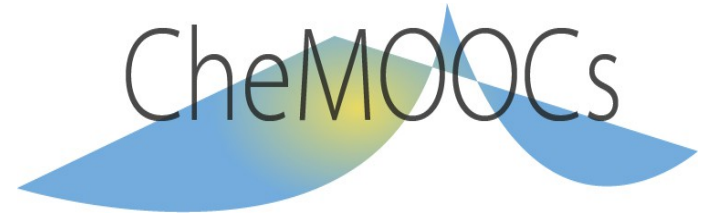


Ce projet est financé par Agropolis Fondation
sous la référence ID 1401-05, depuis le
programme « Investissements d'avenir »
(Labex-Agro : ANR-10-LABX-0001-01)



Projet collaboratif

Présenté par :
Jean-Claude BOULET
INRA, UMR1083 SPO
Plateforme Polyphénols

La chimiométrie pour tous : principes et outils.

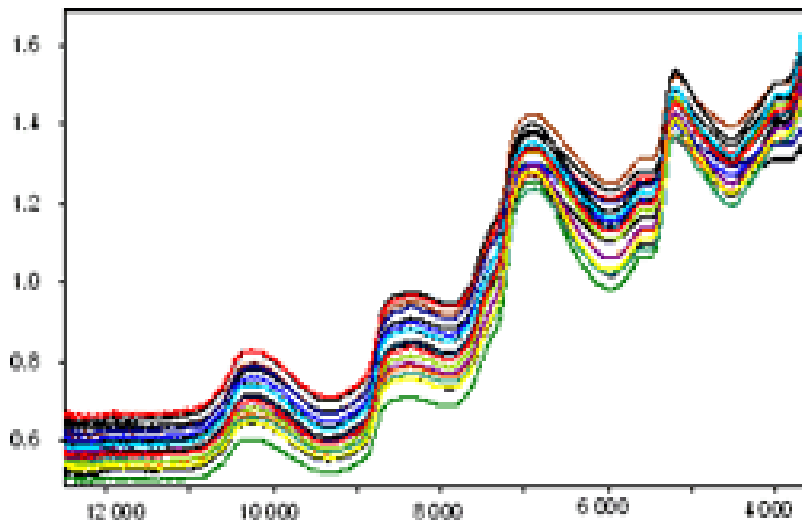


La chimiométrie

Definition (International Chemometric Society):

La chimiométrie est la science mettant en relation des mesures faites sur un système chimique ou un procédé avec l'état de ce système ou procédé, en utilisant des méthodes mathématiques ou statistiques.

Pratique :



Prédictions :

- quantitatives :

ex. taux de gluten

- qualitatives :

ex. pays de provenance

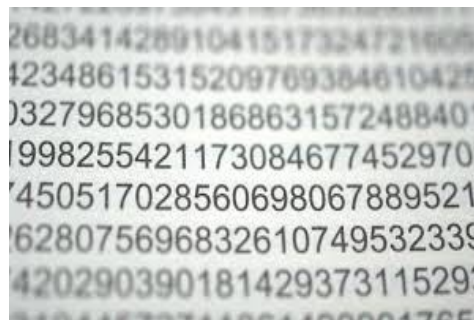
Le projet Chemooocs : 3 piliers



> Un MOOC

*Massive open
online courses*

> Une boîte
à outils



> Une base
de données

FUN : plateforme de diffusion de MOOCs



https://www.france-universite-numerique-mooc.fr/about



france université numérique



CONNEXION

DÉCOUVRIR, APPRENDRE ET RÉUSSIR

QU'EST-CE QUE FUN ?

ACTUALITÉS

LES COURS

LES ÉTABLISSEMENTS

S'INSCRIRE MAINTENANT

QU'EST-CE QUE FUN ?



