

# MIXMOD<sup>®</sup>

## Un ensemble logiciel de classification des données par modèles de mélange

Florent Langrognat



**(Lm<sup>B</sup>)**  
laboratoire de mathématiques de besançon  
UNIVERSITÉ DE FRANCHE-COMTÉ • CNRS • UMR 6623



# PLAN

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

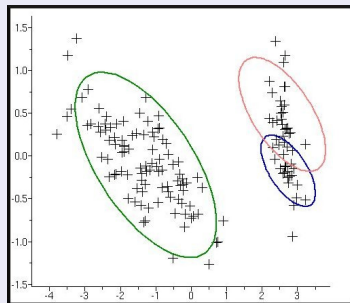
- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

# Classification des données

## Problématiques - Objectifs

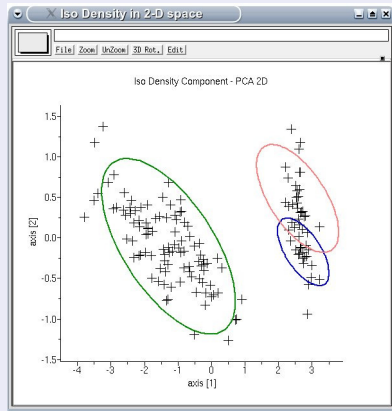
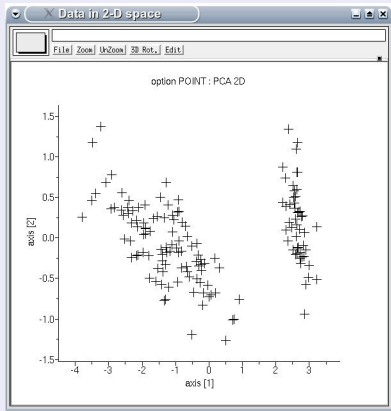
- Trouver une **partition** dans un jeu de données
  - ▶ Trouver le 'bon' **nombre de classes**
  - ▶ **Classer** les individus
- **Caractériser** les classes
- Définir une **règle de classement** à partir d'un échantillon d'apprentissage



# Classification des données

## La question du nombre de classes

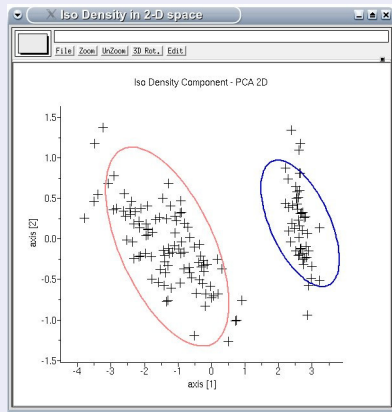
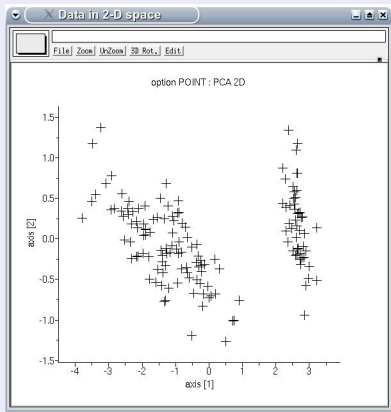
3 classes



# Classification des données

## La question du nombre de classes

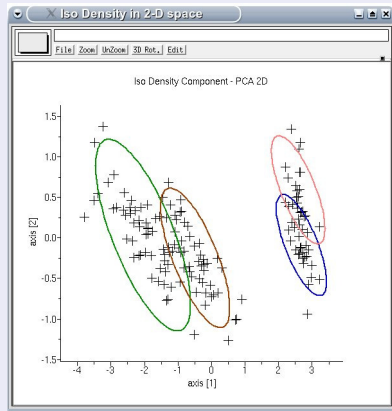
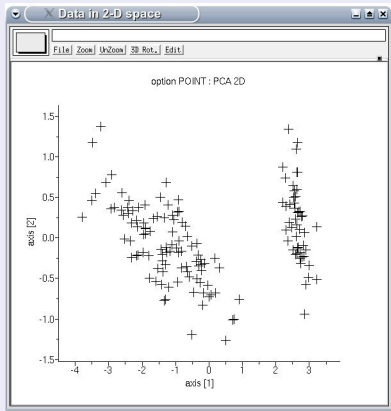
2 classes ?



# Classification des données

## La question du nombre de classes

4 classes ?





