Orthogonal Projections in the Row and the Column Spaces

JM Roger, Irstea, France

Jean-michel.roger@irstea.fr

Hypothesis, notations

- The data :
  - **X**(n,p) the predictors
  - **Y**(nxq) the responses, to be estimated from **X**
  - We will specifically address the highly multivariate case, like spectra
  - *p* is currently greater than *n* : ill-dimensionning
  - **X** variables are highly correlated : ill-conditionning

R^p

– The true data space is very much smaller than

Theory: row and column spaces

$$Col(\mathbf{X}) \subseteq \mathbb{R}^n$$

Columns or linear combinati

Space of the individuals , of the scores (PCA, PLS, etc.)

Used to compute statistics : Mean, Variance, Correlations

Col(**X**) and *Col*(**Y**) are included in the same space

$$\left[\mathbf{X} \right]_{(n,p)}$$

Theory: row and column spaces

$$Row(\mathbf{X}) \subseteq \mathbb{R}^p$$

Rows or linear combination of the rows

Space of :

- the spectra
- spectral features
- the loadings (PCA, PLS, etc.)

The space a new sample belongs to

→ Space of the model

$$\left[\mathbf{X} \right]_{(n,p)}$$

Theory : Orthogonal projections

Scores $\mathbf{T} \in Col(\mathbf{X})$

$$\mathbf{X}^* = \left(\mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\right)\mathbf{X}$$

Keeps the information related to **T**

$$\mathbf{X}^* = \left(\mathbf{I} - \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\right)\mathbf{X}$$

Removes the information related to **T**

Loadings $\mathbf{P} \in Row(\mathbf{X})$

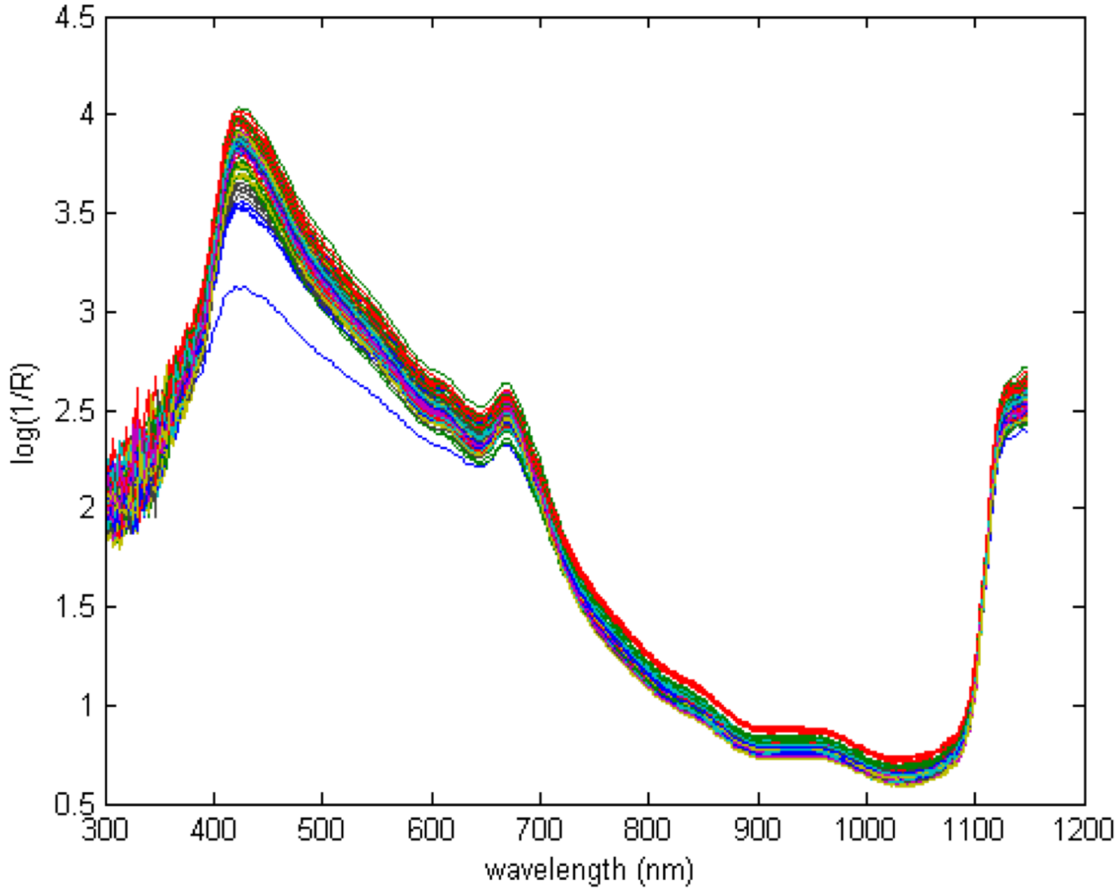
$$\mathbf{X}^* = \mathbf{X}\left(\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)$$