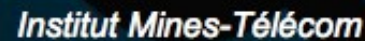
FUN est une plateforme de **MOOC (Massive Open Online Courses, en français « Cours en ligne ouverts à tous »)** mise à disposition des établissements de l'enseignement supérieur français et de leurs partenaires académiques dans le monde entier. Lancée par le Ministère de l'Enseignement Supérieur et de la Recherche en octobre 2013, cette initiative vise à fédérer les projets des universités et écoles françaises pour leur donner une visibilité internationale, et **permettre à tous les publics d'accéder à des cours variés et de qualité où qu'ils soient dans le monde.**

Tous les cours présents sur FUN sont conçus par des professeurs d'universités et écoles françaises et leurs partenaires académiques internationaux. Les étudiants et les internautes peuvent suivre ces cours **de manière interactive et collaborative, à leur rythme.**

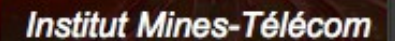
Bases de données relationnelles : Comprendre pour maîtriser



Soyez acteurs du web!



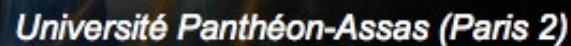
Challenges et enjeux de la mobilité 2.0



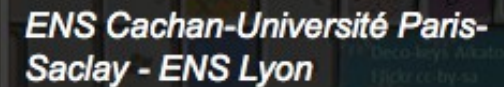
Pratiques du Dimensionnement en Mécanique - Partie 2



Problèmes économiques contemporains



Education aux médias et à l'information à l'ère du numérique (eFAN)



Composition du MOOC



- des grains
 - une vidéo (10')
 - des exercices
- une animation (ex : forum) pour créer une dynamique
- une évaluation
- durée limitée : 8 semaines

Les inscrits se comptent en milliers

CHEMOOCS

TRONC COMMUN

Grain 1 : Introduction au Mooc - Jean-Claude Boulet

Grain 2 : Stats simples - Dominique Bertrand

Grain 3 : ACP 1 - Sébastien Preys

Grain 4 : Prétraitement 1 - Jean-Michel Roger

Grain 5 Prétraitement 2 - Jean-Michel Roger

Grain 6 : ACP 2 - Sébastien Preys

Grain 7 : Classification non supervisée - Philippe Courcoux

Grain 8 : Régressions linéaires - Robert Sabatier

Grain 9 : PLS 1 - Douglas Rutledge

Grain 10 : PLS 2 - Douglas Rutledge

Grain 11 : bonnes pratiques de modélisation - Dominique Bertrand

Grain 12 : Discrimination 1 - Robert Sabatier

Grain 13 : Discrimination 2 - Robert Sabatier

Grain 14 : Optimisation - Jean-Michel Roger

attestation de suivi avec succès

EN SAVOIR PLUS

Grain 15 : Sélection de variables 1
Dominique Bertrand

Grain 16 : Sélection de variables 2
Robert Sabatier

Grain 17 : Régressions locales ?

Grain 18 : Multicurve resolution
Ludovic Duponchel

Grain 19 : Independent component analysis
Douglas Rutledge

Grain 20 : Structure des multitableaux
Mohamed Hanafi

Grain 21 : ACP concaténée
Mohamed Hanafi

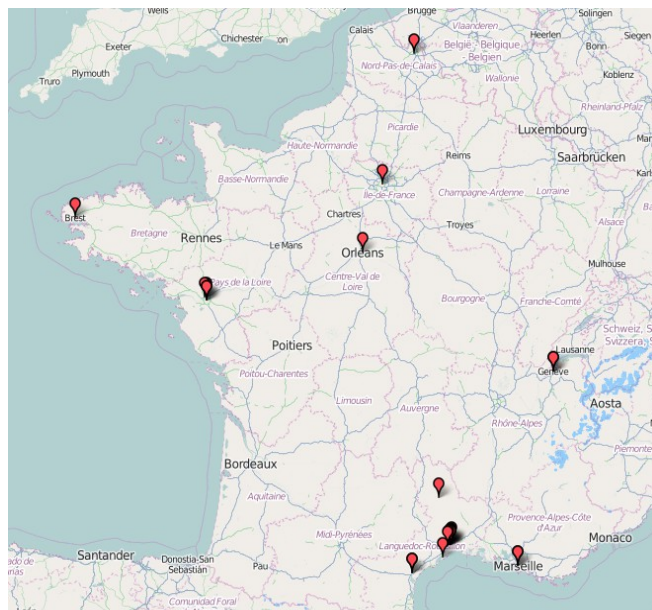
Grain 22 : ACOM
Mohamed Hanafi

Contributeurs au MOOC



Contributeurs :

