PLS

Theory, Algorithm, Practical work

Marina Cocchi

Dipartimento di Scienze Chimiche e Geologiche

marina.cocchi@unimore.it

Outline

- Recall Regression
- What's PLS
- PLS in practice
- Algorithms (just a bit)
- Examples, Practical work

Linear Modeling

Linear modelling has been developed in “pre-computer” era (minor computational complexity)

Anyhow there are good reasons to use them “today”:

- ✓ *they are simple*
- ✓ *less prone to overfitting*
- ✓ *predictive capability can be better w.r.t non-linear, e.g. when data sets are characterised by limited number of samples, high noise, missing data*
- ✓ *They could be applied after data transformation to moderate non-linear raw data (eg: taking log)*

$$y = f(x)$$

- **UNILINEAR** ONE y as a function of ONE x

x	y
3	6
4	8
2,5	5
1	2

Want to find a general expression to obtain y from x

$$y = ?\,x$$

We see that

$$3 \cdot b = 6$$
$$4 \cdot b = 8$$

easy to see that

$$b = 2$$

and so on ...

CHEMIOMETRIA

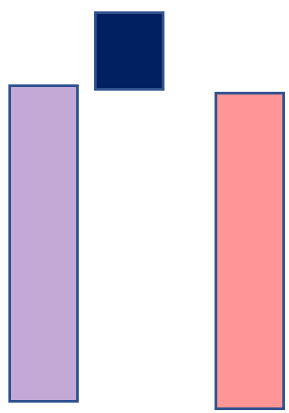
$$y = f(x)$$

- **UNILINEAR** ONE y as a function of ONE x

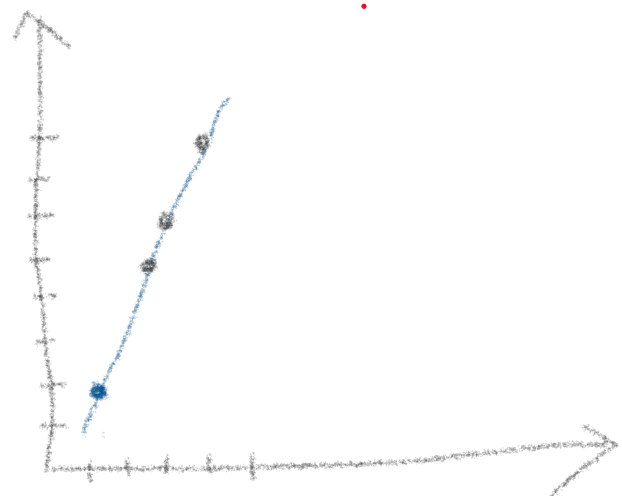
a bit of notation

x	y
3	6
4	8
2,5	5
1	2

x **y**



$\mathbf{x}\,b = \mathbf{y}$



CHEMIOMETRIA

$$y = f(x)$$

- **UNILINEAR** ONE y as a function of ONE x

What about noise

x	y	
3	6	+1
4	8	− 0,5
2,5	5	− 0,5
1	2	+1,5

x y

$$x b + e = y$$

?

Now we have some residuals

Minimize them ... > Least squares

a generic multilinear model (MLR):

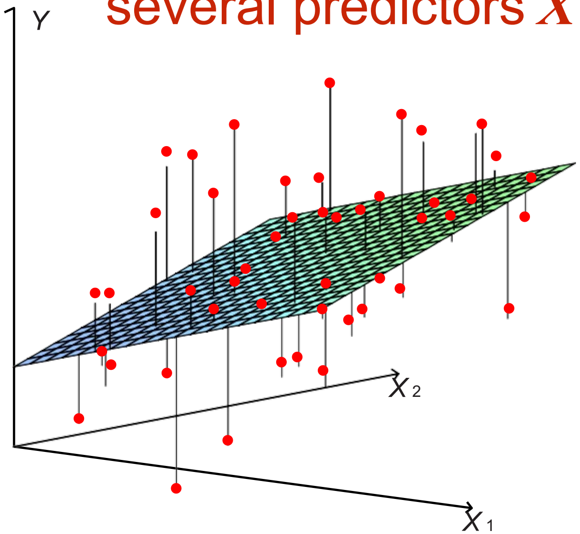
$$y = b_0 + \Sigma_p b_p X_p$$

where **there are** p variables and b_p are the " unknown"
(the model parameters to be determined)

X :

- usually, the measured quantitative variables
- their transformations, *eg.* log, √, ..
- dummy coding of qualitative variables
- may also include interaction (X_1X_2...) or quadratic (X_1^2 …) terms

several predictors X



$$\mathbf{y} = b_0 + \Sigma_p^N b_p X_p$$

$$RSS = \Sigma_i (y_i - b_0 - \Sigma_p^P b_p x_{ip})^2$$

Matrix Notation

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

$$RSS = (y - Xb)^T (y - Xb)$$

Finding minimum RSS

$$\frac{\partial RSS}{\partial b} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}b)$$

1. $\mathbf{X}^T(\mathbf{y} - \mathbf{X}b) = 0$
2. $\mathbf{X}^T\mathbf{y} - \mathbf{X}^T\mathbf{X}b = \mathbf{0}$
3. $\mathbf{X}^T\mathbf{X}b = \mathbf{X}^T\mathbf{y}$
4. $b = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{y}$

Geometric Interpretation

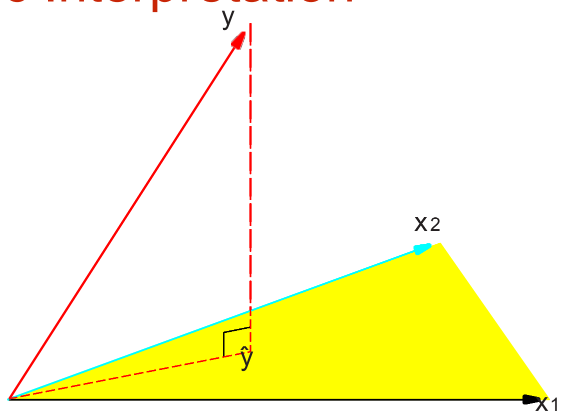


FIGURE 3.2. *The N-dimensional geometry of least squares regression with two predictors. The outcome vector* y *is orthogonally projected onto the hyperplane spanned by the input vectors* x_1 *and* x_2*. The projection* ŷ *represents the vector of the least squares predictions*

$$4.\quad b = (\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T\mathbf{y}$$

$$\hat{\mathbf{y}} = \mathbf{X}b = \underbrace{X(\mathbf{X}^T\mathbf{X})^{-1}\,\mathbf{X}^T}\mathbf{y}$$

called also H it is a projection operator

ŷ is orthogonal to (y - ŷ) and it lays in *X* space

$$y = f(x)$$

REGRESSION

- **UNIVARIATE** ONE y as a function of ONE x
- **MULTIVARIATE** ONE or MORE y's as a function of MORE than one X

- **LINEAR**
- **NON LINEAR**

MODEL

- THEORETICAL: based on theory, use functions derived from basic principles or laws
- EMPIRICAL: not based on theory, function based on “fit”

CHEMIOMETRIA

AIMS

- TO PREDICT the dependent variable/es
- TO INTERPRET the obtained functional relationship

To interpret the model it should necessary be statistical significant and validated

THEORETICAL → Theory is verified

EMPIRICAL → Local Approximation common latent factors

main MLR limitations

- X- variables collinearity

X_1	X_2
1	2
2	4
3	6
4	8

y
2
4
6
8

```
>> inv(x'*x)*x'*y
Warning: Matrix is singular to working precision.
```

X_1	X_2
1	2
2	4
3,001	6,001
4	8

```
>> inv(x'*x)*x'*y

ans =

    2.0000
   -0.0000
```

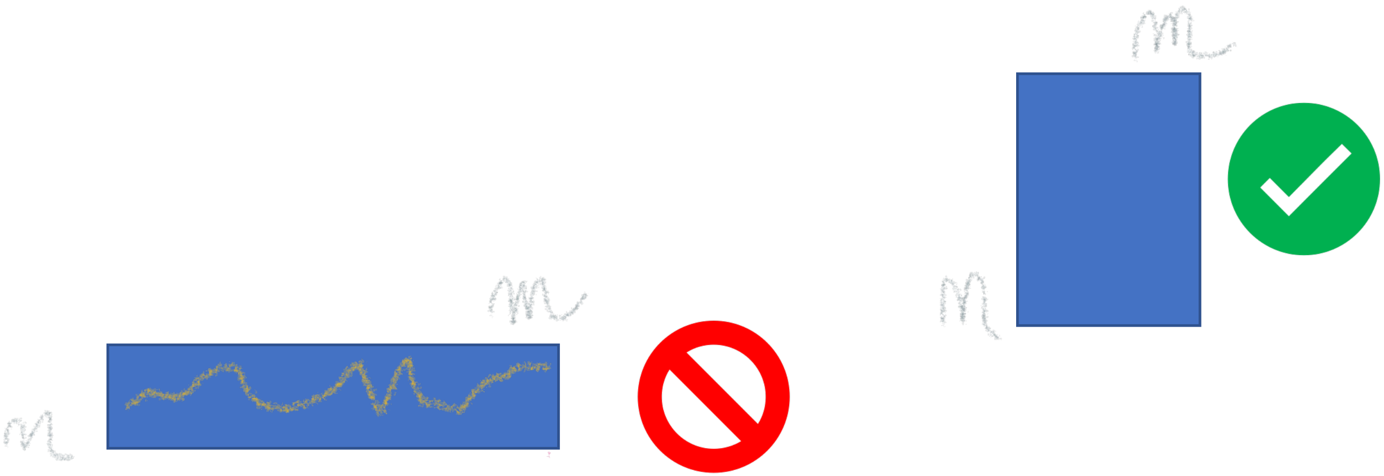
```
>> inv(x'*x)*x'*y

ans =

   -0.0000
    1.0000
```

```
>> inv(x'*x)*x'*y

ans =

   -2.0000
    2.0000
```

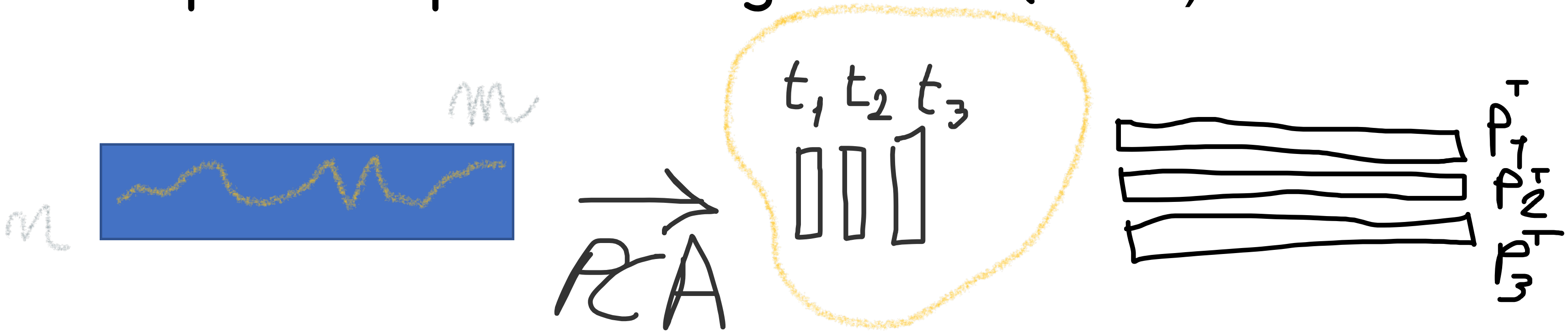
Very unstable !!!

