

MIXMOD - Développement, diffusion, valorisation d'un ensemble logiciel de classification de données quantitatives et qualitatives par modèles de mélanges

www.mixmod.org



Florent Langrognat^a, Benjamin Auder^b

^aLaboratoire de Mathématiques de Besançon (Université de Franche-Comté - CNRS)

^bLaboratoire de Mathématiques d'Orsay (Université de Paris Sud 11 - CNRS)

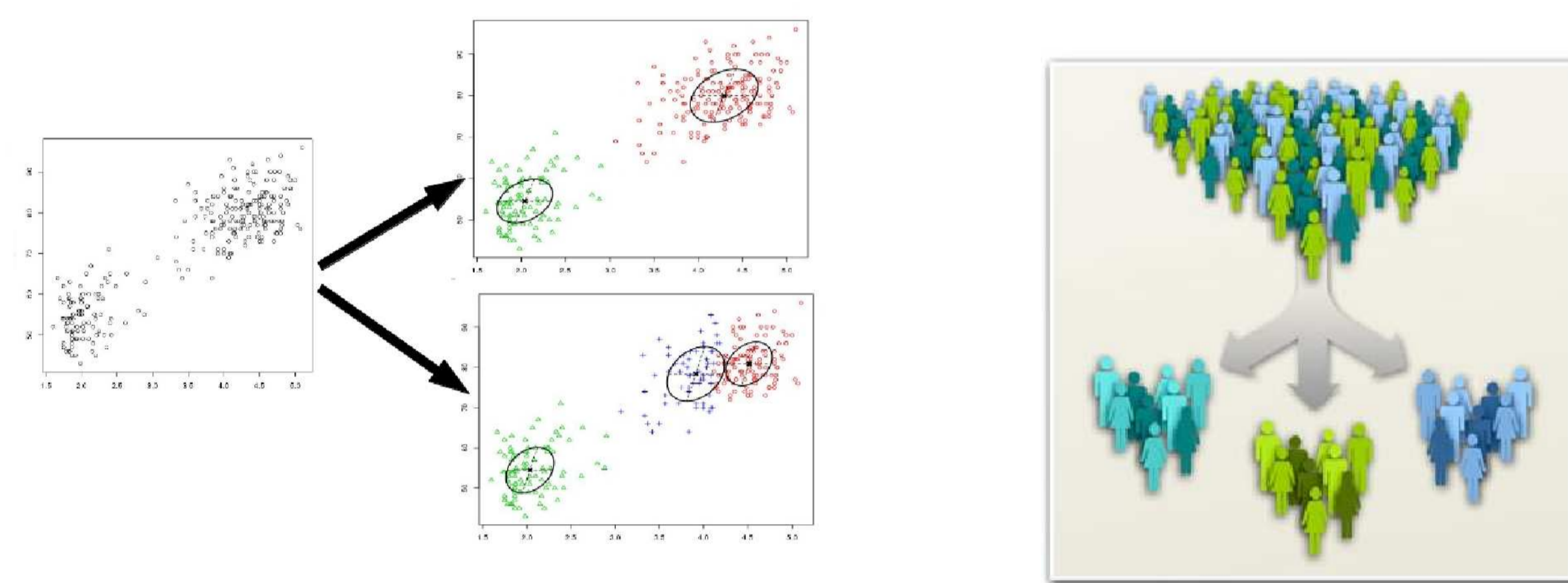


Classification des données

- Regrouper des **individus** dans des **classes homogènes**

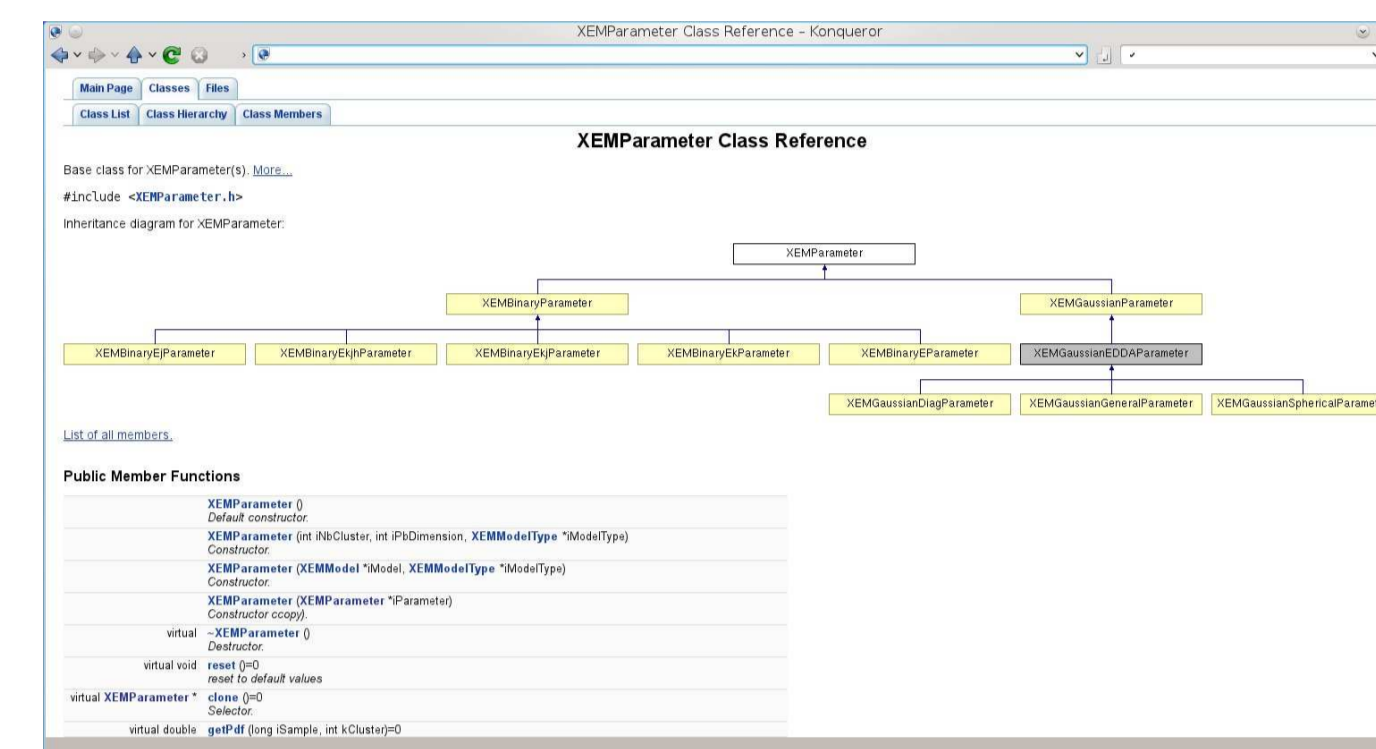
- ▶ Choisir le nombre de classes
- ▶ Caractériser les classes