Ludovic Duponchel
Douglas Rutledge
Yves Lijour
Mathieu Boiret
Dominique Bertrand
Mohamed Hanafi
Philippe Courcoux
Benoit Jaillais
Julien Boccard
Serge Rudaz
Nathalie Dupuy
Sandrine Amat
Sébastien Preys
Robert Sabatier
Christelle Reynes
Myrtille Vivien
Jean-Michel Roger
Martin Ecartot
Jean-Claude Boulet
Eric Latrille



Coordination pédagogique :
Cécile Tredaniel

Réalisation/collaboratif :
Christian Resche
+ l'équipe de
Supagro-Florac

Expert MOOC :
Christophe Lebegue

Expert utilisateurs :
Bernard Barthès

L'outil initial :



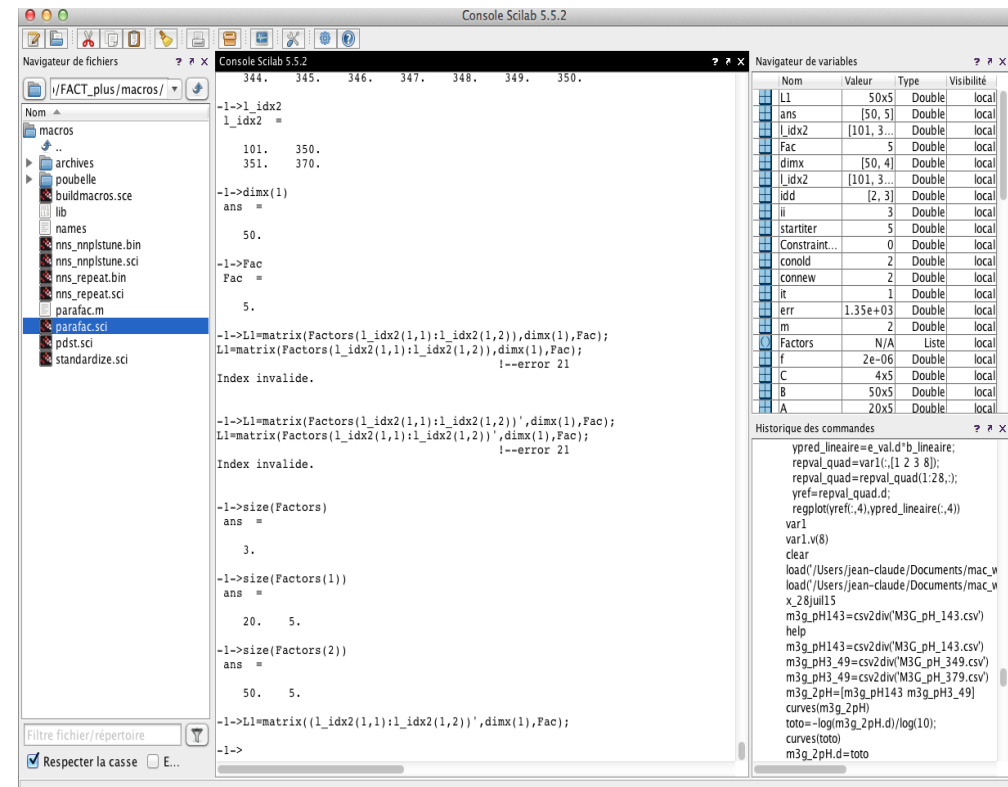
**boîte à outils FACT
sous SCILAB**

Groupe de travail :
Jean-Michel Roger
Jean-Claude Boulet
Eric Latrille
Virginie Rossard
Fabien Gogé

Expert informatique:
Pascal Neveu

Cahier des charges :

- gratuit
- contient les méthodes présentées dans le MOOC
- facile à utiliser (compréhension)
- facile à utiliser (accès, installation)