# Classification supervisée et non supervisée

- Classification non supervisée

2 objectifs :

- ▶ Choisir une (/ la meilleure) **partition** d'un jeu de données
- ▶ Classifier les individus

- Classification supervisée

2 étapes :

- ▶ **Echantillon d'apprentissage** (individus étiquetés (avec label))  
Construire/choisir une règle de classement
- ▶ **Individus non étiquetés**  
Appliquer cette règle pour classifier ces individus

# Classification des données

Pour qui ?

Large spectre de domaines concernés

Reconnaissance de formes, biologie, climatologie, marketing, finance, banque, ...

Exemples :

- Classification d'actifs financiers
- Segmentation de clientèle
- Définition de zones climatiques



# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

## Notations

- **Individus** :  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$   $n$  vecteurs de :
  - ▶  $\mathbb{R}^d$  (données quantitatives)
  - ▶  $\{1, \dots, m_1\} \times \dots \times \{1, \dots, m_d\}$  (données qualitatives)
- **Partition** :  $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$  une matrice binaire où

$$\mathbf{x}_i \in G_k \iff z_{ih} = \mathbb{I}_{\{h=k\}}$$

## Illustrations

Individus $\mathbf{x}$			Partition $\mathbf{z}$		$\Leftrightarrow$	Classe
0.2	0.5	0.8	0	1	$\Leftrightarrow$	$G_2$
0.3	0.1	0.2	1	0	$\Leftrightarrow$	$G_1$
0.9	2.0	-1.2	0	1	$\Leftrightarrow$	$G_2$
0.4	1.2	-0.2	0	1	$\Leftrightarrow$	$G_2$

# Minimisation de l'inertie intra-classe

Sélectionner la partition  $\mathbf{z}$  **minimisant**

$$W_{\mathbf{M}}(\mathbf{z}) = \sum_{i=1}^n \sum_{k=1}^K z_{ik} \|\mathbf{x}_i - \bar{\mathbf{x}}_k\|_{\mathbf{M}}^2$$

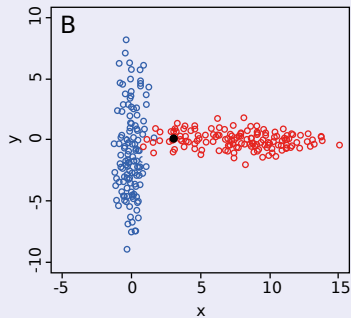
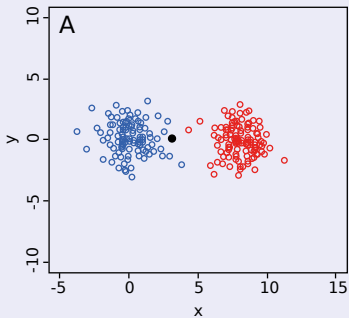
- $\|\cdot\|_{\mathbf{M}}$  est la distance euclidienne avec **métrique**  $\mathbf{M}$  dans  $\mathbb{R}^d$
- $\bar{\mathbf{x}}_k$  est la **moyenne** de la classe  $G_k$

$$\bar{\mathbf{x}}_k = \frac{1}{n_k} \sum_{i=1}^n z_{ik} \mathbf{x}_i$$

et  $n_k = \sum_{i=1}^n z_{ik}$  indique le **nombre d'individus** dans  $G_k$

# Le choix de la métrique

La métrique **M** identité : un choix courant mais risqué



La distance euclidienne n'est pas toujours adaptée

# D'une démarche géométrique vers un modèle de mélange

## Questions en suspens dans l'approche géométrique

- Quelle est la **meilleure métrique**  $\mathbf{M}_{(k)}$  ?
- Choix du **nombre  $K$  de classes** :  $W_{\mathbf{M}}(\mathbf{z})$  décroît avec  $K$ ...

Se doter d'un cadre rigoureux pour répondre à ces questions...

## Idee des modèles de mélange

Individus dans une même classe  $\iff$  Il s'agit d'un échantillon de la même loi de probabilité

$$f(\mathbf{x}_i; \theta) = \sum_{k=1}^K \pi_k \phi(\mathbf{x}_i; \alpha_k)$$

**Outils mathématiques** pour répondre **objectivement** aux problèmes de classification

# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

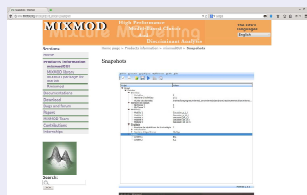


## Fiche d'identité

**Objectif** : diffuser auprès d'un large public un ensemble logiciel de classification par modèles de mélanges

- Projet débuté en 2001, soutenu par 5 organismes
- 4 composants logiciels (fin 2015)  
Licence GNU GPL - [www.mixmod.org](http://www.mixmod.org)