Keeps the features related to **P**

$$\mathbf{X}^* = \mathbf{X}\left(\mathbf{I} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)$$

Removes the features related to **P**

Example

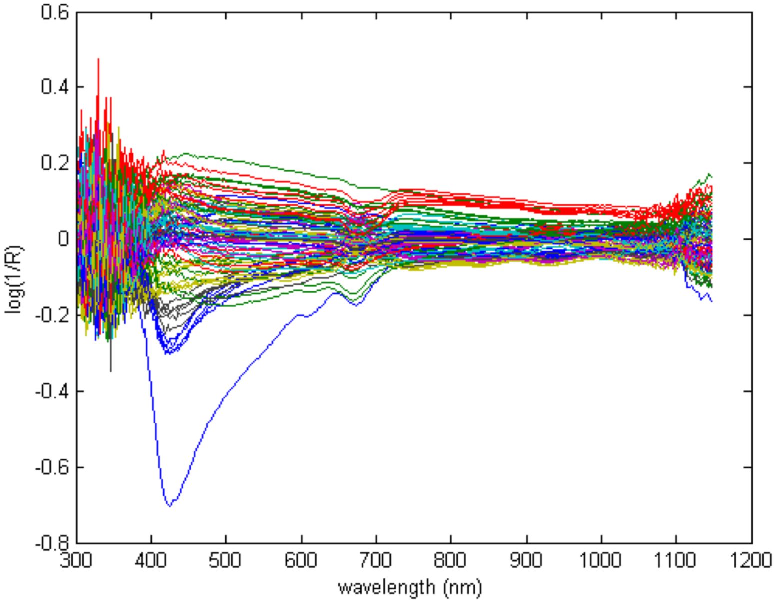


Visible / VNIR spectra of sillage

Examples of projections

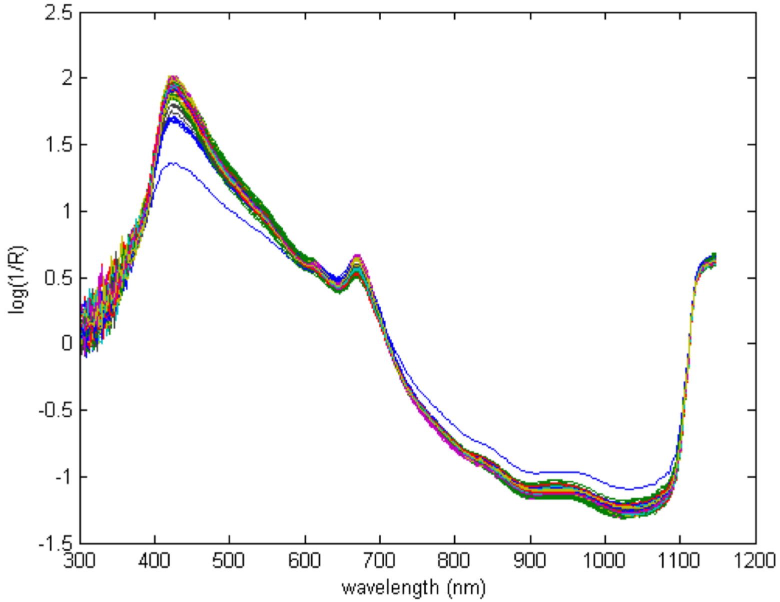
Classical mean centering data is a projection in *Col*(**X**) used to remove constant information and analyze data around the center of mass:

$$\mathbf{X}^* = (\mathbf{I} - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T)\mathbf{X}$$