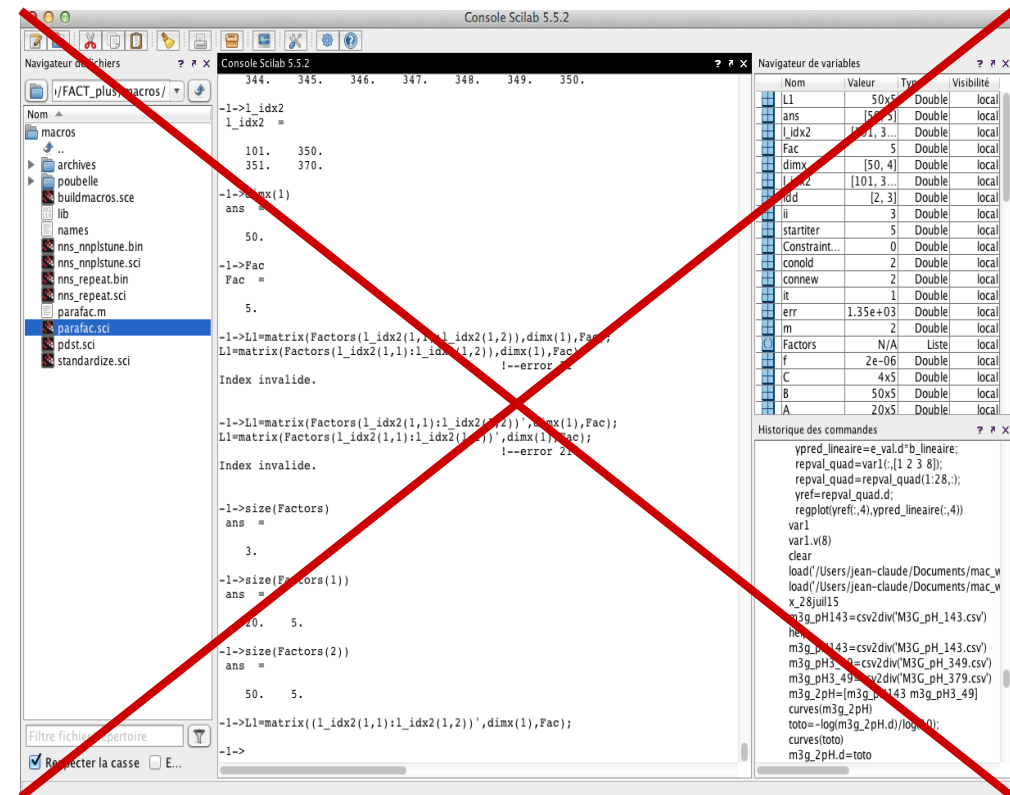


L'outil initial : boite à outils FACT sous SCILAB



Réponse au cahier des charges :

- gratuit
- contient la plupart de méthodes présentées dans le MOOC
- facile à utiliser (accès, installation)
- facile à utiliser (code)



Choix de Galaxy : de grosses possibilités



- programmer sans ligne de commande
- re-utilisation de fonctions existantes (en Scilab, Octave, R)
- gratuit

Plus :

- interface web
- connectivité avec bases de données
- workflows
- grosse communauté

Notre version de Galaxy au 01/11/15:



Firefox Fichier Édition Affichage Historique Marque-pages Outils Fenêtre Aide

localhost:8080 image matrix

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 124.8 MB

Tools

search tools

Get Data

PCA

- calculates a PCA

Pretreatments

- applies Detrend
- applies Savitsky-Golay
- applies Standard Normal Variate

Data visualization

- see spectra
- see observations 2D
- Bar chart for multiple columns
- Boxplot of quality statistics

Orthogonal Projections

Deconvolutions

Multitable/Multiway Analysis

Classifications

Regressions

- calculates a PLS Regression
- applies a regression model to a new set of spectra
- see references vs predictions

Workflows

- chemometric workflows

Chemoocs:

a Galaxy version dedicated to chemometrics

Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology department at Johns Hopkins University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins.

History

search datasets

Unnamed history
6 shown, 25 deleted, 6 hidden

11.4 MB

- 37: plot of observations
- 36: eigenvalues, p.cent
- 35: eigenvectors
- 34: scores
- 8: YNIR.csv
- 7: XNIR.csv

Notre version de Galaxy au 01/11/15:



Chemooocs : LesGrains x Multi-column and multi-r... x ProgSéminaireVidéo1 - G... x Galaxy x +

localhost:8080 Rechercher

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 337.9 MB

Tools

search tools

- Get Data
- Data management
- Plot figures
- PCA
- Pretreatments
- Orthogonal Projections
- Regressions
 - calculates a PLS Regression
 - applies a regression model to a new set of spectra
 - see references vs predictions
- Discriminations
- Deconvolutions
- Multitable/Multiway Analysis
- Fasta
- Workflows
 - chemometric workflows

calculates a PLS Regression (Galaxy Tool Version 0.0.1)

Options

Select X data

7: XNIR.csv
Dataset (n x p) containing the n spectra of p variables.