- ▶ **mixmodLib** (C++) : bibliothèque de calcul
- ▶ **mixmodGUI** : interface graphique (QT)
- ▶ **mixmodForMatlab**
- ▶ **RMixmod**
- ▶ **PyMixmod**



# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- **Principales caractéristiques**
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

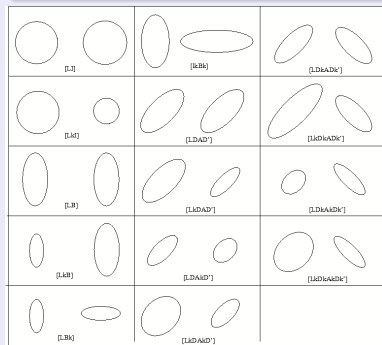
# Caractéristiques (1)

## Modèles et métriques

### Données quantitatives

#### 14 modèles gaussiens

basés sur la décomposition en valeur singulière de la matrice de variance



### Données quantitatives en grande dimension

8 modèles spécifiques pour la grande dimension

### Données qualitatives

#### 5 modèles multinomiaux

basés sur une reparamétrisation de la distribution Multinomiale

### Données mixtes

#### 20 modèles hétérogènes

pour les données quantitatives/qualitatives

# Caractéristiques (2)

## Algorithmes d'estimation

Estimation des paramètres par **maximisation de la vraisemblance**

- **EM**, CEM, SEM
- **6 stratégies** complémentaires pour **initialiser** ces algorithmes

## Une information riche

- **Caractérisation** des classes  
Proportions, moyennes, dispersions
- Labels et **probabilité d'appartenance** aux classes

# Caractéristiques (3)

## Des critères de sélection

Trouver le **bon nombre classes** et la **meilleure métrique** (le meilleur modèle)

$$\underbrace{\text{Critère}}_{\text{à maximiser}} = \underbrace{\text{maximum log vraisemblance}}_{\text{adéquation modèle-données}} - \underbrace{\text{pénalité}}_{\text{"coût" du modèle}}$$

critère	pénalité	interprétation	objectif
BIC	$0.5\nu \ln(n)$	complexité modèle	convergence
ICL	$0.5\nu \ln(n) - \sum_{i,k} \hat{z}_{ik} \ln t_{ik}(\hat{\theta})$	complexité modèle + entropie partition	classes bien séparées

Autres critères disponibles : ICL et CV

# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- **Illustrations**
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

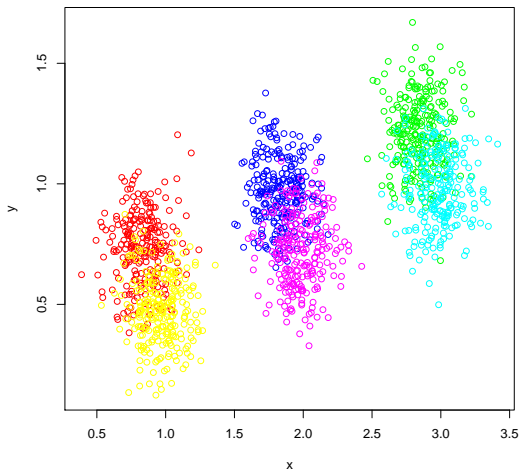
- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

# Nombre de classes

# Nombre de classes ?

6 classes avec chevauchement ( $\omega=0.3$ )

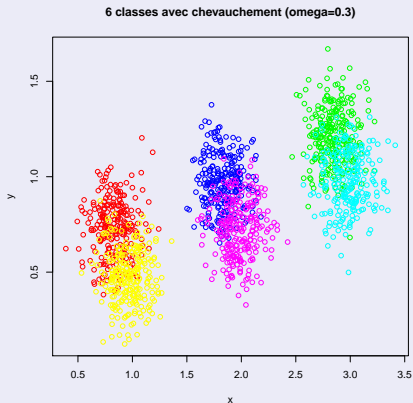




# Nombre de classes ?

Combien de classes ? 3 ou 6

- En théorie : **6 classes**
- Mais l'utilisateur peut souhaiter une solution avec des classes bien séparées : **3 classes**



Que peut faire Mixmod dans cette situation ?