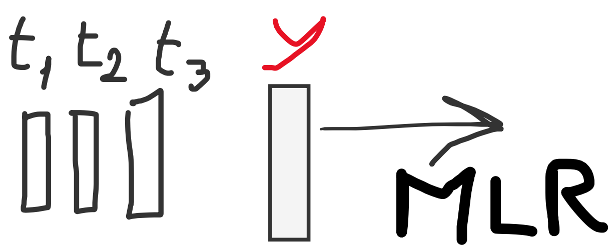
Main MLR limitations

- more samples than variables are needed



- Principal Component Regression (PCR)





$$y = Tc + e$$

$$c = (T^T T)^{-1} T^t y$$

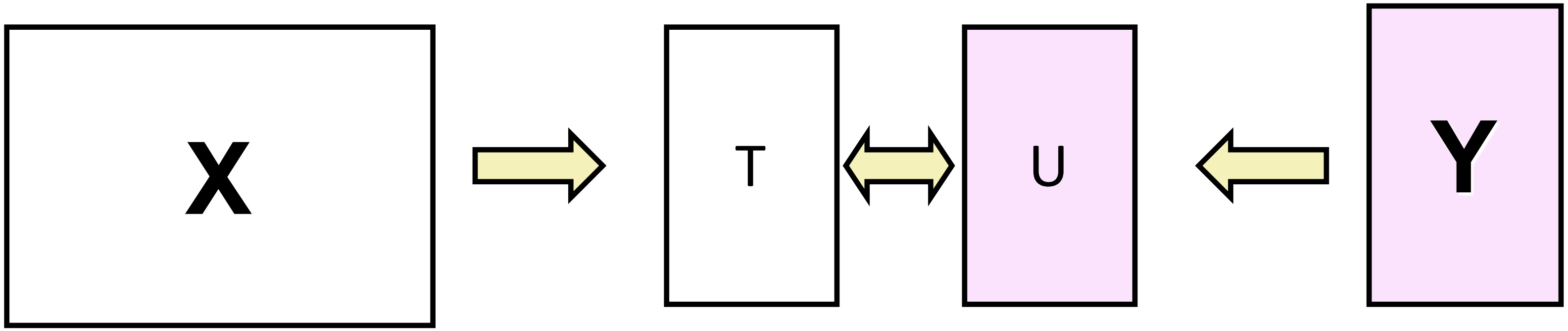
$$T = XP$$

$$y = X\,Pc + f$$

b

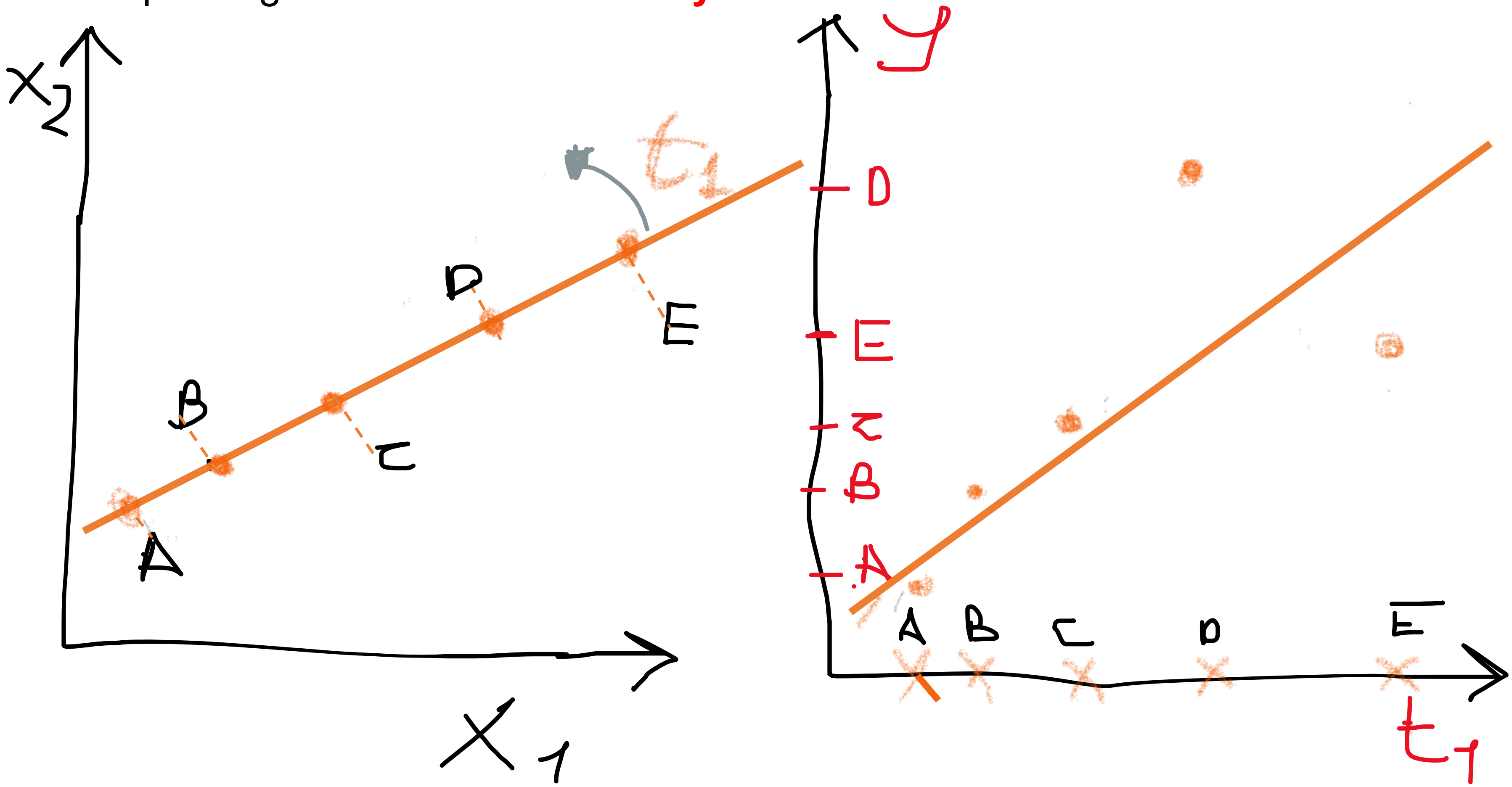
PLS

- Considers **y** when decomposing **x**
- Can handle multiple y (**Y**) simultaneously