Select y data

8: YNIR.csv
Dataset (n x q) containing the reference values.

Column of Y chosen for the calculation

1

Number of blocs for cross-validation

10

Number of latent variables

20

Centering

1
1=centering 0=no centering

Execute

History

search datasets

Unnamed history
2 shown, 53 deleted, 6 hidden
224.5 MB

8: YNIR.csv	eye	pencil	x
7: XNIR.csv	eye	pencil	x

PLS regression with the standard (Wold's) algorithm

The figure represents the root-mean square error of calibration, RMSEC (blue) and the root-mean square error of cross-validation, RMSECV (green).

Wold et al. The multivariate calibration problem in chelistry solved by the PLS method. Matrix pencils. 1983

Notre version de Galaxy au 01/11/15:



Firefox Fichier Édition Affichage Historique Marque-pages Outils Fenêtre Aide

localhost:8080 image matrix

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Using 124.8 MB

Tools

search tools

- Get Data**
- PCA**
 - calculates a PCA
- Pretreatments**
 - applies Detrend
 - applies Savitsky-Golay
 - applies Standard Normal Variate
- Data visualization**
 - see spectra
 - see observations 2D
 - Bar chart for multiple columns
 - Boxplot of quality statistics
- Orthogonal Projections**
- Deconvolutions**
- Multitable/Multiway Analysis**
- Classifications**
- Regressions**
 - calculates a PLS Regression
 - applies a regression model to a new set of spectra
 - see references vs predictions
- Workflows**
 - chemometric workflows

Chemoocs:

a Galaxy version dedicated to chemometrics

Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology department at Johns Hopkins University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins.

python

Scilab

R

Fichiers :

- .xml
- .csv
- .png
- .mat

History

search datasets

Unnamed history
6 shown, 25 deleted, 6 hidden

11.4 MB

- 37: plot of observations
- 36: eigenvalues, p.cent
- 35: eigenvectors
- 34: scores
- 8: YNIR.csv
- 7: XNIR.csv

```
import sys,random
```