# Nombre de classes

## Mixmod : de 2 à 6 classes

La meilleure solution proposée par Mixmod est avec 6 classes

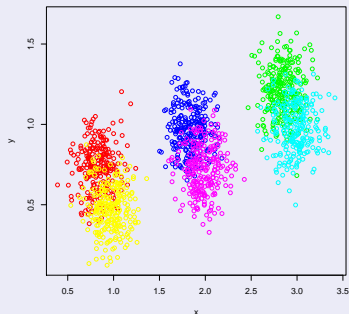
```
> res_mixmod <- mixmodCluster(data6, model=mod, nbCluster=2:6)
> summary(res_mixmod)
*****
* Number of samples      = 1500
* Problem dimension      = 2
*****
*       Number of cluster = 6
*       Model Type = Gaussian_p_L_I
*       Criterion = BIC(1206.5190)
*       Parameters = list by cluster
*       Cluster 1 :
*           Proportion = 0.1667
*           Means = 0.9621 0.4730
*           Variances = | 0.0201 0.0000 |
*                       | 0.0000 0.0201 |
*       Cluster 2 :
*           Proportion = 0.1667
*           Means = 2.8463 1.2216
*           Variances = | 0.0201 0.0000 |
*                       | 0.0000 0.0201 |
*       Cluster 3 :
*           Proportion = 0.1667
*           Means = 1.9756 0.7203
*           Variances = | 0.0201 0.0000 |
*                       | 0.0000 0.0201 |
*       Cluster 4 :
*           Proportion = 0.1667
*           Means = 1.8285 0.8703
```

# 6 classes

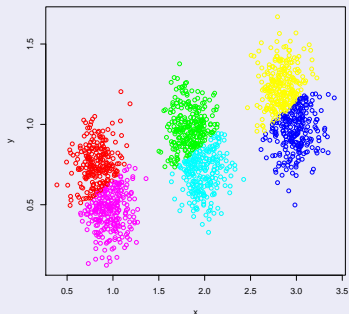
## Résultats (par défaut) de Mixmod

### Résultats avec 6 classes (BIC)

6 classes avec chevauchement ( $\omega=0.3$ )



Classification Mixmod (6 classes)



# Nombre de classes

## Mixmod : de 2 à 6 classes et tri selon ICL

La meilleure solution proposée par Mixmod avec ICL est avec 3 classes

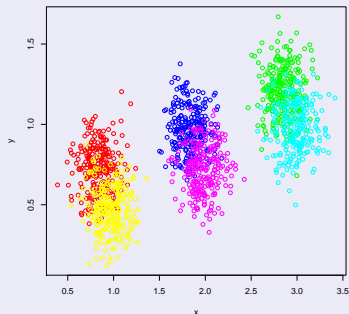
```
> res_mixmod <- mixmodCluster(data6, model=mod, nbCluster=2:6, criterion=c("BIC","ICL"))
>
> res_icl<-sortByCriterion(res_mixmod,"ICL")
>
> summary(res_icl)
*****
* Number of samples   = 1500
* Problem dimension   = 2
*****
*
*   Number of cluster = 3
*   Model Type        = Gaussian_p_L_I
*   Criterion         = BIC(1375.5983) ICL(1379.2134)
*   Parameters        = list by cluster
*
*   Cluster 1 :
*       Proportion = 0.3333
*       Means      = 0.8998 0.6007
*       Variances  = | 0.0304 0.0000 |
*                   | 0.0000 0.0304 |
*
*   Cluster 2 :
*       Proportion = 0.3333
*       Means      = 2.9113 1.0999
*       Variances  = | 0.0304 0.0000 |
*                   | 0.0000 0.0304 |
*
*   Cluster 3 :
*       Proportion = 0.3333
*       Means      = 1.9053 0.8483
*       Variances  = | 0.0304 0.0000 |
*                   | 0.0000 0.0304 |
*
*   Log-likelihood = -662.2029
*****
```

# 3 classes

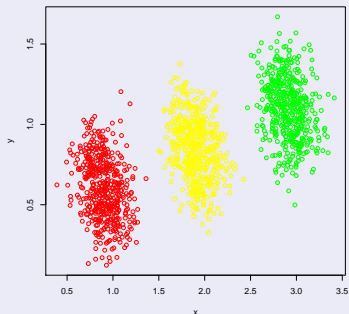
## Résultats obtenus par Mixmod

### Résultats avec 3 classes (ICL)

6 classes avec chevauchement ( $\omega=0.3$ )



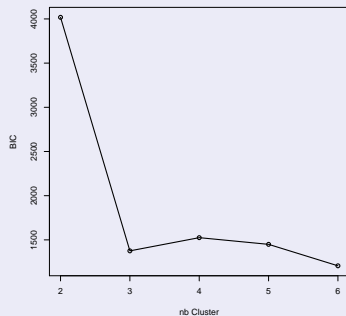
Classification Mixmod (3 classes)