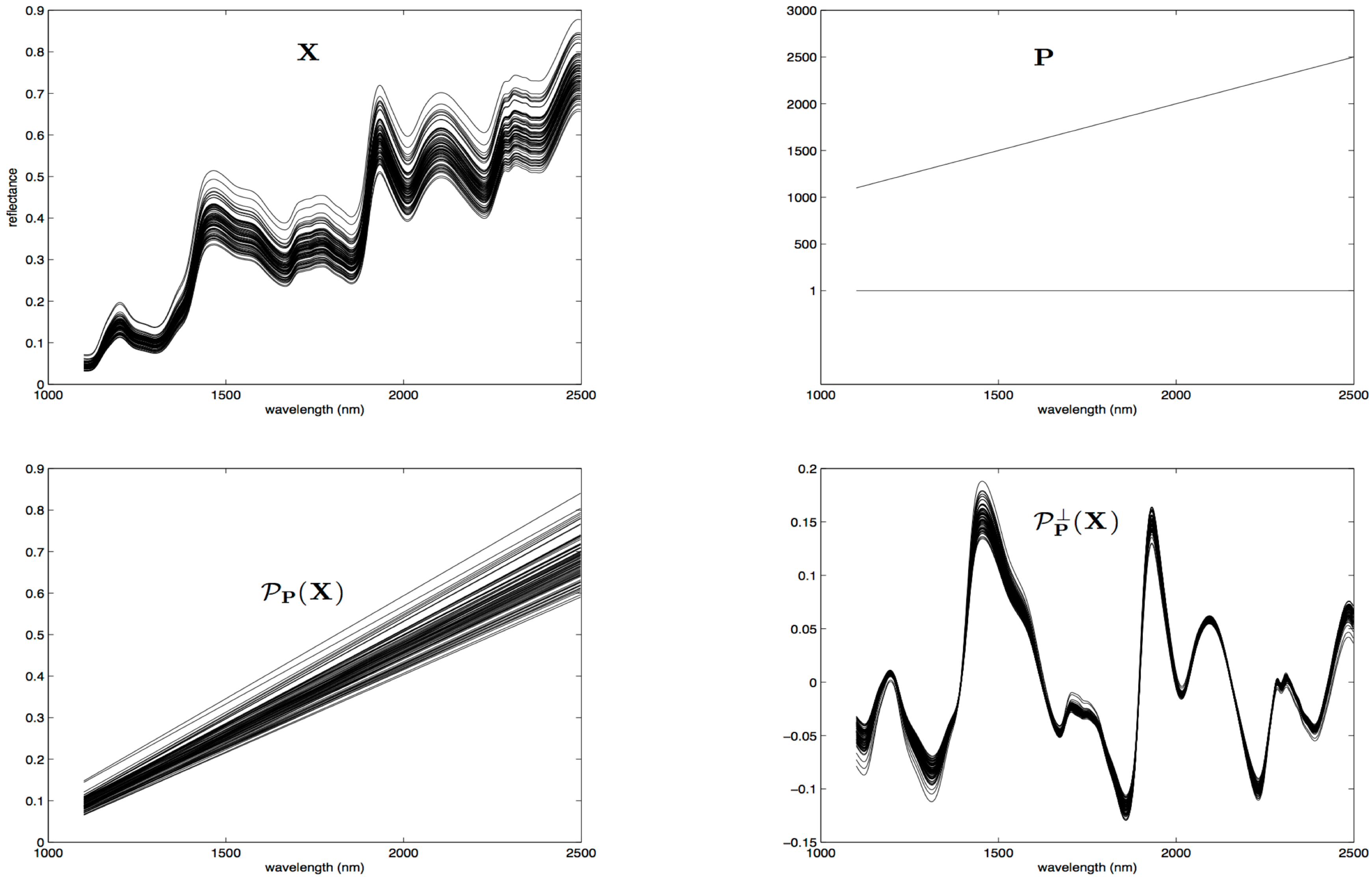


Spectra centering is a projection in *Row*(**X**) used to remove horizontal baselines
(vert translations):

$$\mathbf{X}^* = \mathbf{X}(\mathbf{I} - \mathbf{1}(\mathbf{1}^T\mathbf{1})^{-1}\mathbf{1}^T)$$



Examples of projections



First application:

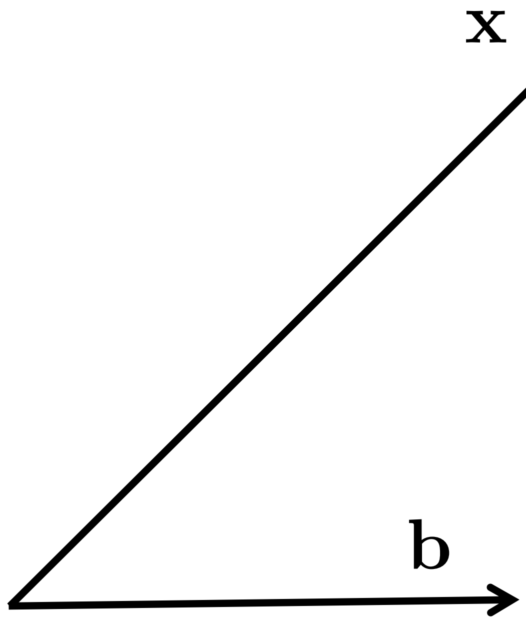
Calibration robustness

Removal of a dementrical effect in the spectral space by means of an orthogonal projection

