

CHIMIOMETRIE XVII

CheMOOCs

la chimométrie pour tous, principes et outils



Namur, Belgique

du 17 au 20 Janvier 2016

Un projet collaboratif

Présenté par : Jean-Claude Boulet
INRA - UMR1083 SPO - Plateforme Polyphénols

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)

et co-porté par



data_frame



Chemooocs : 3 piliers pour diffuser la chimiométrie



Un MOOC
Massive Open Online Courses

gratuit et
accessible à tous



Une boîte à outils

de conception
collaborative



Une base de données

outil moderne de
communication

Le MOOC



Un MOOC
Massive Open Online Courses

FUN : plate-forme de diffusion

The screenshot shows the FUN website interface. At the top, there is a navigation bar with the FUN logo and the tagline "Se former en liberté". On the right side of the navigation bar, there are buttons for "Inscription" and "Connexion". Below the navigation bar is a large hero image featuring a young woman and a young man looking at a laptop together in a library setting. Overlaid on this image is the text "FUN : L'excellence de l'enseignement supérieur pour des cours en ligne, gratuits et ouverts à tous" and a large blue circle with the word "FUN" inside.

Below the hero image, the text "Les cours à la une" is displayed. Underneath, there are three course cards:

- SESSION 5**
Introduction à la statistique avec R
Université Paris-Sud
- NOUVEAU COURS**
PCSC - Modélisation des composites
ENS Cachan - Université Paris Saclay
- NOUVEAU COURS**
Introduction à la logique informatique - Partie 2 : calcul des prédicats
ENS Cachan - Université Paris Saclay

Exemples de MOOCs

NOUVEAU

Bases de données relationnelles : Comprendre pour maîtriser →



Inria

NOUVEAU

Soyez acteurs du web! →



Institut Mines-Télécom

NOUVEAU

Challenges et enjeux de la mobilité 2.0 →



Institut Mines-Télécom

NOUVEAU

Pratiques du Dimensionnement en Mécanique - Partie 2 →



ENS Cachan - Université Paris-Saclay

NOUVEAU

Problèmes économiques contemporains →



Université Panthéon-Assas (Paris 2)

NOUVEAU

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



ENS Cachan-Université Paris-Saclay - ENS Lyon

La composition du Mooc

TRONC COMMUN

Grain 1 : Introduction au Mooc - Jean-Claude Boulet

Grain 2 : Stats simples - Dominique Bertrand

Grain 3 : ACP 1 - Sébastien Preys

Grain 4 : ACP 2 - Sébastien Preys

Grain 5 : Prétraitement 1 - Jean-Michel Roger

Grain 6 : Classification non supervisée - Philippe Courcoux

Grain 7 : Régressions linéaires - Robert Sabatier

Grain 8 : PLS 1 - Douglas Rutledge

Grain 9 : PLS 2 - Douglas Rutledge

Grain 10 : Prétraitement 2 - Jean-Michel Roger

Grain 11 : bonnes pratiques de modélisation - Dominique Bertrand

Grain 12 : Discrimination 1 - Robert Sabatier

Grain 13 : Discrimination 2 - Robert Sabatier

Un tronc commun de 13 grains et 3 parcours « pour en savoir plus »

- des grains : une vidéo (10') et des exercices
- un forum
- une évaluation
- durée limitée : 8 semaines