Ex :PARAFAC

```
""" configuration d'Octave """
```

```
from oct2py import Oct2Py;
```

```
oct=Oct2Py('/usr/bin/octave')
```

```
""" ajout du path pour la boite a outils """
```

```
oct.eval("addpath('/home/jcb/galaxy-  
master/tools/chemoocs/matlab/nway331')")
```

```
""" configuration de Scilab """
```

```
from scilab2py import Scilab2Py
```

```
sci=Scilab2Py('scilab')
```

```
""" execution de la bonne fonction csv: separateur champ = ';' separateur  
decimal = '.' """
```

```
sci.eval("exec('~'/galaxy-  
master/tools/chemoocs/scilab/glx_xstruct2xml.sci', -1)")
```

```
sci.eval("exec('~'/galaxy-  
master/tools/chemoocs/scilab/glx_xml2xstruct.sci', -1)")
```

```
def stop_err( msg ):
```

```
    sys.stderr.write( msg )
```

```
    sys.exit()
```

```
def __main__():
```

```
    """ lecture des parametres et ajustement du format """
```

```
    input2=int(sys.argv[2])
```

```
    input3=float(sys.argv[3])
```

```
    input4=int(sys.argv[4])
```

```
    input5=int(sys.argv[5])
```

```
    input6=int(sys.argv[6])
```

```
    """ passage des arguments dans Scilab et execution """
```

```
    """ scilab charge le fichier .xml de Galaxy et le sauvegarde au format  
Matlab """
```

```
    sci.push('file1',sys.argv[1])
```

```
    sci.push('out1', sys.argv[7])
```

```
    sci.eval('x=glx_xml2xstruct(file1);')
```

```
    sci.eval("savematfile('/home/jcb/galaxy-  
master/tools/chemoocs/matlab/killme.mat','x', '-v7') ")
```

```
""" passage des arguments dans Octave """
```

```
oct.push('file', sys.argv[1])
```

```
oct.push('nbrcomp', input2)
```

```
oct.push('converg', input3)
```

```
oct.push('init', input4)
```

```
oct.push('max_iters',input5)
```

```
oct.push('constr', input6)
```

```
oct.push('def_params',[0.000001,0,0,0,0,2500])
```

```
""" chargement des donnees et des arguments dans Octave """
```

```
oct.eval("load('/home/jcb/galaxy-master/tools/chemoocs/matlab/killme.mat','-  
v7');")
```

```
oct.eval('def_params(1)=converg;')
```

```
oct.eval('def_params(2)=init;')
```

```
oct.eval("def_params(6)=max_iters;")
```

```
oct.eval("constr2=constr*ones(1,3);")
```

```
oct.eval("if constr==1, constr2(1)=0; end ")
```

```
""" x est une structure ; mise au format hypermatrice => x2 """
```

```
oct.eval("n=length(fieldnames(x));")
```

```
oct.eval("[ni,qi]=size(x.plan1);")
```

```
oct.eval("x2=zeros(ni,qi,n);")
```

```
oct.eval("for i=1:n; name=sprintf('%s%s','plan',num2str(i)); name2=sprintf('x.  
%s',name); eval(sprintf('x2(:,i)=%s',name2));end")
```

```
""" application de Parafac """
```

```
oct.eval("res=parafac(x2,nbrcomp,def_params,constr2); clear x;")
```

```
""" valable uniquement pour les matrices 3D: """
```

```
oct.eval(" for i=1:3; name=sprintf('%s%s','plan',num2str(i)); name2=sprintf('x.  
%s',name); eval(sprintf('%s=res{ i}',name2)); end ")
```

```
""" sauvegarde par Octave au format Matlab """
```

```
oct.eval("save /home/jcb/galaxy-master/tools/chemoocs/matlab/killme.mat x")
```

```
""" ecriture au format xml par Scilab """
```

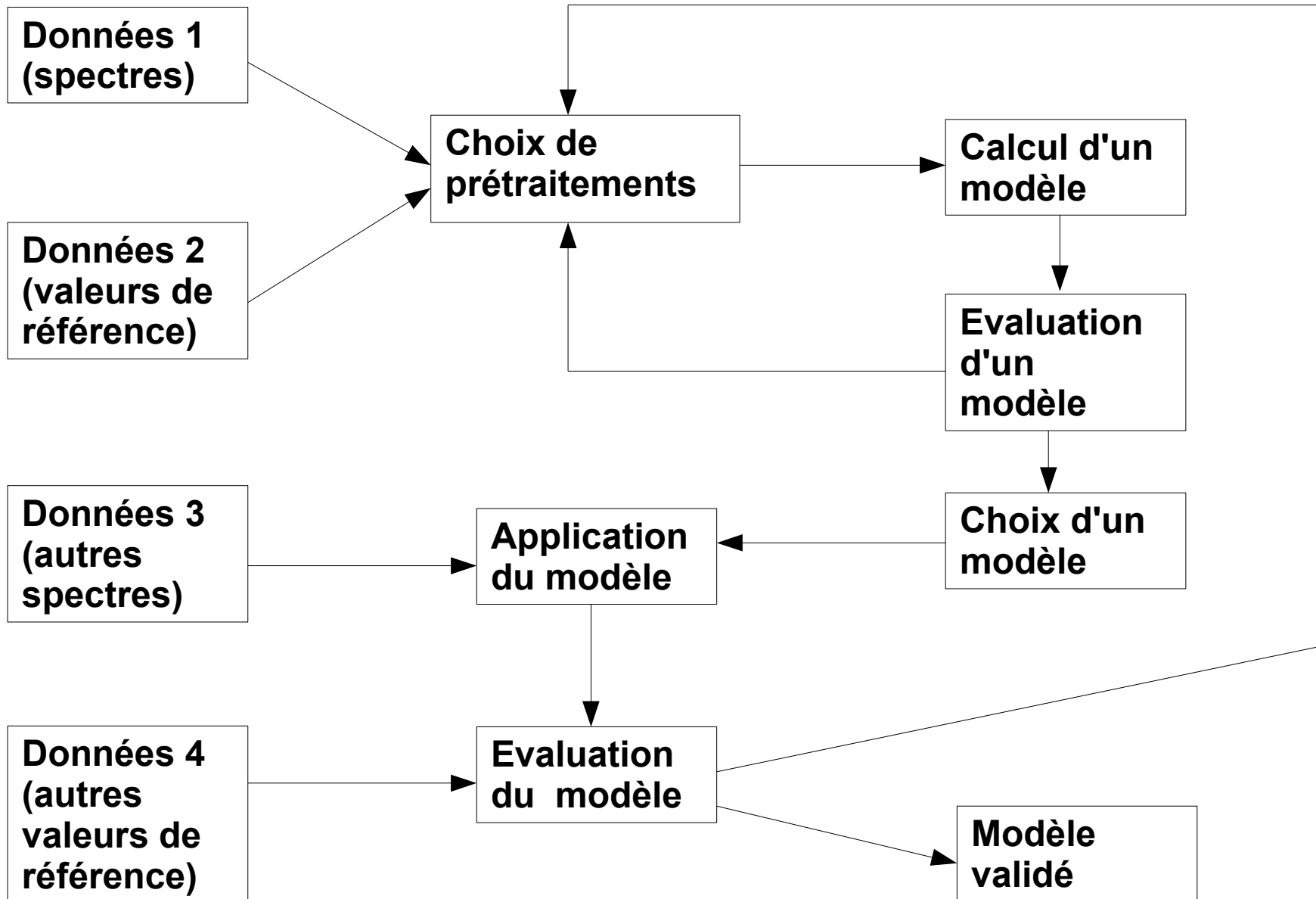
```
sci.eval("loadmatfile('/home/jcb/galaxy-  
master/tools/chemoocs/matlab/killme.mat','x')")
```

```
sci.eval("glx_xstruct2xml(x,out1)")
```

```
sci.eval("deletefile('/home/jcb/galaxy-  
master/tools/chemoocs/matlab/killme.mat')")
```

```
if __name__ == "__main__" : __main__()
```