Application to robustness

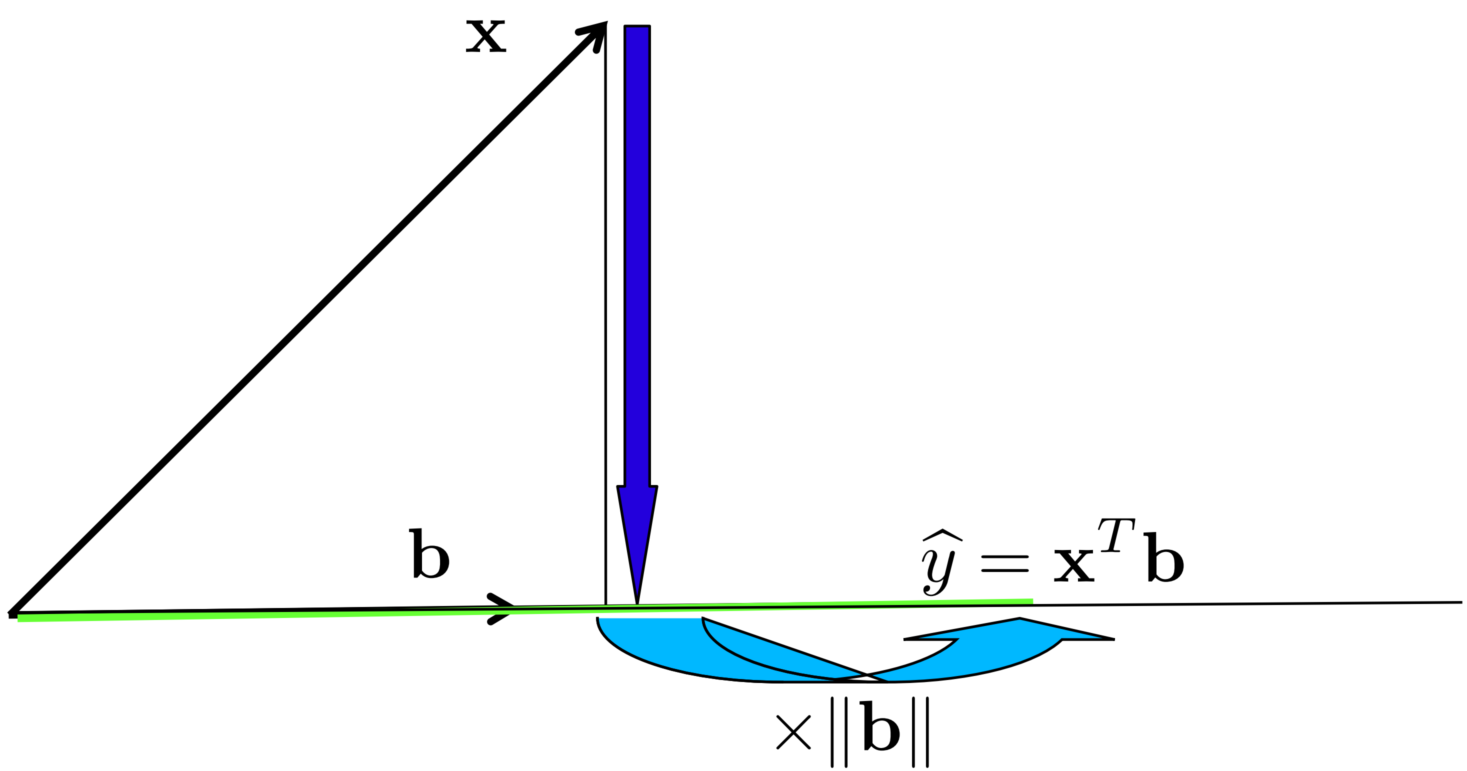
- The main context:
  - We have a calibration database and a model
  - We want to make this model robust against an influence factor variation not included in the database
- OP based methods remove the subspace responsible for the problem from *Row*(**X**).
- Examples of application:
  - Temperature of fruits
  - Moisture of soils
  - Calibration transfer