EN SAVOIR PLUS

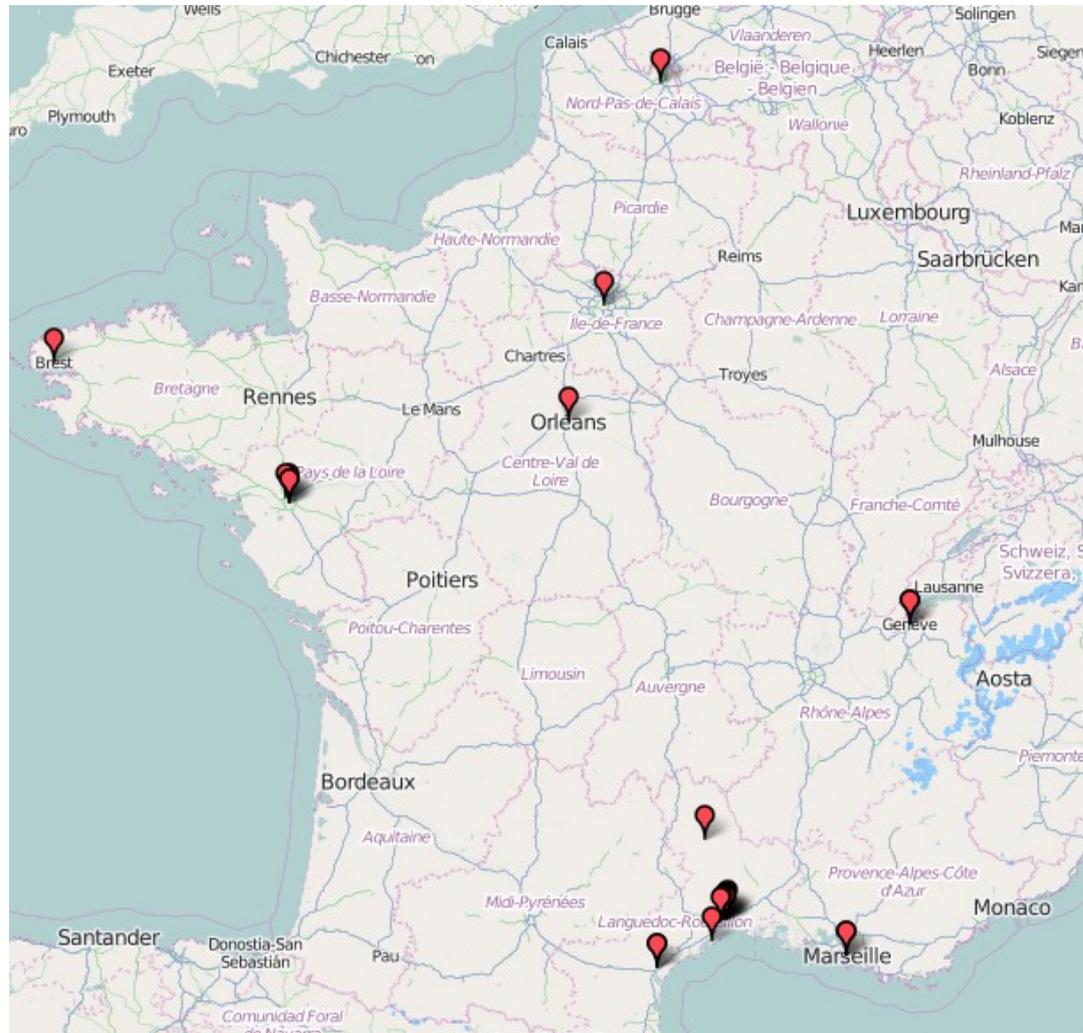
Grain 14 : Sélection de variables 1
Christelle ReynesGrain 15 : Sélection de variables 2
Dominique BertrandGrain 16 : Robustesse
Jean-Michel RogerGrain 17 : Multicurve resolution
Ludovic DuponchelGrain 18 : Independent component analysis
Douglas RutledgeGrain 19 : Structure des multitableaux
Mohamed HanafiGrain 20 : ACP concaténée
Mohamed HanafiGrain 21 : ACOM
Mohamed Hanafiattestation de suivi
succès

Les 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 Ecarnot
 Jean-Claude Boulet
 Eric Latrille

→ 20 contributeurs



Coordination pédagogique :
Cécile Tredaniel

Expert MOOC :
Christophe Lebegue

Expert utilisateurs :
Bernard Barthès

Réalisation/collaboratif :
Christian Resche
Giliane Granjean
Marc Laussens
Alain Prudhomme

L'outil



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

Expert informatique:
Pascal Neveu

L'outil initial : boite à outils FACT sous SCILAB

JC Boulet, JM Roger, D Bertrand
Issu de SAISIR / Matlab

Cahier des charges :

