



---

# Approches génétiques pour le clustering avec pondération locale d'attributs

Pierre Gançarski

# Approches génétiques, clustering et sélection d'attributs

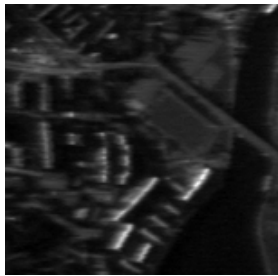
---

## ➤ **Problèmes**

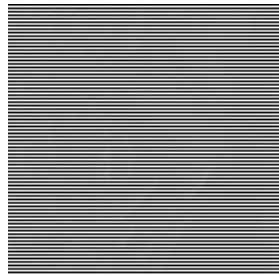
- **Attributs fortement corrélés, bruités et non pertinents**

# Approches génétiques, clustering et sélection d'attributs

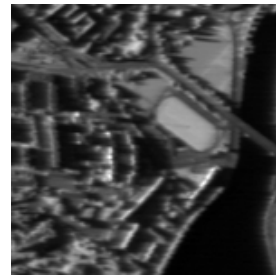
## Images hyperspectrales



Bande  
1



Bande  
14



Bande  
15



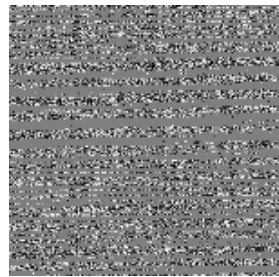
Bande  
26



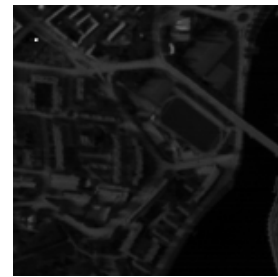
Bande  
41



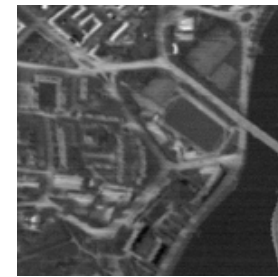
Bande  
43



Bande  
64

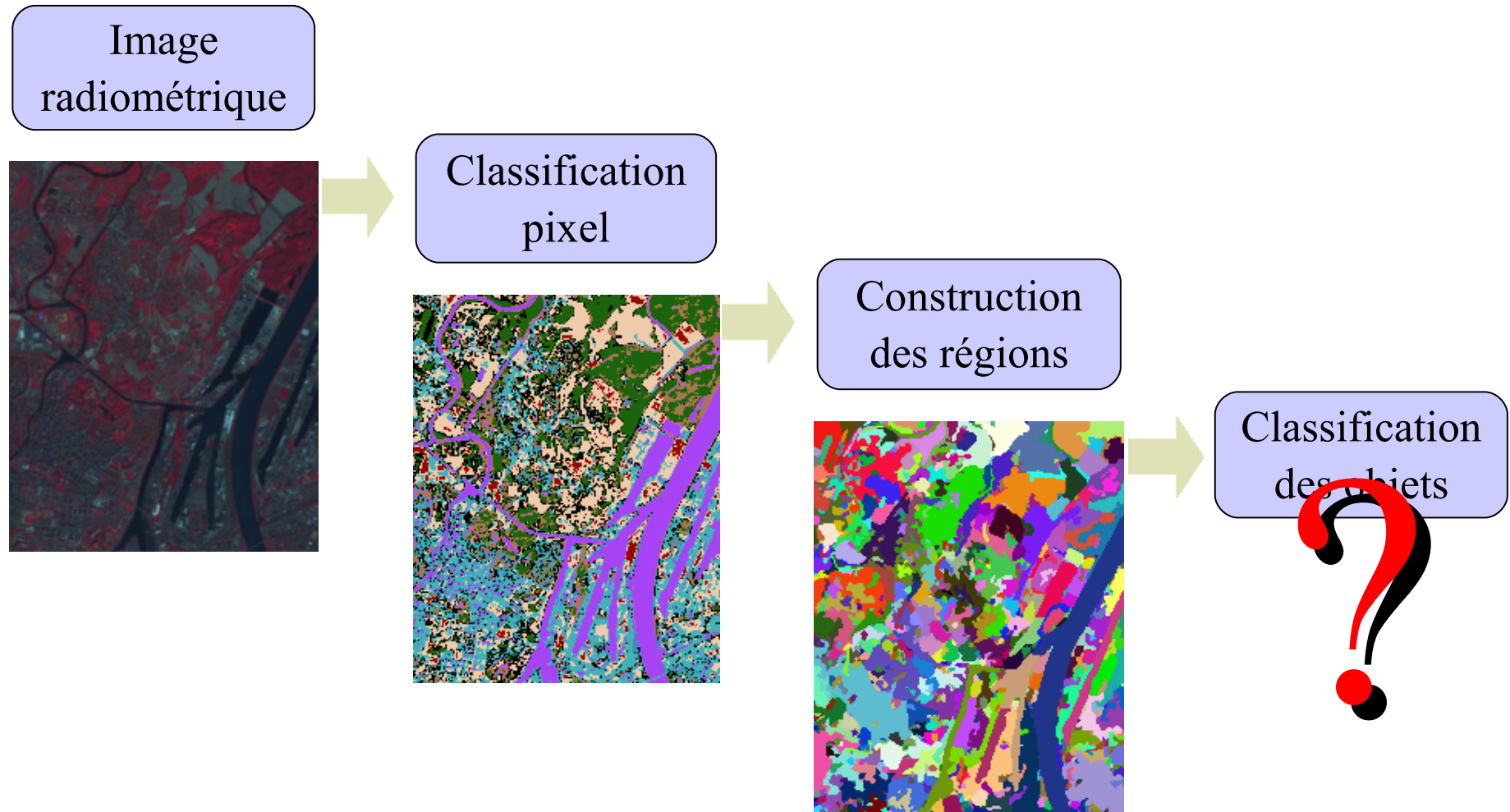


Bande  
73



Bande  
79

# Approches génétiques, clustering et sélection d'attributs



# Approches génétiques, clustering et sélection d'attributs

## ➤ Problèmes

- **Attributs fortement corrélés, bruités et non pertinents**
- **Attributs réellement discriminant pour chaque classe :**
  - **L'importance d'un attribut dépend de la classe que l'on cherche à mettre en évidence : texture pour le bati, radiométrie pour l'eau, ...**

# Approches génétiques, clustering et sélection d'attributs

## ➤ **Problèmes**

- **Attributs fortement corrélés, bruités et non pertinents**
- **Attributs réellement discriminant pour chaque classe :**
  - L'importance d'un attribut dépend de la classe que l'on cherche à mettre en évidence : texture pour le bati, radiométrie pour l'eau, ...

## ➔ **Pondération locale des attributs**

# Approches génétiques, clustering et sélection d'attributs

## ➤ **Deux approches**

- **Hard clustering et optimisation d'une fonction de coût basée sur une distance**
- **Soft clustering et optimisation basée sur la qualité de partitionnement**

# Approches génétiques, clustering et sélection d'attributs

## ➤ **Deux approches**

- **Hard clustering et optimisation d'une fonction de coût basée sur une distance**
- **Soft clustering et optimisation basée sur la qualité de partitionnement**



# Optimisation basée distance

- **Weighting Kmeans** (E. Chan, W. Ching, M. Ng, and J. Huang, "An optimization algorithm for clustering using weighted dissimilarity measures," Pattern Recognition, vol. 37, pp. 943–952, 2004.)

- **Idée : optimiser une fonction de coût**

$$WCost(c, W) = \sum_{k=1}^K \sum_{o \in S} \sum_{i=1}^n C_k(o) w_{(k,i)}^b d_i(o, c_k)$$

$$C = \{C_k\}_{k \in [1, K]}$$

Les deux inconnus

$$W = \{w_{(k,i)}\}_{i \in [1, n], k \in [1, K]}$$

# Weighting K-means

➤ **L'optimisation se fait "à la K-means" :**

- 1) Affectation des objets aux centres  
(utilisation d'une distance pondérée par  $W$ )**
- 2) Calcul des nouveaux centres**
- 3) Calcul des nouveaux poids :**

$$W_{(k,i)} = \begin{cases} 1/z_k & \text{si } s_k^i = 0 \\ 0 & \text{si } s_k^i \neq 0 \text{ et } z_k^i \neq 0 \\ 1 / \sum_{t=1}^n \left[ \frac{s_k^i}{s_k^t} \right]^{1/(b-1)} & \text{si } z_k = 0 \end{cases} \quad \left| \quad \begin{array}{l} \text{avec} \\ s_k^i = \sum_{o \in S} C_k(o) d_i(c_k, o)^2 \\ z_k = \left| \{ t \mid s_k^t = 0 \} \right| \end{array} \right.$$

# Optimisation par évolution génétique

## ➤ Deux dimensions

### ➤ Stratégies d'évolution

- Drawinienne
- Lamarckienne
- Baldwinienne

### ➤ Populations

- Une seule : « évolution classique »
- Plusieurs : co-évolution

# Stratégies d'évolution

## ➤ **Darwin** (approche la plus utilisée)

### ➤ **Principe :**

- **Les individus s'adaptent à leur environnement**
- **Ces adaptations sont « sous contrôle » des gènes**
- **La descendance qui présente la meilleure adaptation survit et se reproduit : «survival of the fittest »**

# Stratégies d'évolution

## ➤ Darwin

### ➤ Principe :

- Les individus s'adaptent à leur environnement
- Ces adaptations sont « sous contrôle » des gènes
- La descendance qui présente la meilleure adaptation survit et se reproduit : «survival of the fittest »

### ➤ Traduit par :

- **A chaque génération, les solutions sont calculées à partir des chromosomes**
- **Chaque solution (chaque individu) est évaluée**
- **Les meilleurs individus sont sélectionnés et leur matériel génétique est recombinaison pour former la génération suivante**

# Stratégies d'évolution

## ➤ **Lamarck** (approche « fausse » dans la vraie vie)

### ➤ **Principe :**

- **Les individus s'adaptent à leur environnement**
- **Ces adaptations influent sur les gènes**
- **La descendance qui présente la meilleure adaptation survit et se reproduit à partir de ces nouveaux gènes : «inheritance of acquired characteristics »**

# Stratégies d'évolution

## ➤ Lamarck

### ➤ Principe :

- Les individus s'adaptent à leur environnement
- Ces adaptations influent sur les gènes
- La descendance qui présente la meilleure adaptation survit et se reproduit à partir de ces nouveaux gènes : «inheritance of acquired characteristics »

### ➤ Traduit par :

- **A chaque génération, les solutions sont calculées à partir des chromosomes : le calcul d'une solution modifie le chromosome associé (lifetime learning method)**
- **Chaque solution (chaque individu) est évaluée**
- **Les meilleurs individus sont sélectionnés et leur nouveau matériel génétique est recombiné**

# Stratégies d'évolution

## ➤ **Baldwin**

### ➤ **Principe :**

- **Les individus s'adaptent à leur environnement**
- **Ces adaptations dépendent des gènes et de la « phenotypic plasticity » qui décrit la capacité à s'adapter**
- **La descendance qui présente la meilleure capacité à s'adapter survit et se reproduit (comme pour Darwin)**



# Stratégies d'évolution

## ➤ Darwin

### ➤ Principe :

- Les individus s'adaptent à leur environnement
- Ces adaptations dépendent des gènes et de la « phenotypic plasticity » qui décrit la capacité à s'adapter
- La descendance qui présente la meilleure adaptation survit et se reproduit (comme pour Darwin)

### ➤ Traduit par :

- **A chaque génération, les solutions sont calculées à partir des chromosomes : le calcul d'une solution modifie l'individu SANS modifier ses gènes**
- **Chaque solution (chaque individu) est évaluée**
- **Les meilleurs individus sont sélectionnés et leur (ancien) matériel génétique est recombinaison**

# Evolution vs co-évolution

## ➤ **Evolution**

- **Une seule population**
- Chaque individu est apte à construire une solution globale qui peut être évaluée
- **Chaque chromosome porte tous les gènes nécessaires au calcul d'une solution globale**
  - **Exemple : un chromosome = ensemble des pondérations pour toutes les classes.**

# Evolution vs co-évolution

## ➤ **Co-évolution**

- **Plusieurs populations : un individu évolue dans une population**
- **Chaque population cherche une solution locale « optimale »**
- **Un individu ne peut construire qu'une solution locale qui peut être évaluée :**
  - Localement : par rapport aux autres individus de la population
  - Globalement : par rapport à ce qu'elle amène à la solution globale

# Evolution vs co-évolution

## ➤ **Co-évolution**

- **Un individu ne peut construire qu'une solution locale**
  - **Un gène ne porte que les gènes nécessaires au calcul d'une solution locale**
    - **Exemple : un chromosome = la pondération pour extraire une classe**
- Nécessité d'un mécanisme de construction et d'évaluation des solutions globales**

# Optimisation par évolution génétique

## ➤ **Stratégies d'évolution**

- **Drawinienne**
- Lamarckienne
- Baldwinienne

## ➤ **Populations**

- **évolution**
- co-évolution

# Approche évolutive darwinienne

- **Chaque individu calcule une solution à l'aide d'un K-means**
  - initialisé avec la meilleure solution courante
  - et en utilisant la pondération donnée par son chromosome
- **Les individus sont évalués en utilisant la fonction Wcost :**
  - initialisé avec la solution trouvée
  - et en utilisant le chromosome
  - **la meilleure solution est conservée**
- **Une reproduction « classique » est faite**

# Optimisation par évolution génétique

## ➤ **Stratégies d'évolution**

- Drawinienne
- **Lamarckienne**
- Baldwinienne

## ➤ **Populations**

- **évolution**
- co-évolution

# Approche évolutive lamarckienne

- **Chaque individu calcule une solution à l'aide d'un Weighting-Kmeans**
  - initialisé avec la meilleure solution courante
  - et en utilisant en pondération initiale celle donnée par son chromosome
- **Les nouveaux poids sont affectés au chromosome de l'individu**



# Approche évolutive lamarckienne

- **Chaque individu calcule une solution à l'aide d'un Weighting-Kmeans**
- **Les nouveaux poids sont affectés au chromosome de l'individu**
- **Les individus sont évalués par Wcost :**
  - initialisé avec la solution trouvée
  - et en utilisant le nouveau chromosome
  - **la meilleure solution est conservée**
- **Une reproduction « classique » est faite avec le nouveau matériel génétique**

# Optimisation par évolution génétique

## ➤ **Stratégies d'évolution**

- Drawinienne
- Lamarckienne
- **Baldwinienne**

## ➤ **Populations**

- **évolution**
- co-évolution

# Approche évolutive baldwinienne

## Identique à Lamarck

- **Chaque individu calcule une solution à l'aide d'un Weighting-Kmeans**
  - initialisé avec la meilleure solution courante
  - et en utilisant en pondération initiale celle donnée par son chromosome
- **Les nouveaux poids NE sont PAS affectés au chromosome de l'individu**