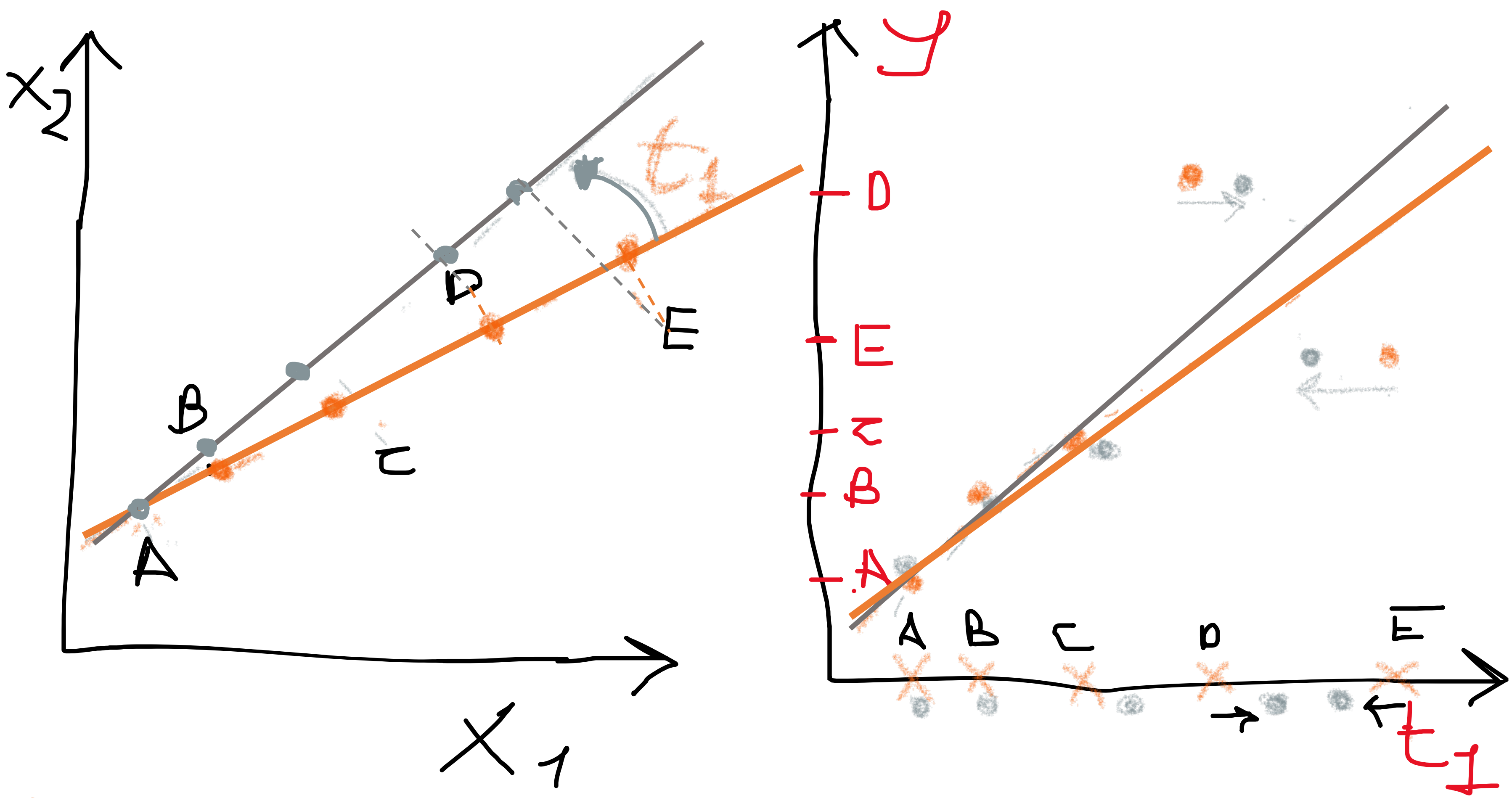


What's PLS

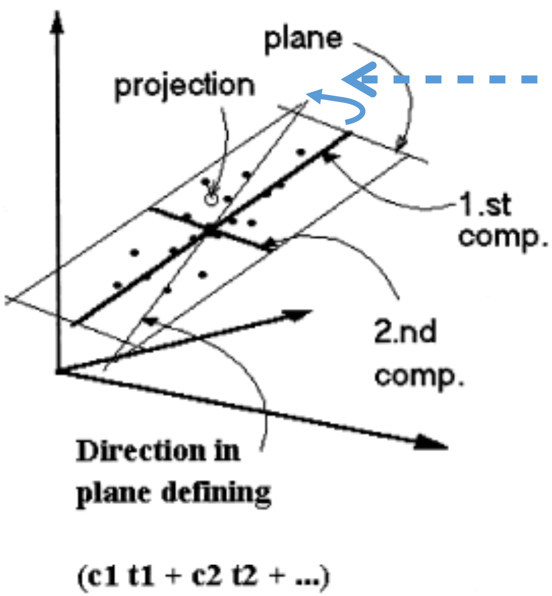
PLS strikes a compromise among explaining **X** and correlation with **y**



What's PLS



PLS strikes a compromise among explaining **X** and correlation with **y**

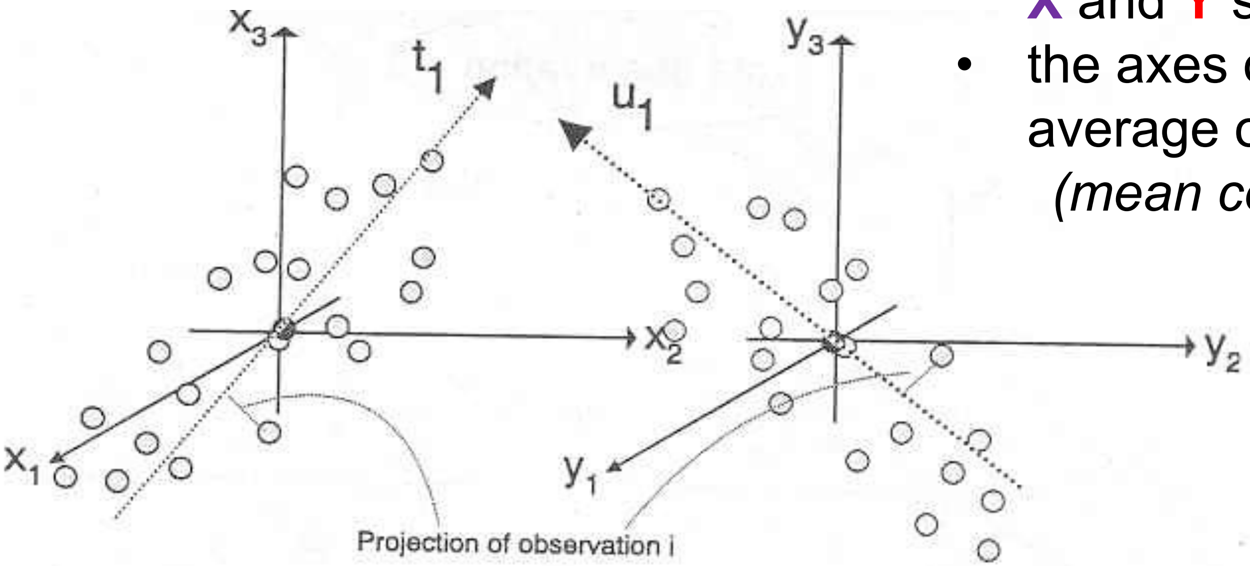


To find a direction of maximal covariance (**X**,**y**) a vector of weights ***w*** is defined component wise

Criterion: find a ***w*** such as:

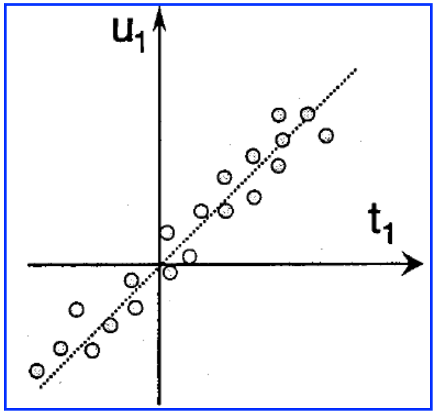
$$\max \left[\mathrm{cov}(t,y) \,|\, Xw=t \text{ and } ||w||=1 \right]$$

What's PLS 2



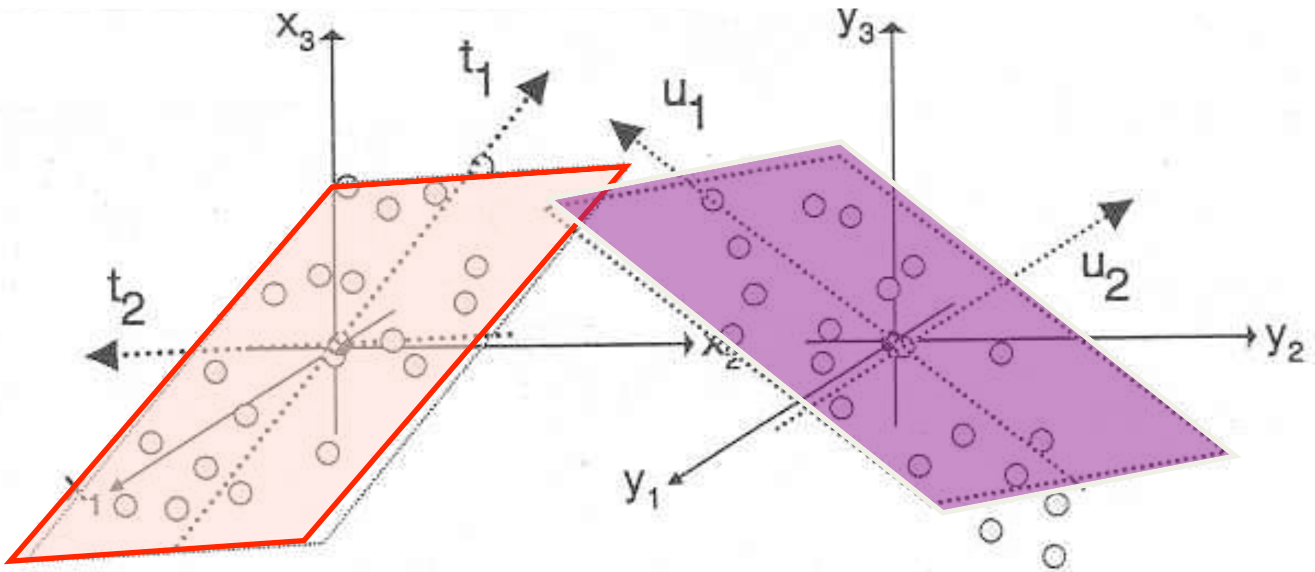
- Each sample is a point both in **X** and **Y** space
- the axes origin is in the average of **X** and **Y**
  (mean centering of X e Y)

PLS
"inner-relation"



covariance between scores in **X** (**t_1**) and scores in **Y** (**u_1**) is maximized component wise

What's PLS2

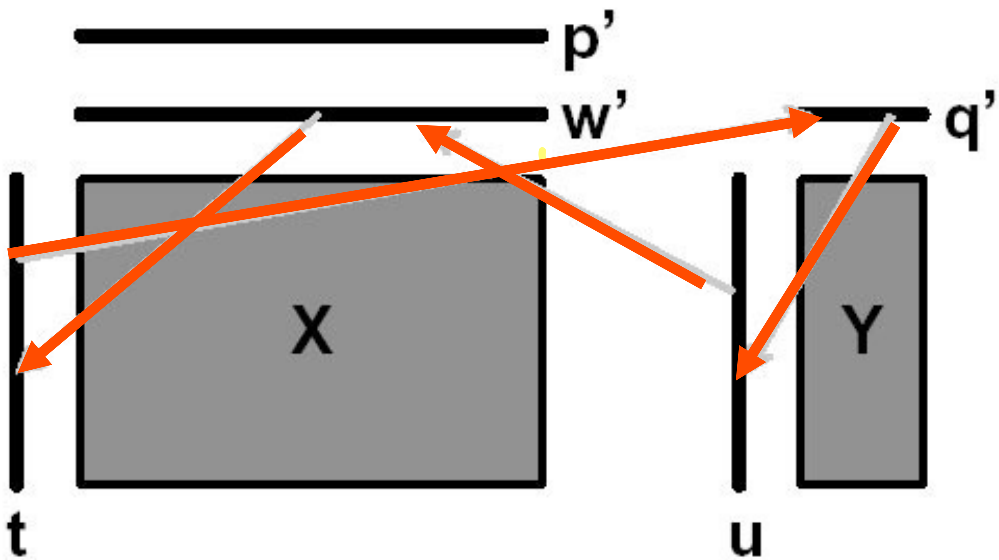


- each **X** (n x j) and **Y** (n x k) matrices defines a space in j and k dimensions respectively