- **Apprentissage** puis classement

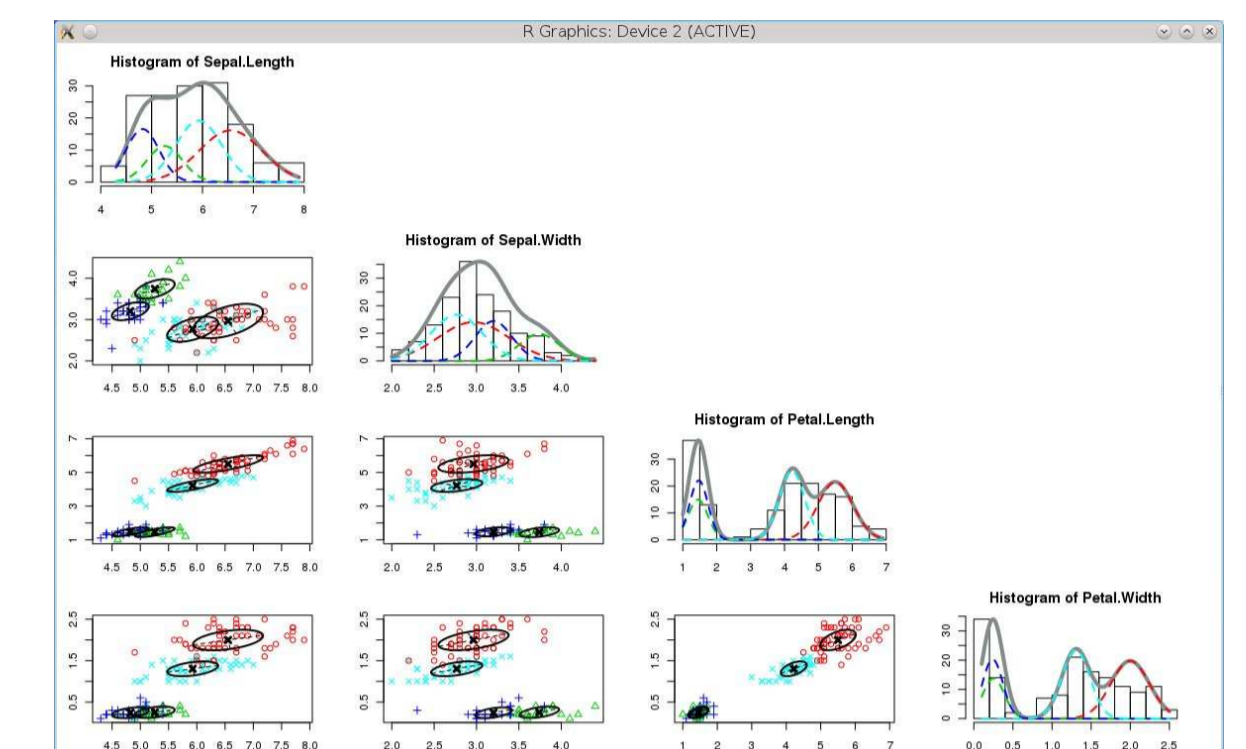


4 composants logiciels

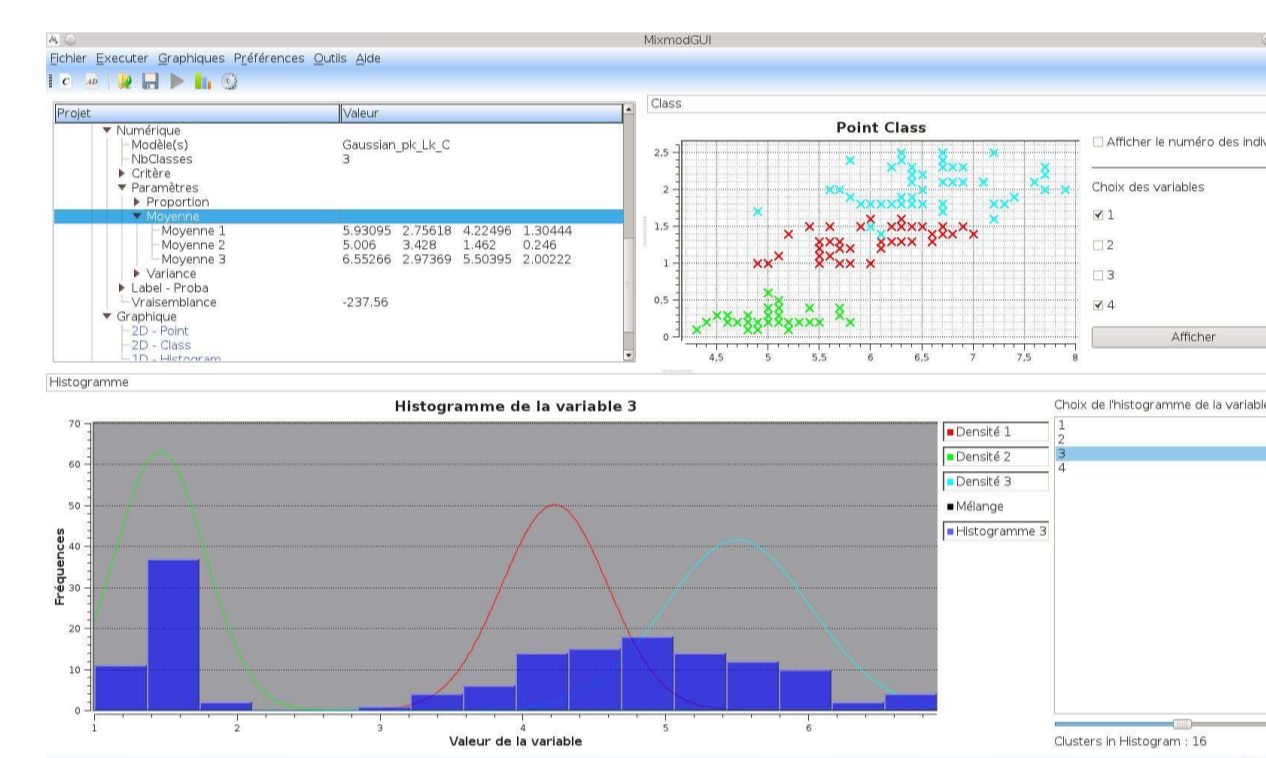
Bibliothèque de Calcul (mixmodLib)