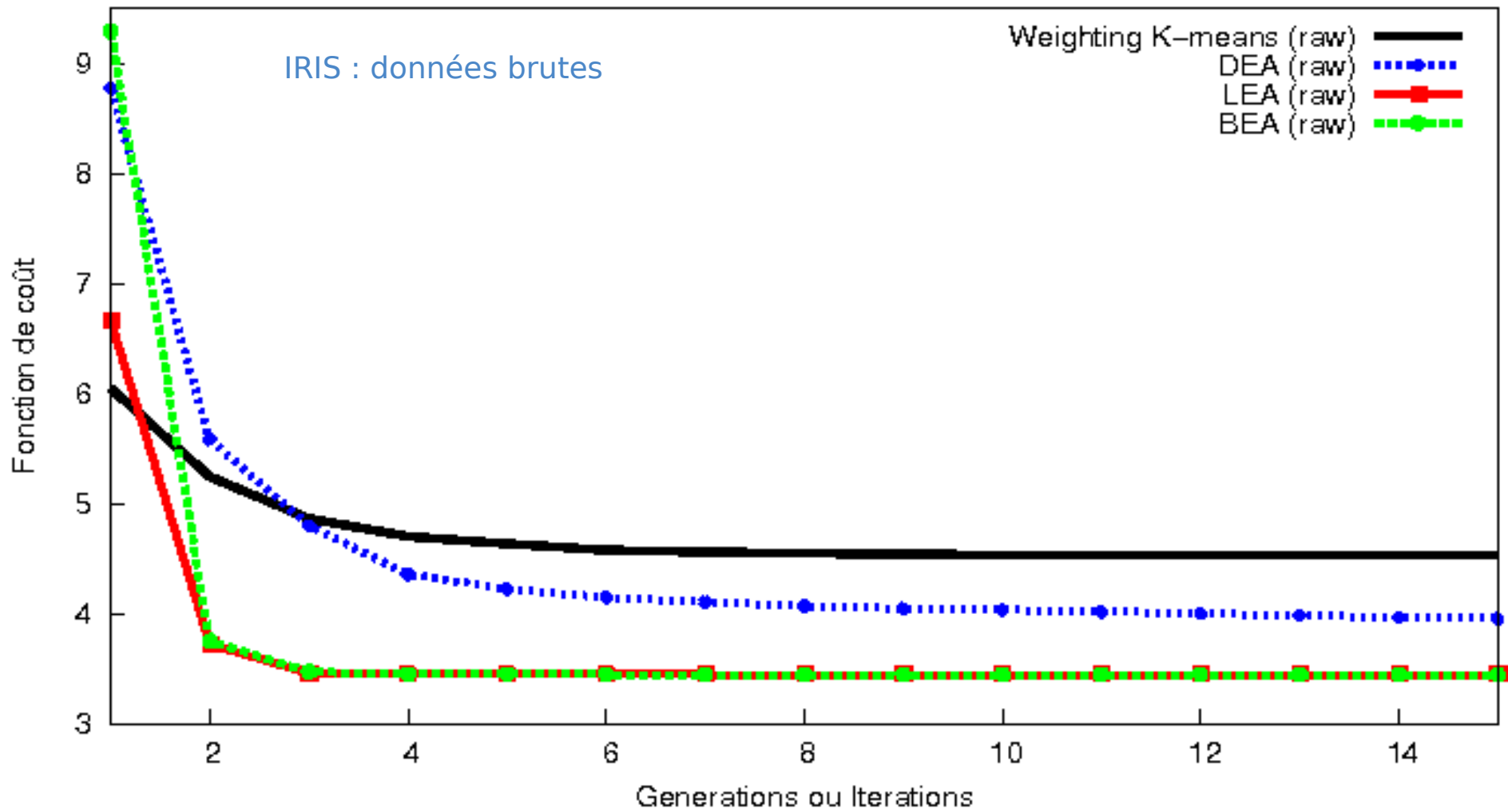
# Approche évolutive baldwinienne

- **Chaque individu calcule une solution à l'aide d'un Weighting-Kmeans**
- **Les nouveaux poids ne sont pas affectés au chromosome de l'individu**
- **Les individus sont évalués par Wcost :**
  - initialisé avec la solution trouvée
  - et en utilisant le nouveau chromosome
  - **la meilleure solution est conservée**
- **Une reproduction « classique » est faite à partir des chromosomes initiaux**

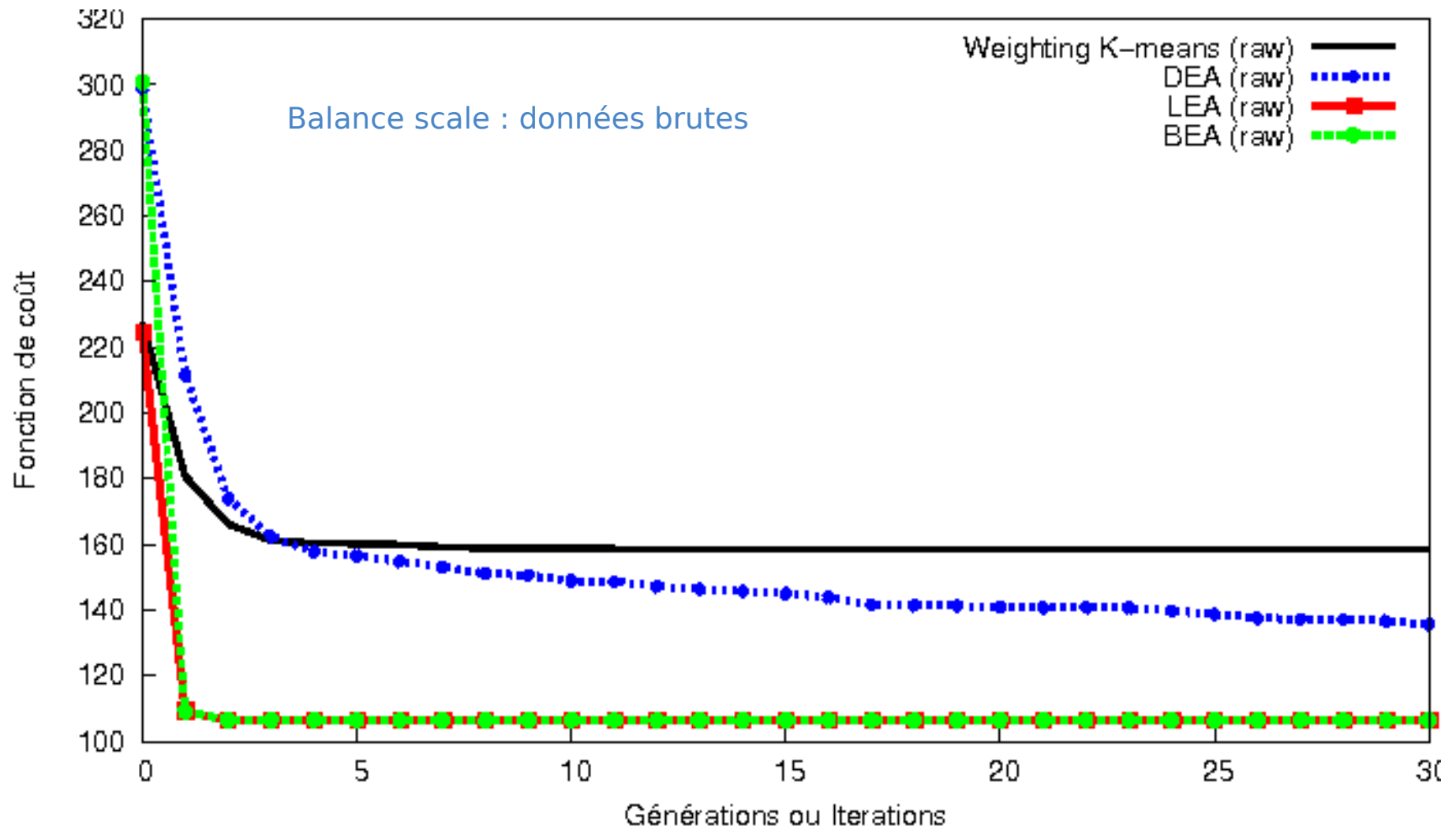
# Résultats des approches évolutives

- **Données : Iris et Balance-scale**
  - **Iris : 4 attributs, 150 objets 3 classes**
  - **Balance scale : 4 attributs numériques, 3 classes , 768 objets (46,1%; 7,8%; 46,1%)**
- **Deux critères de qualité**
  - **Évaluation de la fonction de coût**
  - **Comparaison avec la vérité-terrain (accuracy)**
- **Moyenne sur 100 tests**
  - **Iris : 60 individus;**
  - **Balance scale : 80 individus**