- eg. 2 PLS components define a plane both in **X** and in **Y**

t2 is orthogonal (90°) to t1 while it is not necessarily so for u2 and u1

PLS is iterative e.g. NIPALS for first LV



Take u start as the single y with max variance

$w = X'u / u'u$

$t = Xw / ||w||$

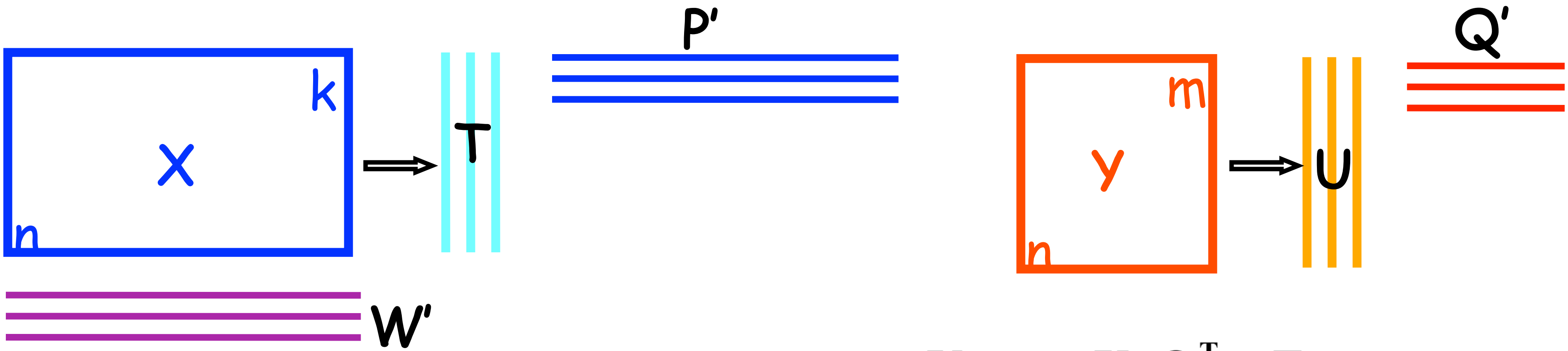
$q = Y't / t't$

$u_{updated} = Yq / ||q||$

At convergence:

$p = X't / t't$
$q = Y'u/u'u$

$X_{res} = X\text{-}tp'$ go for next LV



$$\mathbf{X} = \bar{\mathbf{x}} + \mathbf{T} \times \mathbf{P}^{\mathrm{T}} + \mathbf{E}$$

T = matrix of X-scores

P = matrix of X-loadings

W = matrix of PLS X-weights

$$\mathbf{Y} = \bar{\mathbf{y}} + \mathbf{U} \times \mathbf{Q}^{\mathrm{T}} + \mathbf{F}$$

$$\boxed{\text{"inner relation": } \mathbf{U} = \mathbf{bT}}$$

$$\mathrm{Y} = \bar{\mathrm{y}} + \mathrm{TbQ}^{\mathrm{T}} + \mathrm{F}$$

U = matrix of Y-scores

Q= matrix of Y-loadings

Re-expressing as a regression model:

$$\boxed{\mathbf{B}_{PLS} = \mathbf{W}(\mathbf{P}^{\mathrm{T}}\mathbf{W})^{-1}\mathrm{diag}(\mathbf{b})\mathbf{Q} \qquad \hat{\mathbf{Y}} = \mathbf{X}\mathbf{B}_{PLS}}$$

PLS in practice

- **How many latent variables ?** (model dimensionality)
- **Is PLS model adequate ?**
- **Are there "anomalous" or "influential" samples ?**
- **Which results to look at ? What plots to display ?**
- **Which are the most "important" X variables to model Y ?**

- When/what preprocessing ?
- when/why do I need variables selection ?

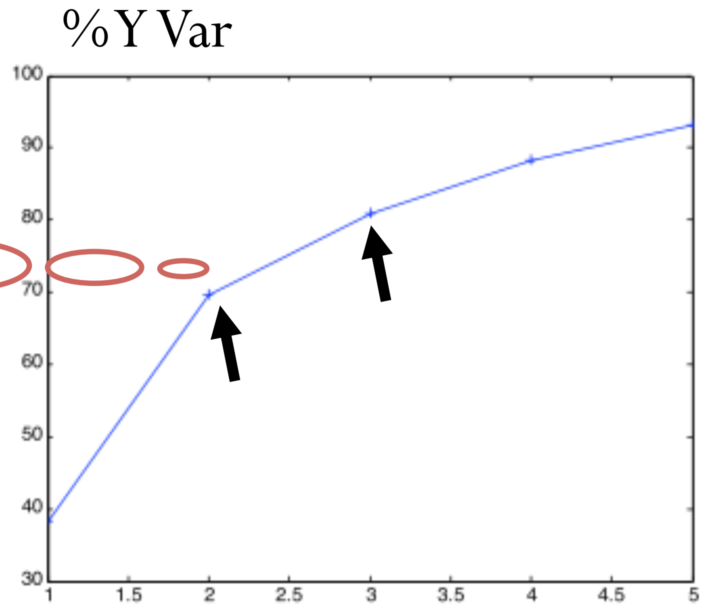
PLS in practice: how many latent variables?

Empirical rules

1. Rule of thumb LV number <= 1/3 min(n,m)

2. Slope change of %Y vs. N LV

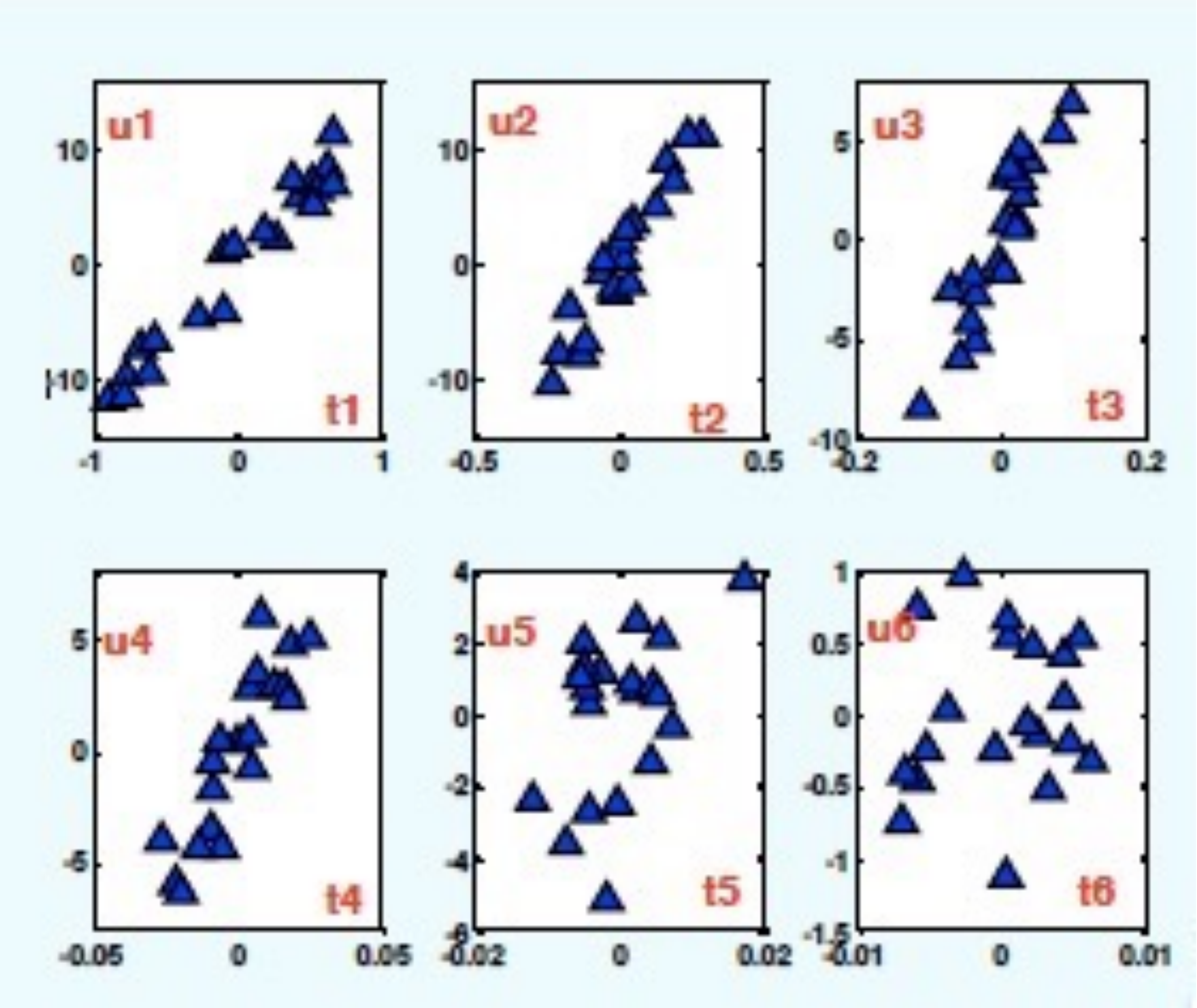
If components account for noise they explain similar % of variance, thus a plateau is reached



3. look for end of structure/information
In inner-relationships plots

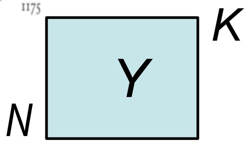
PLS in practice: how many latent variables?

PLS inner relation for subsequent components



- t and u scores are correlated until there is structure in X related to Y
- may choose 4-5 LVs on this consideration
- There are exceptions to this trend with spectral data

PLS in practice: how many latent variables?