Geometrical point of view



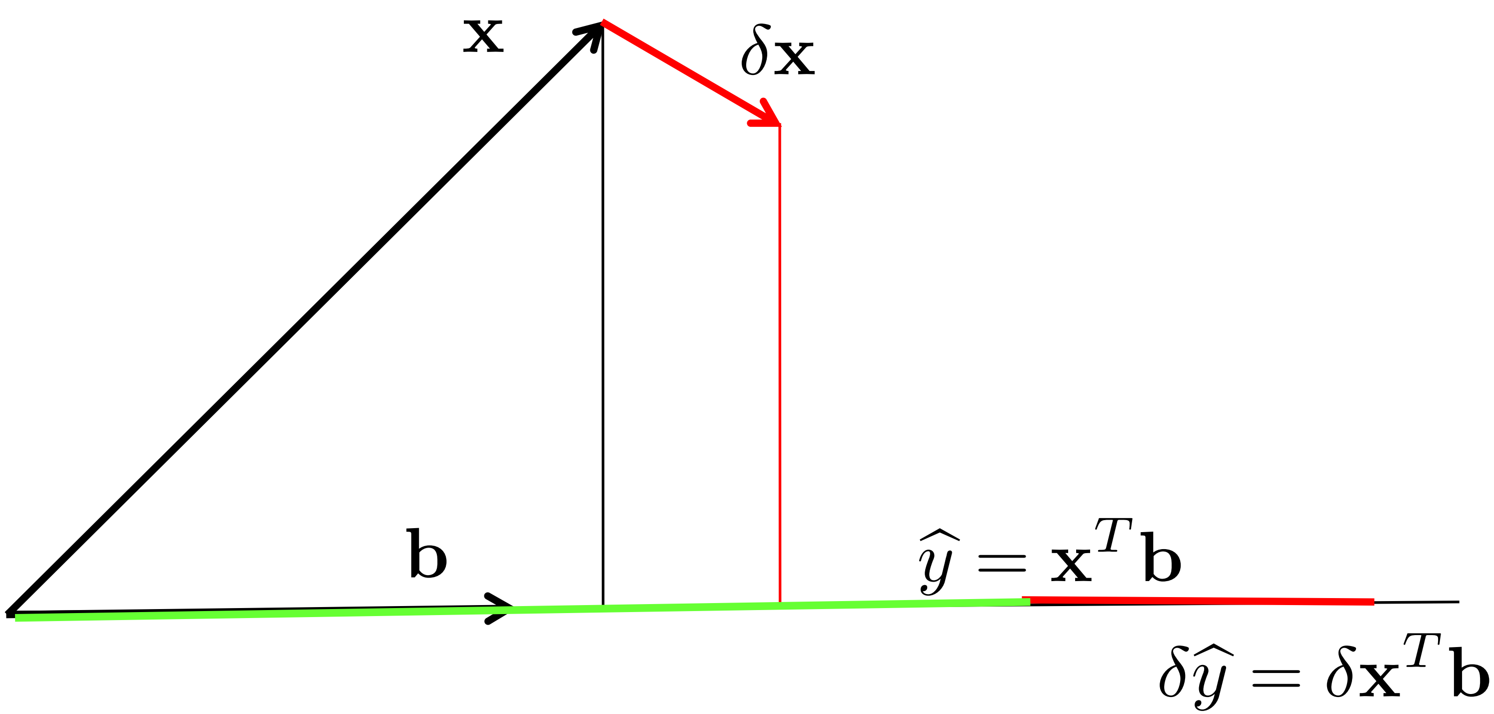
Let **b** be a model and **x** a spectrum

Geometrical point of view



Applying **b** on **x** relies on the inner product $\mathbf{x}^T\mathbf{b}$

Geometrical point of view



Any deviation $\delta\mathbf{x}$ produces a absolute error:

$$|\delta\widehat{y}| = \|\delta\mathbf{x}\| \times \|\mathbf{b}\| \times |\cos(\delta\mathbf{x}, \mathbf{b})|$$

Application to robustness

- 3 ways to lower $|\delta\widehat{y}| = \|\delta\mathbf{x}\| \times \|\mathbf{b}\| \times |\cos(\delta\mathbf{x}, \mathbf{b})|$
  - To lower $\|\delta\mathbf{x}\|$

    : prep $\|\mathbf{b}\|$ essing, standardisation
  - To lower $|\cos(\delta\mathbf{x}, \delta\mathbf{b})|$

    : parcimony of the model
  - To lower

    : orthogonalisation

$$\delta\mathbf{X}$$

- Orthogonalisation, the principle:
  - To express the loadings **P** spanned by
  - To project **X** orthogonal to **P** (in *Row*(X))