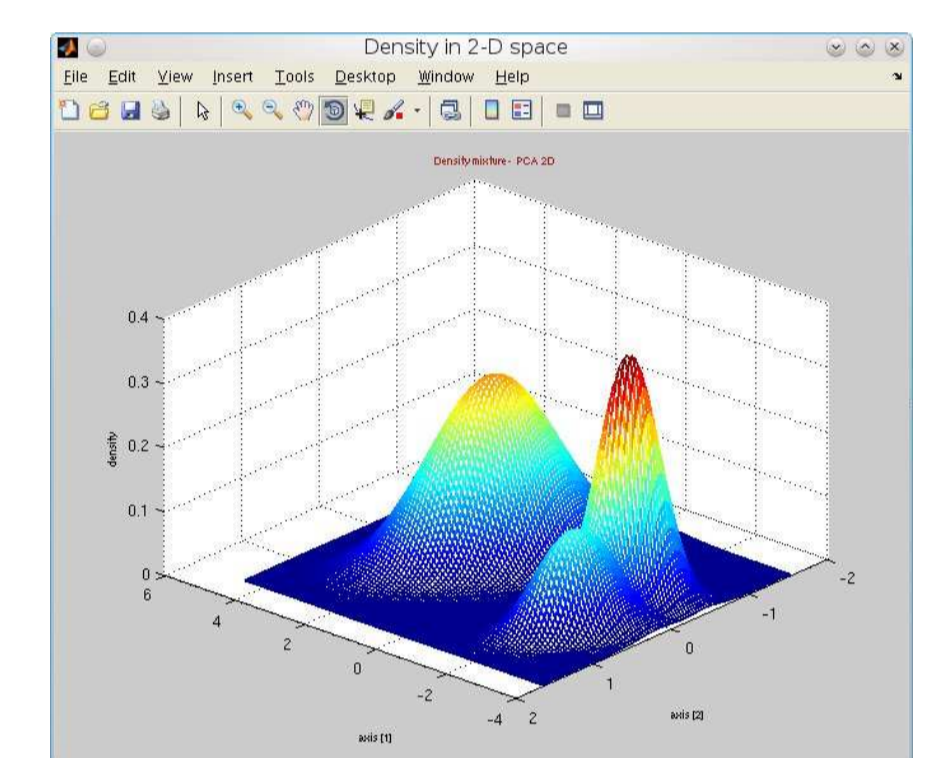
Package pour R (RMixmod)



Graphical User Interface (mixmodGUI)



Package pour Matlab (mixmodForMatlab)