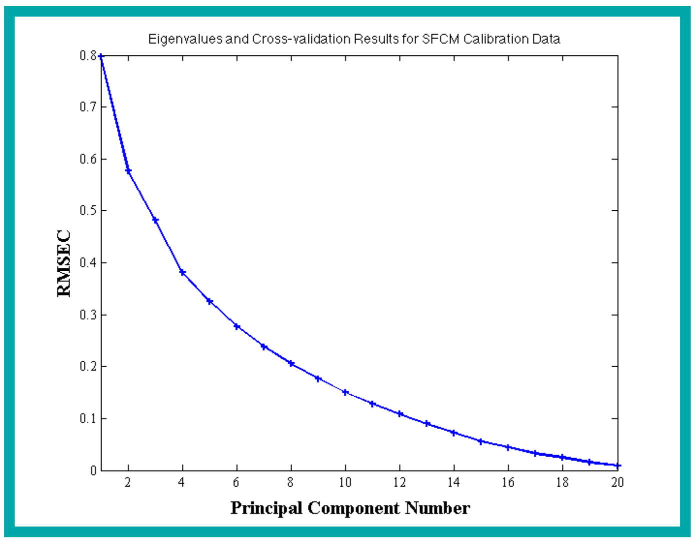


The maximum number of components that can be calculated is equivalent to X-rank (in this case PLS converges to multilinear regression)

$$\sqrt{\frac{\sum_{i=1}^{n_{TR}} (y_i - \hat{y}_i)^2}{n_{TR}}}$$

RMSEC (error in fit)

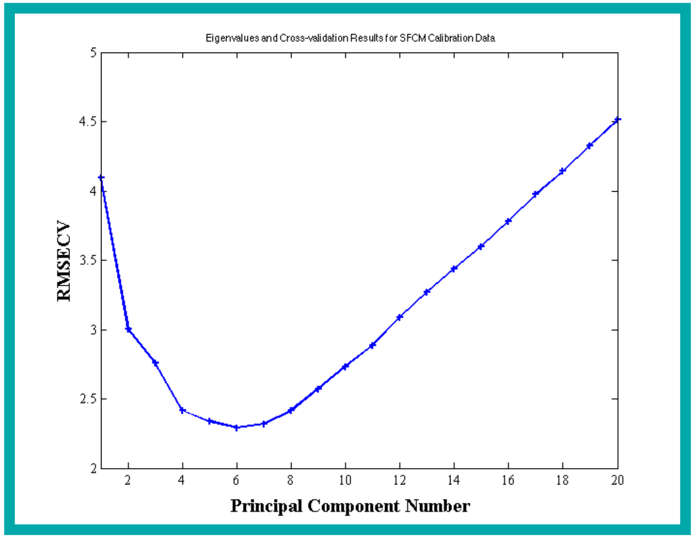
Using all samples



As in any regression model by adding more parameters fit increases

RMSECV

(error internal validation, excluded samples in turn)



the predictive capability (estimation of new/future sample) decreases

PLS in practice: how many latent variables?

Look at prediction errors

Regression – Error measures

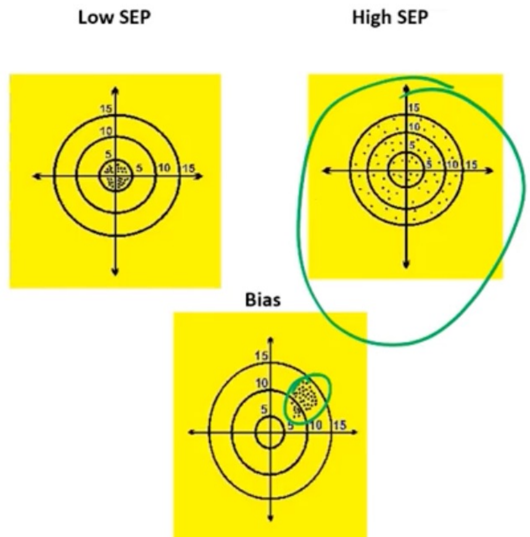
RMSE – total error
- SEP – random error
- Bias – systematic error

RMSE² ≈ SEP²+ Bias²

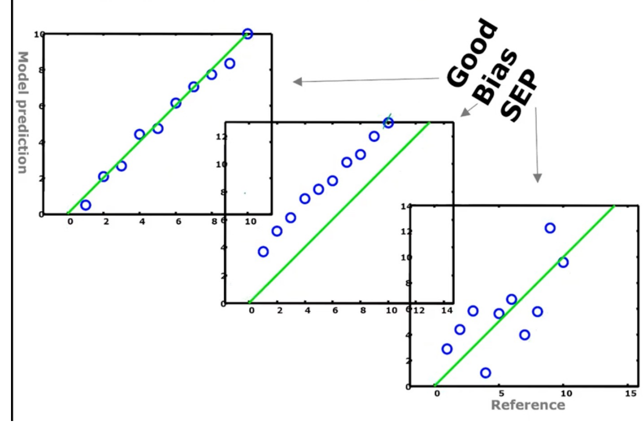
Can be calculated for the calibration set (fit: RMSEC , SEC)

For the validation set (Prediction: RMSEP, SEP)

Root Mean Squared Error of Prediction



Predicted versus Measured plot



Figures from R. Bro kvl dk

$$RMSEP = \sqrt{\frac{\Sigma(\hat{y}_{val} - y_{val})^2}{n}}$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i - bias)^2}{n-1}}$$

$$bias = \sum_{i=1}^{n}\frac{(\hat{y}_i - y)}{n}$$

About R^2

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n_{TR}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{TR}} (y_i - \bar{y}_{TR})^2}$$

* RMSEC is in the same scale of the variable/s y
* R^2 varies between 0 e 1

RMSEC and R2 have to show inverse proportion, the lower the error, the higher R2

- ✓ always true if we compare the same data set (same samples, same Y)
- ✓ if R2 is used to compare different models same Y different number of samples; different Y;

we have to take into account **that R2 depends on y dispersion**, the more $(y - \bar{y})^2$ is small the larger R2 even if RMSEC is equal

About R^2

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n_{TR}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{TR}} (y_i - \bar{y}_{TR})^2}$$

✓ Test set have different number of samples
Ytest dispersion may differ form Ycal dispersion

$$\sum_{i=1}^{n\,tr} (y_i - \bar{y}_{TR})^2 = \sum_{i=1}^{n\,ts} (y_i - \bar{y}_{Ts})^2$$

so which TSS to use in R_{TEST} ? Test or Training ?

Todeschini et al… show a correction which works using TSS training

$$R^2_{TEST} = 1 - \frac{PRSS/n_{TEST}}{TSS/n_{TR}}$$

PLS in practice: how many latent variables?

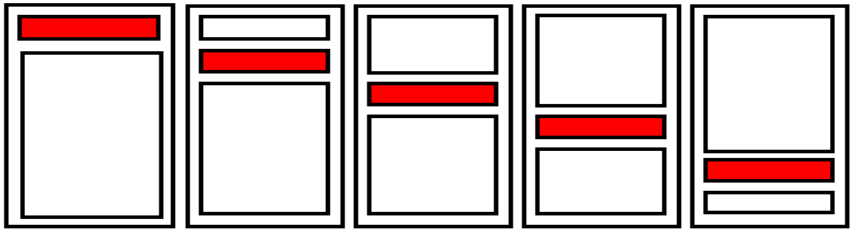
4. Estimate predictive capability (recommended)

    1. Cross validation
    2. Double CV
    3. Bootstrap
    4. Permutation test

PLS in practice: how many latent variables?

In this case the significance is tested on predictive capability

1. one (LOO) or more objects (raws) are deleted from the data matrix



2. a PLS model (1 component) is calculated. The left out objects are projected on LV1, and the scores, t1, are estimated for the "out" objects. From the inner relation Y scores u1 are estimated. From the "PCA-like" model of Y knowing u1 the y of left out can be predicted

   (or use the $Y=XB_{pls}$)

2. The PLS model is applied to the left "out" objects and their y squared residuals are calculated.
3. Iterate 1-3 untill each objects (of the data set) has been left out once.
4. Calculates the Predicted Residuals Sum squares for all objects (PRESS1)
5. Iterate 1-5 for a model with 2 components, thus calculating PRESS2 and so on …

PLS in practice: how many latent variables?

Cancelation schemes different from LOO:

Leave More Out

LOO is unique , LMO is not; so possible alternatives are:

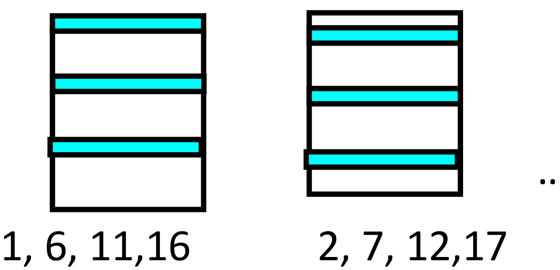
- **Random groups**, e.g. of 5 objects each, are formed
- Apply CV procedure
- Repeat 1-2 many times, e.g. 15 iterations
- Average sum of prediction error over iterations

Venetian blind

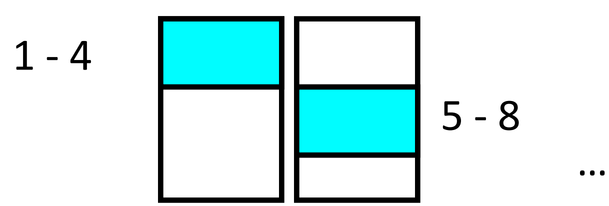
- Decide number of splits (e.g. 20 object 5 splits = 4 samples taken out at time)
- Take out every, e.g. 4th objects from 1 to n

(better to sort y first)