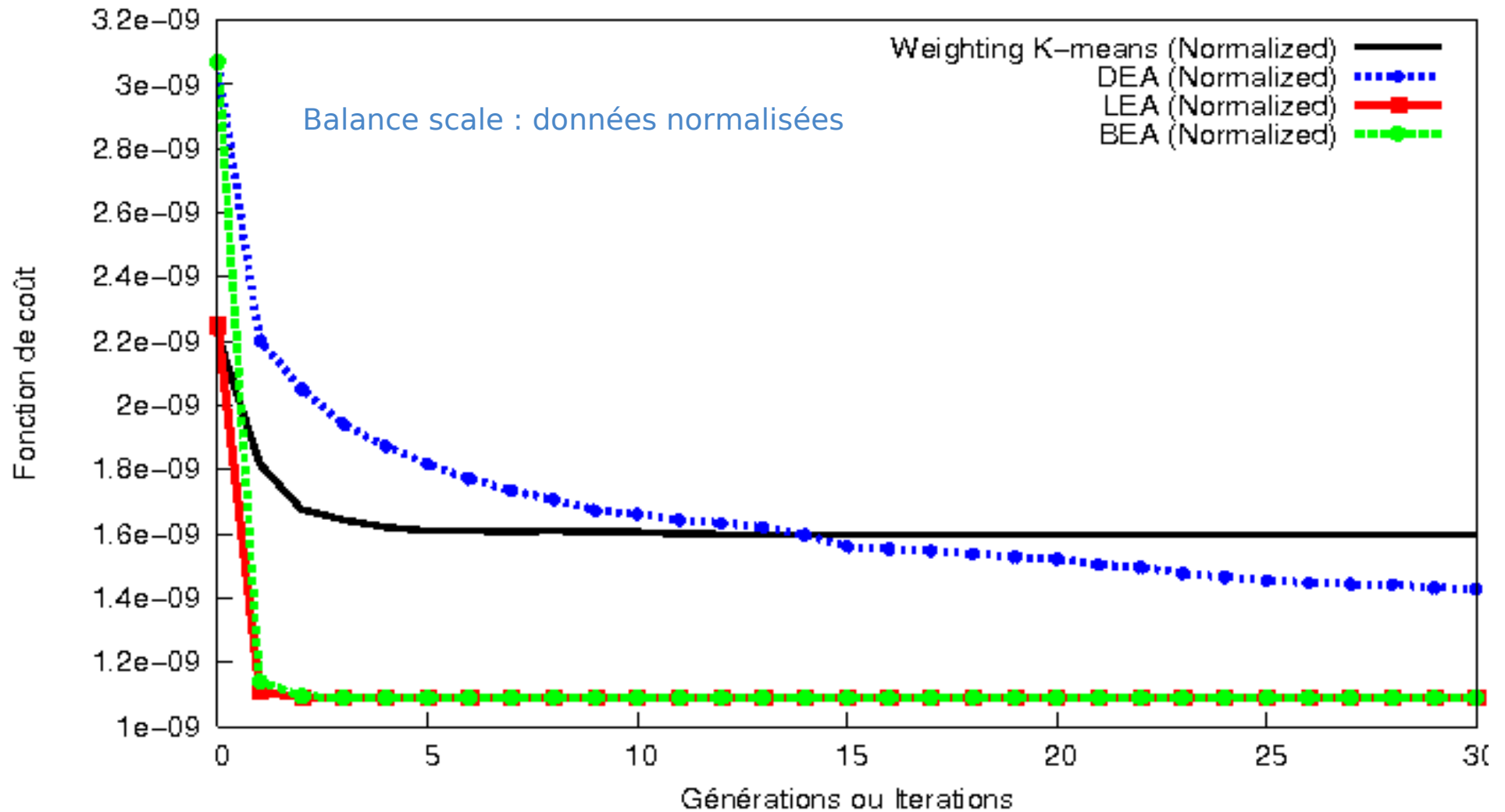
# Résultats des approches évolutives



# Résultats des approches évolutives



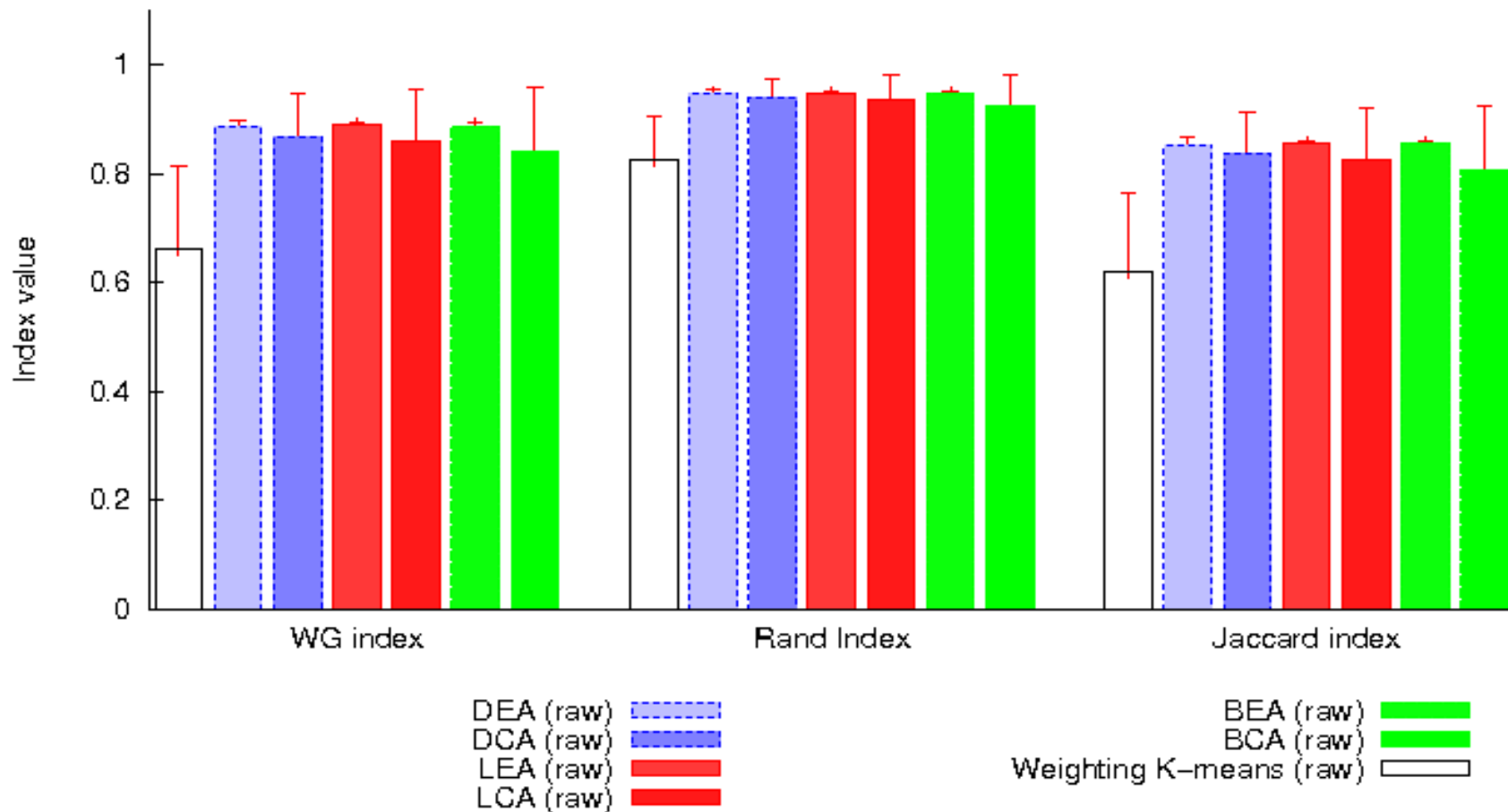
# Résultats des approches évolutives





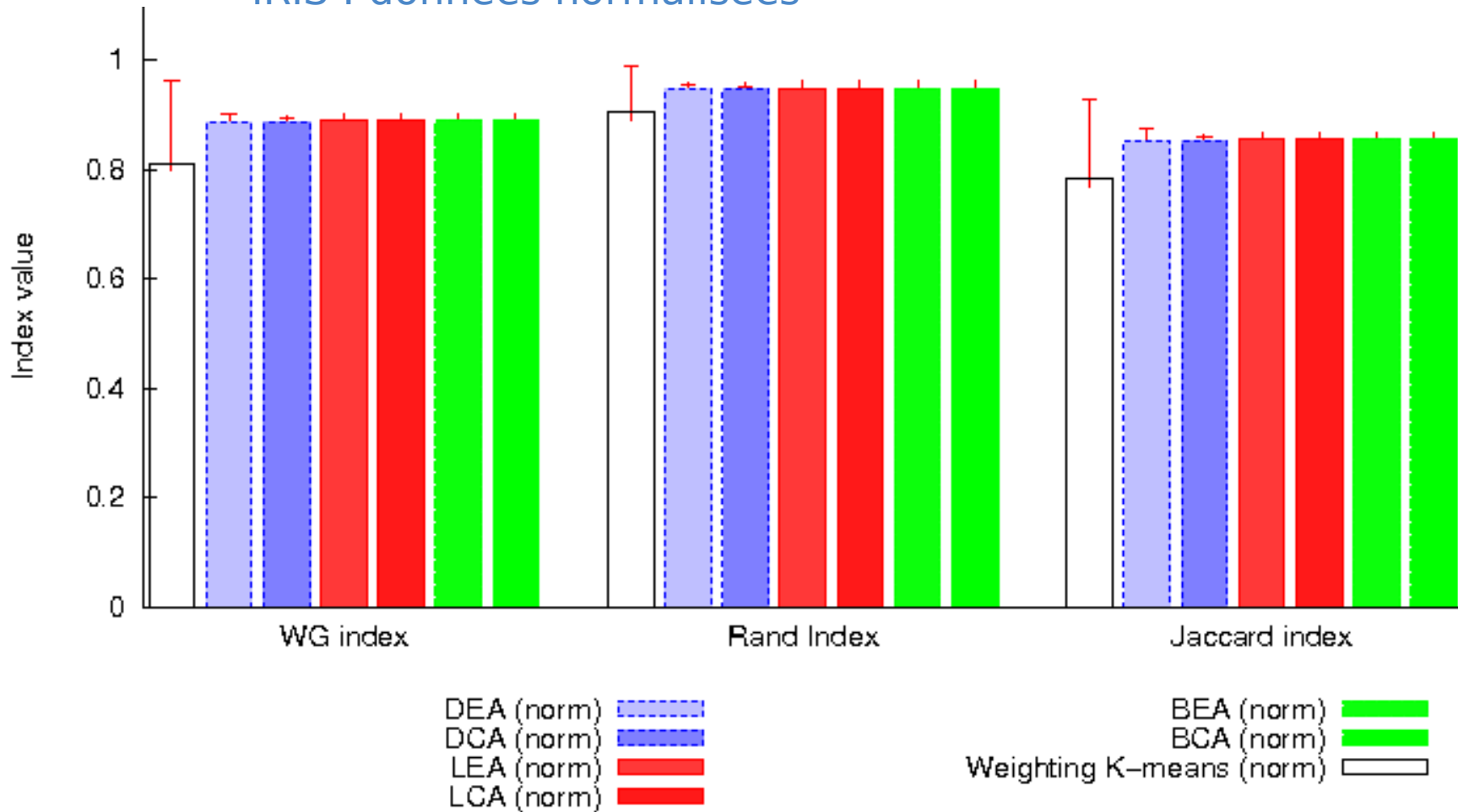
# Résultats des approches évolutives

- IRIS : données brutes



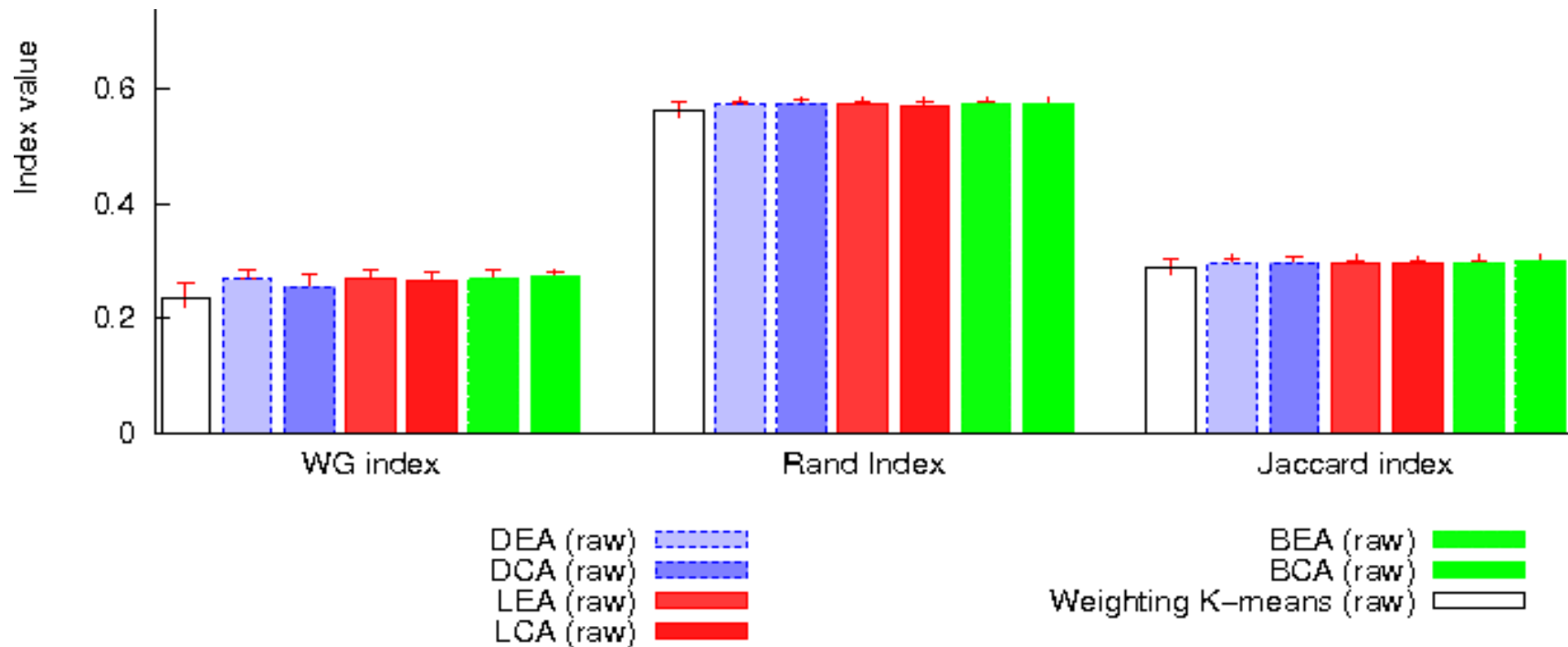
# Résultats des approches évolutives

- IRIS : données normalisées



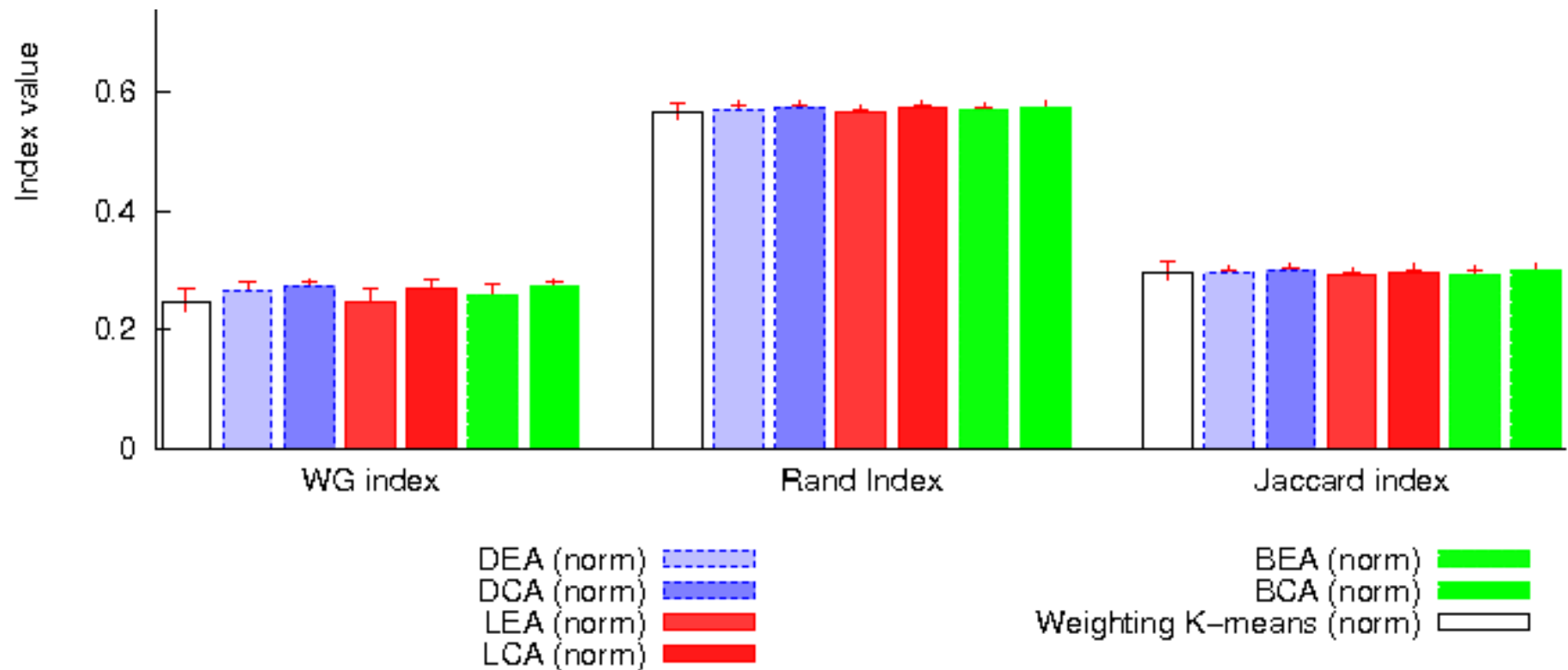
# Résultats des approches évolutives

- Balance scale : données brutes



# Résultats des approches évolutives

- Balance scale : données normalisées



# Optimisation par évolution génétique

## ➤ **Stratégies d'évolution**

- **Drawinienne**
- Lamarckienne
- Baldwinienne

## ➤ **Populations**

- évolution
- **co-évolution**

# Approches co-évolutives

## ➤ **Plusieurs populations**

- **$p$  individus par population**
- **Chaque population va, in fine, se spécialiser dans l'extraction d'un cluster**
  - **Individu = extracteur de cluster**
  - **Un chromosome = pondération locale à un cluster**

# Approches co-évolutives

## ➤ **Plusieurs populations**

- $p$  individus par population
- Chaque population se spécialise dans l'extraction d'un cluster
- **Nécessité d'un mécanisme de construction de la solution globale :**
  - **Une solution globale est construite en prenant une pondération locale dans chaque population**
    - Mais le nombre de solutions globales possibles en prenant un individu par population est :  $p^K$
  - **puis en appliquant un Kmeans pondéré**
    - Problème de l'initialisation des centres

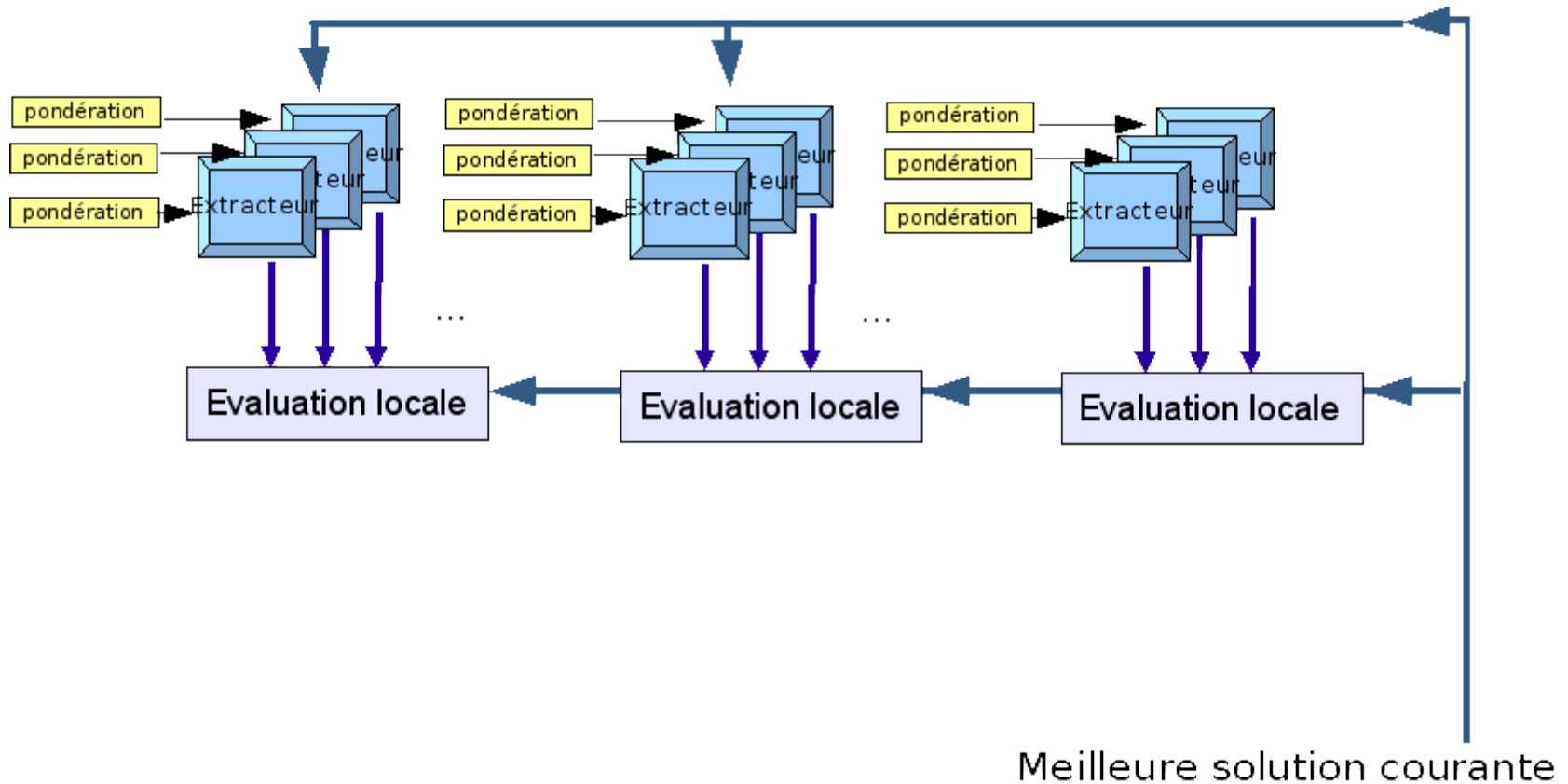
# Approches co-évolutives

## ➤ Plusieurs populations

- $p$  individus par population
  - Chaque population se spécialise dans l'extraction d'un cluster
  - Nécessité d'un mécanisme de construction de la solution globale
  - **Evaluation globale**
    - **Meilleure solution courante :**
      - K centres (partition courante)
      - K pondérations locales (pondération courante)
    - **Un individu est évalué par WCost**
      - en prenant sa pondération locale (pour son cluster) et la pondération courante pour les autres clusters
      - en initialisant par la partition courante
- ➔ **P x K évaluations**



# Approches co-évolutives



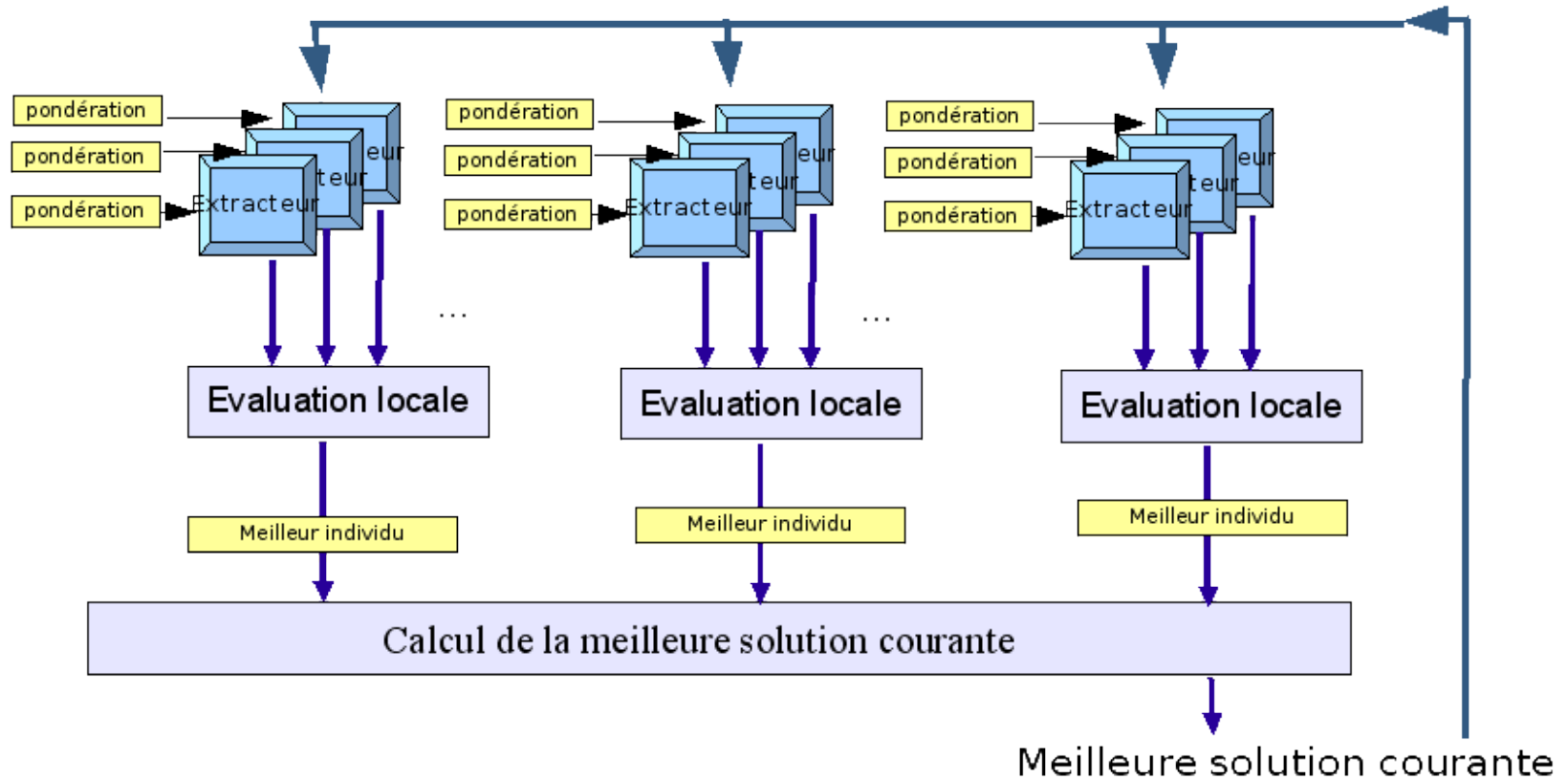
# Approches co-évolutives

## ➤ Plusieurs populations

- $p$  individus par population
- Chaque population se spécialise dans l'extraction d'un cluster
- Nécessité d'un mécanisme de construction de la solution globale
- **Evaluation globale**
  - Meilleure solution courante
  - Un individu est évalué par  $W_{cost}$
  - Nouvelle solution courante :
    - Sélection dans chaque population du meilleur individu
    - Evaluation de toutes les combinaisons possibles en prenant pour chacun de ces meilleurs individus :
      - soit son ancien chromosome (anti Red-Queen)
      - soit le nouveau

