

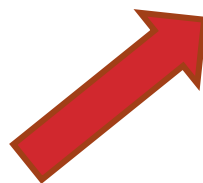
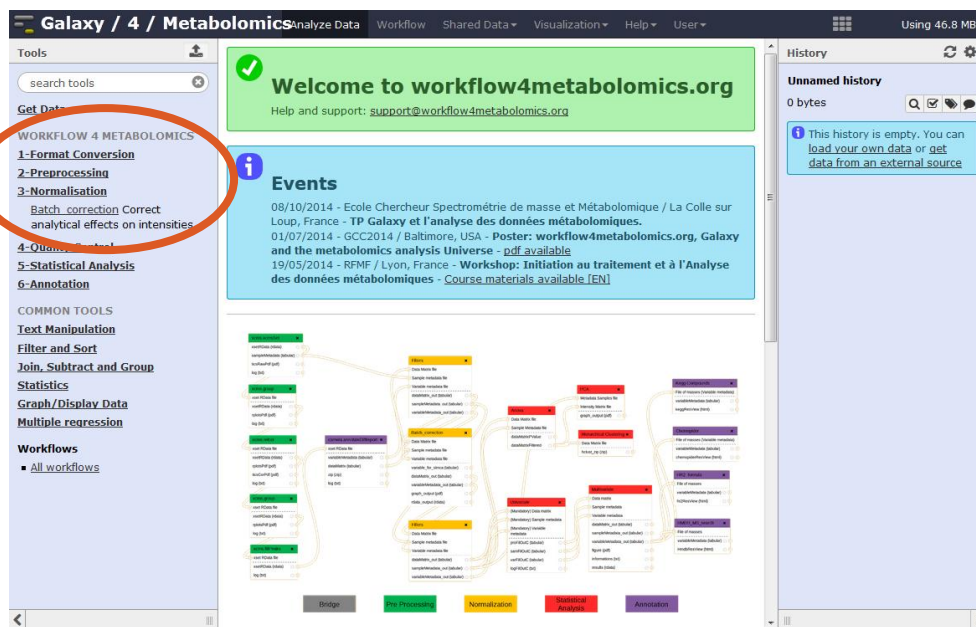
# **HOW TO PERFORM DRIFT AND BATCH CORRECTION?**

**W4M Core Team**

The Batch correction module provides correction of analytical effects inter and intra batch on intensities, using quality-control pooled samples (QC-pools), according to the algorithm mentioned by Van der Kloet (J Prot Res 2009).

This module is accessible via the left panel:

- > WORKFLOW 4 METABOLOMICS
  - > 3-Normalisation
    - > Batch correction

To run this module, your data must fit particular specifications:

- The 3 input .tsv data files (dataMatrix, sampleMetadata and variableMetadata) must be in the required **W4M post-processing format** (see *How to format your data for post-processing?*)
- The data matrix must **not contain missing values**
- The sample meta-data file **must contain 3 specific columns with the following names:**
  - **batch**
  - **injectionOrder**
  - **sampleType**

See the Help section of the module for more information.



Note: if you have only one batch, you still need to provide the « batch » column, that will contain a constant value.

## Input files

Parameter : num + label	Format
1 : Data Matrix file	tabular
2 : Sample metadata file	tabular
3 : Variable metadata file	tabular

### Data Matrix file must contain the intensity values of variables.

First line must contain all the samples names


First column must contain all the variables id

### Sample metadata file must have at least the three following columns :

"batch" to identify the batches of analyses ; need at least 3 pools for linear adjustment and 8 for lo(w)ess adjustment

"injectionOrder" integer defining the injection order of all samples : QC-pools and analysed samples

"sampleType" indicate if defining a samples or a pool

 NO MISSING DATA are allowed

Parameters are described in the Help section.

## Parameters

### Factor of interest

factor name (column header) that will be used as a categorical variable for plots and PCA.  
(often a biological factor ; if none, leave "batch")  
This factor does not affect correction calculation.

### Type of regression model :

To choose between *linear*, *lowess* and *loess* regression.  
Define which model type should be used in Van der Kloet correction algorithm concerning QC-pools regression.

