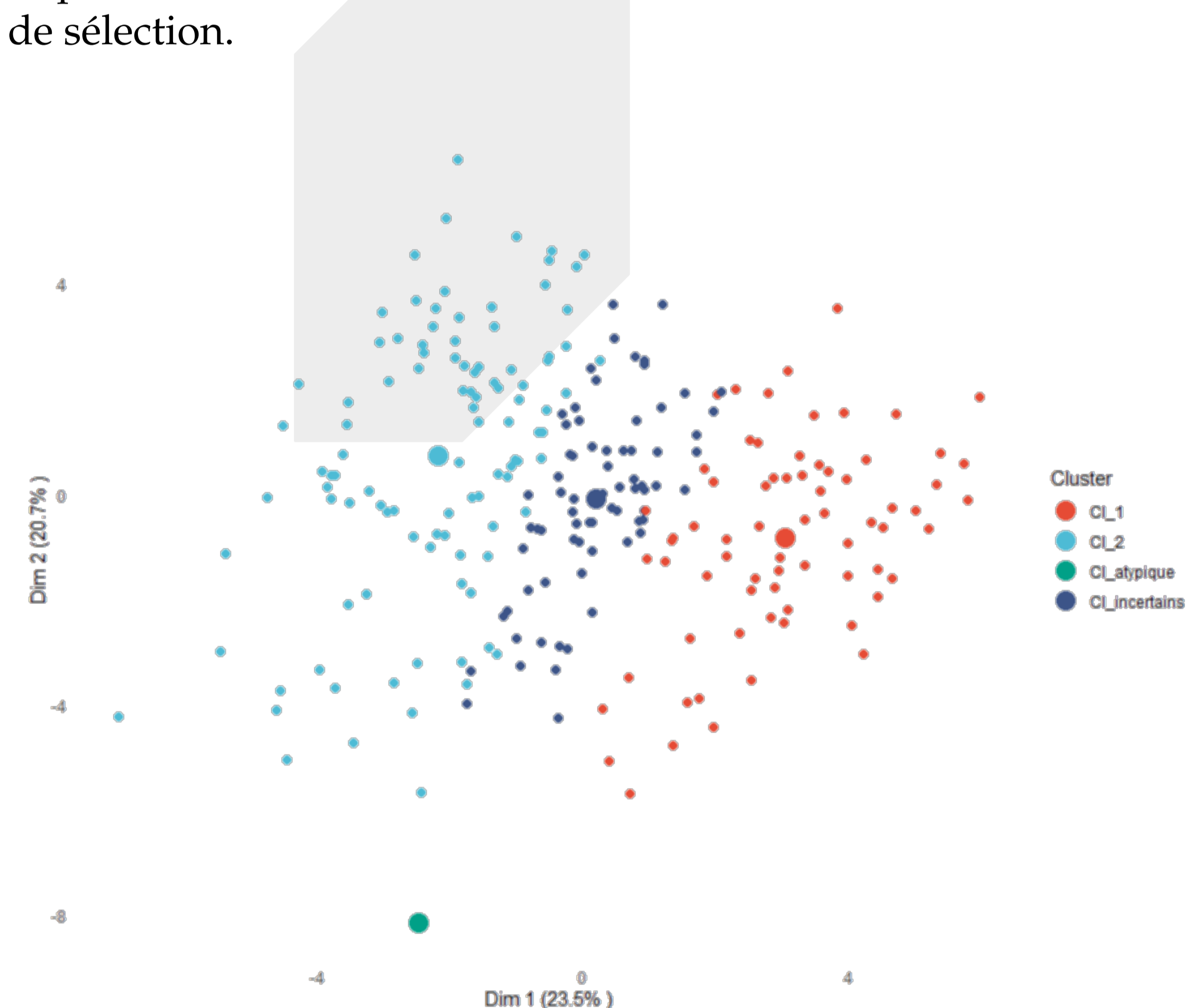


FIGURE 4 – Typologie de la partition crédale nette.

Le nombre optimal de clusters, qui maximise la silhouette moyenne est deux clusters : un cluster caractérisé par les douleurs liées au stress et un cluster regroupant des patients dont la douleur est liée à la fatigue et des troubles de sommeil. Deux sous-groupes sont également identifiés : un groupe qui représente les patients incertains entre les deux clusters et un groupe de patients atypiques. L'explicabilité des clusters déterminés à travers une analyse descriptive et des tests statistiques nous a permis de déterminer le profil de ces patients douloureux chroniques.

## Conclusions

Cette approche nous a permis d'identifier deux typologies de parcours avec des profils différents. En plus de ces parcours, nous avons identifié des groupes de patients incertains et atypiques.

## Références

[1] X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. Technical report, Department of Computer Science, University of Chicago, 2008.  
 [2] N. Kerckhove and al. eDOL mhealth app and web platform for self-monitoring and medical follow-up of patients with chronic pain : Observational feasibility study. *JMIR Form Res*, 6(3) :e30052, Mar 2022.  
 [3] M. H. Masson and T. Dencoux. ECM : An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41 :1384–1397, 4 2008.