→  $2^K$  évaluations

# Approches co-évolutives

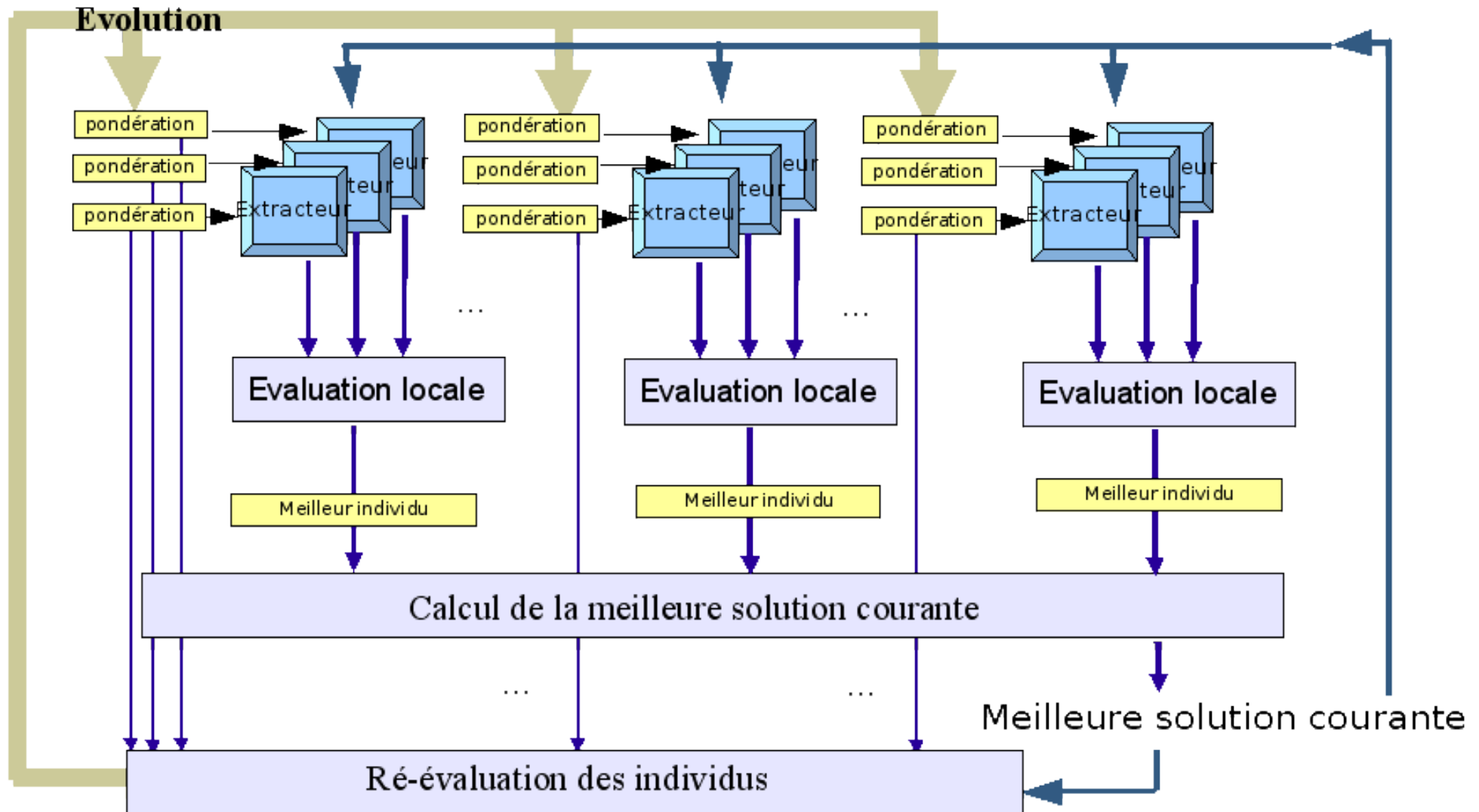


# Approches co-évolutives

## ➤ **Plusieurs populations**

- $p$  individus par population
- Chaque population se spécialise dans l'extraction d'un cluster
- Nécessité d'un mécanisme de construction de la solution globale
- **Evaluation globale**
  - Meilleure solution courante
  - Un individu est évalué par  $W_{cost}$
  - Nouvelle solution courante
  - Chaque individu est ré-évalué
    - en remplaçant dans la nouvelle pondération courante, la pondération de sa population par la sienne
    - en utilisant  $W_{cost}$  et la nouvelle meilleure partition
- **Reproduction « classique » à l'intérieur de chaque population**

# Approches co-évolutives



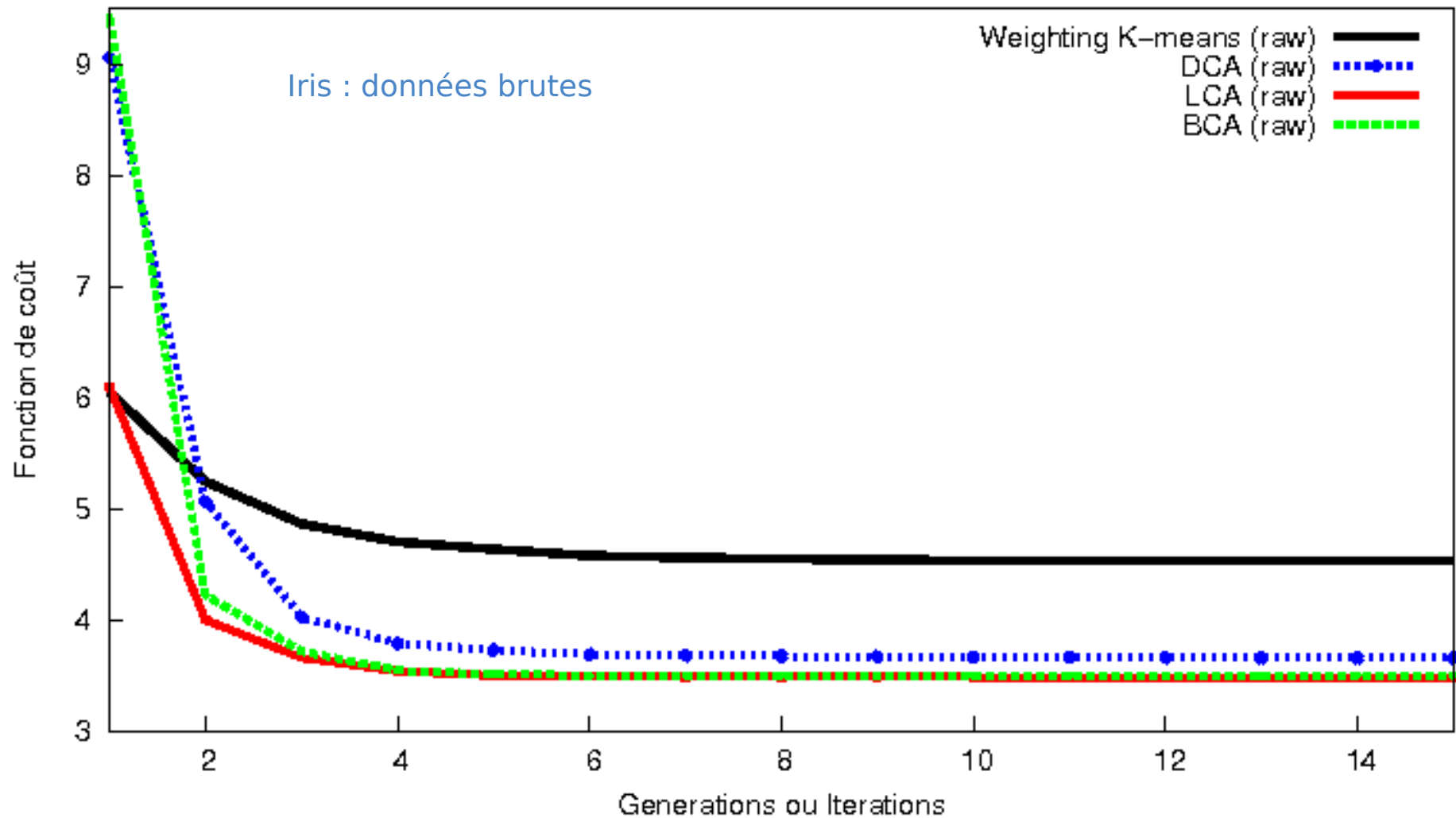
# Approches co-évolutives

## ➤ Lamarck / Baldwin

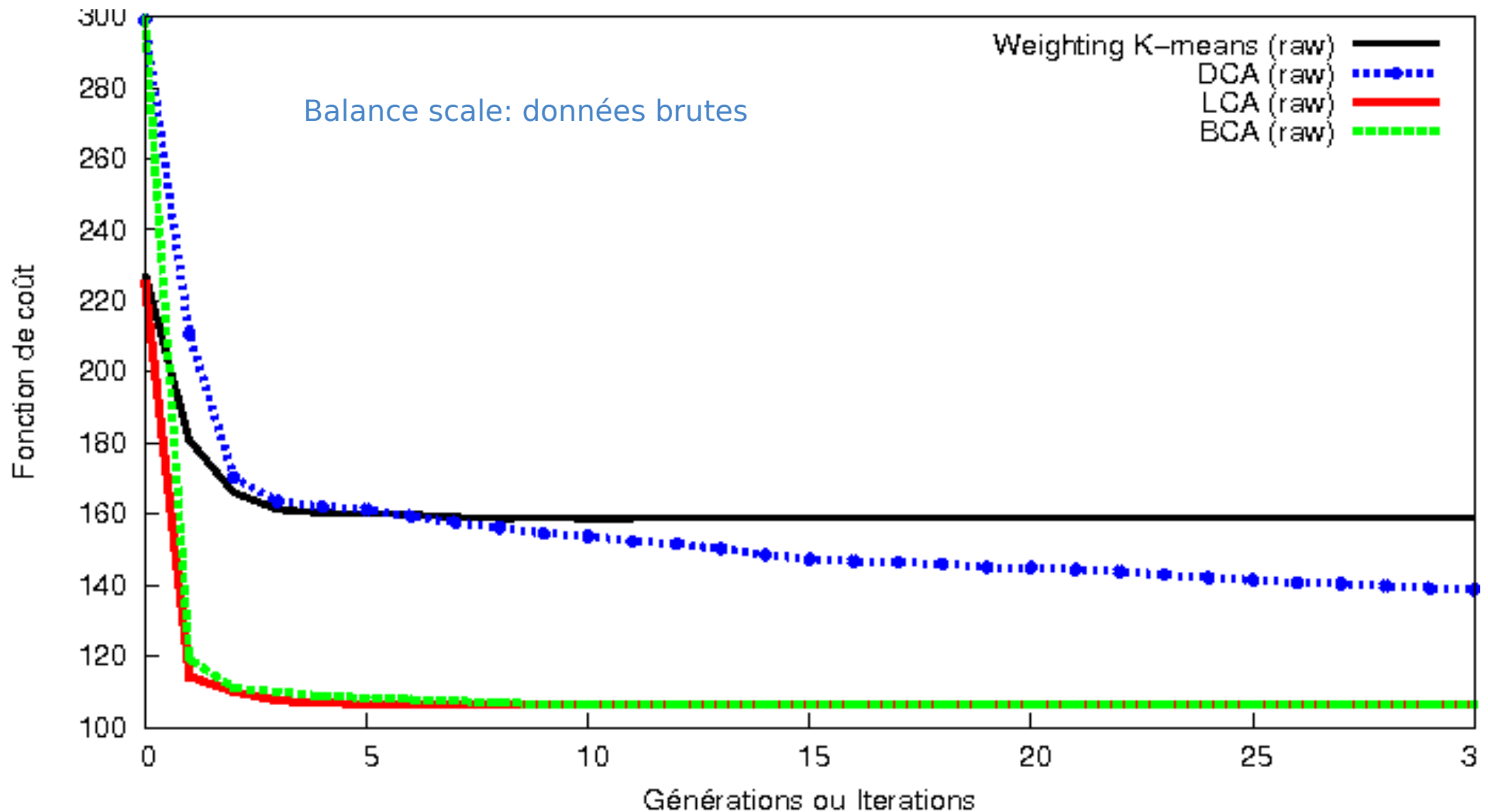
### ➤ Lifetime learning method :

- Variante de Weighting Kmeans : seule la pondération correspondant à l'individu est modifiée