### Level of details for plots

*basic* : PCA + CV boxplot (before and after correction)  
*standard* : 'basic' plots + before/after-correction plots of intensities and design effects for each ion  
*complete* : 'standard' plots + QC-pools regression plots per batch with samples intensities

### Note:

The choice of the type of regression model is left to the expert assessment of the user:

- the module "**Determine batch correction**" flags variables for which correction is not appropriate (see slide #8)
- on an indicative basis, one may consider using « linear » when pools intensities vary linearly according to the injection order, and « loess » otherwise;

Still the user is the only judge regarding which model is the more appropriate for his data.

Galaxy / METABO
Analyze Data Workflow Shared Data Visualization Help User
Using 33.0 MB

**Tools**

search tools

**Get Data**

WORKFLOW 4 METABOLOMICS

**2-Preprocessing**

**3-Normalisation**

[Batch\\_correction](#) Correct analytical effects on intensities

**4-Quality Control**

[Filters](#) Removes columns/rows according to : CV, test pvalues, POOL, Factors

[Determine\\_batch\\_correction](#) to choose between linear, lowess and loess methods

**5-Statistical Analysis**

**6-Annotation**

COMMON TOOLS

**Text Manipulation**

[Filter and Sort](#)

[Join, Subtract and Group](#)

[Statistics](#)

[Graph/Display Data](#)

[Multiple regression](#)

**Workflows**

- All workflows

## Batch\_correction (version 2014-04-24)

**Data Matrix file :**  
4: Filters\_sacuri\_dataMatrix.tsv

**Sample metadata file :**  
5: Filters\_sacuri\_sampleMetadata.tsv  
must have at least the three following columns : 'batch' + 'injectionOrder' + 'sampleType'

**Variable metadata file :**  
6: Filters\_sacuri\_variableMetadata.tsv

**Factor of interest :**  
gender

**Type of regression model :**  
loess

**Level of details for plots :**  
basic

[Execute](#)

**1** **Authors** Franck Giacomoni and Marion Landi (for interface and wrapper) and Jean-Francois Martin and Melanie Petera (for R)

**1** **Please cite** If you use this tool, please cite [F.M. Van Der Kloet, I. Bobeldijk, E.R. Verheij, R.H. Jellema. \(2009\). "Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping." Journal of Proteome Research p5132-5141](#)

### Batch\_correction

#### Description

Correction of analytical effects inter and intra batch on intensities using quality control pooled samples (QC-pools) according to the algorithm mentioned by Van der Kloet (J Prot Res 2009).

#### Workflow position

**History**

**RFMF - TP2 - Statistics**

2.2 MB

6: [Filters\\_sacuri\\_variableMetadata.tsv](#)

5: [Filters\\_sacuri\\_sampleMetadata.tsv](#)

4: [Filters\\_sacuri\\_dataMatrix.tsv](#)

3: [sacuri\\_variableMetadata.tsv](#)

2: [sacuri\\_sampleMetadata.tsv](#)

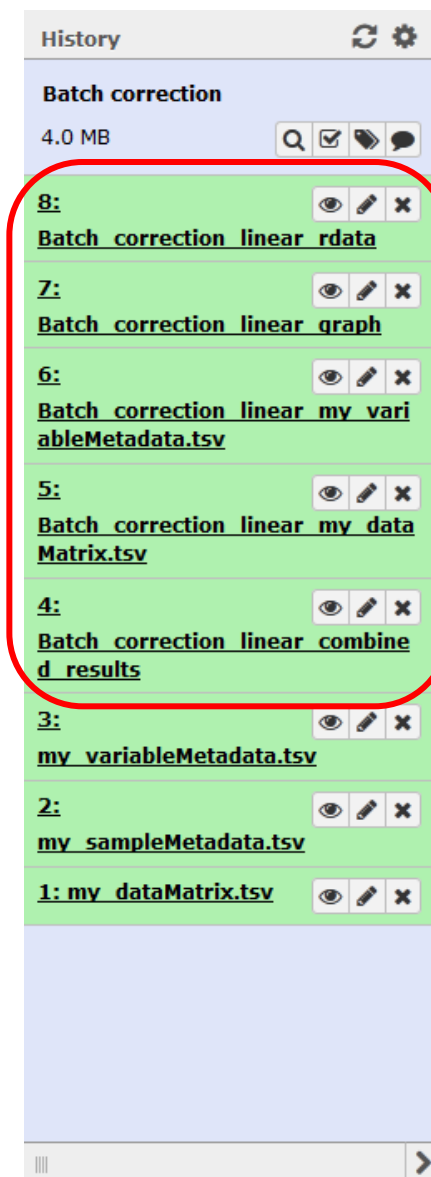
1: [sacuri\\_dataMatrix.tsv](#)