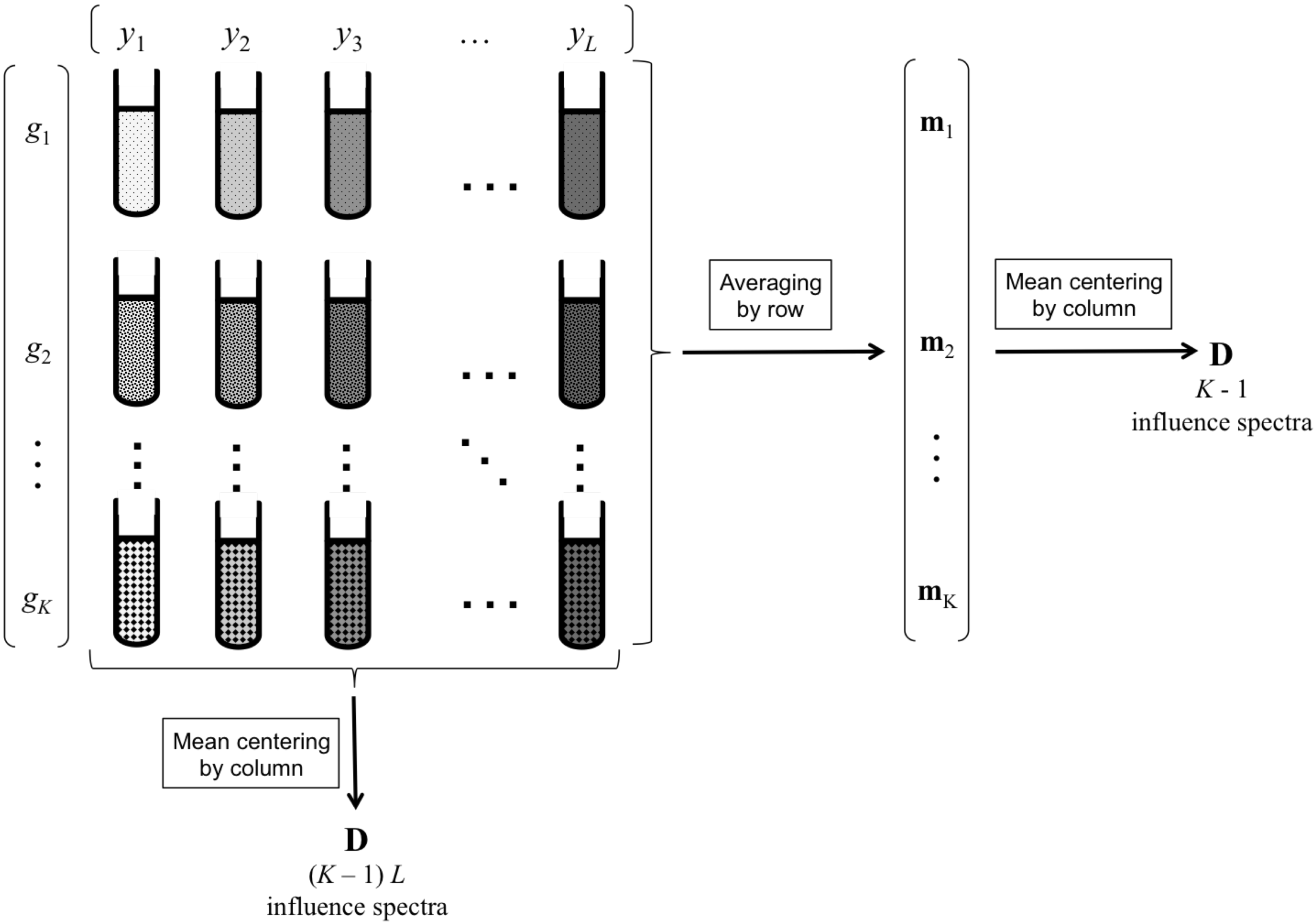
Application to robustness

- General process of OP-based methods :
  - To collect examples of $\delta\mathbf{x}$ into a matrix **D**
  - To identify the space spanned by **D**
    - By a non centered PCA
    - By an SVD
  - To put the *k* first loadings into **P**
  - To orthogonalise **X** vs **P** :

$$\mathbf{X}^{*} = \mathbf{X}\left(\mathbf{I} - \mathbf{P}^{T}\mathbf{P}\right)$$

  - To recalibrate the model
- All the OP based methods differ on **D** building

EPO
External Parameter Orthogonalisation



EPO
External Parameter Orthogonalisation

Application to temperature influence
on apple sugar prediction

Test results without EPO

Test results with EPO (*k*=4)

Pros and cons of OP methods

- Pros
  - The model becomes independent from $\delta \mathbf{x}$
  - The model still works if $\delta \mathbf{x}$ disappears
  - The correction is embedded into the model
  - Several influences could be managed
  - Loadings **P** are meaningful
  - Existing databases could be processed with a few extra spectra
- Cons
  - One parameter (K) to be tuned

Second application:

Variable selection

CovSel method

CovSel objective

- To select X variables which are :
  - the closest to a multivariate Y
  - as indepedent as possible
- Y could contain :
  - q quantitative responses
  - The dummy encoding to q classes