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

The screenshot displays the SCILAB 5.5.2 environment. On the left is a file navigator showing the directory structure of the FACT_plus tool. The central console window shows the execution of several commands, including matrix operations and error messages. On the right, a variable navigator window lists the current workspace variables with their names, values, types, and visibility.

```

344. 345. 346. 347. 348. 349. 350.
-1->l_idx2
l_idx2 =
    101.    350.
    351.    370.

-1->dimx(1)
ans =
    50.

-1->Fac
Fac =
    5.

-1->L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2)),dimx(1),Fac);
L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2)),dimx(1),Fac);
!--error 21
Index invalide.

-1->L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2))',dimx(1),Fac);
L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2))',dimx(1),Fac);
!--error 21
Index invalide.

-1->size(Factors)
ans =
    3.

-1->size(Factors(1))
ans =
    20.    5.

-1->size(Factors(2))
ans =
    50.    5.

-1->L1=matrix((l_idx2(1,1):l_idx2(1,2))',dimx(1),Fac);
-1->
  
```

| Nom | Valeur | Type | Visibilité |
|---------------|------------|--------|------------|
| L1 | 50x5 | Double | local |
| ans | [50, 5] | Double | local |
| l_idx2 | [101, 3... | Double | local |
| Fac | 5 | Double | local |
| dimx | [50, 4] | Double | local |
| l_idx2 | [101, 3... | Double | local |
| idd | [2, 3] | Double | local |
| ii | 3 | Double | local |
| startiter | 5 | Double | local |
| Constraint... | 0 | Double | local |
| conold | 2 | Double | local |
| connew | 2 | Double | local |
| it | 1 | Double | local |
| err | 1.35e+03 | Double | local |
| m | 2 | Double | local |
| Factors | N/A | Liste | local |
| C | 2e-06 | Double | local |
| F | 4x5 | Double | local |
| B | 50x5 | Double | local |
| A | 20x5 | Double | local |

```

Historique des commandes
ypréd_lineaire=e_val*d*b_lineaire;
repval_quad=var1(:,[1 2 3 8]);
repval_quad=repval_quad(1:28,:);
yref=repval_quad.d;
regplot(yref(:,4),ypréd_lineaire(:,4))
var1
var1.v(8)
clear
load('/Users/jean-claude/Documents/mac_v
load('/Users/jean-claude/Documents/mac_v
x_28juil15
m3g_pH143=csv2div('M3G_pH_143.csv')
help
m3g_pH143=csv2div('M3G_pH_143.csv')
m3g_pH3_49=csv2div('M3G_pH_349.csv')
m3g_pH3_49=csv2div('M3G_pH_379.csv')
m3g_2pH=[m3g_pH143 m3g_pH3_49]
curves(m3g_2pH)
toto=-log(m3g_2pH.d)/log(10);
curves(toto)
m3g_2pH.d=toto
  
```

L'outil initial : boite à outils FACT sous SCILAB

JC Boulet, JM Roger, D Bertrand
Issu de SAISIR / Matlab

Cahier des charges :

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

The screenshot shows the SCILAB 5.5.2 environment. On the left is a file navigator showing the directory structure of the FACT_plus tool. The central console window displays the following code and output:

```

-1->l_idx2
l_idx2 =
    344.    345.    346.    347.    348.    349.    350.
    101.    350.
    351.    370.

-1->dimx(1)
ans =
    50.

-1->Fac
Fac =
    5.

-1->L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2)),dimx(1),Fac);
L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2)),dimx(1),Fac);
!--error 21
Index invalide.

-1->L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2))',dimx(1),Fac);
L1=matrix(Factors(l_idx2(1,1):l_idx2(1,2))',dimx(1),Fac);
!--error 21
Index invalide.

-1->size(Factors)
ans =
    3.

-1->size(Factors(1))
ans =
    20.    5.

-1->size(Factors(2))
ans =
    50.    5.

-1->L1=matrix((l_idx2(1,1):l_idx2(1,2))',dimx(1),Fac);
-1->
  