# Résultats des approches co-évolutives

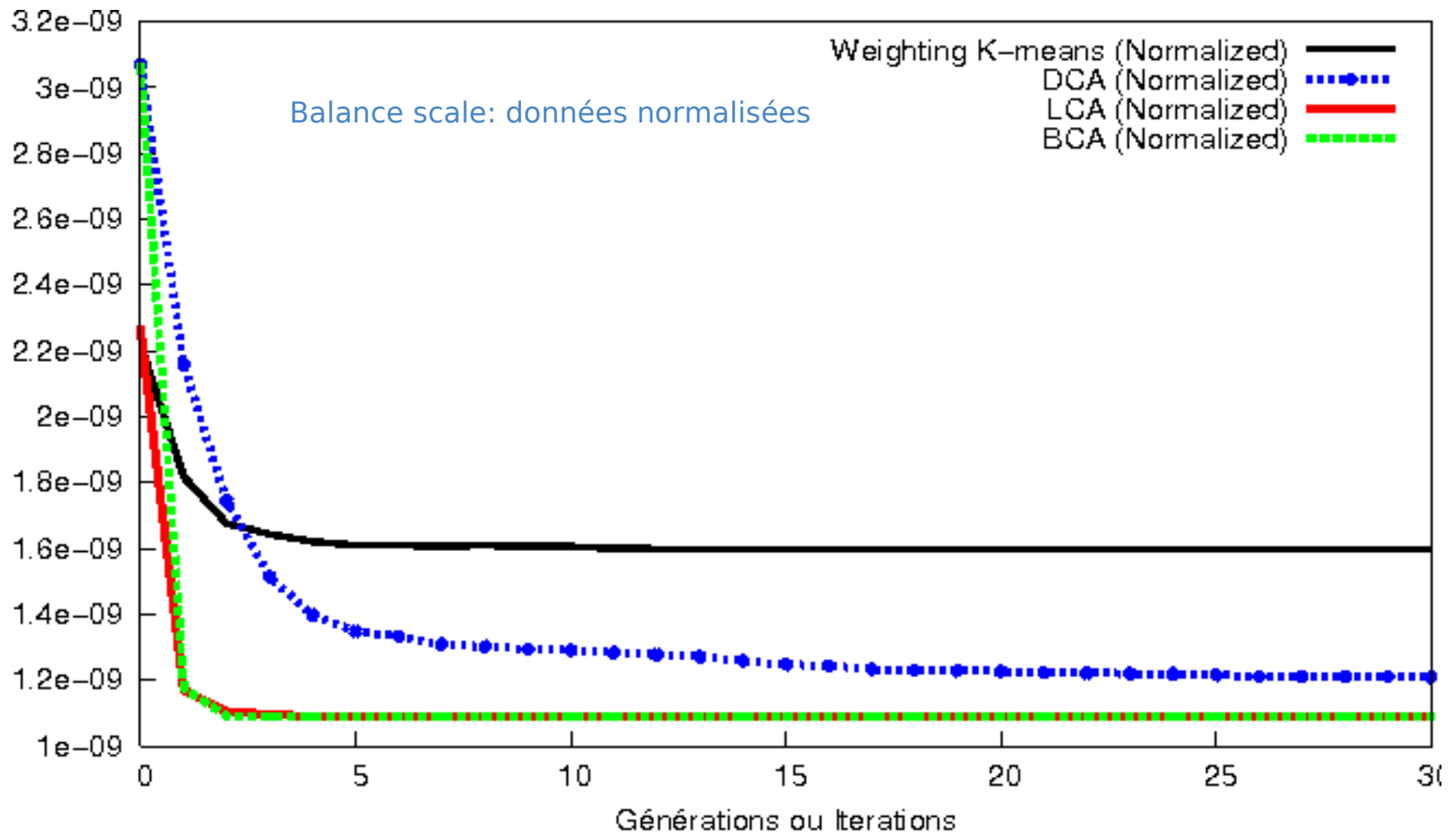


# Résultats des approches co-évolutives



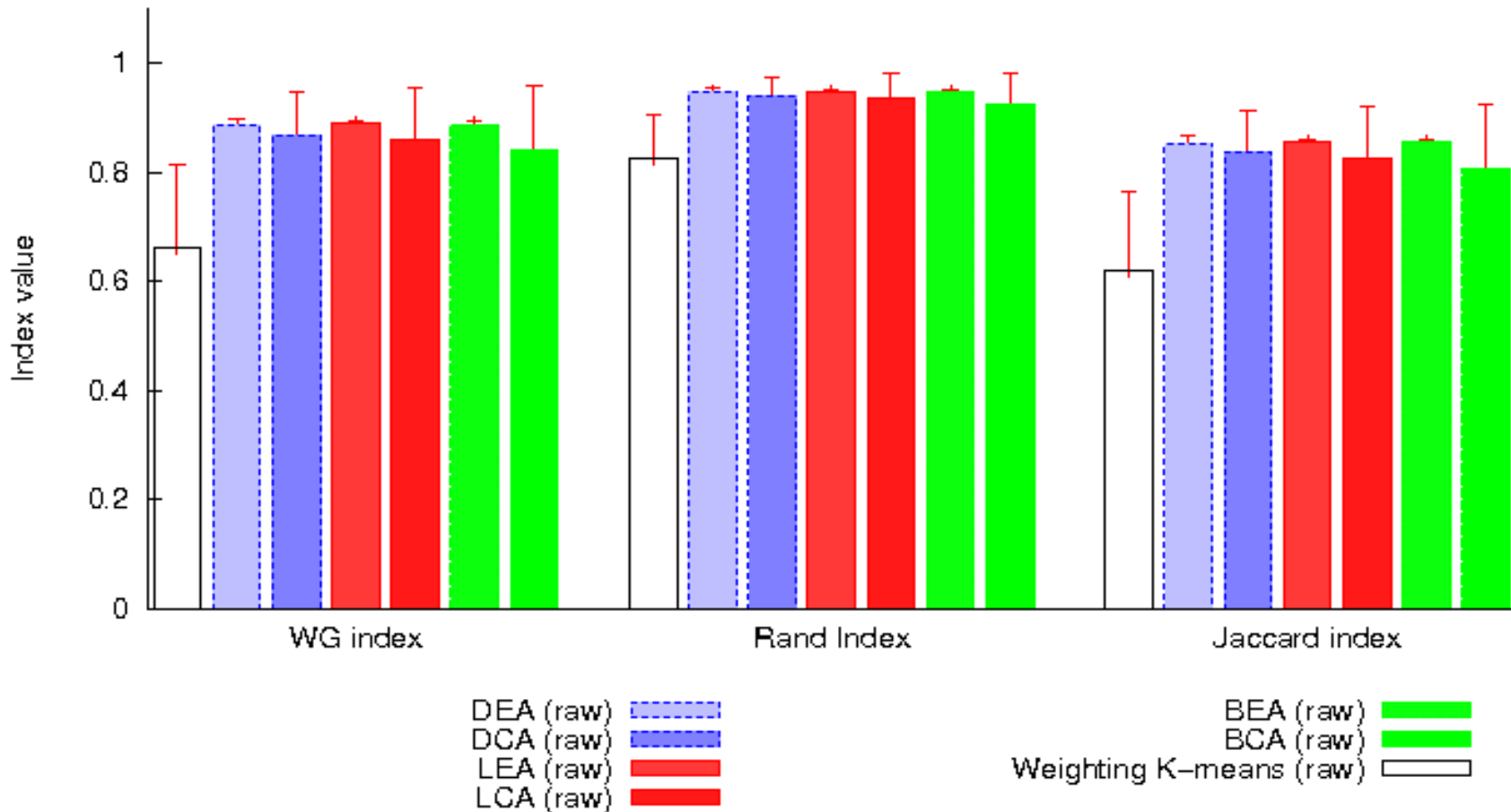


# Résultats des approches co-évolutives

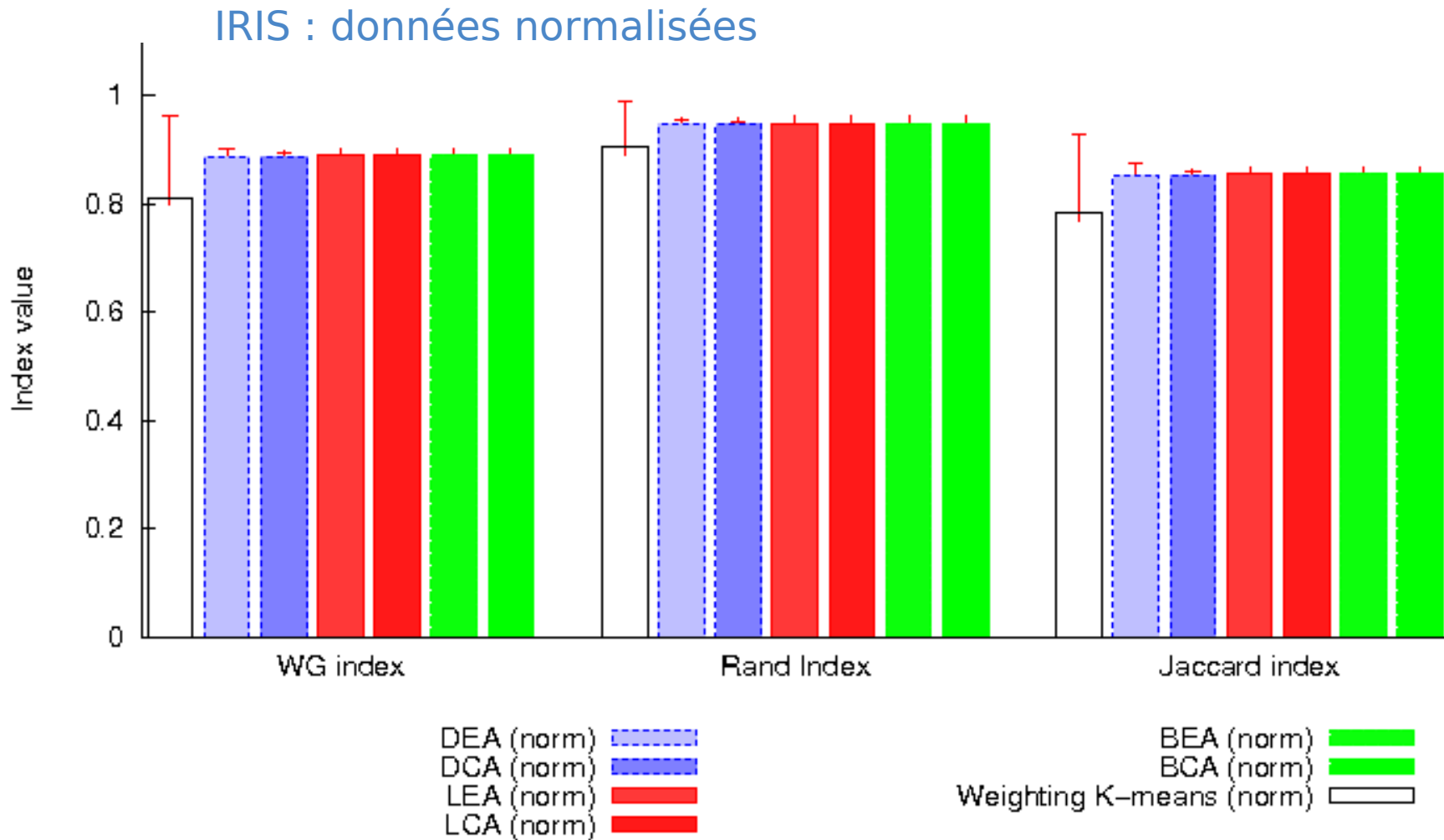


# Résultats des approches (co-)évolutives

IRIS : données brutes

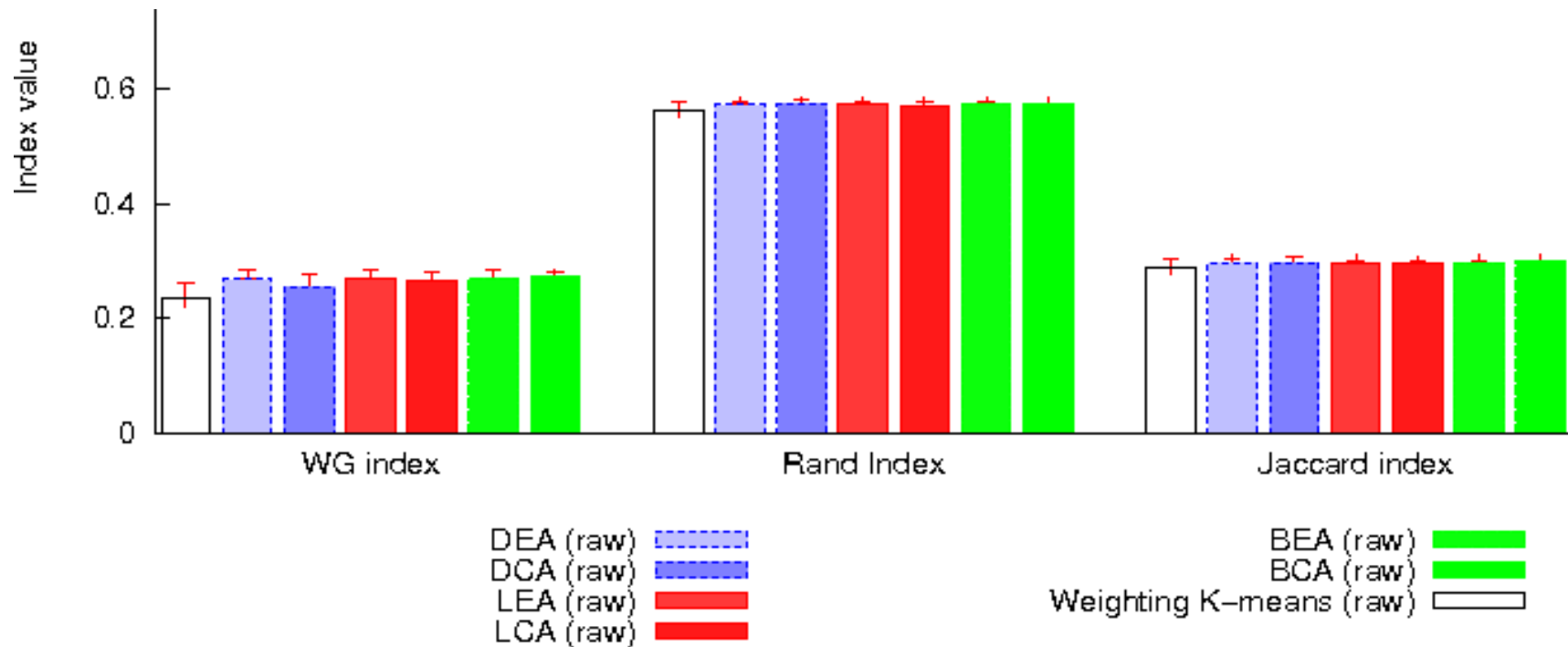


# Résultats des approches (co-)évolutives



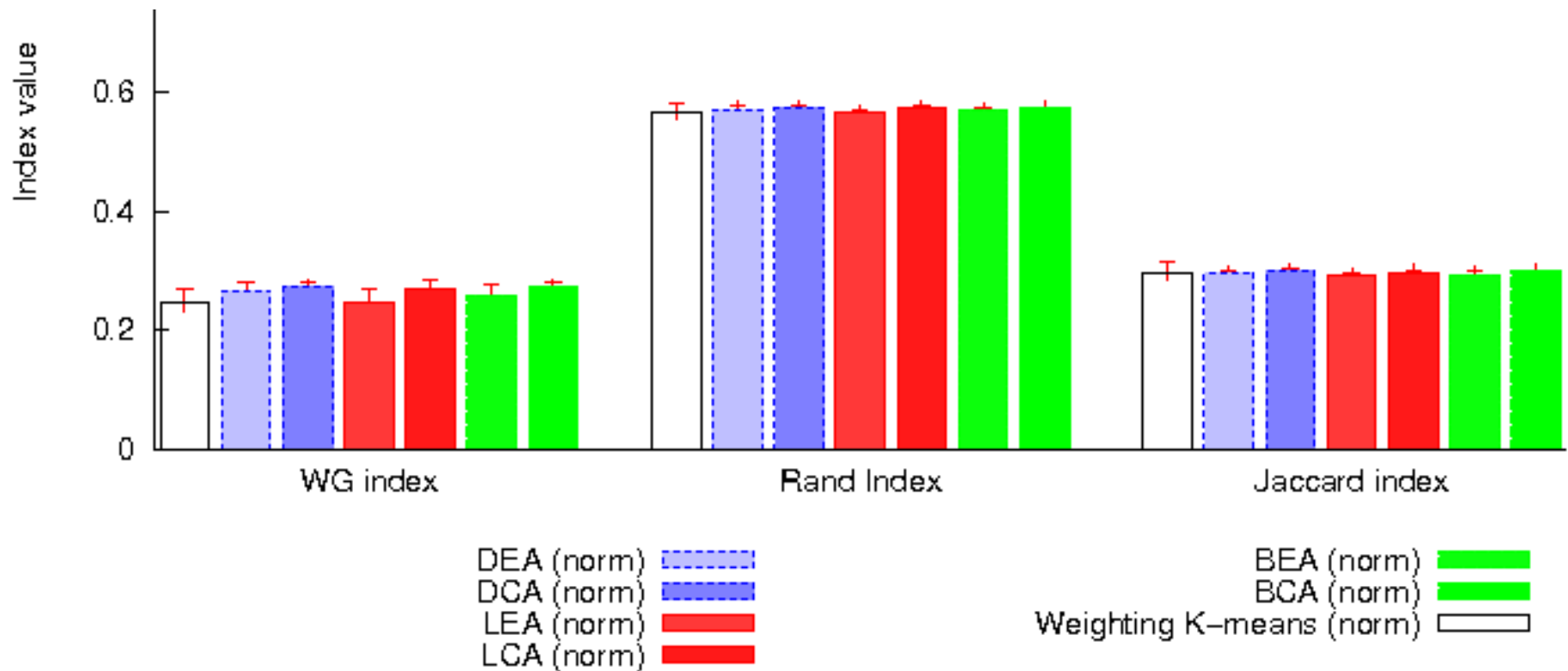
# Résultats des approches (co-)évolutives