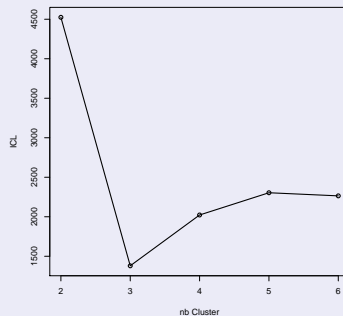
# Critères de sélection

## Valeurs de BIC et ICL en fonction du nombre de classes

Valeurs de BIC en fonction du nombre de classes



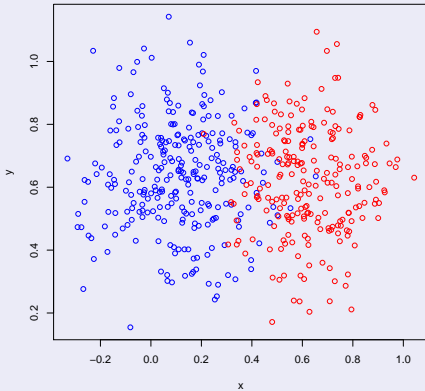
Valeurs de ICL en fonction du nombre de classes



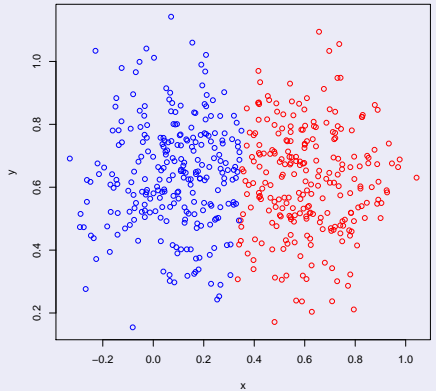
# Probabilités d'appartenance

# Chevauchement de classes

Jeu de données (simulés)



Classification Mixmod





# Probabilité d'appartenance

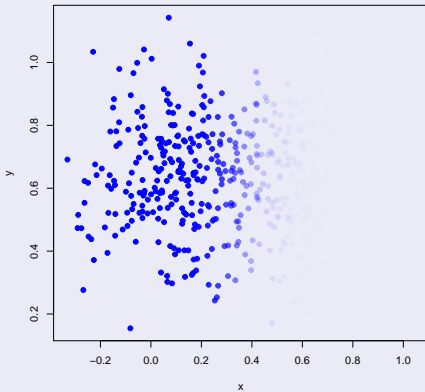
La class. par modèles de mélanges donne une information complémentaire :

la probabilité d'appartenance de l'individu  $i$  à la classe  $k$  :  $p_{ik}$

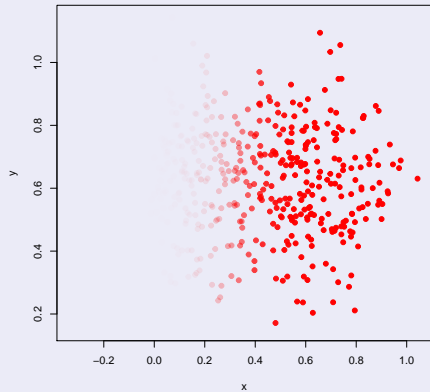
```
> probas <- res_mixmod@bestResult@proba
>
> probas
      [,1]      [,2]
[1,] 8.398762e-01 1.601238e-01
[2,] 7.551083e-01 2.448917e-01
[3,] 9.999366e-01 6.339925e-05
[4,] 2.692252e-01 7.307748e-01
[5,] 9.990483e-01 9.517056e-04
[6,] 9.999801e-01 1.987949e-05
[7,] 9.992282e-01 7.717543e-04
[8,] 9.992276e-01 7.724087e-04
[9,] 9.902374e-01 9.762624e-03
[10,] 3.779903e-01 6.220097e-01
[11,] 9.987837e-01 1.216312e-03
[12,] 9.936269e-01 6.373078e-03
[13,] 9.556769e-01 4.432311e-02
[14,] 9.679888e-01 3.201120e-02
[15,] 9.988387e-01 1.161303e-03
[16,] 6.010903e-02 9.398910e-01
```

# Chevauchement de classes

Classification Mixmod – classe bleue



Classification Mixmod – classe rouge



# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

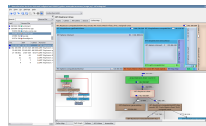
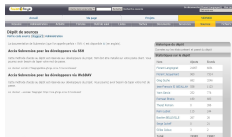
## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