```

On the right, the variable navigator shows a table of variables:

| Nom | Valeur | Type | Visibilité |
|---------------|------------|--------|------------|
| L1 | 50x5 | Double | local |
| ans | [50, 5] | Double | local |
| l_idx2 | [101, 3... | Double | local |
| Fac | 5 | Double | local |
| dimx | [50, 4] | Double | local |
| l_idx2 | [101, 3... | Double | local |
| idd | [2, 3] | Double | local |
| ii | 3 | Double | local |
| startiter | 5 | Double | local |
| Constraint... | 0 | Double | local |
| conold | 2 | Double | local |
| connew | 2 | Double | local |
| it | 1 | Double | local |
| err | 1.35e+03 | Double | local |
| m | 2 | Double | local |
| Factors | N/A | Liste | local |
| C | 2e-06 | Double | local |
| F | 4x5 | Double | local |
| B | 50x5 | Double | local |
| A | 20x5 | Double | local |

At the bottom right, the command history window shows the following commands:

```

ypred_lineaire=e_val*d*b_lineaire;
repval_quad=var1(:,[1 2 3 8]);
repval_quad=repval_quad(1:28,:);
yref=repval_quad.d;
regplot(yref(:,4),ypred_lineaire(:,4))
var1
var1.v(8)
clear
load('/Users/jean-claude/Documents/mac_w
load('/Users/jean-claude/Documents/mac_w
x_28juil15
m3g_pH143=csv2div('M3G_pH_143.csv')
help
m3g_pH143=csv2div('M3G_pH_143.csv')
m3g_pH3_49=csv2div('M3G_pH_349.csv')
m3g_pH3_49=csv2div('M3G_pH_379.csv')
m3g_2pH=[m3g_pH143 m3g_pH3_49]
curves(m3g_2pH)
toto=-log(m3g_2pH.d)/log(10);
curves(toto)
m3g_2pH.d=toto
  
```

L'outil initial : boite à outils FACT sous SCILAB

JC Boulet, JM Roger, D Bertrand
Issu de SAISIR / Matlab

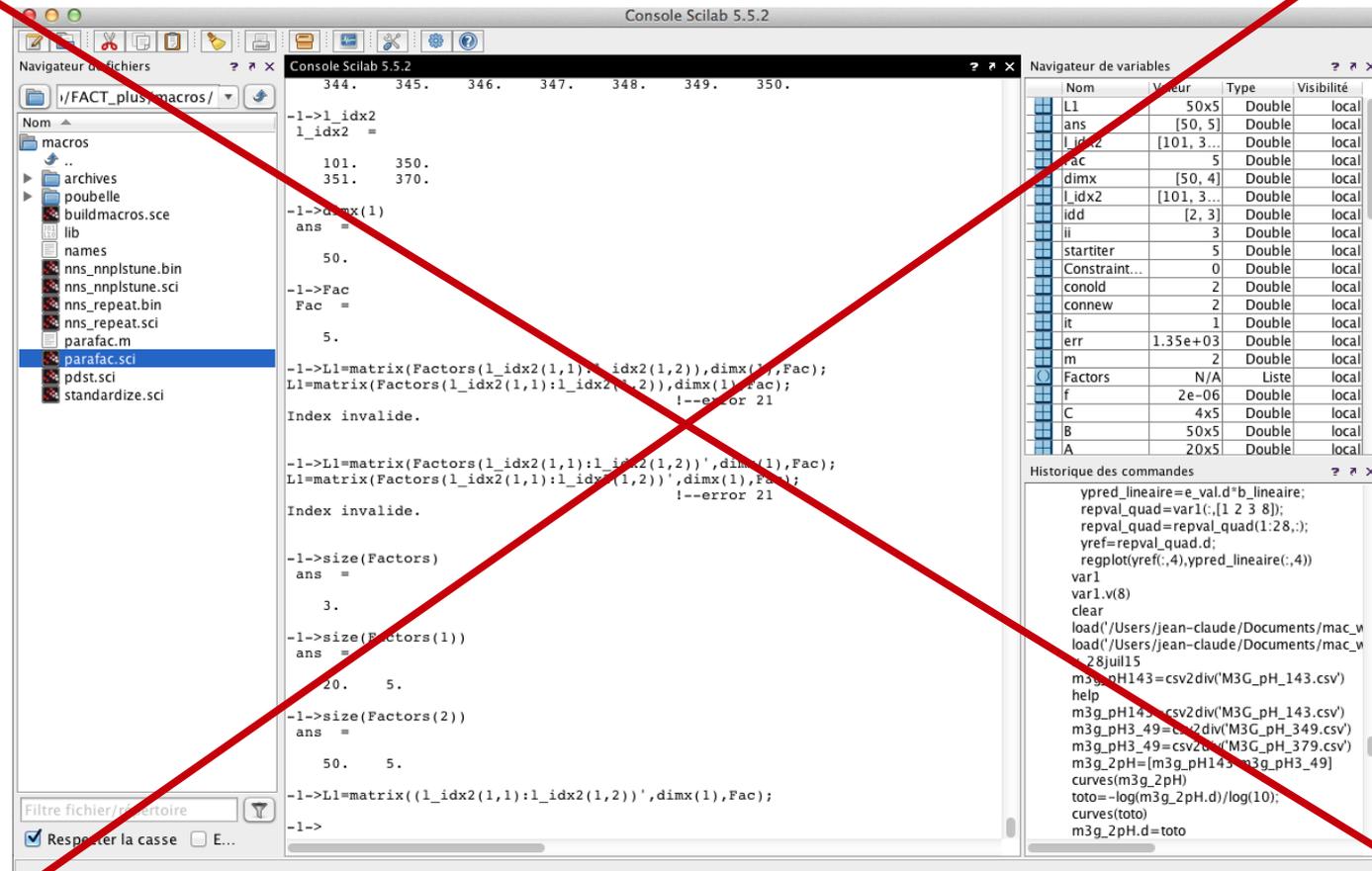
Cahier des charges :