Balance scale : données brutes

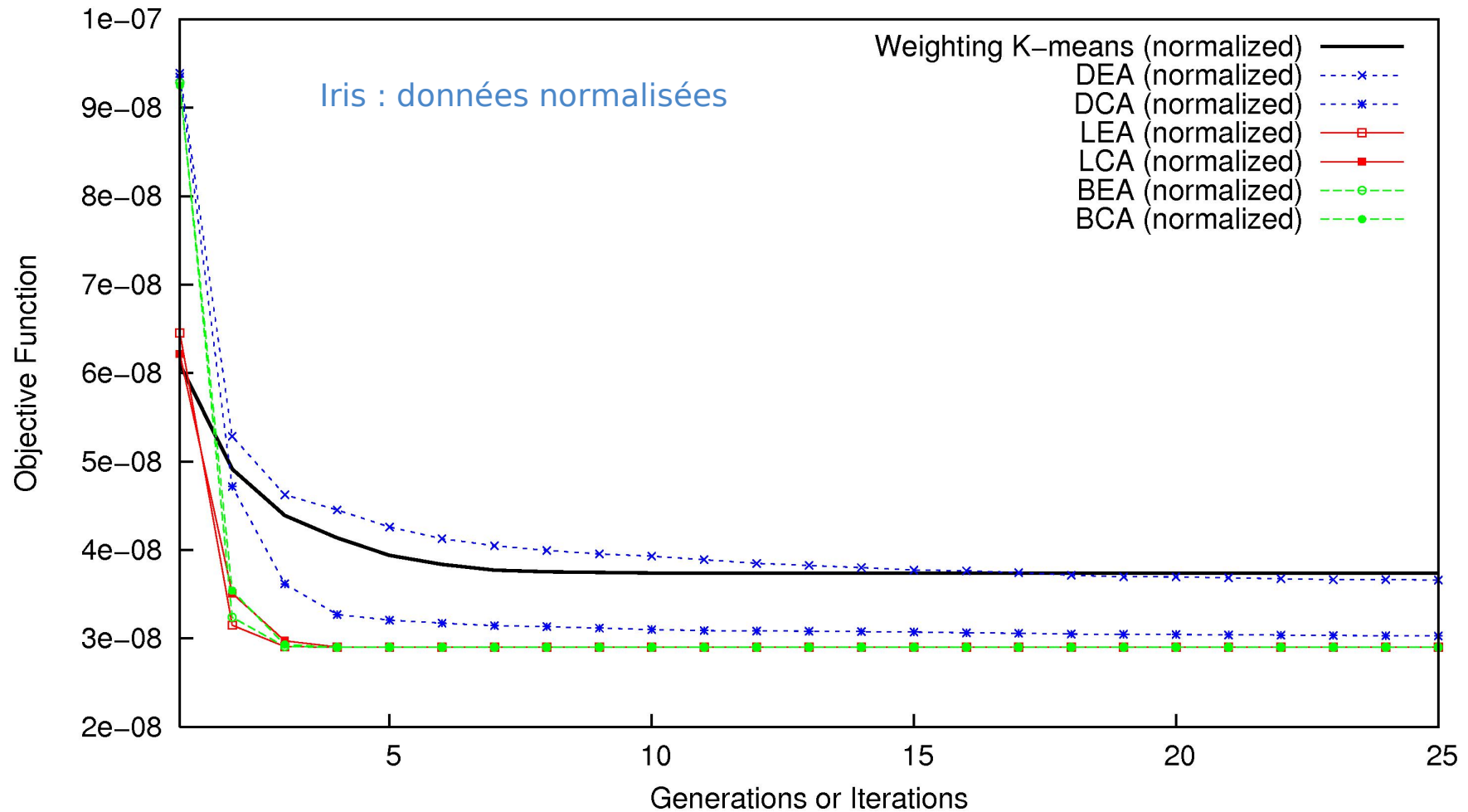


# Résultats des approches (co-)évolutives

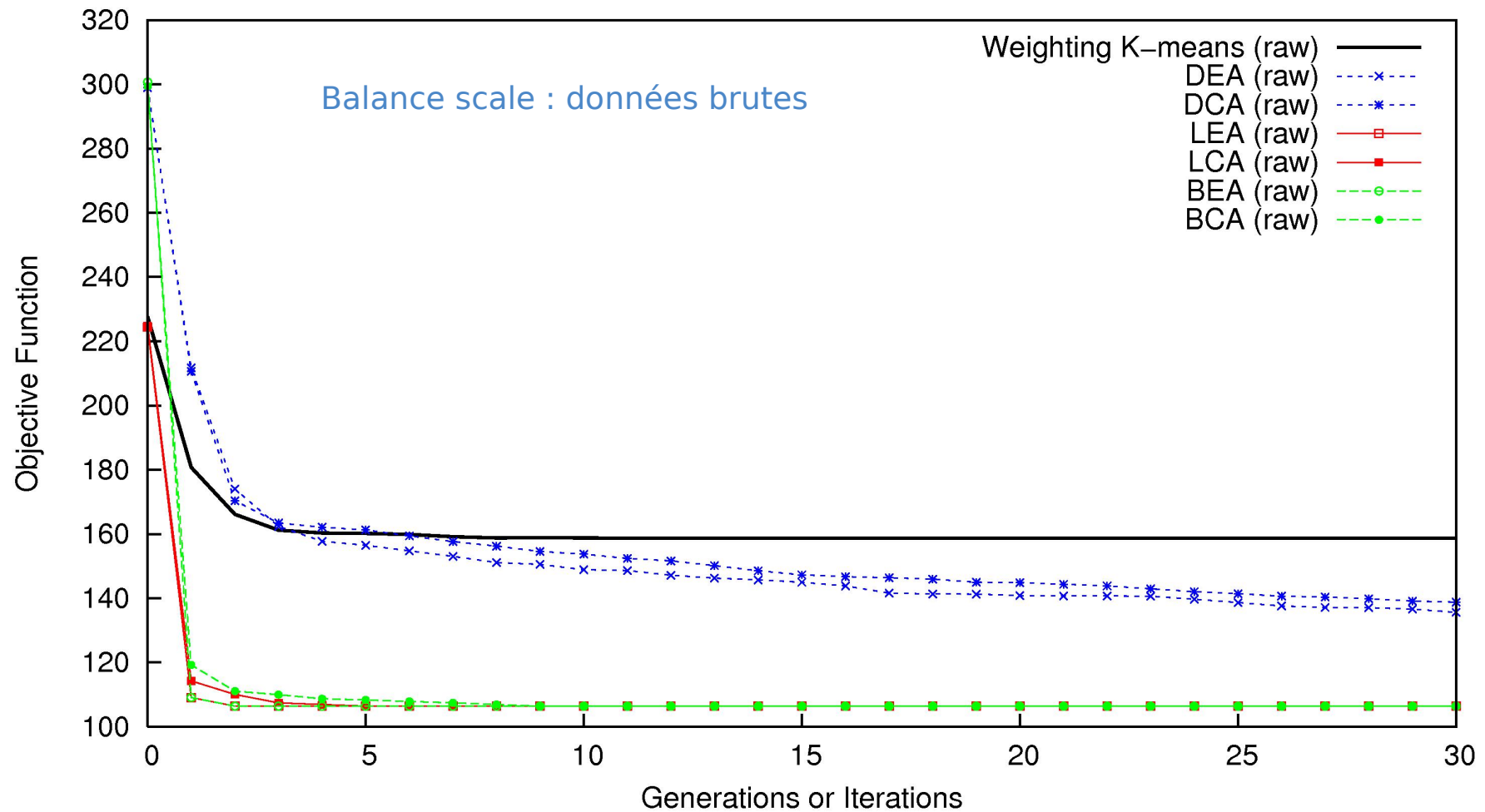
- Balance scale : données normalisées



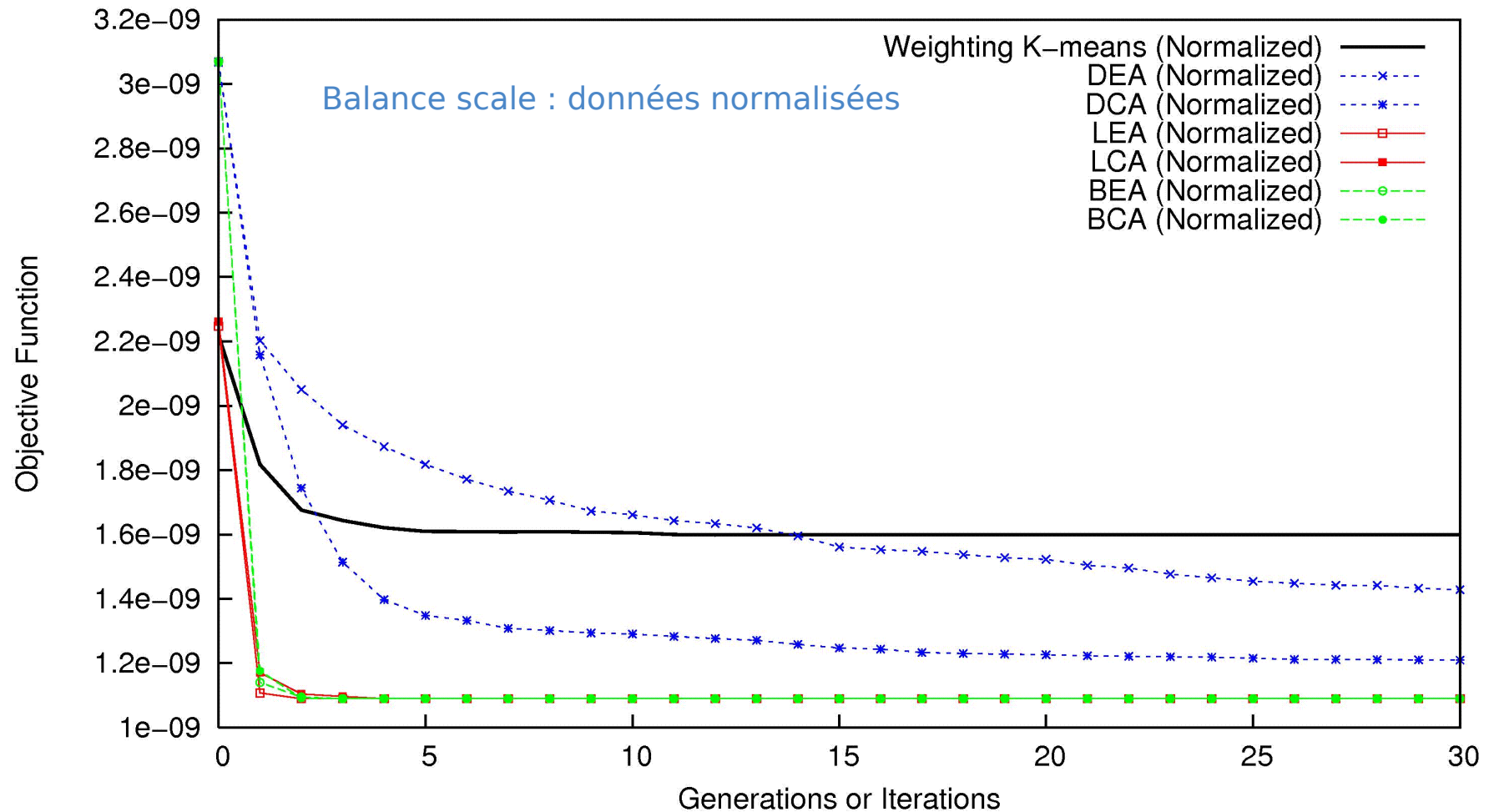
# Résultats des approches (co-)évolutives



# Résultats des approches (co-)évolutives



# Résultats des approches (co-)évolutives





## ➤ Références

**Alexandre Blansché, Pierre Gañçarski and Jerzy J. Korczak**  
***Genetic Algorithms for Feature Weighting: Evolution vs. Coevolution and Darwin vs. Lamarck***  
**Proc. of 4th Mexican International Conf. on AI,  
Monterrey, Mexique, 14-18 Novembre 2005**

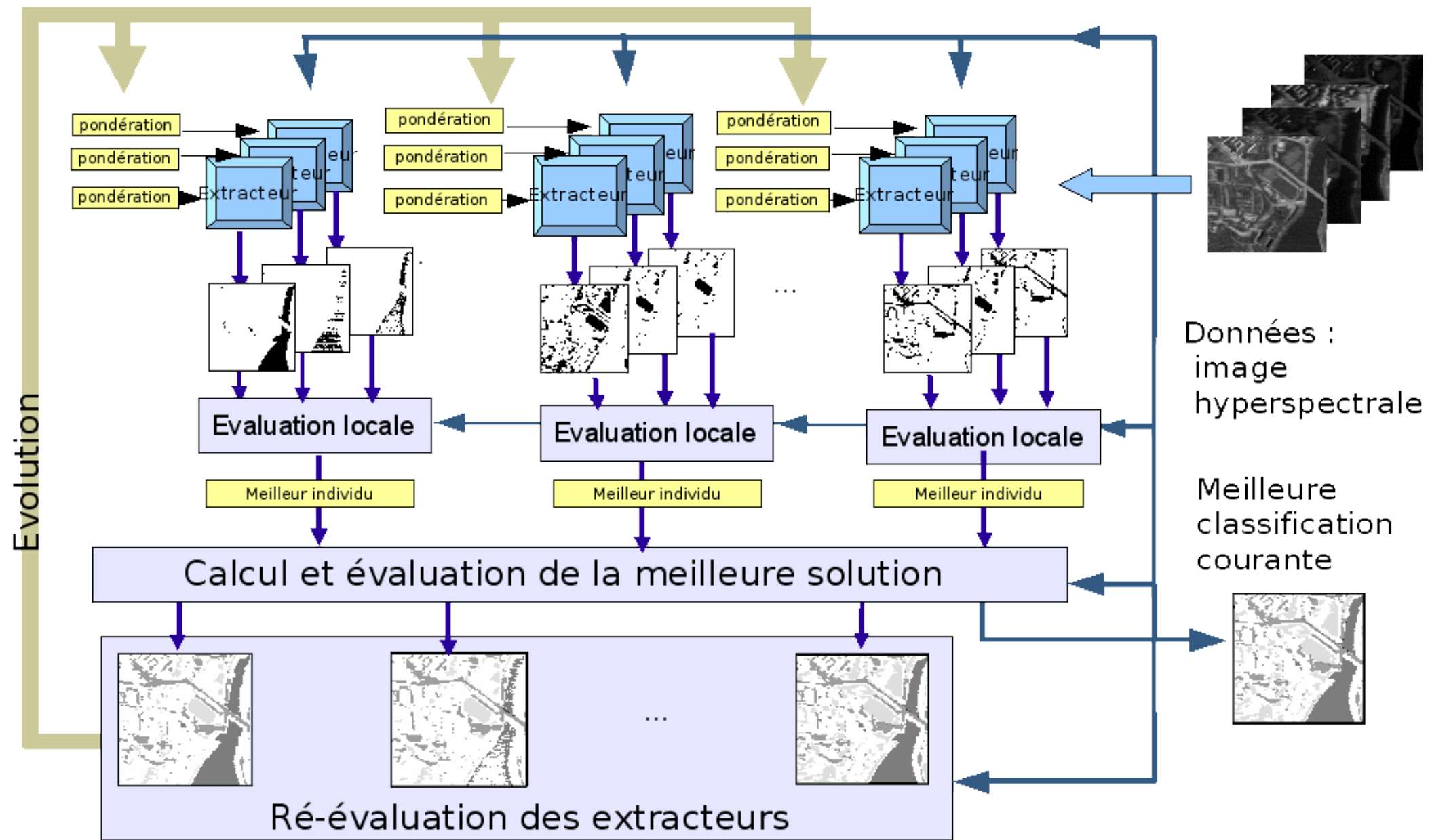
# Approches génétiques, clustering et sélection d'attributs

## ➤ **Deux approches**

- **Hard clustering et optimisation d'une fonction de coût basée sur une distance**
- **Soft clustering et optimisation basée sur la qualité de partitionnement**

# Optimisation basée sur la qualité de partitionnement

- **Méthode darwienne par co-évolution**
- **Chaque extracteur fournit un cluster en extension -> liste d'objets**
- **Un soft clustering partiel est construit :**
  - **Un objet peut appartenir à 0, 1 ou plusieurs clusters**

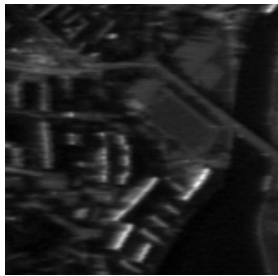


# Optimisation basée sur la qualité de partitionnement

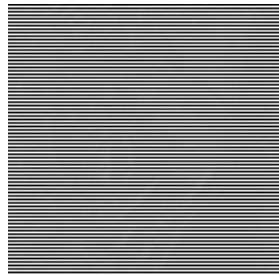
- **Un soft clustering partiel est construit :**
  - Un objet peut appartenir à 0, 1 ou plusieurs clusters
- ➔ **Les critères habituels ne sont plus satisfaisants**
- ➔ **Utilisation d'un critère basé sur le degré de partitionnement ne nécessitant pas (directement) de mesure de distance**
  - **Prise en compte des recouvrements de classes**
  - **Prise en compte du nombre d'objets non classés**
  - **(Eventuellement) prise en compte de la qualité des clusters proposés**