## Data inputs

Data inputs

The module generates 5 files:

- An Rdata file containing the results for further export and processing with R
- A pdf file containing diagnosis plots of the correction
  - The level of detail can be selected as one of the arguments of the module
- Two tables corresponding to the data matrix and the variable meta-data file
  - The data matrix contains the new intensities
  - The variable meta-data table is the original one plus columns indicating if intra-batch correction as been applied for each ion and each batch (0=no, 1=yes)
- A combined table that corresponds to the junction of the sample meta-data file and the data matrix of corrected intensities



Galaxy / METABO

Analyze Data Workflow Shared Data Visualization Help User Using 36.1 MB

Tools

search tools

Get Data

WORKFLOW 4 METABOLOMICS

2-Preprocessing

3-Normalisation

Batch\_correction Correct analytical effects on intensities

4-Quality Control

5-Statistical Analysis

6-Annotation

COMMON TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Statistics

Graph/Display Data

Multiple regression

Workflows

- All workflows

mzmax	rtmed	rtmin	rtmax	npeaks	bio	blank	isotopes	adduct	pcgroup	norm.b1
96.96055802	60.66063731	47.58330236	72.91457193	84	30	0		[M-H]- 97.9672	57	1
98.95569154	60.65002621	47.58330236	68.16608993	85	30	0			57	1
135.0297374	54.09586485	52.0061208	64.67221155	35	30	4	[13][M]-		81	1
136.0329949	54.26549421	52.0061208	54.86304067	23	23	0	[13][M+1]-		81	1
187.037491	53.19310488	51.5001856	53.97122548	29	29	0		[M+Cl]- 152.067	138	1
189.0345776	53.29490329	51.5001856	54.21757457	30	30	0			413	1
256.0589882	54.5664543	53.06384694	55.55977854	19	19	0			81	1
111.0086463	63.7899207	60.71821949	65.02074055	30	30	0	[6][M]-	[M-H]- 112.017	4	1
99.04505157	318.0022226	311.9503382	322.4209864	34	30	0	[3][M]-	[M-H-CO2]- 144.045	164	0
145.061749	349.0726556	342.9844247	365.437733	34	30	2			5	1
405.1915084	463.9998039	462.403398	464.6430272	29	29	0	[219][M]-	[M-H+HCOOH]- 360.193	175	1
132.0301813	55.93629944	53.97744616	56.97515728	30	30	0			109	1
263.1032901	348.9179671	348.0183655	350.3571702	35	30	5	[121][M]-	[M-H]- 264.111	5	1
288.9868076	64.78969478	61.54127858	66.05907556	30	30	0			167	1
264.1064523	348.9415214	347.4534425	350.8255902	34	30	4	[121][M+1]-		5	1
167.0208833	63.67500027	60.71821949	64.9407069	32	27	5		[3M-2H]2- 112.017	4	1
85.02948246	63.79149641	60.71821949	65.14071555	30	30	0			4	1
288.8161154	51.11847144	49.70497776	51.87284865	12	12	0			44	1
194.9271839	66.16606849	48.139931	71.88939742	38	30	0		[M+Cl]- 159.963	219	1
285.0826499	57.38433388	55.65326942	58.52112417	24	24	0			71	1
100.0038593	293.4379771	291.649447	294.8320576	30	30	0			33	1
							[121][M+2]-		5	1
									129	1
							[13][M+2]-		81	1
							[47][M]-		164	0
								[M-H+NaCOOH]- 223.106 [M-H]- 291.095	406	1
							[90][M]-		253	1
									188	1
									103	1
									129	1
								[M+Cl]- 193.881	129	1
							[6][M+1]-		4	1
								[M+Cl]- 57.9559	129	1
									188	1
									14	1
									11	1
							[10][M]-	[M-H-CO2]- 174.016	11	1
								[M+Cl]- 159.08	33	1
									233	1
							[12][M]-	[M-H-CO2]- 179.058	2	1
							[162][M]-	[M-H]- 311.122	104	1