- gratuit

- contient les méthodes
présentées dans le MOOC

- facile à utiliser
(compréhension)

- facile à utiliser
(accès, installation)



Galaxy : de grosses possibilités

GALAXY :

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

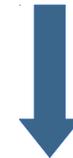
Plus :

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

GALAXY 

MOINS les fonctions de bioinformatique

PLUS les fonctions de chimométrie



ChemFlow

Galaxy au 31.12.2015

Firefox Fichier Édition Affichage Historique Marque-pages Outils Fenêtre Aide 100 % jeu. 10:41

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 Savisky-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 | 👁️ | ✎ | ✕ |

Démonstration de la toute dernière version dans le hall d'entrée

Galaxy au 31.12.2015

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

7: XNIR.csv

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.

Galaxy au 31.12.2015

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

History

search datasets

Unnamed history

6 shown, 25 deleted, 6 hidden

11.4 MB

37: plot of observations

36: eigenvalues, p.cent

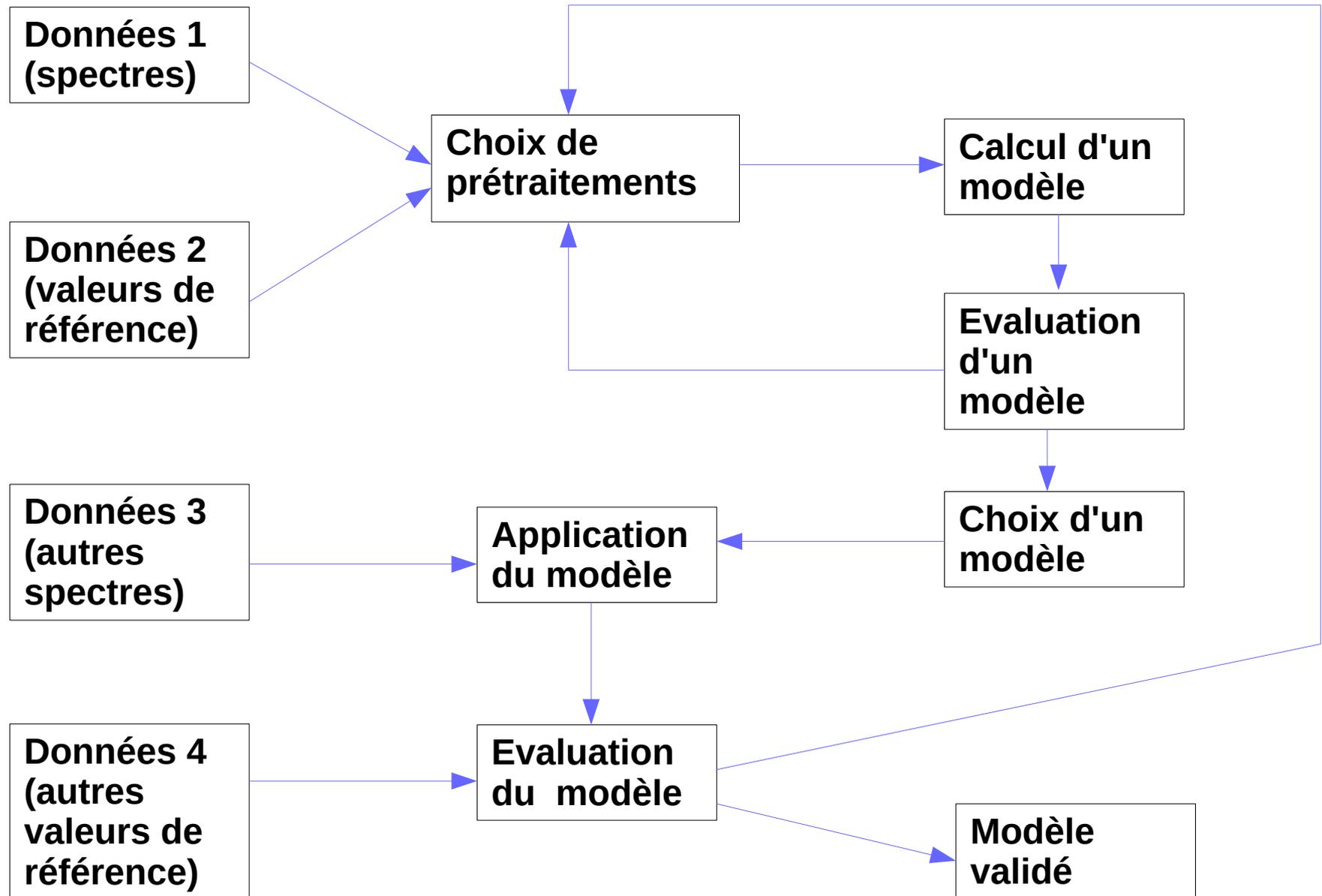
35: eigenvectors

34: scores

8: YNIR.csv

7: XNIR.csv

Exemple de WorkFlow en chimiométrie



Exemple de WorkFlow sous Galaxy

Galaxy Analyze Data **Workflow** Shared Data Visualization Admin Help User Using 131.2 MB

