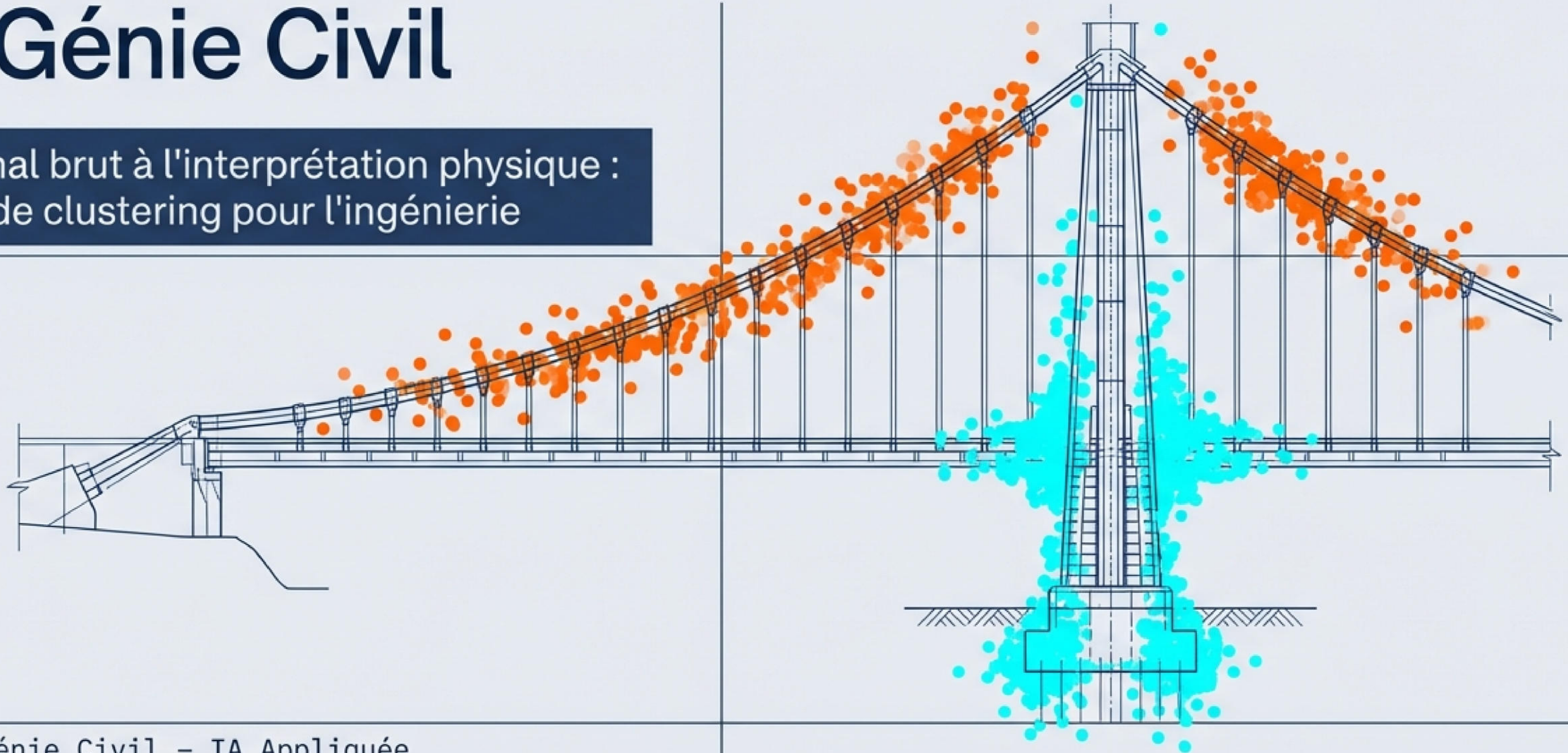


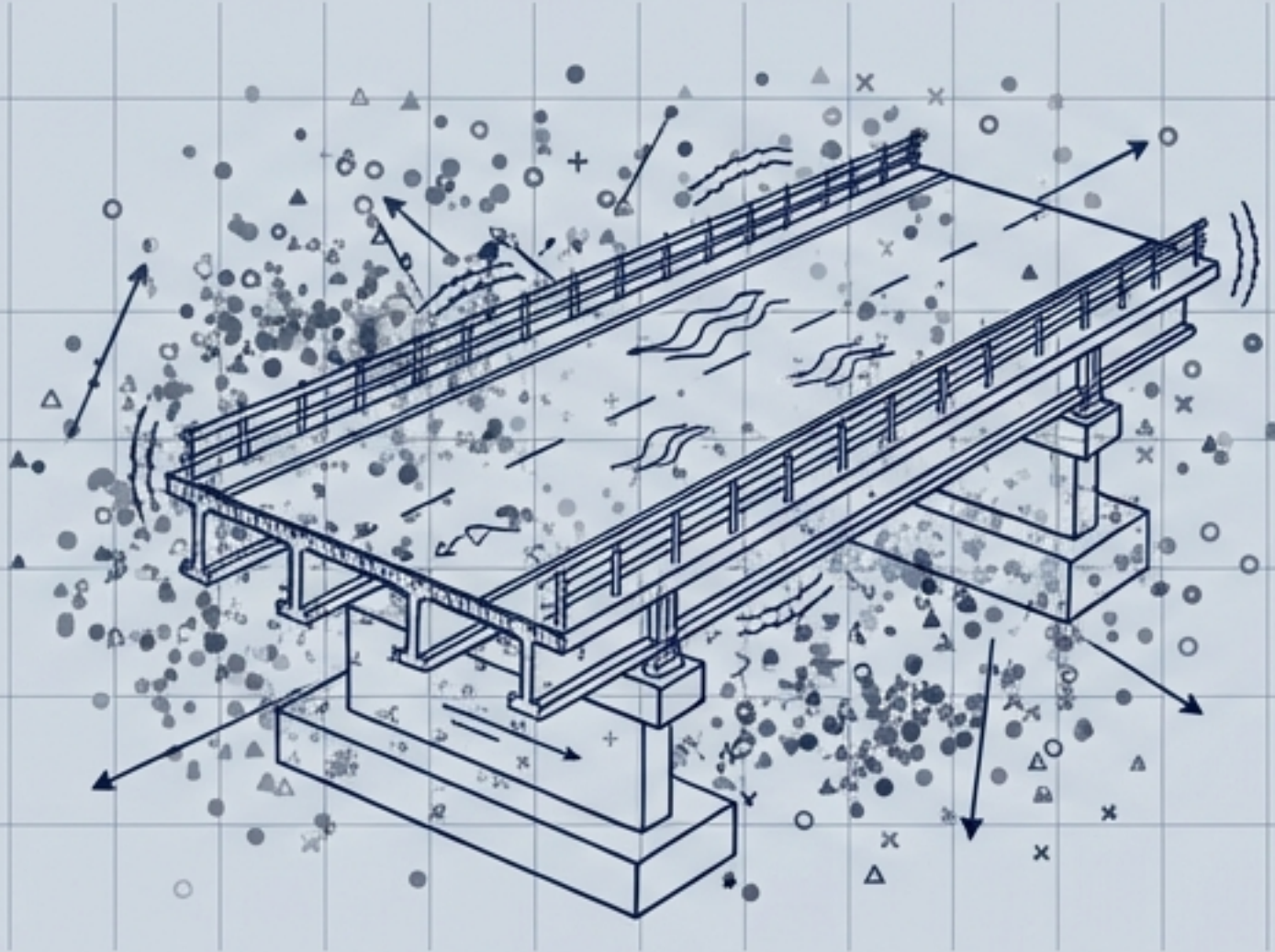
Chapitre 5 Apprentissage Non Supervisé : Découvrir les Structures Cachées en Génie Civil

