

# **MIXMOD : un logiciel pour les modèles de mélange en classification et en analyse discriminante**

**C. Biernacki\* – G. Celeux\*\* – G. Govaert\*\*\***

**F. Langrognet\* – G. Noulin\*\* – Y. Vernaz\*\***

\* Université de Franche-Comté - UMR CNRS 6623

16 route de Gray

F-25030 Besançon Cedex

{Christophe.Biernacki,Florent.Langrognet}@math.univ-fcomte.fr

\*\* INRIA Rhône-Alpes

ZIRST, 655 avenue de l'Europe, Montbonnot

F-38334 Saint-Ismier Cedex

{Gilles.Celeux, Gregory.Noulin}@inrialpes.fr

\*\*\* Université de Technologie de Compiègne - UMR CNRS 6599

BP 20529 - F-60205 Compiègne Cedex

{Gerard.Govaert}@utc.fr



## **Plan**

Présentation générale  
Fonctionnalités scientifiques  
Autres logiciels du domaine  
Perspectives

# PRÉSENTATION GÉNÉRALE

## Objectif

Trois points de vue pour les modèles de mélange multivariés

- estimation de densité
- classification automatique
- analyse discriminante

## Public visé

- la recherche
- l'industrie pour la reconnaissance statistique des formes
- l'enseignement pour des cours d'introduction ou avancés

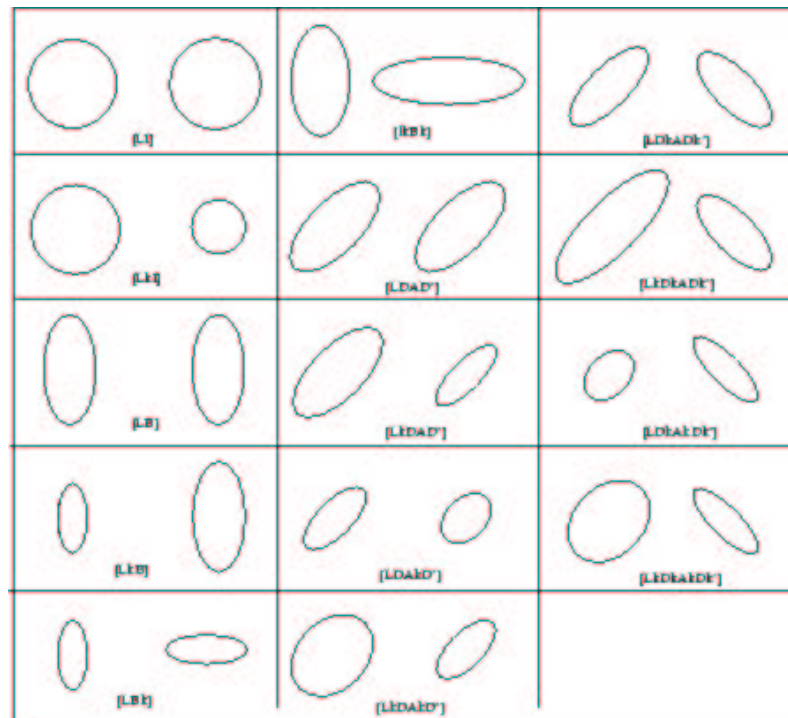
## Aspects techniques

- coproduction INRIA-UFC-UTC
- disponibilité : web (<http://www-math.univ-fcomte.fr/MIXMOD/index.php>), CD INRIA
- langage : C++ (orienté objet)
- gratuit : open source, conditions “the GNU General Public Licence”
- plateformes : Linux, Unix, Windows
- documentation (html, pdf, ps) : guide utilisateur, documentation statistique, manuel de référence sur la programmation
- niveau d'utilisation : interface Matlab/Scilab ou mode expert

# FONCTIONNALITÉS SCIENTIFIQUES

## Modèles

- seuls les mélanges gaussiens multivariés sont implantés
- 14 modèles sur les matrices de variances disponibles