Tools Workflow Canvas | simplePLS Details

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

Input dataset x
output

applies Savitsky-Golay x
Select X data
outfile1

calculates a PLS Regression x
Select X data
Select y data
outfile1 (csv)
outfile2 (csv)
outfile3 (csv)
outfile4 (xml)

Input dataset x
output

applies a regression model to a new set of spectra x
Select the regression model
Select X data
ypredtest (csv)

Input dataset x
output

see references vs predictions x
Select Yref
Select Ypred
outfigure1 (png)

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.

2 missions pour ChemFlow

1. Enseignement via le Mooc
2. Diffusion + évaluation de nouvelles méthodes

Échéancier du projet

Novembre 2015 → Avril 2016

- Création du contenu du Mooc
- Conception et développement de ChemFlow

Mai → Juin 2016

- Bêta-tests du Mooc et des outils
- Lancement des inscriptions pour le Mooc

Juillet → Août 2016

- Dernières corrections

Septembre → Novembre 2016

- Lancement et diffusions sur FUN

Pour en savoir plus

Démonstration de « ChemFlow »

Objectif :

Donner un *aperçu* de ce que sera ChemFlow

Mais nos moyens sur place sont limités :

- ChemFlow : en tout début de développement
- Utilisation de nos outils (réseau, ordinateurs)
- Pas d'utilisation de matériel spécialisé (ex : serveur)

Dans le hall
face à l'accueil

Merci