Du signal brut à l'interprétation physique :
Guide de clustering pour l'ingénierie



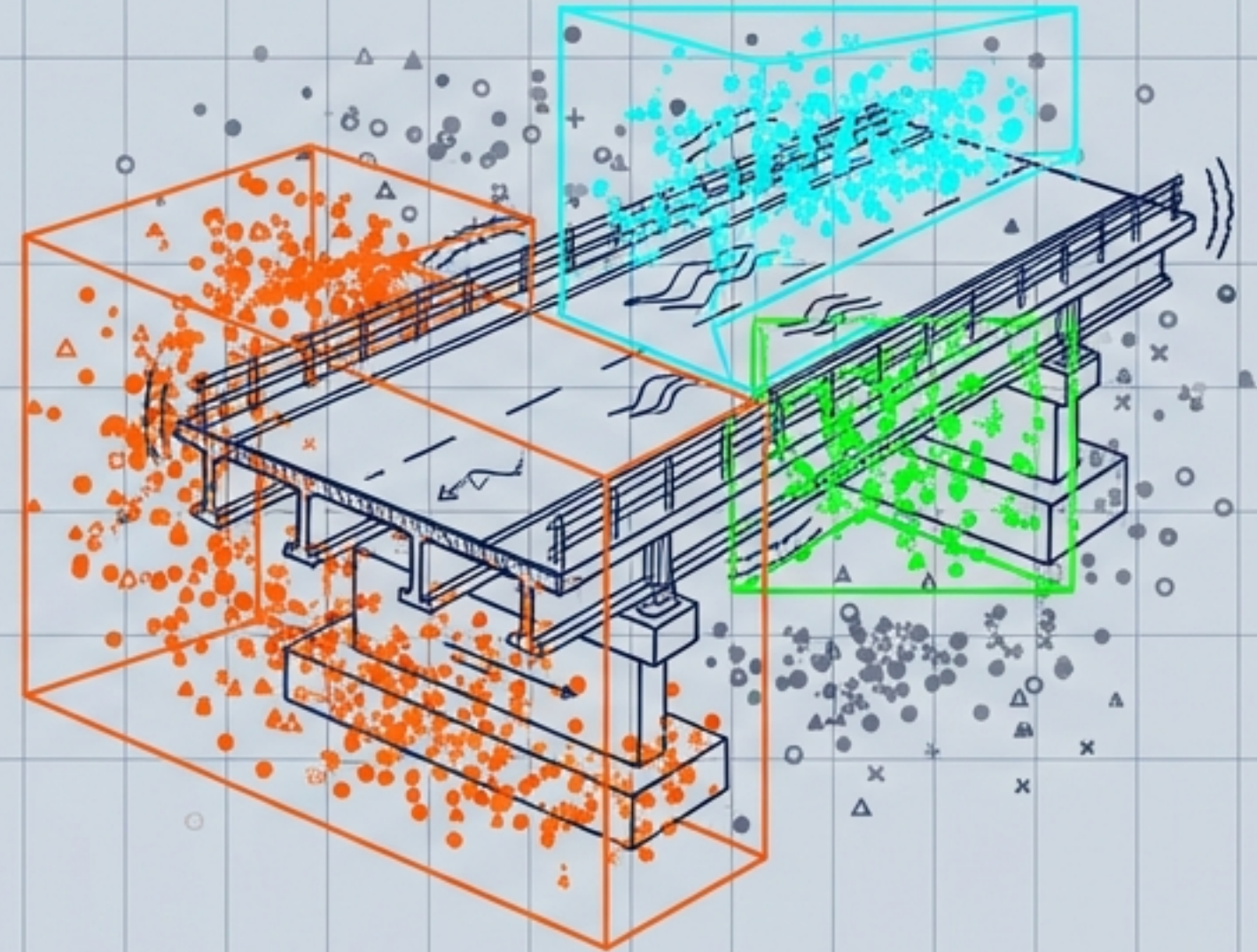
Le Problème des Données Modernes en Génie Civil

Le Chaos



- Capteurs massifs : SHM, données géotechniques, signaux vibratoires.
- Absence d'étiquettes : Aucun label préalable pour définir l'état de la structure.
- Comportements complexes : Les phénomènes émergent des données elles-mêmes, au-delà des équations mécaniques simples.

La Structure



L'objectif : Passer d'un nuage de points complexe à des groupes physiquement homogènes (ex: État Normal, Trafic Lourd, Anomalie Structurelle).

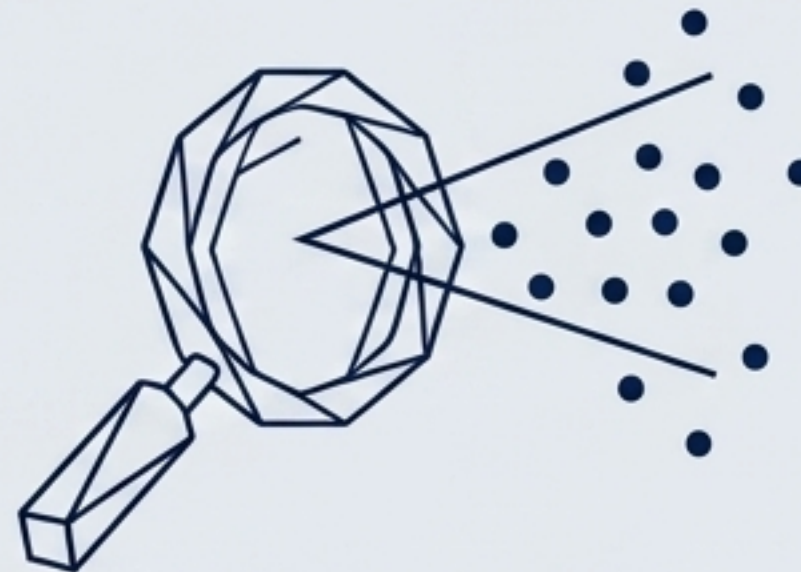
Qu'est-ce que le Clustering ?