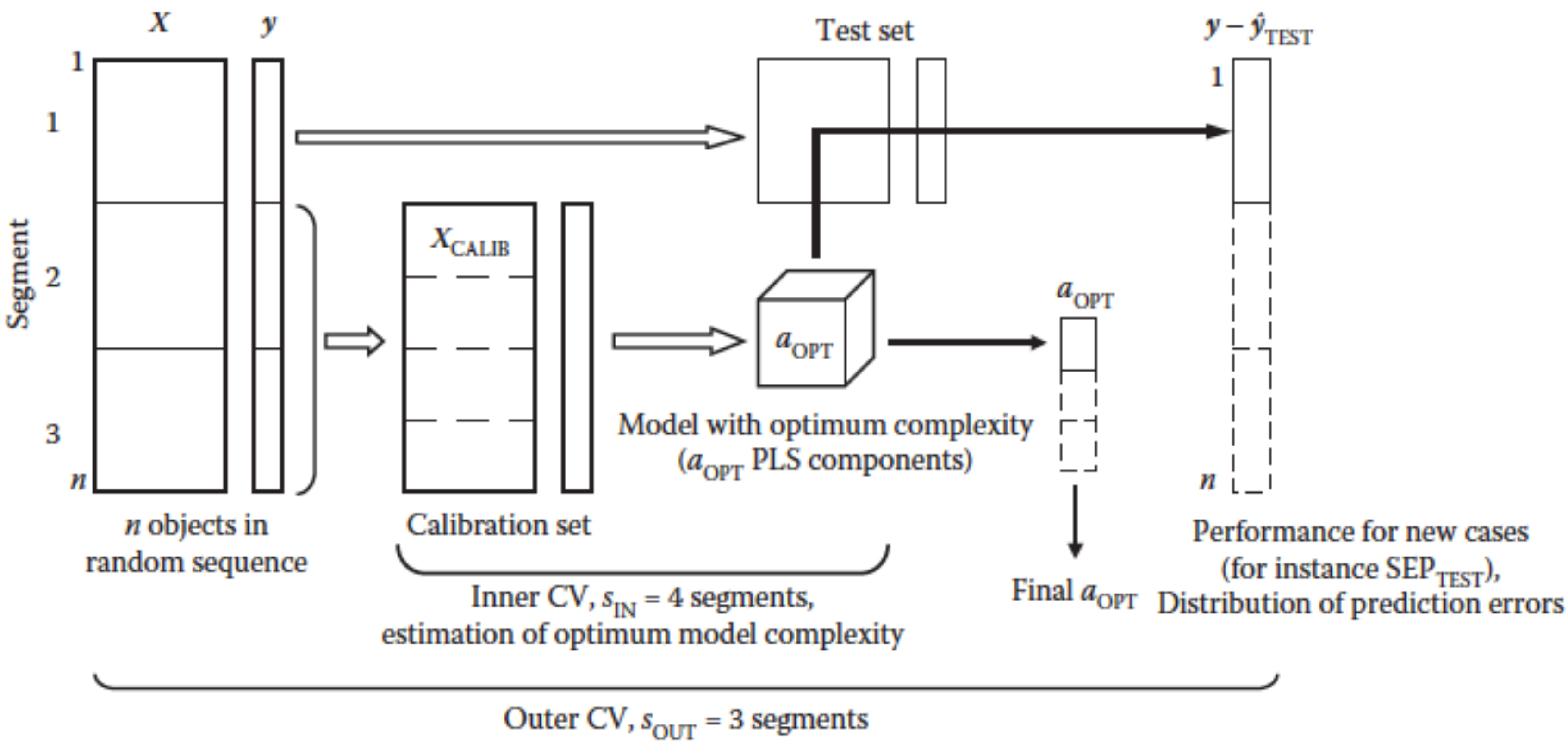
- Apply CV procedure



Contiguous



Double CV



The prediction in “outer loop” may corrispond to a different number of a_{opt}

How to choose a final model (unique)?

- use the median of all a_{opt}
- use the most frequent value of a_{opt}

re-calculate (with all samples included) a model with this number of components

Bootstrap

Resampling with repetition

for z=1: 1000 (>1000)
    for i=1: N samples
        Select randomly a sample
        put it in calibration set

    end

- the calibration set has N samples
- some are repeated
- some are NOT present

    build a model and predict the NOT present samples
end

Probability to select a single sample in a single boot: 1/N
Probability of NOT selecting " "   "   " " "   " : 1 -1/n
Probability of NOT selecting in z boots : $(1\text{-}1/N)^N$
for N large, probability tends to 0.37

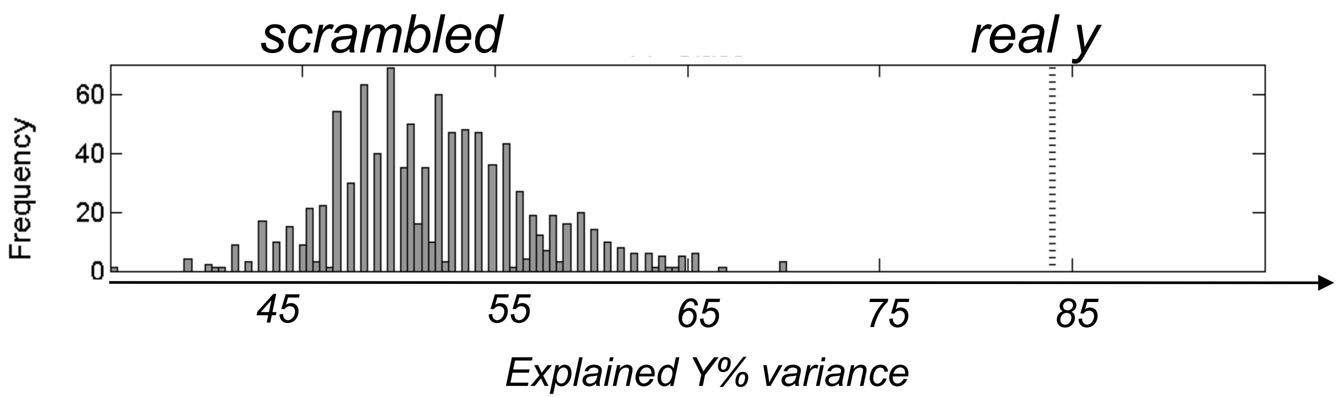
number of total predictions: varies between 0 and z

Permutation test

for p=1: 1000 (>1000)

randomly permute the Y
calculate a model with the “fake” y
store predictions, model parameters

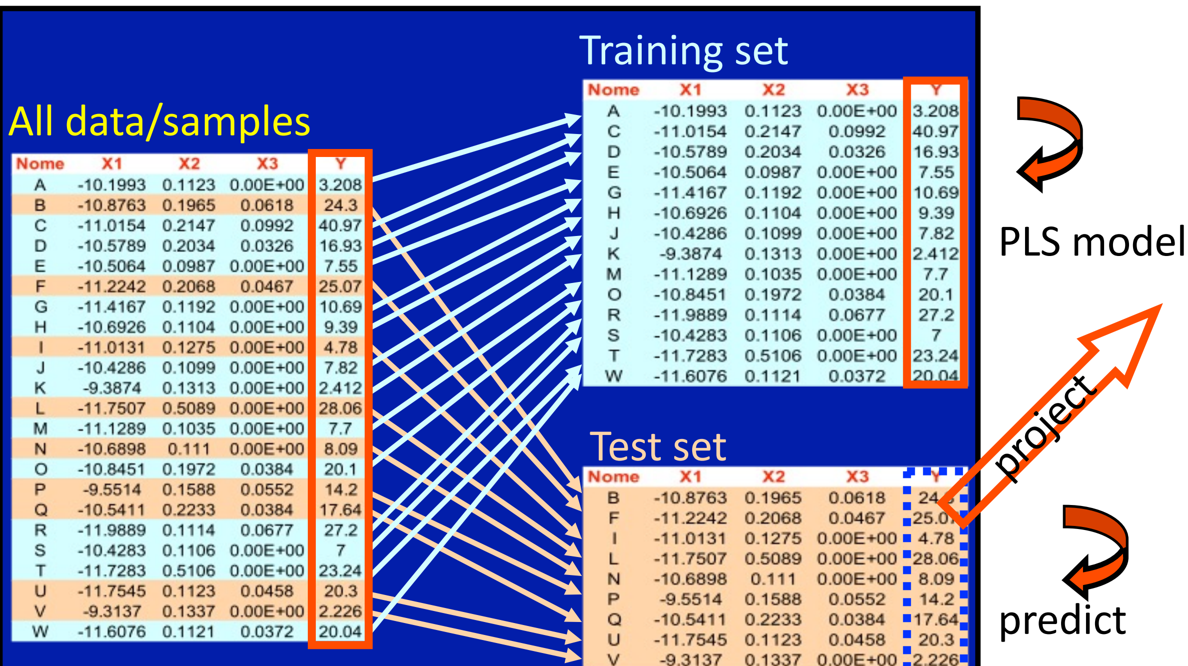
end



PLS in practice: how many latent variables?

Monitoring, Test set

- representative & independent from training
- replicates in the same set
- ¼ | ½ of training



All data/samples

Nome	X1	X2	X3	Y
A	-10.1993	0.1123	0.00E+00	3.208
B	-10.8763	0.1965	0.0618	24.3
C	-11.0154	0.2147	0.0992	40.97
D	-10.5789	0.2034	0.0326	16.93
E	-10.5064	0.0987	0.00E+00	7.55
F	-11.2242	0.2068	0.0467	25.07
G	-11.4167	0.1192	0.00E+00	10.69
H	-10.6926	0.1104	0.00E+00	9.39
I	-11.0131	0.1275	0.00E+00	4.78
J	-10.4286	0.1099	0.00E+00	7.82
K	-9.3874	0.1313	0.00E+00	2.412
L	-11.7507	0.5089	0.00E+00	28.06
M	-11.1289	0.1035	0.00E+00	7.7
N	-10.6898	0.111	0.00E+00	8.09
O	-10.8451	0.1972	0.0384	20.1
P	-9.5514	0.1588	0.0552	14.2
Q	-10.5411	0.2233	0.0384	17.64
R	-11.9889	0.1114	0.0677	27.2
S	-10.4283	0.1106	0.00E+00	7
T	-11.7283	0.5106	0.00E+00	23.24
U	-11.7545	0.1123	0.0458	20.3
V	-9.3137	0.1337	0.00E+00	2.226
W	-11.6076	0.1121	0.0372	20.04

Training set

Nome	X1	X2	X3	Y
A	-10.1993	0.1123	0.00E+00	3.208
C	-11.0154	0.2147	0.0992	40.97
D	-10.5789	0.2034	0.0326	16.93
E	-10.5064	0.0987	0.00E+00	7.55
G	-11.4167	0.1192	0.00E+00	10.69
H	-10.6926	0.1104	0.00E+00	9.39
J	-10.4286	0.1099	0.00E+00	7.82
K	-9.3874	0.1313	0.00E+00	2.412
M	-11.1289	0.1035	0.00E+00	7.7
O	-10.8451	0.1972	0.0384	20.1
R	-11.9889	0.1114	0.0677	27.2
S	-10.4283	0.1106	0.00E+00	7
T	-11.7283	0.5106	0.00E+00	23.24
W	-11.6076	0.1121	0.0372	20.04

Test set

Nome	X1	X2	X3	Y
B	-10.8763	0.1965	0.0618	24
F	-11.2242	0.2068	0.0467	25.0
I	-11.0131	0.1275	0.00E+00	4.78
L	-11.7507	0.5089	0.00E+00	28.06
N	-10.6898	0.111	0.00E+00	8.09
P	-9.5514	0.1588	0.0552	14.2
Q	-10.5411	0.2233	0.0384	17.64
U	-11.7545	0.1123	0.0458	20.3
V	-9.3137	0.1337	0.00E+00	2.226

If used to set model parameters (num LVs, select variables, ..)

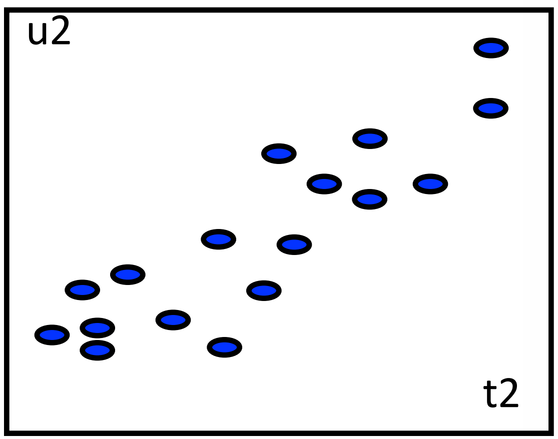
Not usable for estimating predictive capability

is PLS model adequate?