History

RFMF - TP2 - Statistics

5.3 MB

11: Batch\_correction loess\_rdata

10: Batch\_correction loess\_graph\_pdf

9: Batch\_correction loess Filters\_sacuri\_variableMetadata.tsv

8: Batch\_correction loess Filters\_sacuri\_dataMatrix.tsv

Z: Batch\_correction loess\_result\_for\_simca\_tsv

6: Filters\_sacuri\_variableMetadata.tsv

5: Filters\_sacuri\_sampleMetadata.tsv

4: Filters\_sacuri\_dataMatrix.tsv

3: sacuri\_variableMetadata.tsv

2: sacuri\_sampleMetadata.tsv

1: sacuri\_dataMatrix.tsv

Graphical output

Variable meta-data

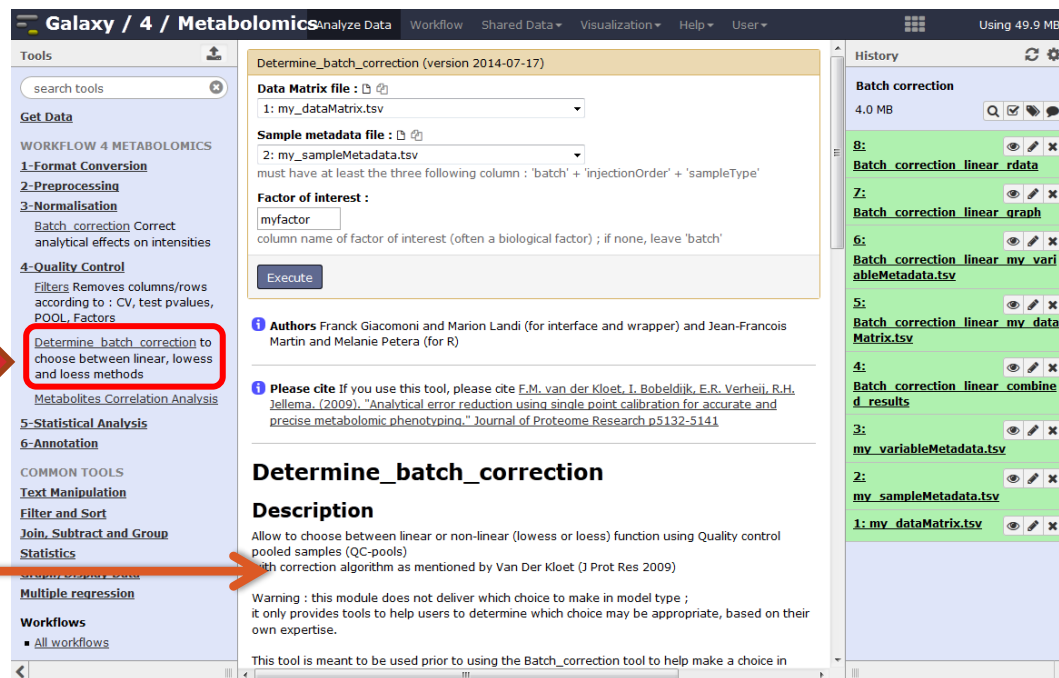
Normalised intensities (data matrix)

Indicator for intra-batch normalisation (added in variable meta-data table)

To help you determine which type of regression model you should use with your data, you can use the “Determine batch correction” module.

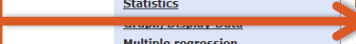
- It generates plots of intensities according to injection order and design effects for each ion.
- Each type of regression is displayed on intensity plots to enable comparison between the methods.
- A table indicating whether intra-batch correction is possible for each ion is generated.

4 – Quality Control  
Determine batch correction

The screenshot shows the Galaxy web interface for the 'Determine\_batch\_correction' tool. The tool is part of the 'Metabolomics' workflow. The configuration panel includes fields for 'Data Matrix file' (1: my\_dataMatrix.tsv), 'Sample metadata file' (2: my\_sampleMetadata.tsv), and 'Factor of interest' (myfactor). The 'Execute' button is visible. The history panel on the right shows a sequence of tools: 'Batch correction', 'Batch correction linear\_rdata', 'Batch correction linear\_graph', 'Batch correction linear\_mv\_variableMetadata.tsv', 'Batch correction linear\_mv\_dataMatrix.tsv', 'Batch correction linear\_combine\_d\_results', 'my\_variableMetadata.tsv', 'my\_sampleMetadata.tsv', and 'my\_dataMatrix.tsv'. The tool description and authors information are also visible.

See the Help section of the module for more information





F.M. Van Der Kloet, I. Bobeldijk, E.R. Verheij, R.H. Jellema (2009).  
"Analytical error reduction using single point calibration for  
accurate and precise metabolomic phenotyping." Journal of  
Proteome Research p5132-5141