Le Clustering (ou partitionnement) consiste à regrouper automatiquement des données similaires en sous-ensembles appelés clusters, sans connaissance a priori.

Données Brutes

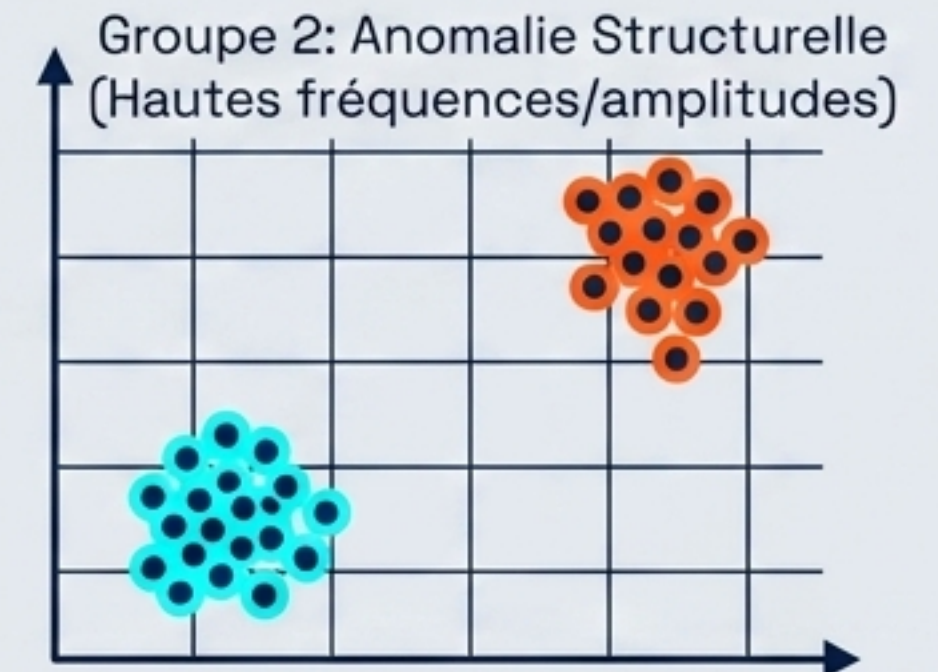
Fréquence	Amplitude
10Hz	0.2
12Hz	0.25
45Hz	1.2
50Hz	1.3