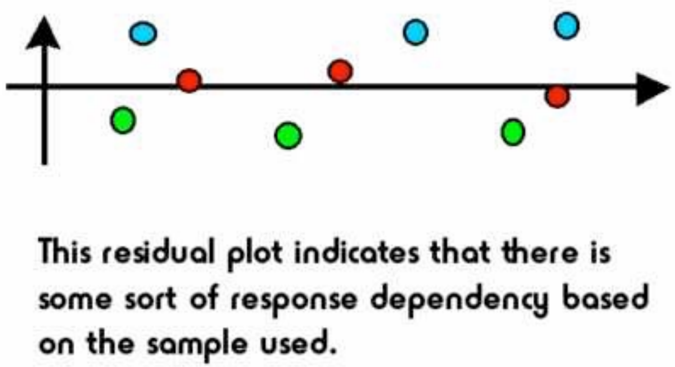
Goodness of fit : R^2 , RMSEC
should be compared with experimental error
DO REPLICATES
(eventually known from historical data …. Method)

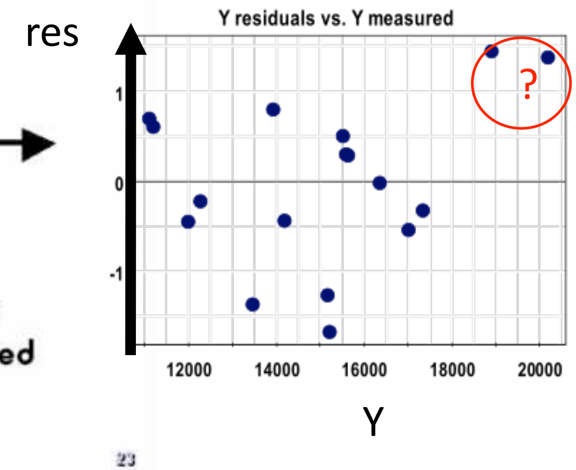
If fit is higher than experimental error on Y then we are modeling noise!

Inspection of inner relation :



Inspection of residuals :

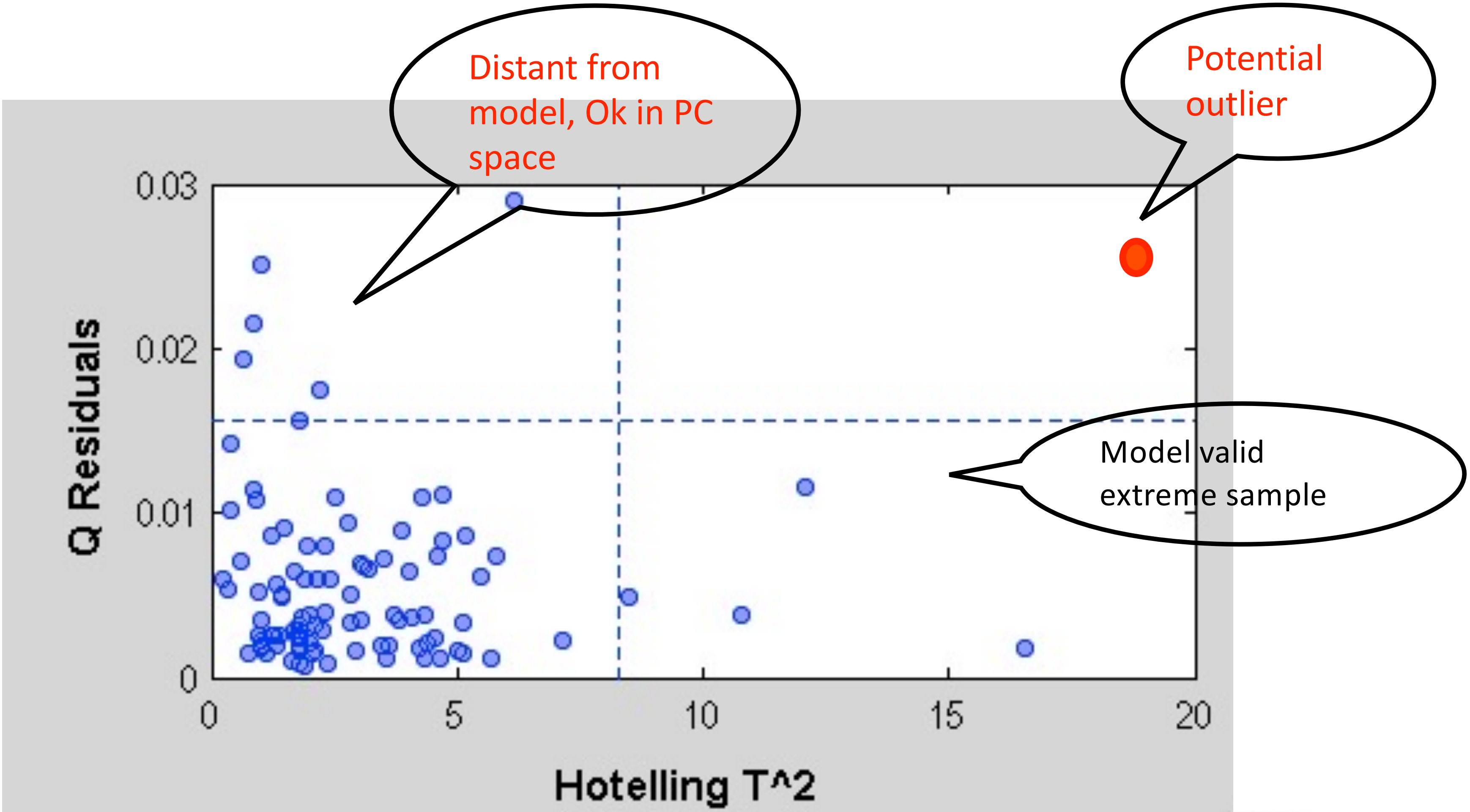




is PLS model adequate?

Are there "anomalous" or "influential" samples ?

Q (DmodX) - T^2 plot (as well for Y: Q (DmodY) - U^2)

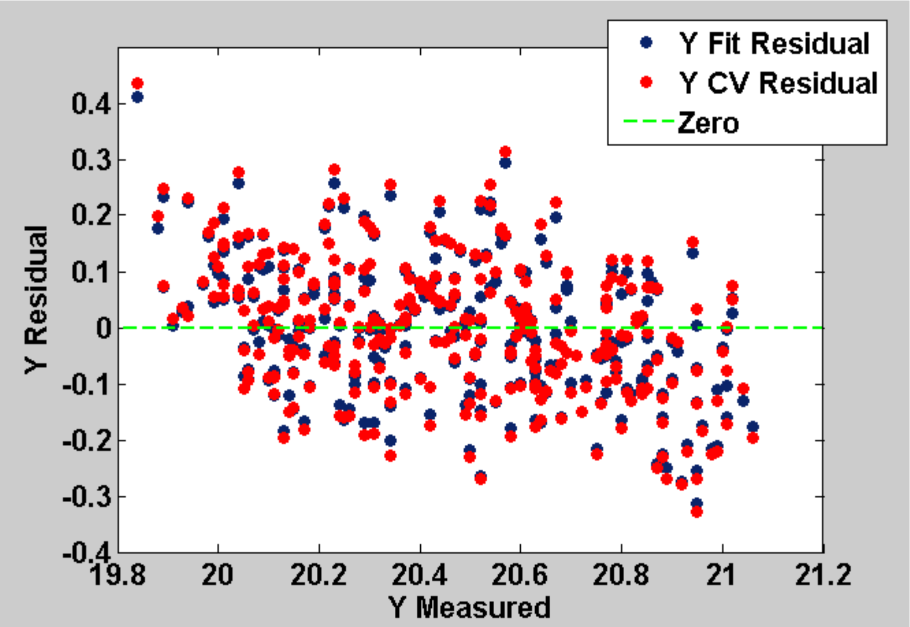


which results to look at ?

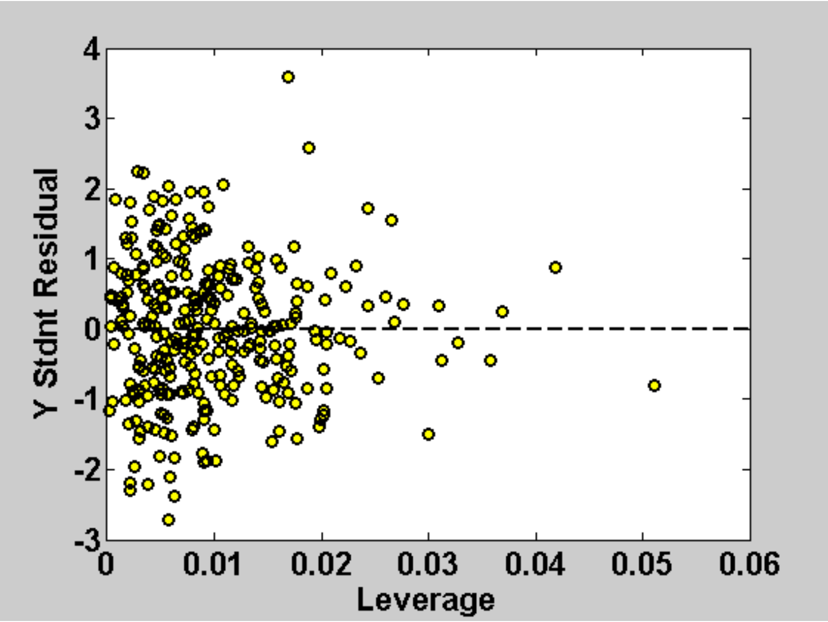
- **check randomness:**

- Plot Y- residuals vs Leverage

X



Y



- **Y Leverage:**

$U(U^TU)^{-1}U^T$ (Y-block)

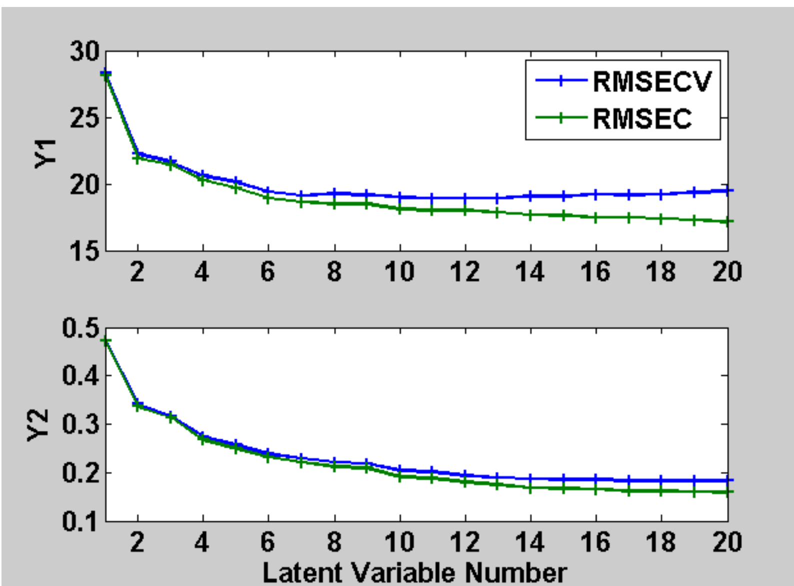
how influential objects are in determining Y PCA-model

- Plot X-residuals **E** vs Y, vs Order of spectra aquisition,..

which results to look at ?