CovSel principle

1Project **X** on **Y**

To focus on the common information between **X** and **Y**

In *Col*(**X**)

CovSel principle

1Project **X** on **Y**

2Calculate the variance covariance matrix

To quantify the information carried by **X**

In *Row*(**X**)

CovSel principle

1Project **X** on **Y**

2Calculate the variance covariance

3Find the maximum of the diagonal

4

To find the more informative variable in **X**

In *Row*(**X**)

CovSel principle

1Project **X** on **Y**

2Calculate the variance covariance

3Find the maximum of the diagonal

4Project **X** and **Y** orthogonally to the selected x_i

To remove the information carried by the selected variable

In *Col*(**X**)

CovSel principle

1Project **X** on **Y**

2Calculate the variance covariance

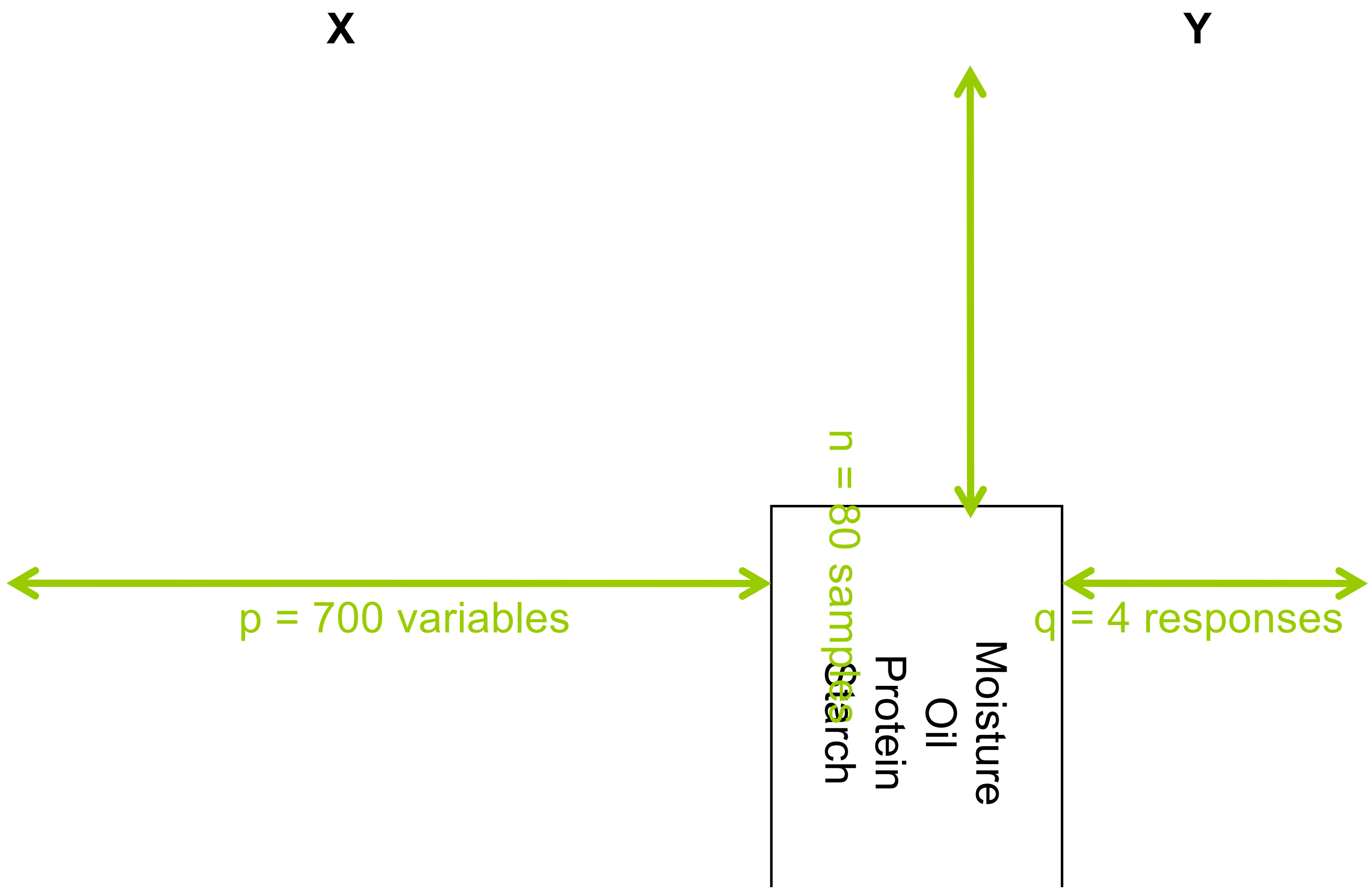
3Find the maximum of the diagonal

4Project **X** and **Y** orthogonally to the selected x_i

5Go to 1

CovSel : Corn dataset

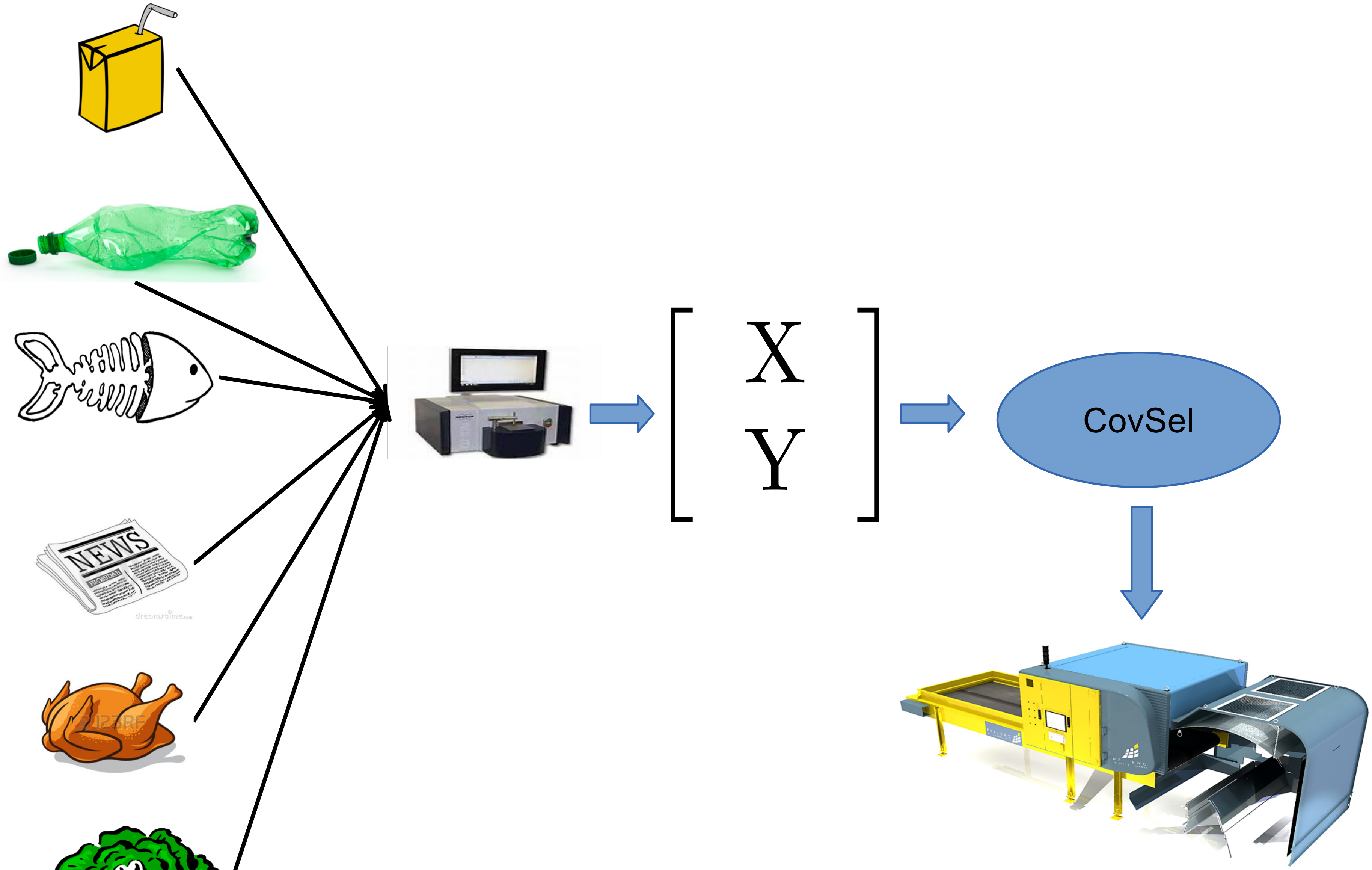
- Data from Eigenvector web site :
  - http://software.eigenvector.com/Data/Corn/index.html



CovSel : Corn

Covsel : Corn

Covsel : waste sorting



Waste sorting machine

Covsel : waste sorting

Some perspectives

- When data are structured in *Col*(**X**) :
  - design of experiment (**Y**)
- And are structured in *Row*(**X**) :
  - spectra

- How to combine projections in *Col*(**X**) and *Row*(**X**) ?

- Extension to
  - temporal data

Thank you
for your kind attention