Transformation



Algorithme de similarité

Découverte

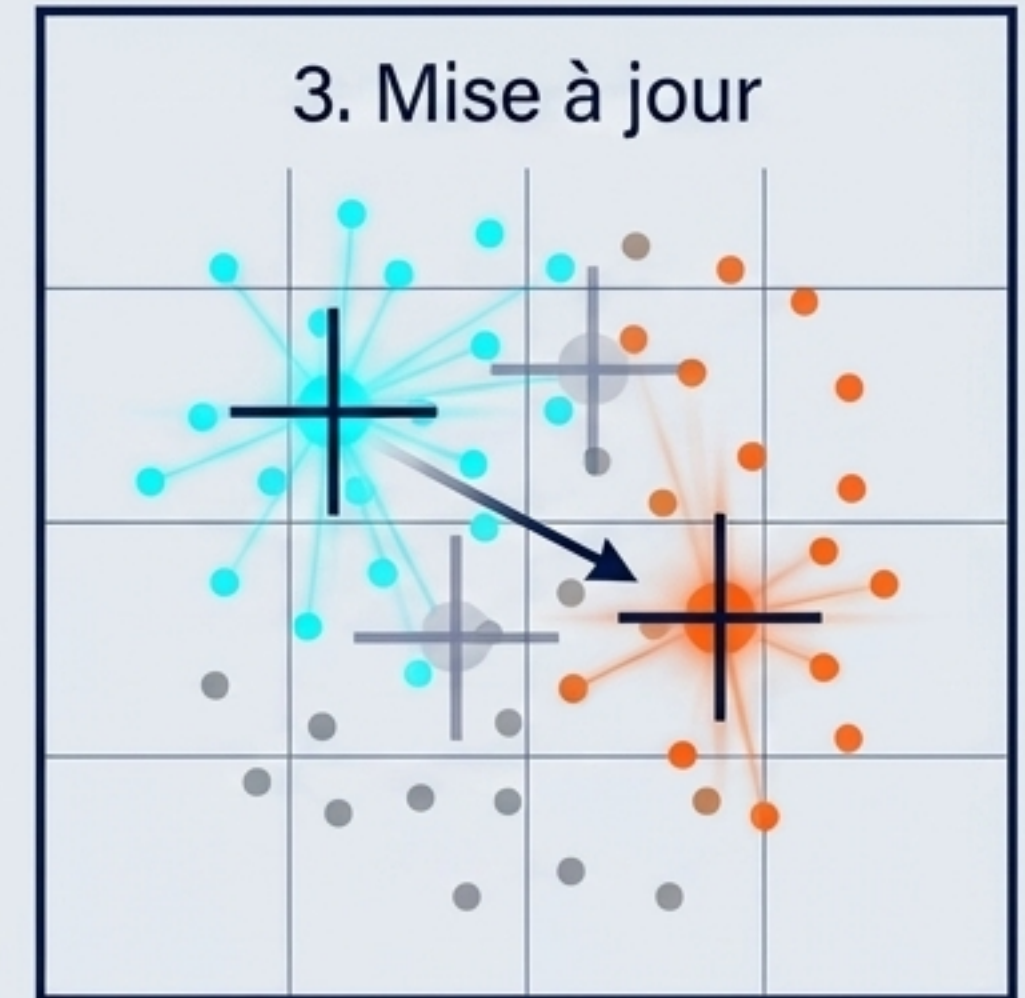
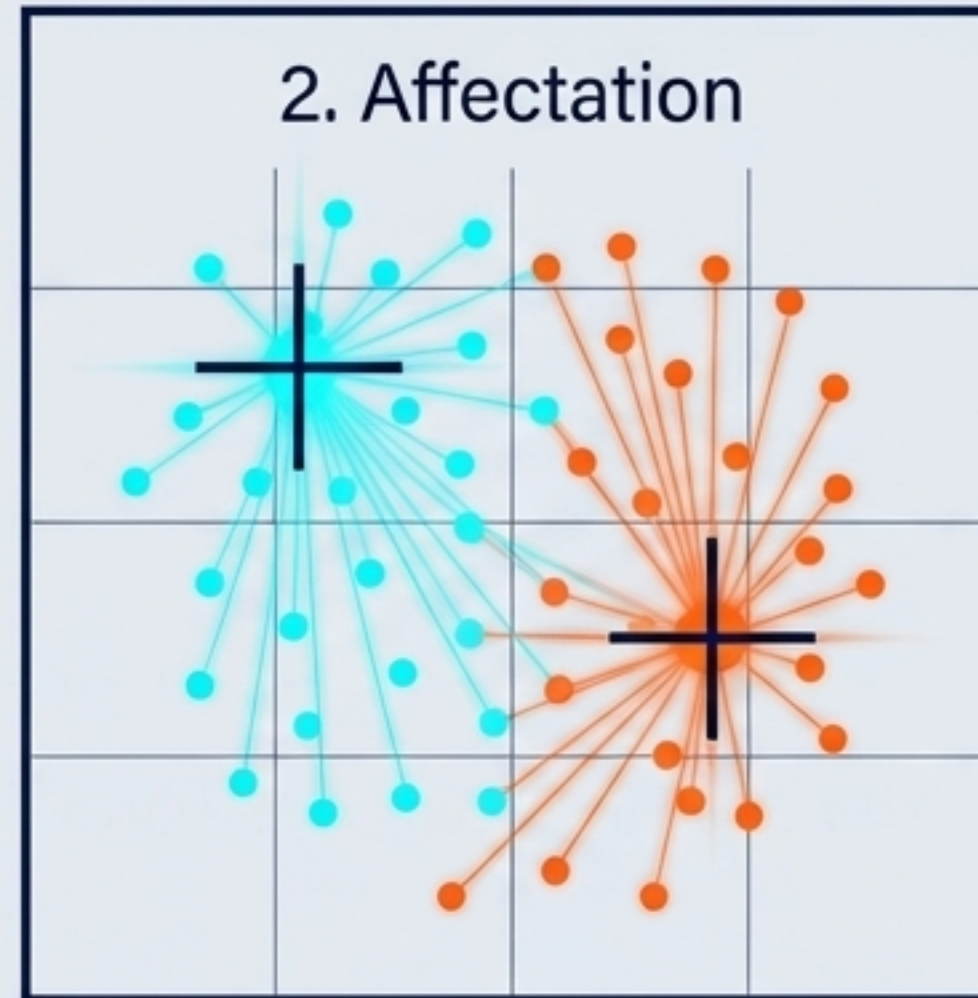
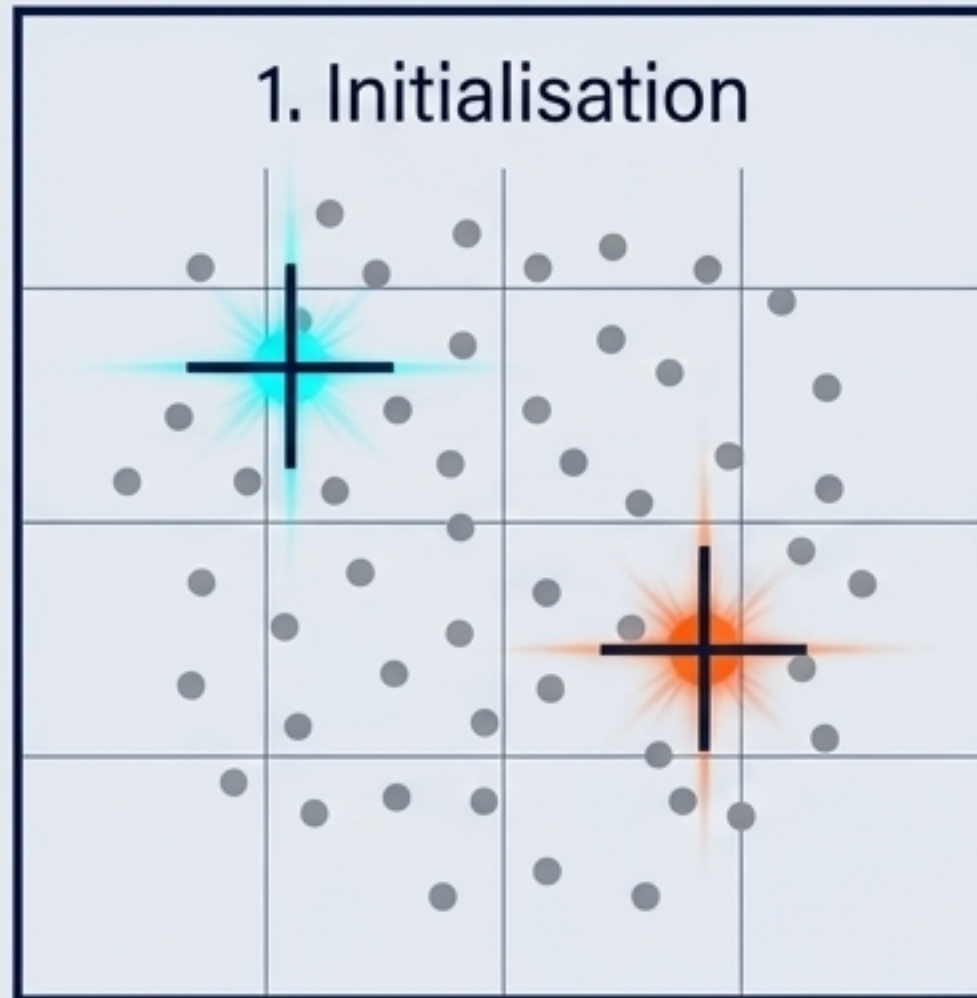


Groupe 1: Comportement Normal
(Basses fréquences/amplitudes)

Taxonomie : Supervisé vs. Non Supervisé

	 Apprentissage Supervisé	 Apprentissage Non Supervisé (Clustering)
Nature des Données	Cible connue (Labels présents)	Cible inconnue (Données brutes)
Objectif Analytique	Prédire une valeur précise	Découvrir une organisation cachée
Approche Mathématique	Minimiser l'erreur de prédiction	Maximiser la similarité intra-groupe
Exemple (Génie Civil)	Prédire la résistance (MPa) d'un béton selon sa composition	Identifier des profils de dégradation sur un pont (SHM)