# Résultats sur des images de télédétection

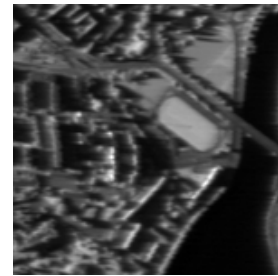
## Images hyperspectrales



Bande  
1



Bande  
14



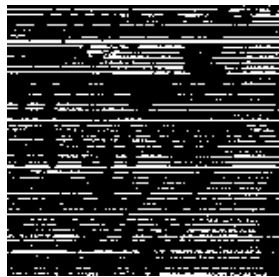
Bande  
15



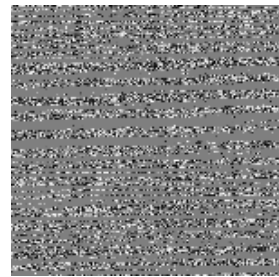
Bande  
26



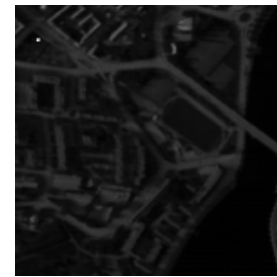
Bande  
41



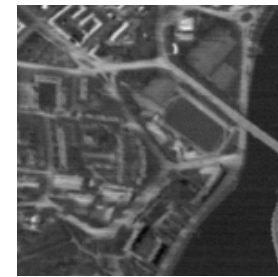
Bande  
43



Bande  
64

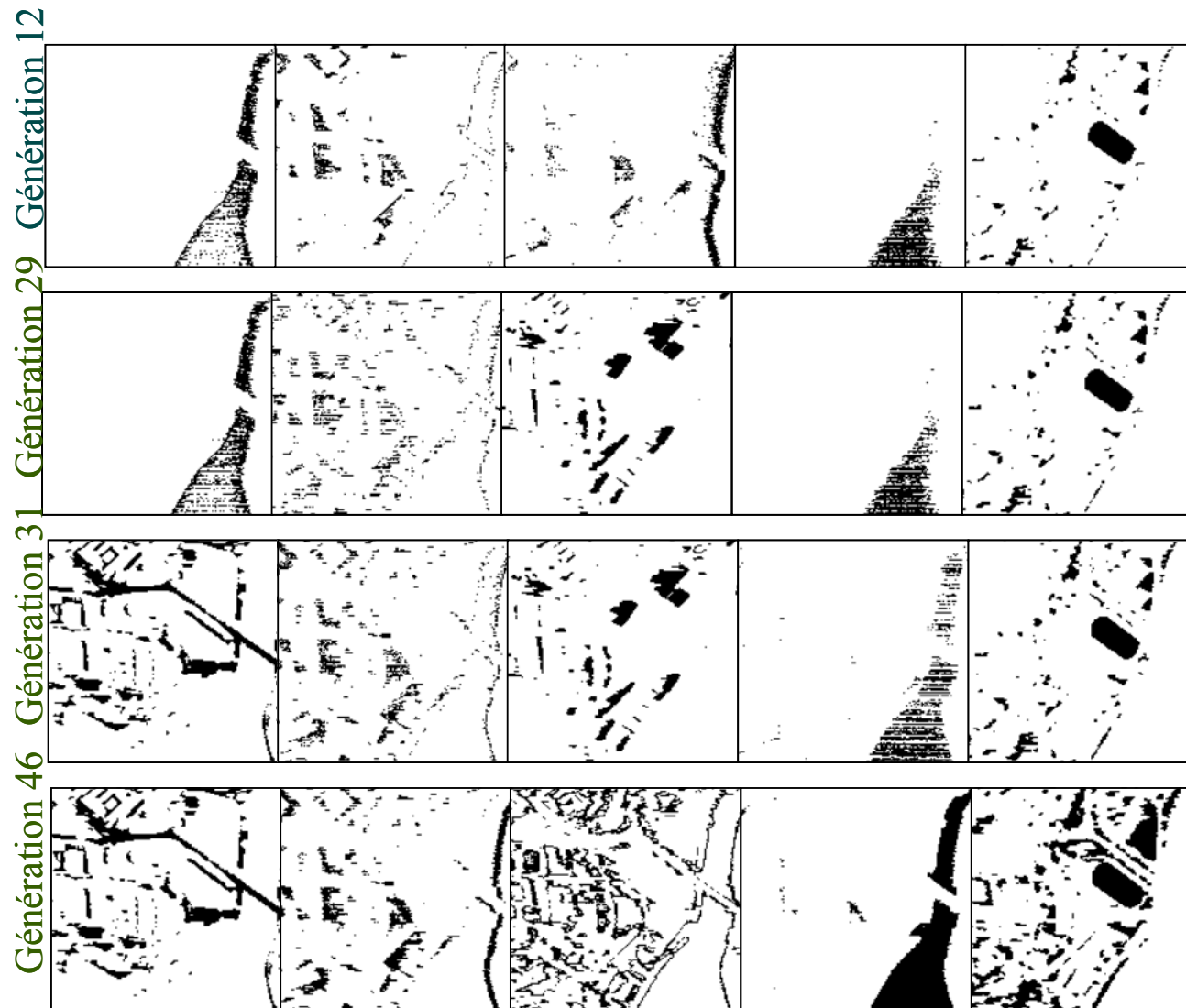


Bande  
73

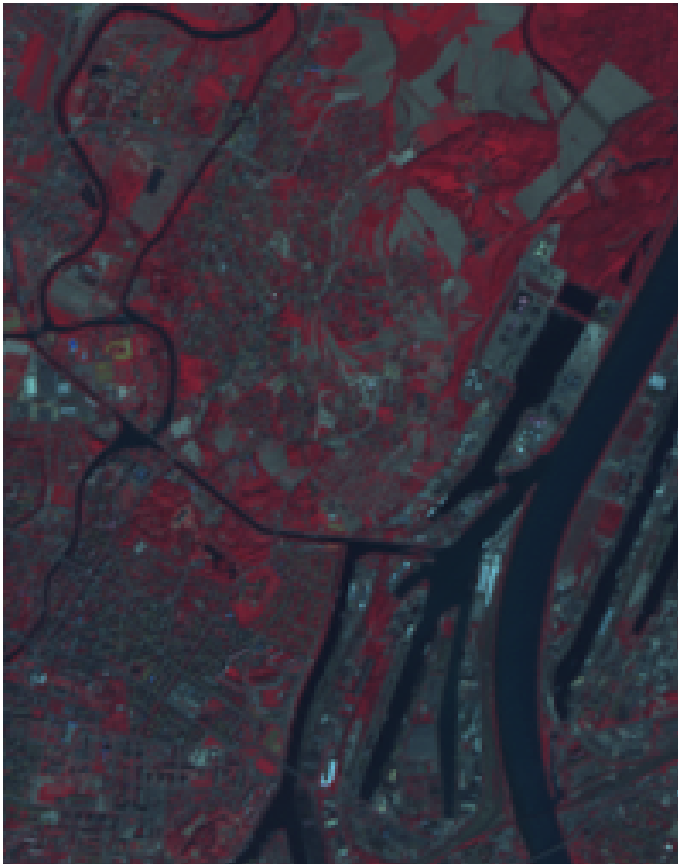


Bande  
79

# Résultats sur des images de télédétection



# Résultats sur des images de télédétection



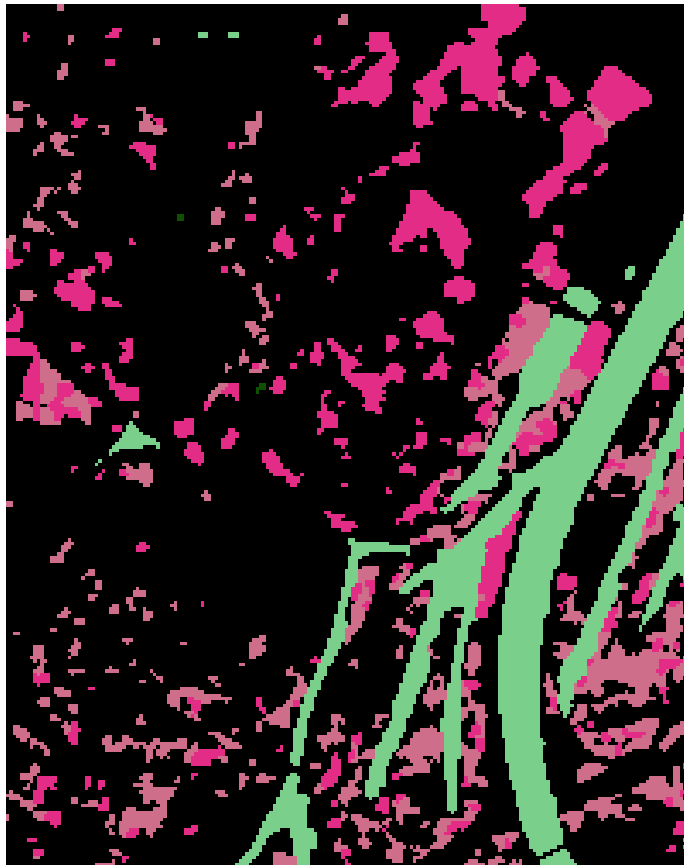
Classification par multiclassifieur sur  
l'image SPOT de la Robertsau  
6 classes demandées  
150 individus par population



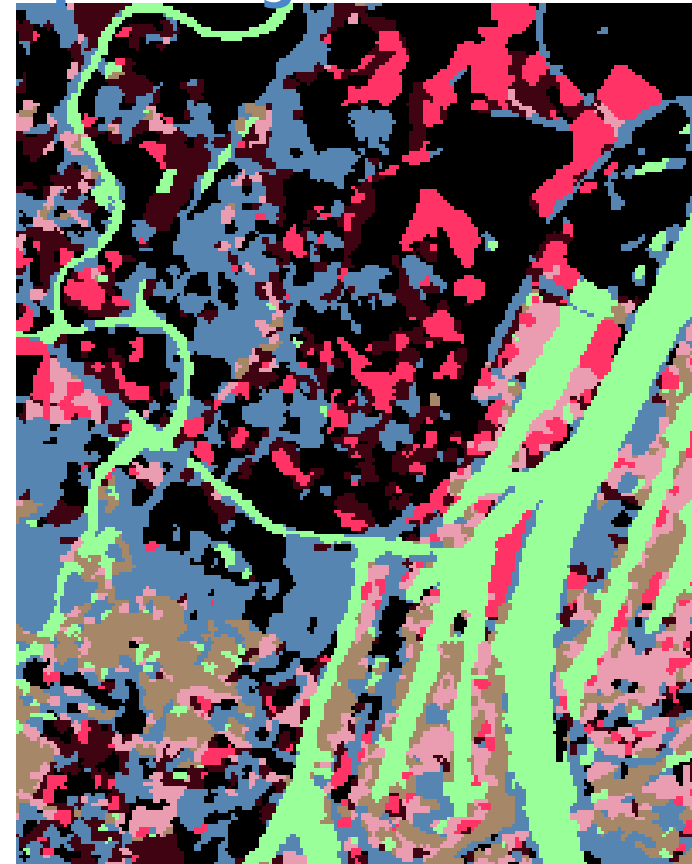
# Résultats sur des images de télédétection

## ➤ Classifications résultat

Initial



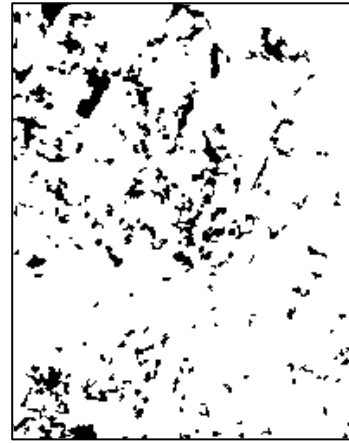
Après 6 générations



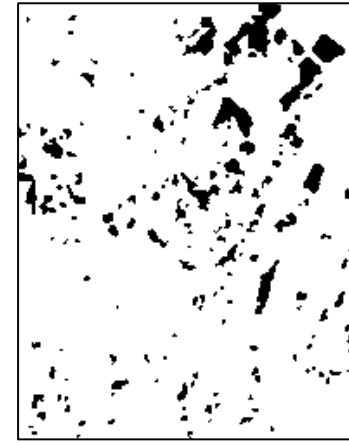
# Résultats sur des images de télédétection