Exemple de workflow en chimiométrie



Exemple de workflow sous Galaxy



Galaxy

Analyze DataWorkflowShared DataVisualizationAdminHelpUser

Using 131.2 MB

Tools

search tools

[Get Data](#)[Data management](#)[Plot figures](#)[PCA](#)[Pretreatments](#)[Orthogonal Projections](#)[Regressions](#)[Discriminations](#)[Deconvolutions](#)[Multitable/Multiway Analysis](#)[Fasta](#)

[Workflow control](#)

[Inputs](#)

Workflow Canvas | simplePLS

Details

Edit Workflow Attributes

Name:
simplePLS

Tags:

Apply tags to make it easy to search for and find items with the same tag.

Annotation / Notes:

Describe or add notes to workflow
Add an annotation or notes to a workflow; annotations are available when a workflow is viewed.

Quelques questions



Formats des fichiers

Download data directly from web or upload files from your disk

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
XNIR.csv	1.2 MB	Auto-det...	unspecified (?)		0%

Type (set all): Auto-detect Genome (set all): unspecified (?)

Choose local file Paste/Fetch data Start Pause Reset Close

**Inutile
A supprimer**

- supprimer de nombreux formats

- rajouter d'autres formats

ex : ENVI, MAT

Quelques questions (suite)



Graphiques

- interactivité possible avec Galaxy/Python

Diffusion

- Machine virtuelle de type Virtual Box
- Serveur / cloud ?

Quelques questions (fin)



Gains de temps

- processus un peu longs
- cause : lecture/écriture de fichiers, ouvertures de logiciels

3 finalités à Galaxy

- 1- enseignement via le MOOC
- 2- diffusion + évaluation de nouvelles méthodes
- 3- garantie de la qualité et l'inter-opérabilité des outils

Echéancier :



Novembre 2015 → Avril 2016

- construction du MOOC
- implémentation de Galaxy
 - + validation des méthodes
 - + robustification

Mai → Juin 2016

- tests du MOOC + Galaxy

Juillet → Août 2016

- dernières corrections

Septembre → Novembre 2016

- lancement sur FUN



Invitation :

MOOC en chimiométrie

à suivre sur FUN

Début : Septembre 2016