L'Outil de Distance : Algorithme K-Means



Formule:
$$J_w = \sum_{k=1}^K \sum_{i \in C_k} d^2(x_i, \mu_k)$$

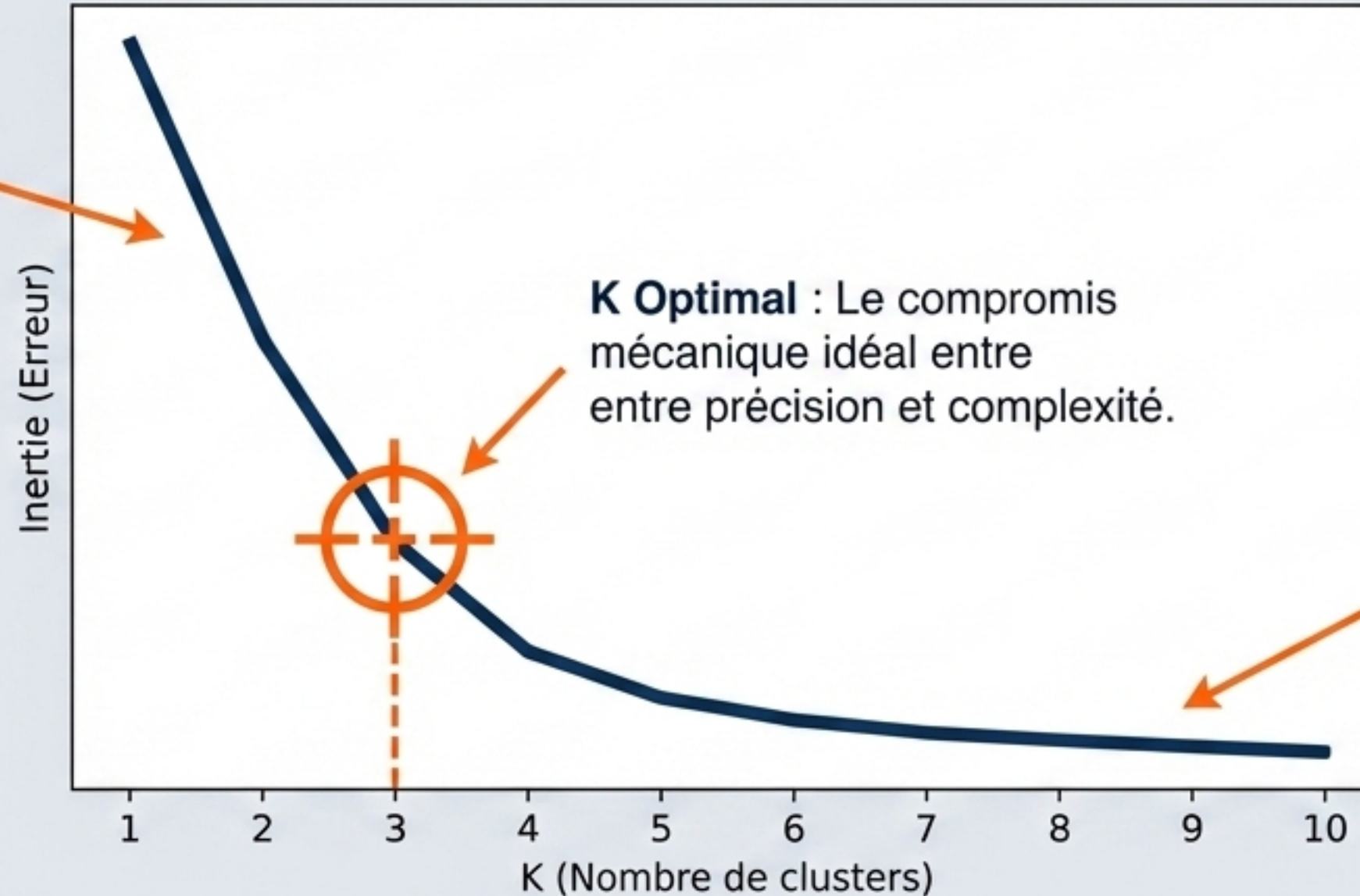
Interprétation : Minimiser l'inertie intra-classe (la variance des points autour de leur centre).

! Limitation critique : Impose des clusters de forme sphérique/convexe. Sensible aux valeurs aberrantes (outliers).

Optimisation de K-Means : La Méthode du Coude

Sous-segmentation :

Les groupes forcent ensemble des comportements physiques distincts.



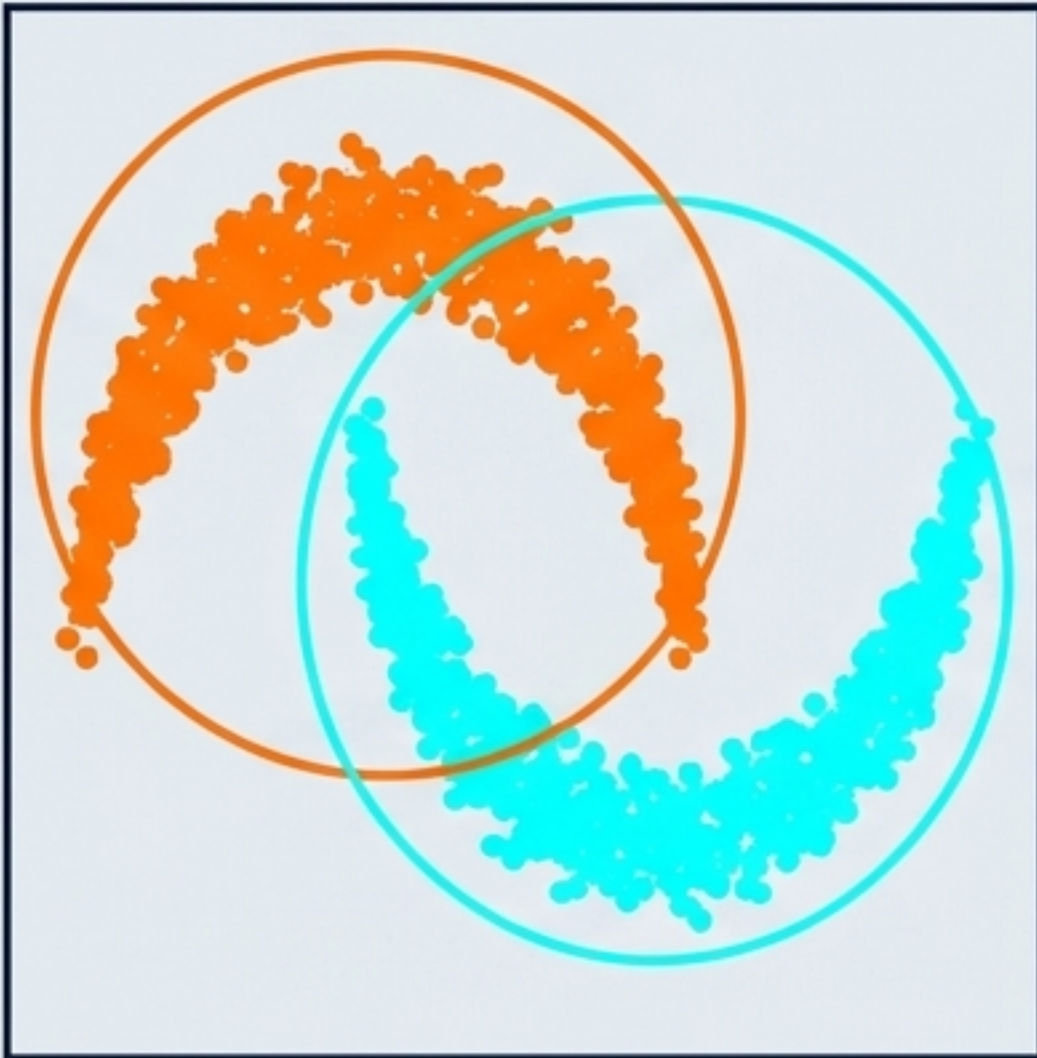
Sur-segmentation : Bruit statistique, division artificielle d'un même comportement physique.