classifieur 1



classifieur 2



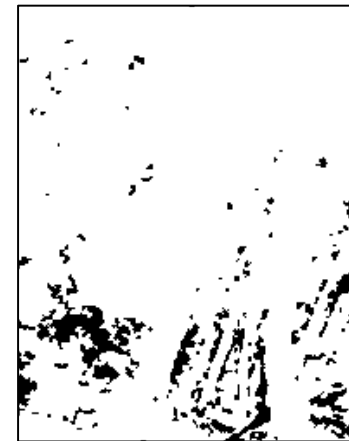
classifieur 3



classifieur 4

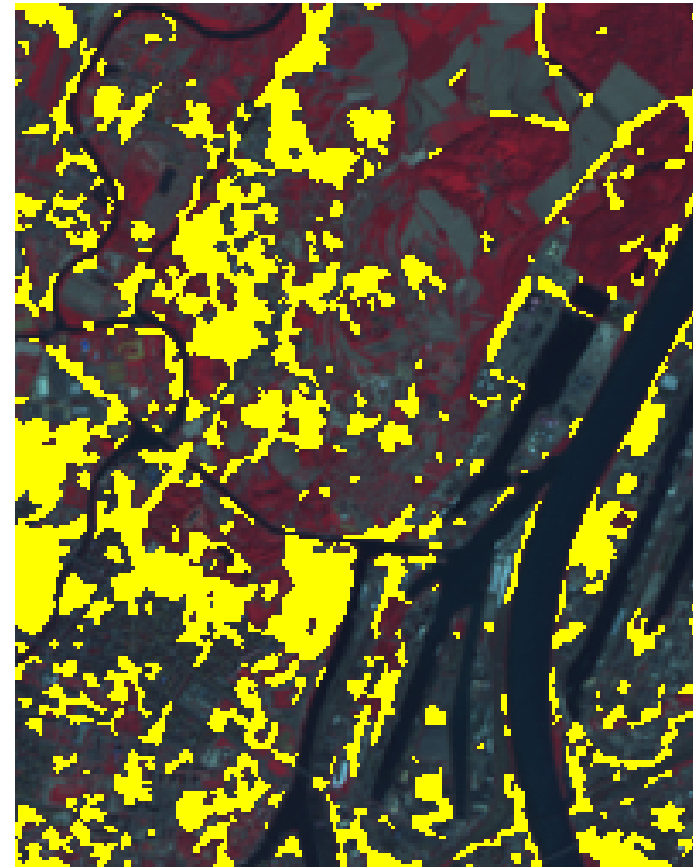


classifieur 5



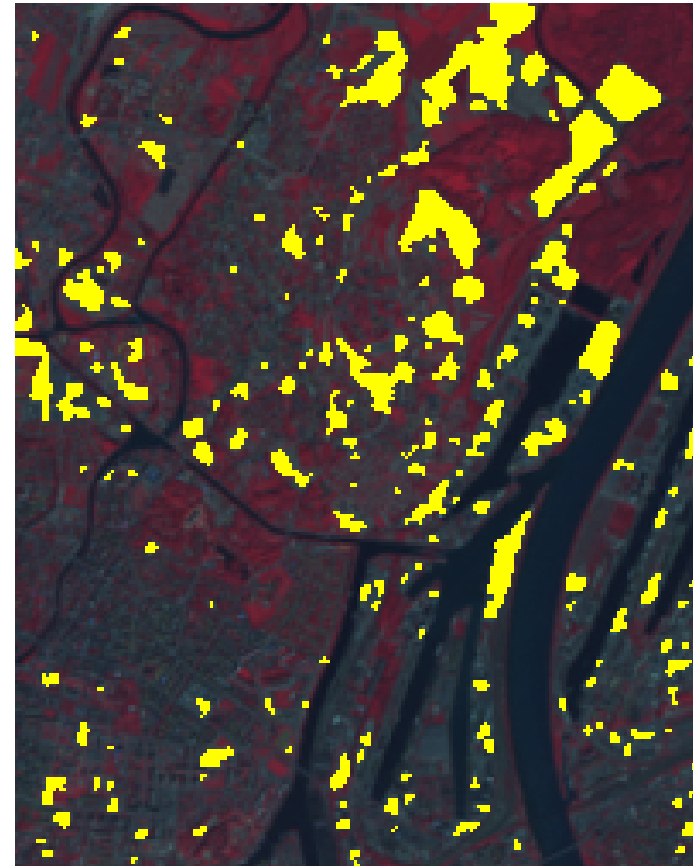
classifieur 6

# Résultats sur des images de télédétection



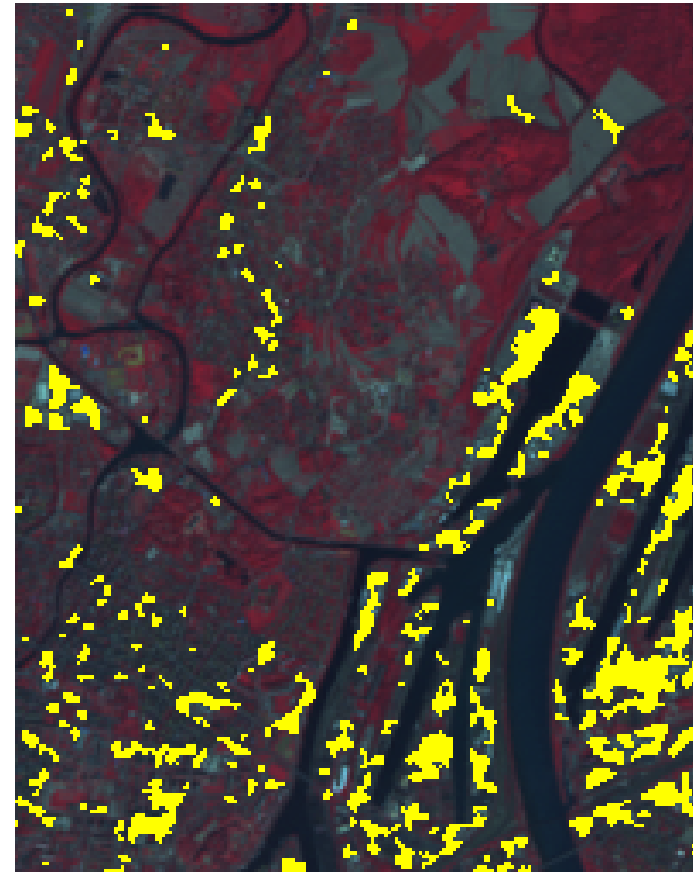
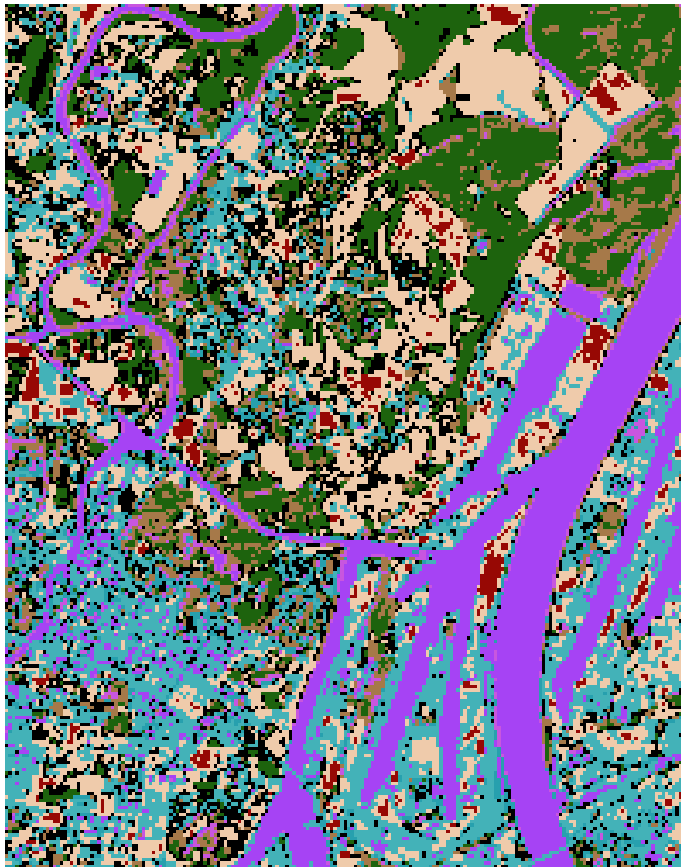
Classifieur 1  
Quartiers pavillonnaires

# Résultats sur des images de télédétection



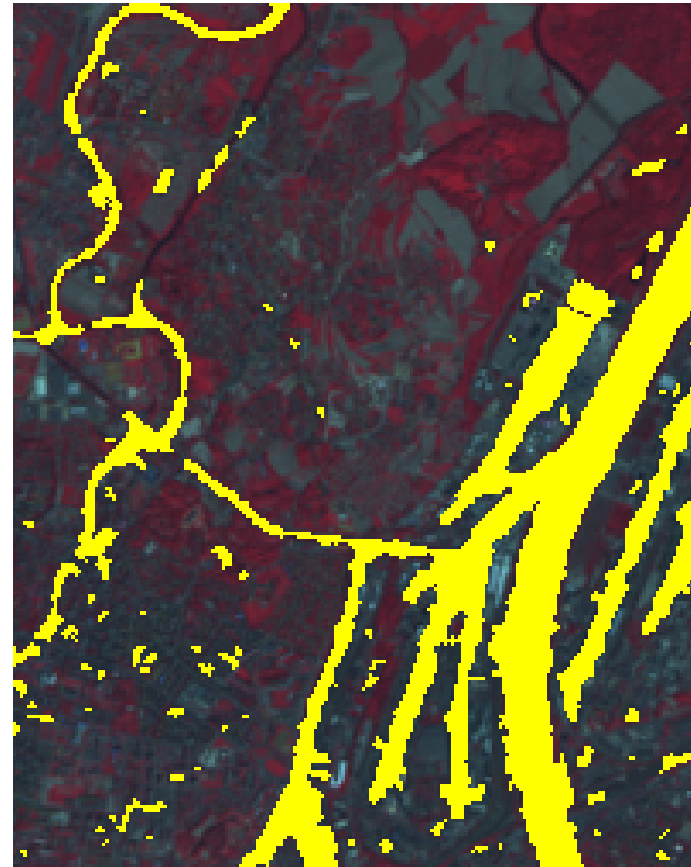
Classifieur 3  
Champs

# Résultats sur des images de télédétection



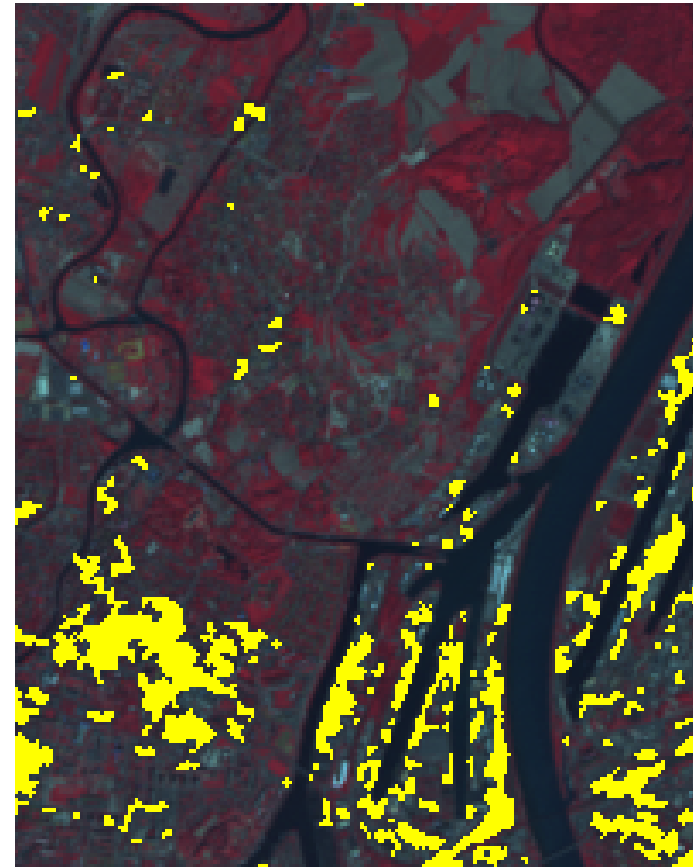
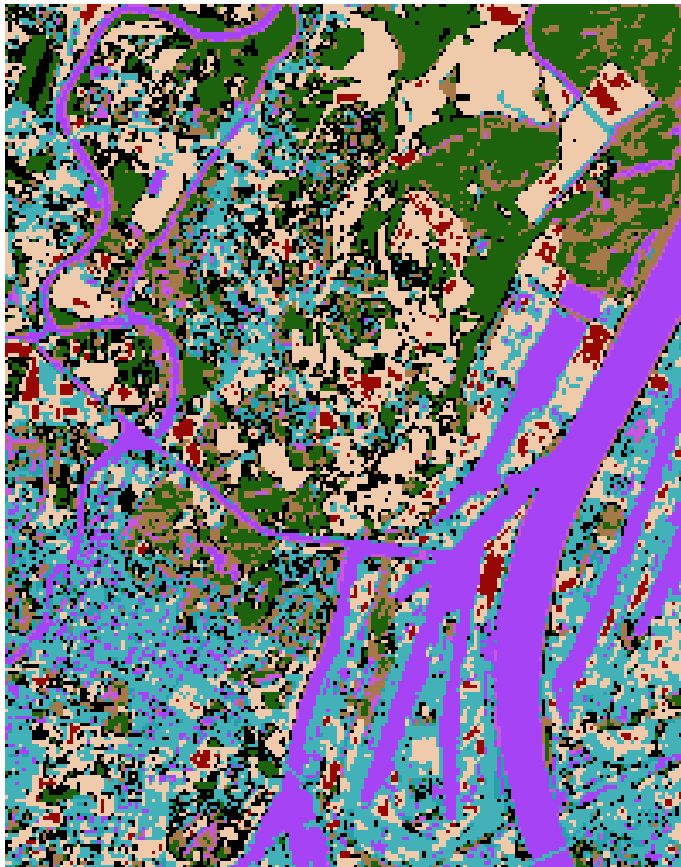
Classifieur 4  
Zones industrielles

# Résultats sur des images de télédétection



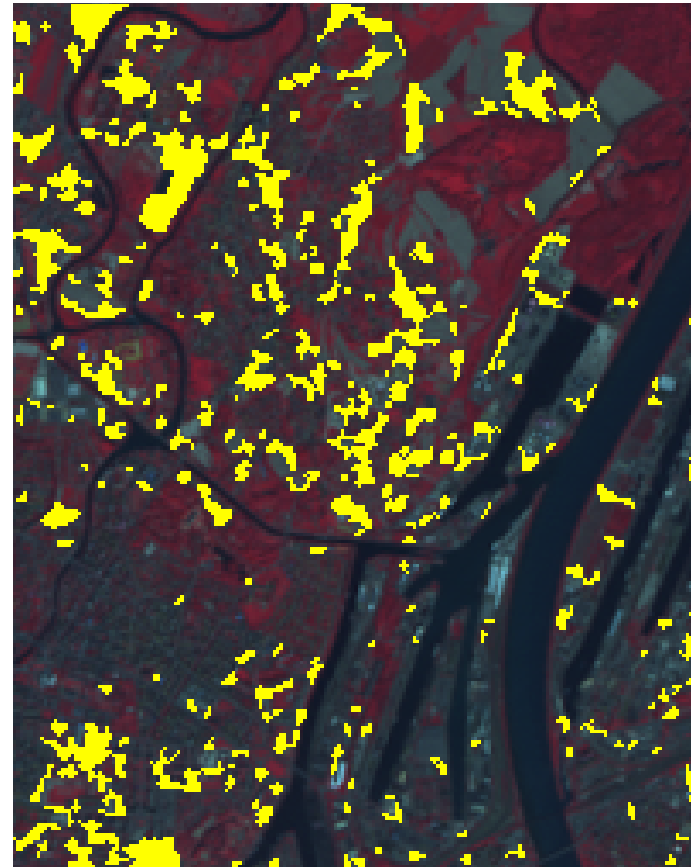
Classifieur 5  
Fleuve et voies d'eau

# Résultats sur des images de télédétection



Classifieur 6  
Bâti dense

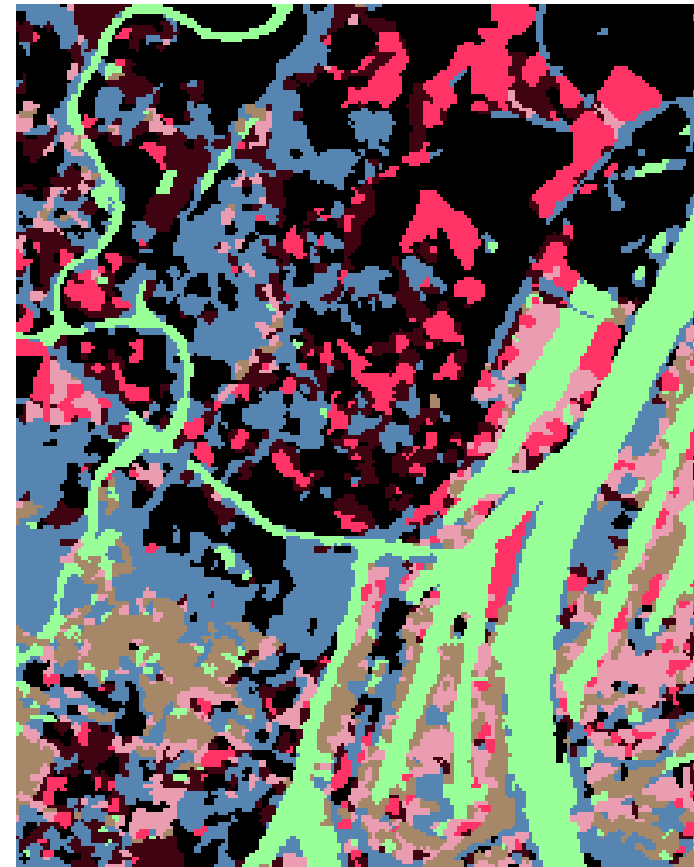
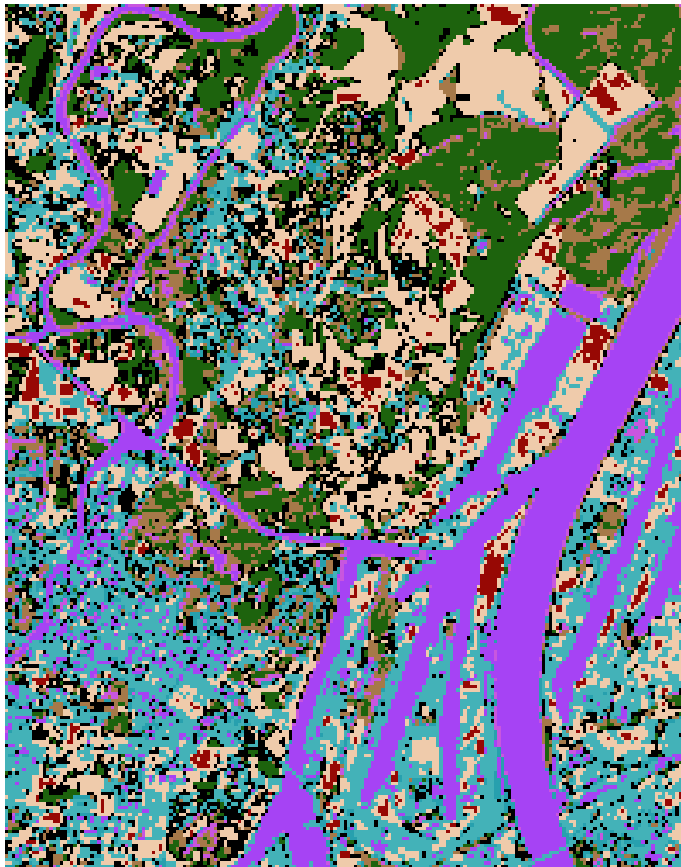
# Résultats sur des images de télédétection



classifieur 2  
classe sans signification  
« sémantique »



# Résultats sur des images de télédétection



Objets non classés  
Végétation arborée

# Résultats sur des images de télédétection



Image haute définition ( $\approx 1\text{m}$ )  
de l'Esplanade  
 $400 \times 400$  pixels,  
5 classes demandées

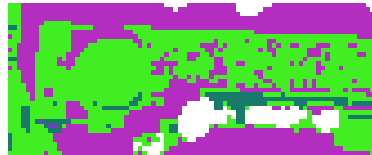
Mon bureau enseignant

# Résultats sur des images de télédétection

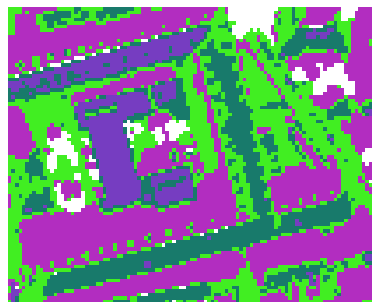
image  
radiométrique



classification  
radiométrique



classification  
des régions



# Conclusion

- **Les méthodes génétiques semblent plus efficaces que les optimisations par hill-climbing**
- **De nombreuses perspectives :**
  - **Plusieurs stratégies d'extraction des classes simultanément :**
    - Dans le domaine de la classification : cobweb, neurone, ...
    - Dans le domaine de l'expert : méthodes propres à son domaine (extracteur de route, de bâtiments, ...)
  - **Plusieurs sources de données**
    - Images à des résolutions différentes ou à des moments différents
    - Images de natures différentes

# Approches génétiques, clustering et sélection d'attributs

## ➤ **Références**

**Alexandre Blansché, Pierre Gançarski and Jerzy J. Korczak**  
***Genetic Algorithms for Feature Weighting: Evolution vs. Coevolution and Darwin vs. Lamarck***

**Proc. of 4th Mexican International Conf. on AI,  
Monterrey, Mexique, 14-18 Novembre 2005**

**Alexandre Blansché, Pierre Gançarski and Jerzy J. Korczak**  
***MACLAW: a Modular Approach for Clustering with Local Attribute Weighting***

**in Pattern Recognition Letter, Special Issue on  
Evolutionary Computer Vision and Image Understanding  
(in press)**

# Approches génétiques, clustering et sélection d'attributs

**C'EST FINI,**

**vous pouvez reprendre  
une activité normale...**