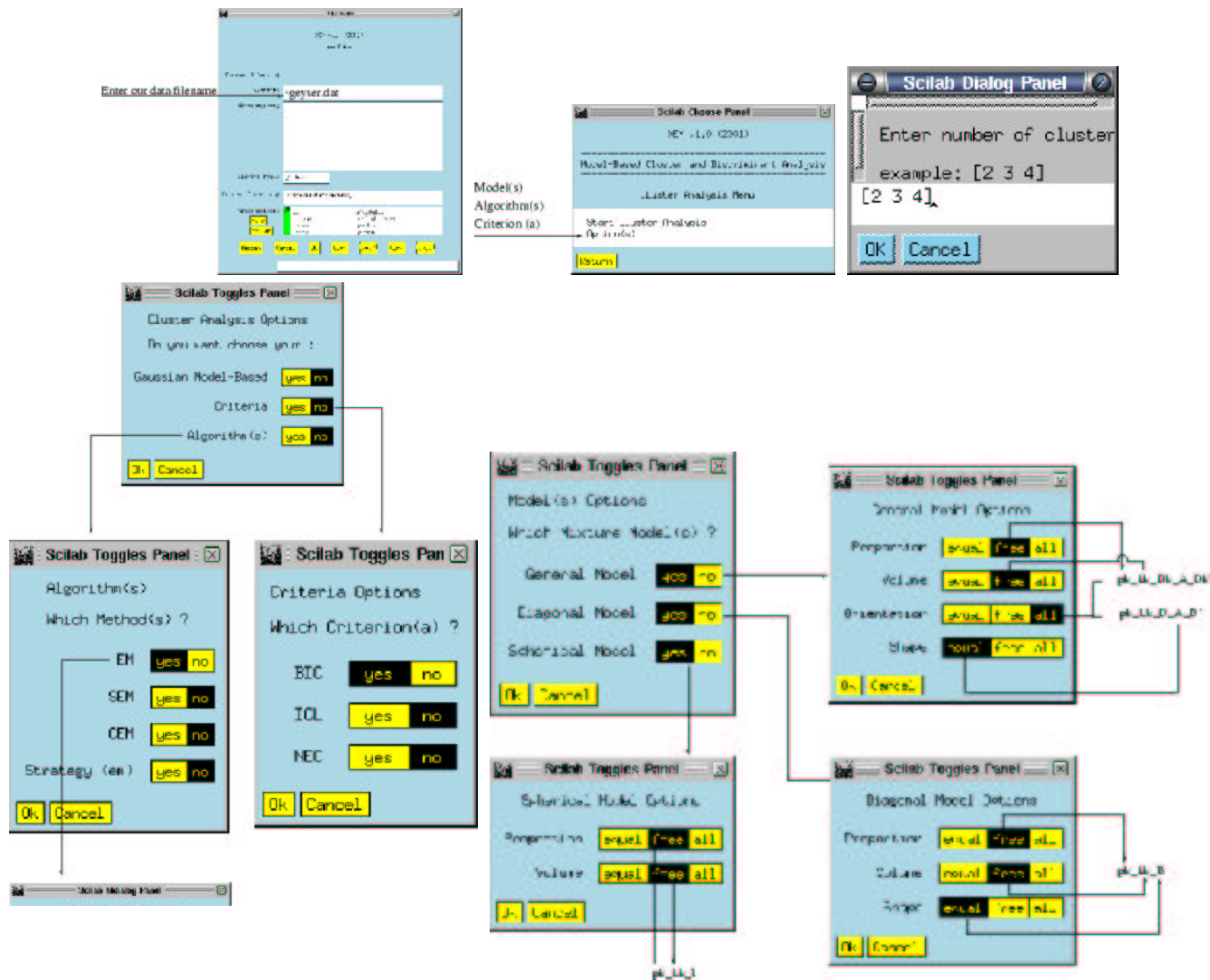
## **Algorithmes et stratégies**

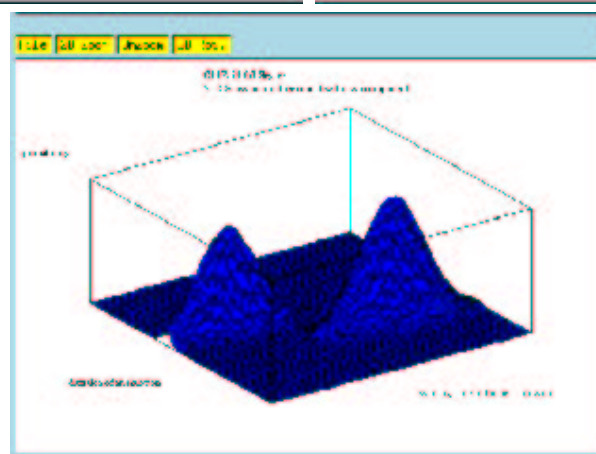
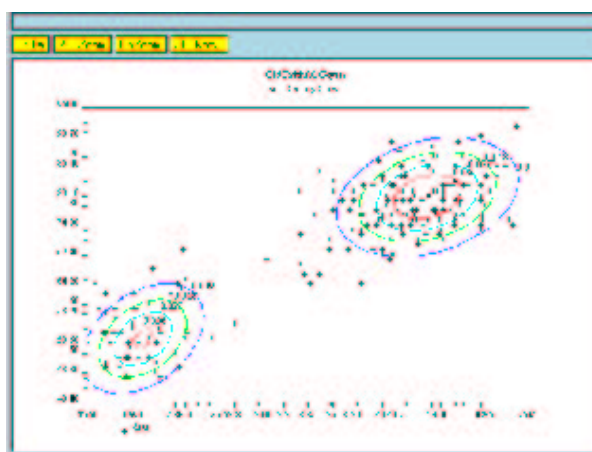
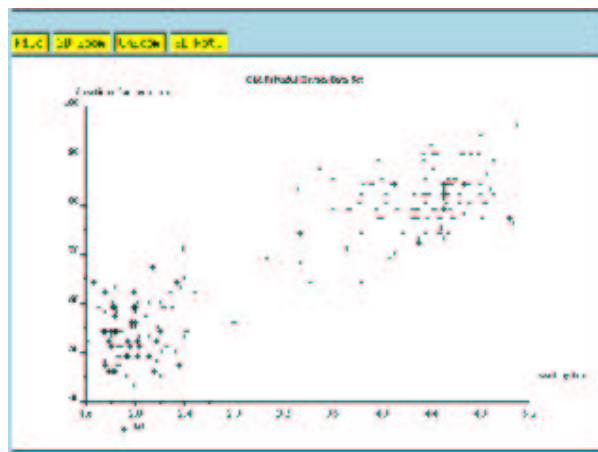
- algorithmes disponibles : EM, SEM, CEM
- possibilité de les enchaîner pour éviter les maxima locaux
- paramètres initiaux : utilisateur, centres aléatoires
- différents arrêts : nombre itérations, critère, partition
- possibilité de connaissance partielle des labels

## **Critères de choix de modèles**

- classification (modèle/nombre de classes) : BIC, ICL, NEC
- discrimination (modèle) : BIC, CV

# Exemple de classification sur le “geyser”







## AUTRES LOGICIELS DU DOMAINE

- EMMIX par G. McLachlan et D. Peel :  
<http://www.maths.uq.edu.au/gjm/emmix/emmix.html>
- MCLUST par C. Fraley et A. Raftery :  
<http://www.stat.washington.edu/fraley/mclust/soft.shtml>
- AUTOCLASS par P. Cheeseman :  
<http://ic-www.arc.nasa.gov/ic/projects/bayes-group/autoclass/>
- SNOB par D. Dowe :  
<http://www.csse.monash.edu.au/dld/Snob.html>
- NORMIX par John H. Wolfe :  
<http://www.alumni.caltech.edu/wolfe/normix.htm>.

## PERSPECTIVES

- développer des *fonctions* Scilab/Matlab (R ?) pour un usage transparent dans chaque environnement
- ajouter des aides à la visualisation en Scilab/Matlab
- site web dynamique
- extension au cas qualitatif : mélange de lois multinomiales

<http://www-math.univ-fcomte.fr/MIXMOD/index.php>