Principe Clé : On cherche le point où l'ajout d'un cluster supplémentaire **ne réduit** plus significativement l'erreur.

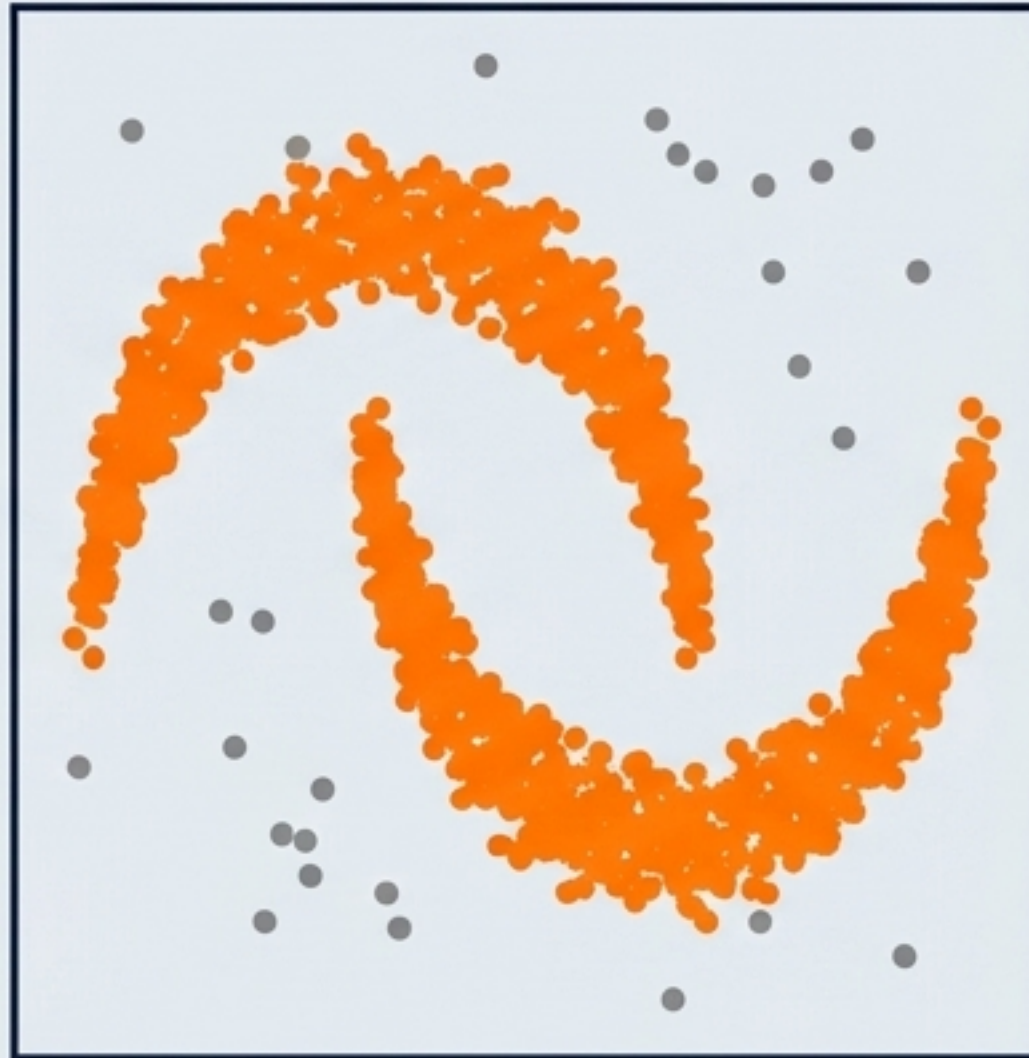
L'Approche par Densité : Algorithme DBSCAN

Contrairement à K-Means (basé sur la distance au centre), DBSCAN détecte les clusters par la continuité de la densité locale.

Échec de K-Means (Distance)



Succès de DBSCAN (Densité)



Les Paramètres Clés 🔍

- **ϵ (Epsilon)** : Le rayon de recherche physique.
- **minPts** : Le seuil critique de points pour valider une zone dense.