# Outils de développement

## Quelques outils...



# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

# Diffusion d'un logiciel

La diffusion (et la valorisation) n'est pas un long fleuve tranquille

Diffusion et valorisation peuvent évoluer dans le temps en fonction :

- des **objectifs**
- des **envies**
- des **rencontres** (congrès, réseaux, ...)
- des **connaissances** en terme de
  - ▶ Actions de valorisation
  - ▶ Acteurs de la valorisation
  - ▶ Licences
- Des demandes des utilisateurs
- Du succès (ou non) de la diffusion



# Diffuser : oui mais pour qui ?

## Quelques questions à se poser

- A t-on identifié les utilisateurs potentiels (type et nombre) ?
- Connait-on leurs attentes ?
- A t-on pris suffisamment en compte l'activité de support ?

## La réalité

- On découvre des utilisateurs qui n'étaient pas la cible de départ
- Le nombre d'utilisateurs peut être (très) différent des estimations

## Adapter la diffusion aux utilisateurs

Moyens de diffusions, produits logiciels, type de support,...



# 1<sup>re</sup> période à partir de 2000

## Objectif

Développer un logiciel de classification des données **efficace**, **robuste** et **le diffuser au sein de la communauté scientifique**

## Composants logiciels

- Une bibliothèque de calcul : **mixmodLib**
- Des packages pour des environnements tiers : **mixmodForMatlab** et **mixmodForScilab**

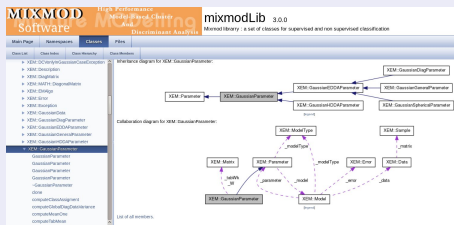
## Moyens de diffusion

- Site web [www-math.univ-fcomte.fr](http://www-math.univ-fcomte.fr) devenu **[www.mixmod.org](http://www.mixmod.org)**
- Licence GNU GPL

# 1<sup>er</sup> composant logiciel : mixmodLib

## Bibliothèque de calcul

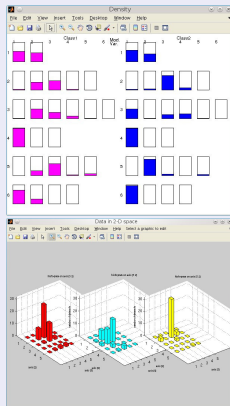
- Ensemble de classes C++
  - ▶ Architecture évolutive
  - ▶ Env. 100 classes et 40 000 lignes de code C++
- Qualités
  - ▶ Rapidité
  - ▶ Robustesse
  - ▶ Evolutivité
- Env. 500 téléchargements par an depuis 2001



## 2<sup>e</sup> composant logiciel : mixmodForMatlab (/Scilab)

### Package pour Matlab

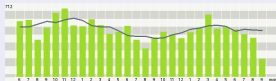
- Interface de mixmodLib pour Matlab
- Ensemble de fonctions Matlab :
  - ▶ Classification supervisée et non supervisée
  - ▶ Outils de visualisation
- Les calculs sont effectués dans mixmodLib
- Disponible depuis 2003  
env. 300 téléchargements par an



# 2<sup>e</sup> période à partir de 2003

### La situation évolue

- La “renommée” de Mixmod croît  
Environ 500 visites et 150 téléchargements par mois



- Mixmod touche de plus en plus d’**utilisateurs “non spécialistes”**

### Conséquences

**Demande croissante** en terme de

- Support
- Demandes d’évolution
- Utilisation de Mixmod  
Les outils initialement mis à disposition ne conviennent pas toujours à cette nouvelle population d’utilisateurs

# Conséquences

## Composants logiciels

- **Enrichissement** des packages pour Matlab et Scilab
  - ▶ Utilisation plus aisée (pour les utilisateurs non spécialistes )
  - ▶ Fonctions graphiques enrichies

## Diffusion

- **Dépôt à l'APP** (Agence de Protection des Programmes)
  - ▶ A conseiller vivement
  - ▶ Procédure **simple** (en théorie) mais **longue** (en pratique) avec 5 organismes concernés
- Licence **GNU GPL**
  - ▶ Favorise l'utilisation au sein de la communauté scientifique
  - ▶ Encourage les développements s'appuyant sur Mixmod et les contributions
  - ▶ Améliore la qualité logicielle (peer review, remontée de



# 3<sup>e</sup> période

## à partir de 2007-2008



## Constats

- **Hétérogénéité** encore plus grande des utilisateurs  
Comment proposer des outils pour le **spécialiste** et le **novice** ?
- Nécessité de répondre à certaines demandes (**package pour R**)