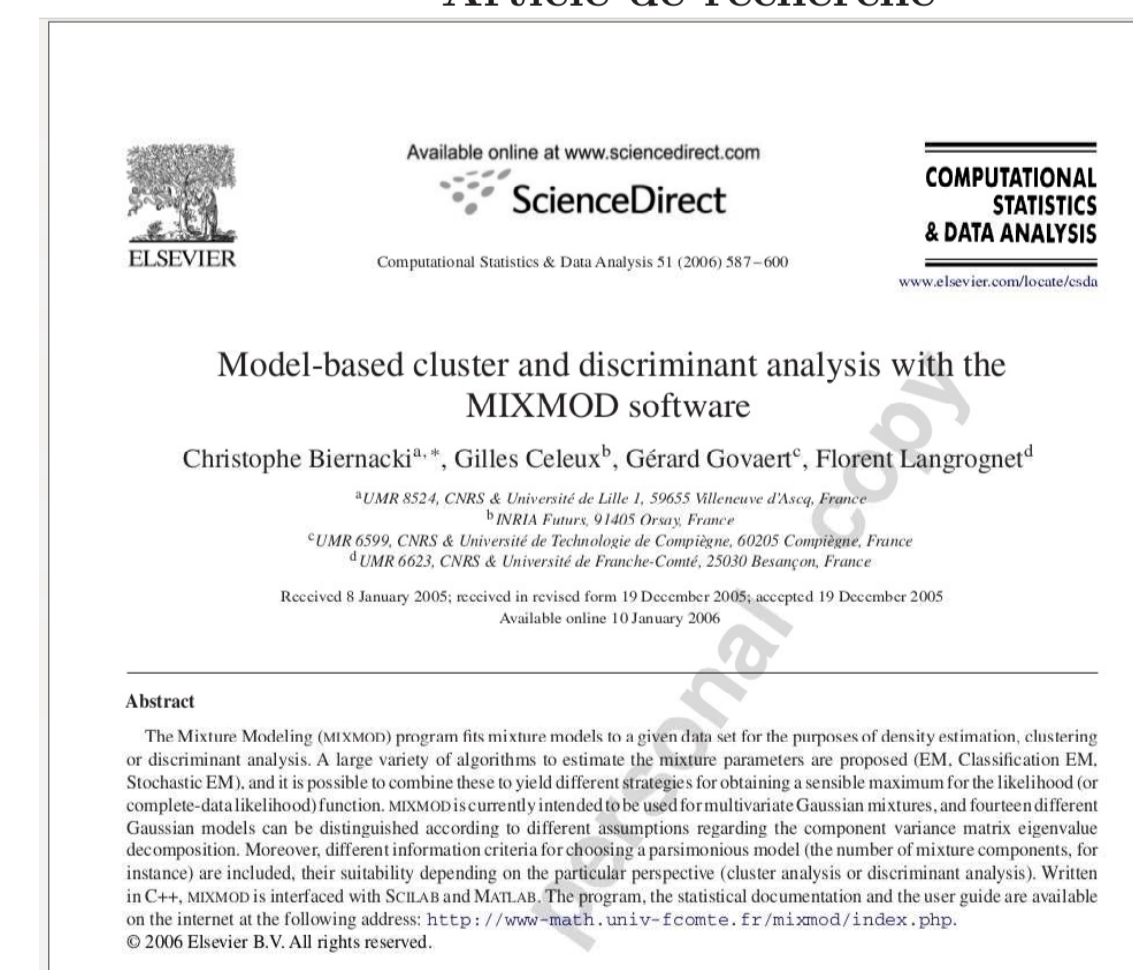
MIXMOD et les modèles de mélanges

- Approche **probabiliste**
- **Modèle explicatif** souple
- Données **quantitatives** ou **qualitatives**
- **Big data** (nombre d'individus et/ou nombre de caractéristiques)
- **Principales fonctionnalités**
 - ▶ **3 Algorithmes** : EM, CEM, SEM
 - ▶ **6 Méthodes d'initialisation**
 - ▶ **4 Critères de sélection** : BIC, ICL, NEC, CV
 - ▶ **14 modèles gaussiens** pour les données quantitatives
 - ▶ **5 modèles multinomiaux** pour les données qualitatives
 - ▶ **8 modèles** pour les données quantitatives en **Grande Dimension**

Diffusion, valorisation

- **Licence GNU GPL** (ou licence propriétaire sur demande)
- **Packages** sources ou binaires pour Windows, Linux, MacOS
- Env. **250 téléchargements** par mois
- **Rencontres Mixmod** bisannuelles
- **Communauté d'utilisateurs** (forum)
- **Conférences, articles**

Article de recherche



www.mixmod.org



Utilisations, collaborations, intégrations

- **Domaines très variés** : Banque, marketing, environnement, sociologie, épidémiologie, génomique, biochimie, ...
- **Utilisateurs hétérogènes**
 - ▶ Communauté scientifique
 - ▶ Monde socio-économique : EdF, Orange, La Redoute, ArcelorMittal, PGXIS, Banque Accord, CEA, ...
- **Collaborations** : Climatologie, géosciences, économie, banque, ...
- **Intégration** de la bibliothèque de Calcul MIXMOD (ex : OpenTurns)

Développement (méthodes, outils)

- **Langages, bibliothèques** : C++, R, QT, Qwt, Matlab, XML, UML
- **Outils** : Forge (INRIA), Doxygen, cmake, SVN, KDevelop, Code::Blocks, netbeans, Eclipse, Valgrind, Jenkins, innosetup
- **Tests, Validation** : Tests unitaires (Boost::test), de non régression, de performance, intégration continue (<https://ci.inria.fr/>)
- **mixmodLib** (C++) : 100 classes, 40 000 lignes