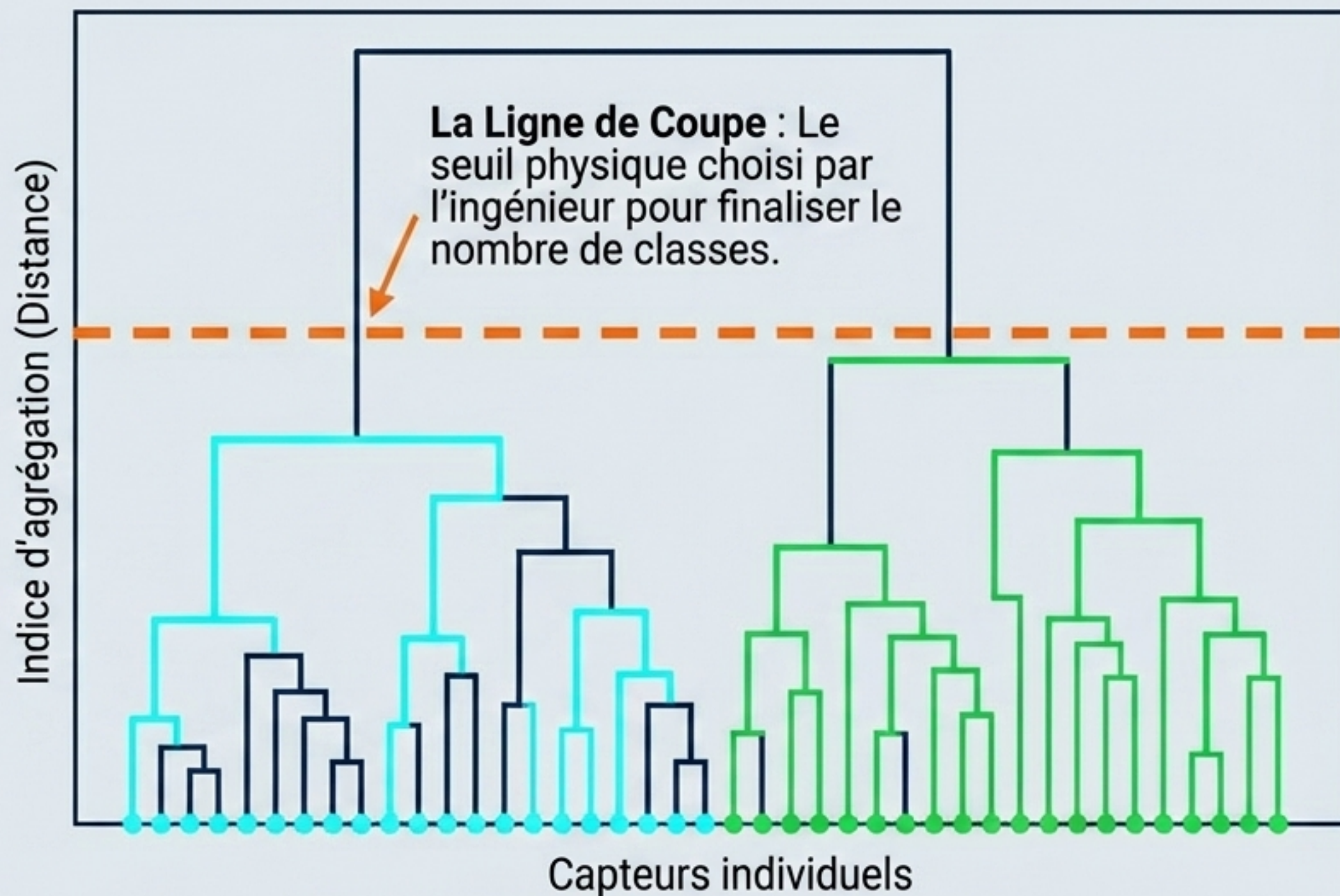
Avantages Ingénieur

- Détecte les formes complexes (ex : propagation de fissures).
- Isole le bruit de mesure (outliers) sans fausser les groupes.
- Pas besoin de deviner le nombre de clusters (K).

L'Approche Ascendante : Classification Hiérarchique (CAH)

Le Processus Mécanique :

1. **Initialisation** : Chaque capteur/point est un cluster isolé.
2. **Fusion Itérative** : Les groupes les plus proches (distance de Ward) s'assemblent.
3. **Hiérarchie** : Création d'un arbre binaire complet décrivant la structure globale.

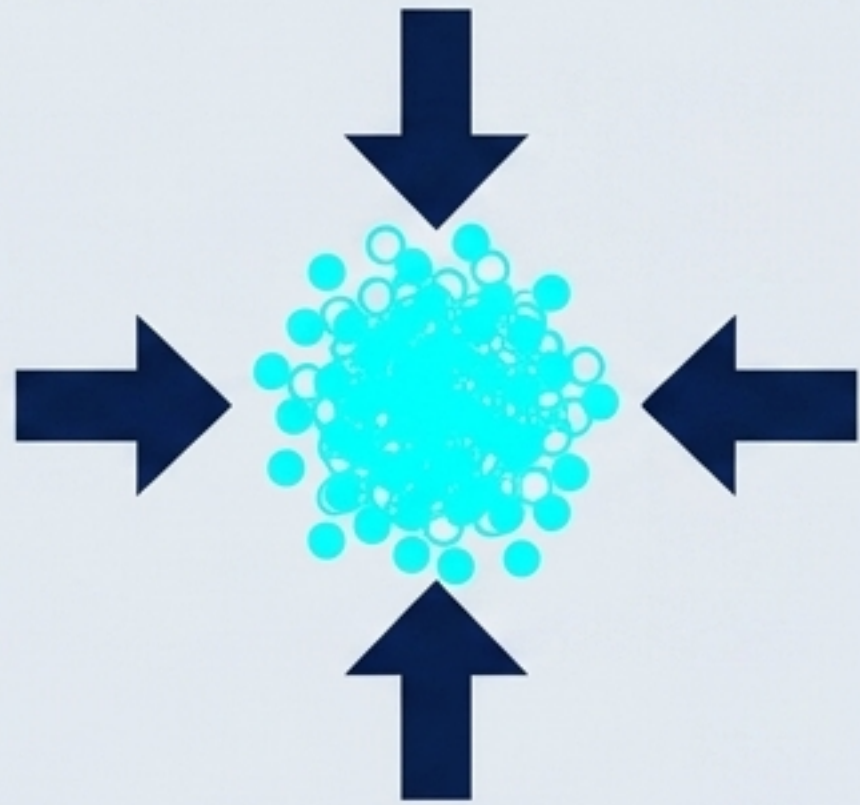


Matrice de Diagnostic Algorithmique

Critère	K-Means	DBSCAN	Hiérarchique (CAH)
Forme des Clusters	Sphérique / Convexe	Formes géométriques complexes	Hiérarchique (Arborescence)
Choix des Paramètres	Fixer K a priori	Fixer ϵ et minPts	Choisir la ligne de coupe
Gestion du Bruit (Outliers)	Mauvaise (Sensible) 	Excellente  (Isole le bruit)	Moyenne
Cas d'usage Génie Civil	Profils de matériaux homogènes	Détection d'anomalies (fissures, chocs)	Taxonomie des types de sols

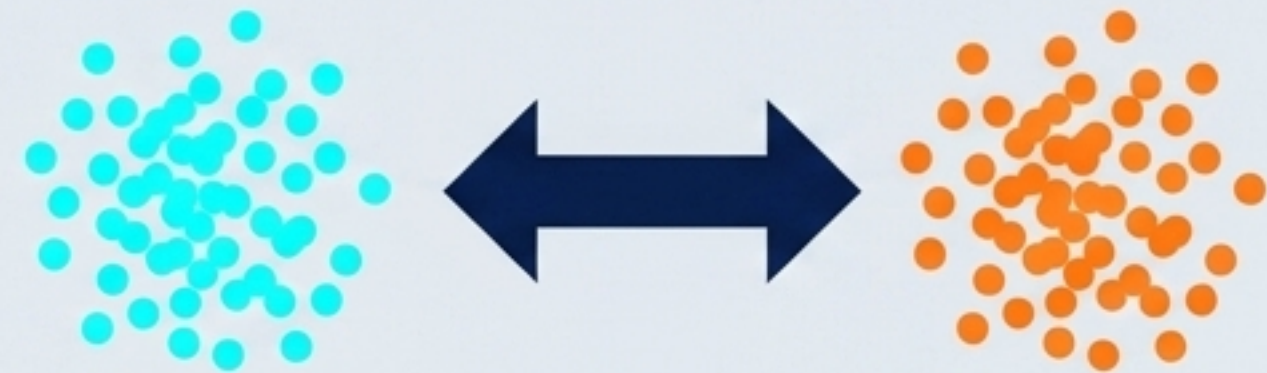
Philosophie d'Évaluation : Mesures Internes

Compacité (Intra-cluster)



Cohésion interne. Les individus d'un même groupe doivent se ressembler le plus possible. (Minimiser la variance).