Mais aussi :

- Difficulté à développer, maintenir plusieurs composants logiciels très différents

## Evolutions des composants logiciels

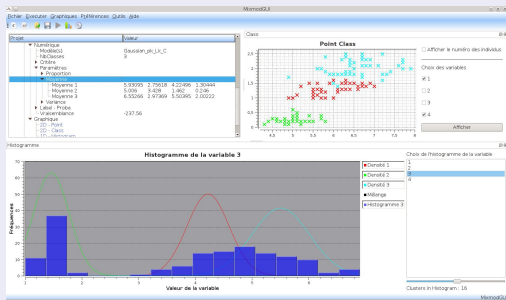
- Abandon de **mixmodForScilab**
- Développement d'un logiciel avec interface graphique : **mixmodGUI**
- Développement d'un package pour R : **RMixmod**

# 3<sup>e</sup> composant logiciel : mixmodGUI

## Interface graphique

Pour une utilisation sans environnement tiers (Matlab, R, ...)

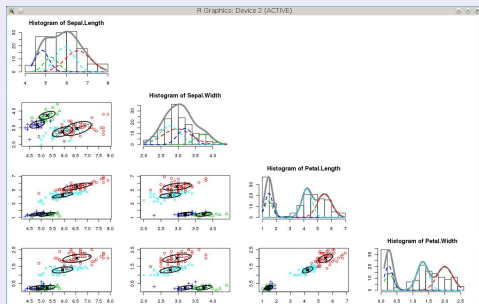
- Conviviale
- Entrées/Sorties XML
- Utilisation des bibliothèques QT et Qwt
- Disponible depuis 2011 env. 500 téléchargements sur un an



# 4<sup>e</sup> composant logiciel : Rmixmod

## Package pour R

- Interface de mixmodLib pour R
- Ensemble de fonctions R :
  - ▶ Classification supervisée et non supervisée
  - ▶ Outils de visualisation
- Les calculs sont effectués dans mixmodLib
- Disponible depuis 2012



# 4<sup>e</sup> période à partir de 2014

## Constats

- **RMixmod** devient le composant le plus utilisé
- Demande croissante pour un composant Mixmod pour **Python**
- Trop de composants à maintenir, faire évoluer

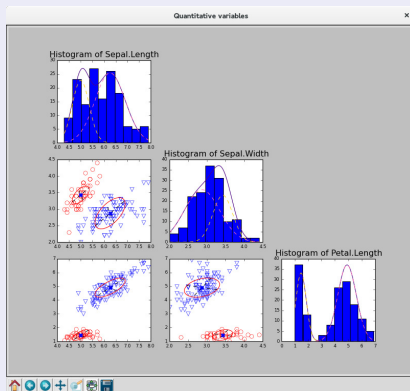
## Evolutions des composants logiciels

- "Abandon" de **mixmodForMatlab**
- Développement de **PyMixmod** (pour Python)

# 5<sup>e</sup> 4<sup>e</sup> composant logiciel : PyMixmod

## Mixmod pour Python

- Interface de mixmodLib pour Python
  - ▶ Classification supervisée et non supervisée
  - ▶ Outils de visualisation
- Les calculs sont effectués dans mixmodLib
- Disponible fin 2015



Utilisation de Cython, NumPy, SciPy, Pandas, Matplotlib

# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

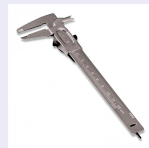
# Outils de mesure

Comment mesurer la diffusion, la valorisation ?



## Outils

- Utilisations
- Contributions
- Enrichissements et redistributions
- Collaborations
- Intégrations





# Utilisations

## Mesures

- Nombre de **téléchargements** (mesure biaisée) - env. 200 par mois
- Activité du **forum**
- **Contacts** directs (mail)
- **Citations** (articles)
- **Remerciements** !!

## Typologie/Qualité

Des **composants logiciels** pour des utilisateurs différents

- Outils pour la recherche
- Utilisation dans des projets réels, industriels

## Comment en savoir plus ?

Sans être trop intrusif (formulaire à remplir, ...)