Séparabilité (Inter-cluster)



Isolation externe. Les groupes doivent se démarquer les uns des autres de manière univoque. (Maximiser la distance entre barycentres).

Le Tableau de Bord des Métriques

Indice de Calinski-Harabasz (CH)

Rapport direct entre la dispersion inter-groupes et intra-groupes.

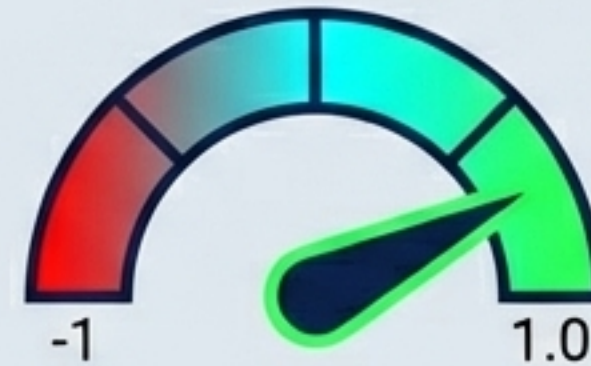


Règle d'ingénieur : Maximiser. Plus CH est grand, meilleure est la partition.

$$CH = \frac{(n-K) \times B}{(K-1) \times W} \quad (CH \geq 0)$$

Critère Silhouette

Évalue à la fois la compacité et l'écartement d'un point par rapport aux autres clusters (Score de -1 à 1).



Règle d'ingénieur : Approcher 1.

$S > 0.7$ indique une structuration forte.

Indice de Davies-Bouldin (DB)

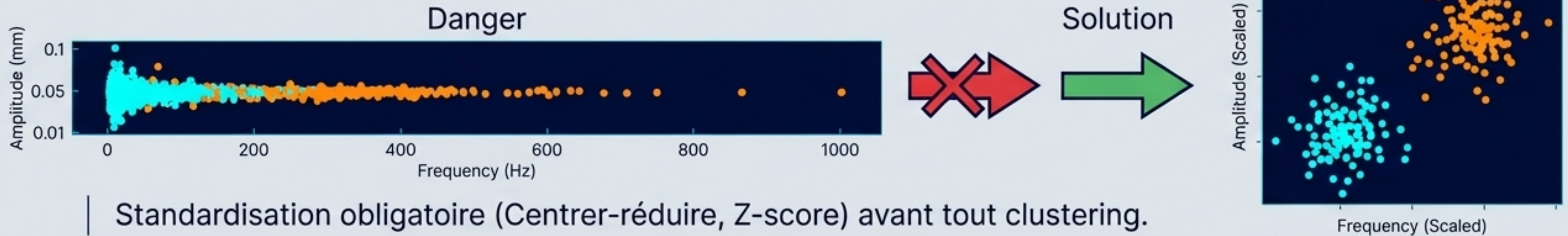
Ratio de la somme des dispersions intra-classes sur la distance inter-classes.



Règle d'ingénieur : Minimiser. Un indice faible indique des clusters denses et bien séparés.

Pièges Classiques et Bonnes Pratiques

Piège n°1 : Le danger des échelles (Mauvaise normalisation)



Piège n°2 : Variables non-physiques



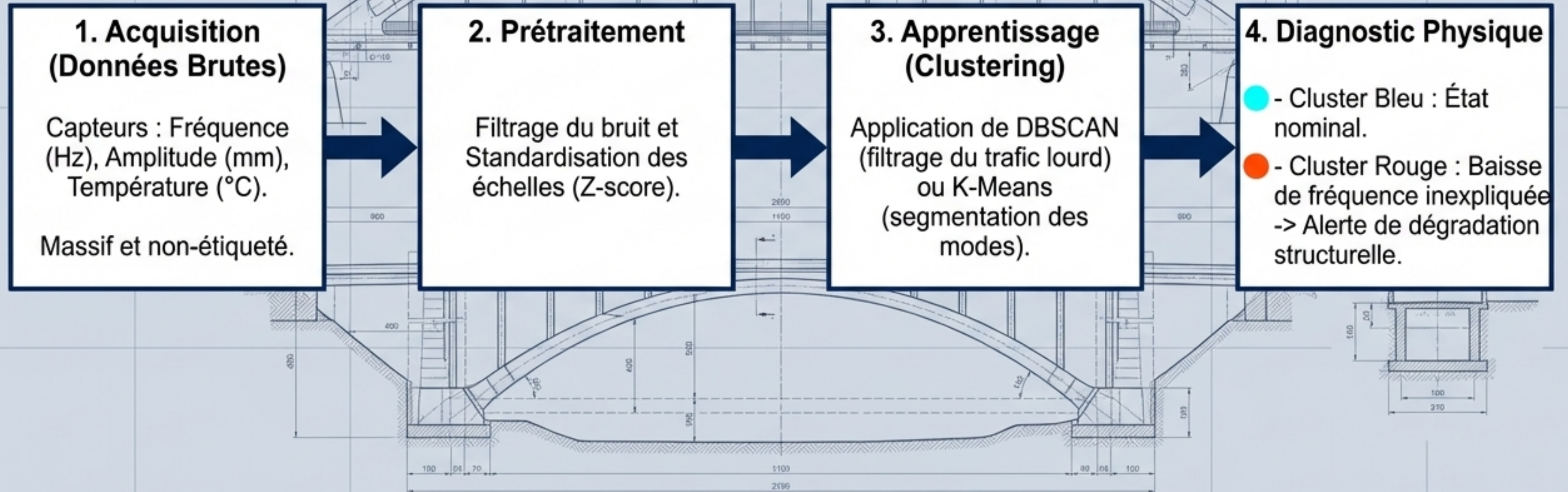
Intégrer des identifiants de capteurs ou des horodatages comme variables de calcul détruit la validité mécanique.

Piège n°3 : La Surinterprétation



Règle d'or : Un cluster mathématique n'est pas une vérité absolue. Il nécessite toujours une validation mécanique par l'ingénieur.

Étude de Cas : Workflow SHM sur un Pont



La Vision de l'Ingénieur

“Le clustering ne remplace pas la mécanique des structures. Il agit comme un **projecteur**, révélant des comportements cachés que les équations seules ne montrent pas.” ”

- L'apprentissage non supervisé transforme le bruit en structure.
- Le choix de l'algorithme (K-Means, DBSCAN) dépend de la physique du problème.
- L'interprétation finale repose toujours sur l'expertise de l'ingénieur civil.

Prochaine Étape : L'Intelligence Artificielle Profonde (Deep Learning)