# Contributions

- **Très peu** de contributions (en dehors du cercle de connaissance)
- **Complexité** intrinsèque
- Certaines contributions sont cependant moins *techniques*
  - ▶ *Traduction FR/US des documentations*
  - ▶ *Participation au forum, Support*
  - ▶ *Construction des packages*
- Cession des **droits**

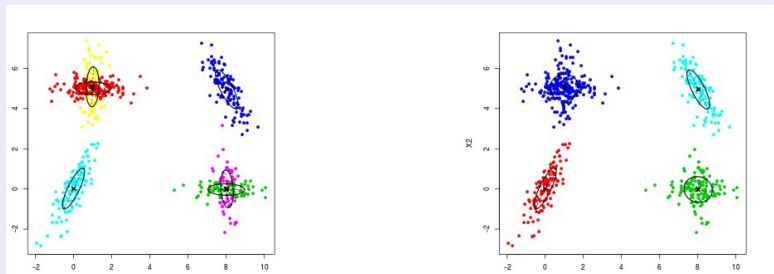


# Packages R basés sur Mixmod

- Mise à disposition de **packages complémentaires**
- **Intérêt scientifique** de (R)Mixmod reconnu par la communauté

## Exemple

### RMixmodCombi



# Intégration de mixmodLib dans un logiciel

## Motivations

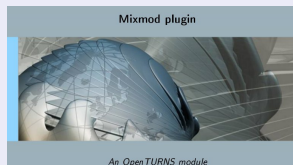
- **Intérêt scientifique** de MIXMOD
- Bibliothèque **éprouvée, robuste, rapide**

## Quelques questions, contraintes

- **Licence**
  - ▶ Mixmod sous licence GNU GPL => Logiciel sous licence GNU GPL
  - ▶ Autre type de licence sur demande (**double licence**)
- **Aide** à l'intégration

## Exemples

- **SelVarClust**  
Sélection de variables en classification non supervisée et supervisée
- **Plugin MIXMOD/ OpenTurns**  
Traitement des incertitudes (EdF, EADS, Phimeca)



# Collaborations

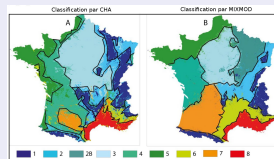
## Utilisation avancée de MIXMOD

- **Situations complexes** ou **originales**
  - ▶ Taille des données (nombre d'individus, grande dimension)
  - ▶ Classification difficile (convergence lente, chevauchement important, ...)
- Nécessite parfois des (petites) **évolutions** de MIXMOD
- Contact **priviliégié** et **enrichissant** avec des utilisateurs (dans des domaines variés)

## Exemple

### Classification des climats en France

- Très gros jeu de données  
15 millions d'individus, 14 variables
- Nombre de climats ?
- Zones climatiques homogènes
- Limites franches



# MIXMOD

## 1 Pourquoi les modèles de mélanges en classification des données ?

- La classification des données
- Intérêts des modèles de mélanges

## 2 Mixmod

- Principales caractéristiques
- Illustrations
- Outils de développement

## 3 Retour sur 15 ans de développement et diffusion

- Un long fleuve tranquille ?
- Comment mesurer la diffusion, la valorisation ?

## 4 Perspectives

## Volet statistique

### Nouvelles fonctionnalités

- Enrichissement des méthodes d'initialisation
- Traitement des données bruitées
- Traitement des données manquantes

## Volet informatique

- Intégration des "meilleurs" concepts **C++14**
- **Optimisation de calcul** (parallélisation, changement de bibliothèque d'algèbre linéaire)
- **Précision** de calcul

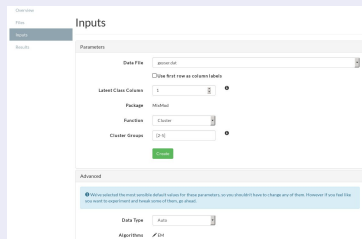
# Composants logiciels

## Des composants logiciels pour des utilisateurs hétérogènes

- mixmodLib
- mixmodGUI
- RMixmod
- PyMixmod

Et ... pour 2016/2017

- **Une interface web**
  - ▶ Toucher un public encore plus large
  - ▶ Accès multi-supports (tablettes, smartphone)





## Disponibles sur [www.mixmod.org](http://www.mixmod.org)

- MIXMOD Statistical Documentation (rubrique documentation)
- Exposés des journées Mixmod
- *Rmixmod : The R Package of the Model-Based Unsupervised, Supervised and Semi-Supervised Classification Mixmod Library (JSS 2015)*
- *Model-Based Cluster and Discriminant Analysis with the MIXMOD Software (CSDA - 2006)*
- *MIXMOD : un logiciel de classification supervisée et non supervisée pour données quantitatives et qualitatives (La Revue de Modulad - 2009)*

MIXMOD<sup>®</sup>

Un ensemble logiciel de classification des données  
par modèles de mélange

F. Langrognet

Merci de votre attention