Model

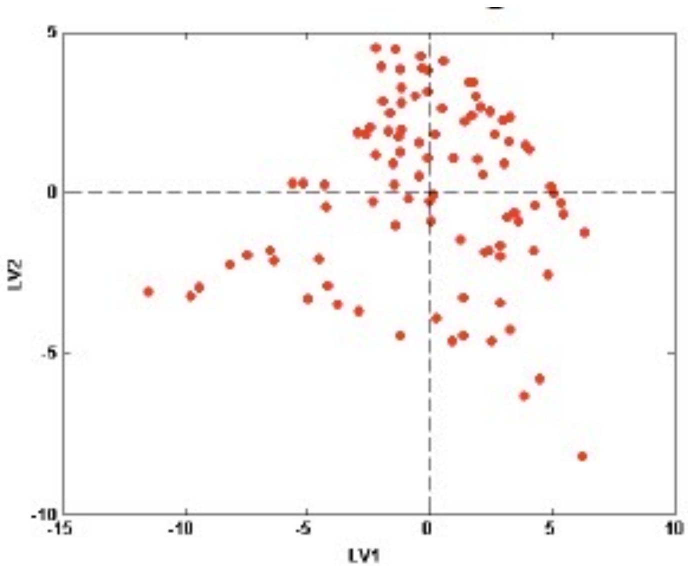
	X-Block LV	X-BLock Cumulative	Y-Block LV	y-Block Cumulative
1	83.16	83.16	6.69	6.69
2	4.97	88.13	40.98	47.67
3	7.58	95.71	4.42	52.09
4	2.11	97.82	8.37	60.46
5	0.99	98.81	3.23	63.68
6	0.37	99.18	3.55	67.23
7	0.24	99.43	1.64	68.86
8	0.11	99.54	1.13	69.99
9	0.21	99.75	0.30	70.29
10	0.03	99.78	2.24	72.54



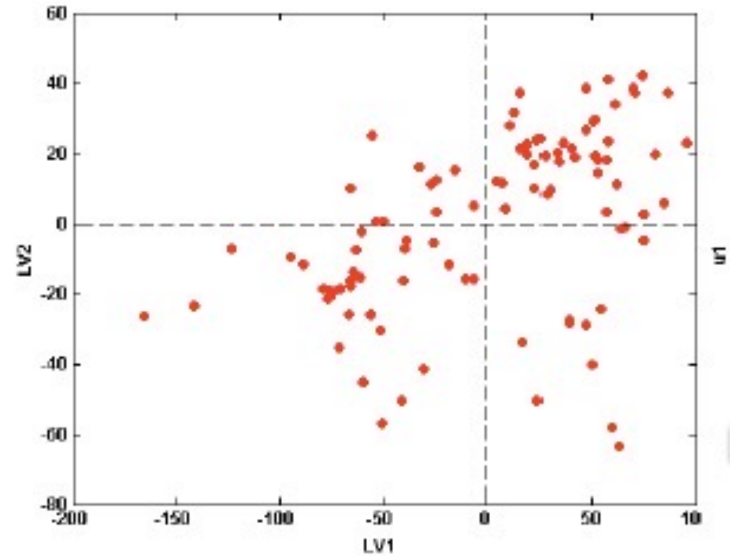
- **Fit**: **R^2**, % Explained Variance of **Y (as well, for each y-variable)**
- **Internal validation**: **RMSECV**
- **Contribution to the PLS model**: % Explained Variance of **X (as well for each x-variable)**
- ***Prediction capability:*** ***RMSEP*** *test set (truly independent)*

which results to look at ?

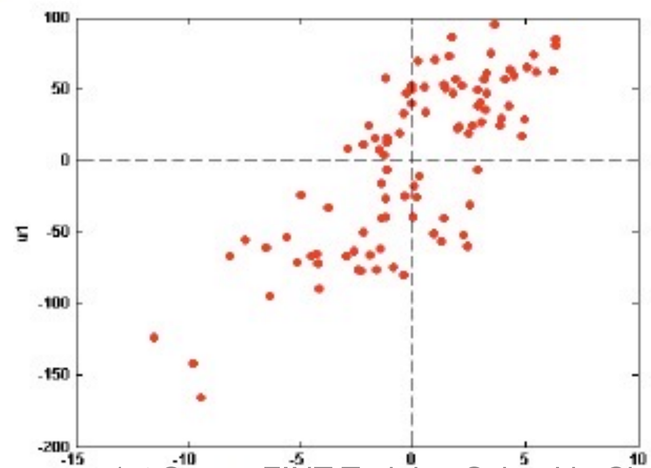
2. Objects (samples, systems)



X scores



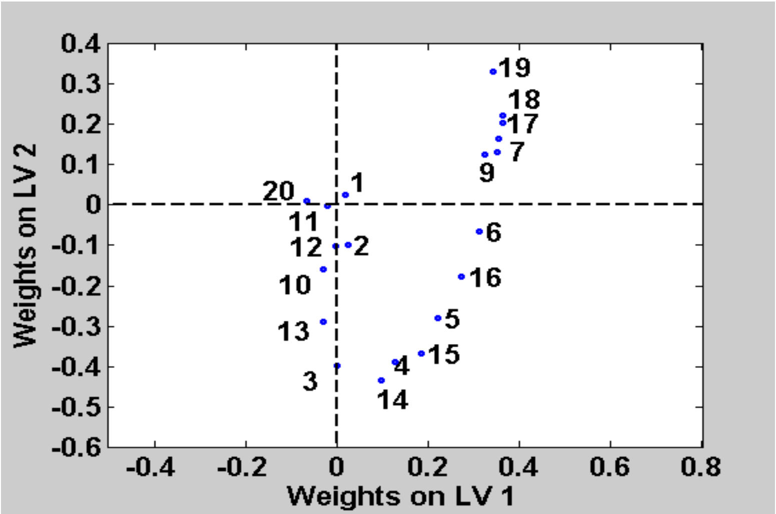
Y scores



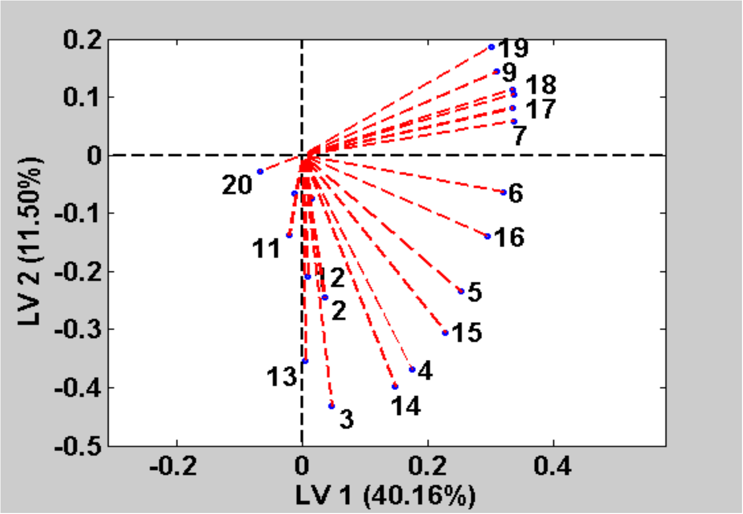
- inner relation U/T

which results to look at ?

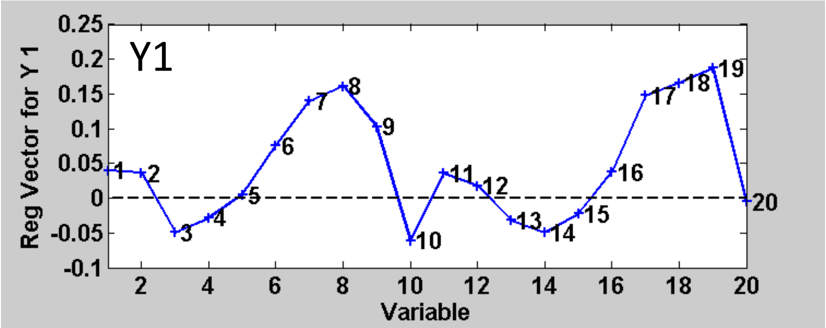
- **Correlation among X and Y**: PLS weights, **w**1, **w**2, ..;



- **Trends- Correlation among X variables**: X- Loadings plot **p**1, **p**2; ...



- **Variable importance: regression coefficients $\mathbf{B}_{PLS}$**



- **Trends –Correlation among Y variables**: Y- Loadings plot **q**1, **q**2;.

which are most important variables?

- **Variable importance: Variable Influence on projection, VIP** [1]

One of the parameter for ranking variables:

DEFINITION:

$\mathbf{VIP^2_k = \sum_a w^2_{ak} *SSY_a*K/(SSY_{tot,expl.}*A)}$ **A**= number of LVs; K = number of X variables

VIP is derived from PLS weights weighted by how much of Y is explained in each model dimension;
Since $\sum_k VIP_k^2 = K$ the proposed threshold is 1.

- **Variable importance: Selectivity Ratio, SR** [2]

$$t = Xw_{TP} = X\frac{b_{PLS}}{\|b_{PLS}\|}$$

$$p^T = \frac{t^T X}{(t^T t)}$$

Project on target component, y-correlated

$$\widehat{X} = tp^T$$

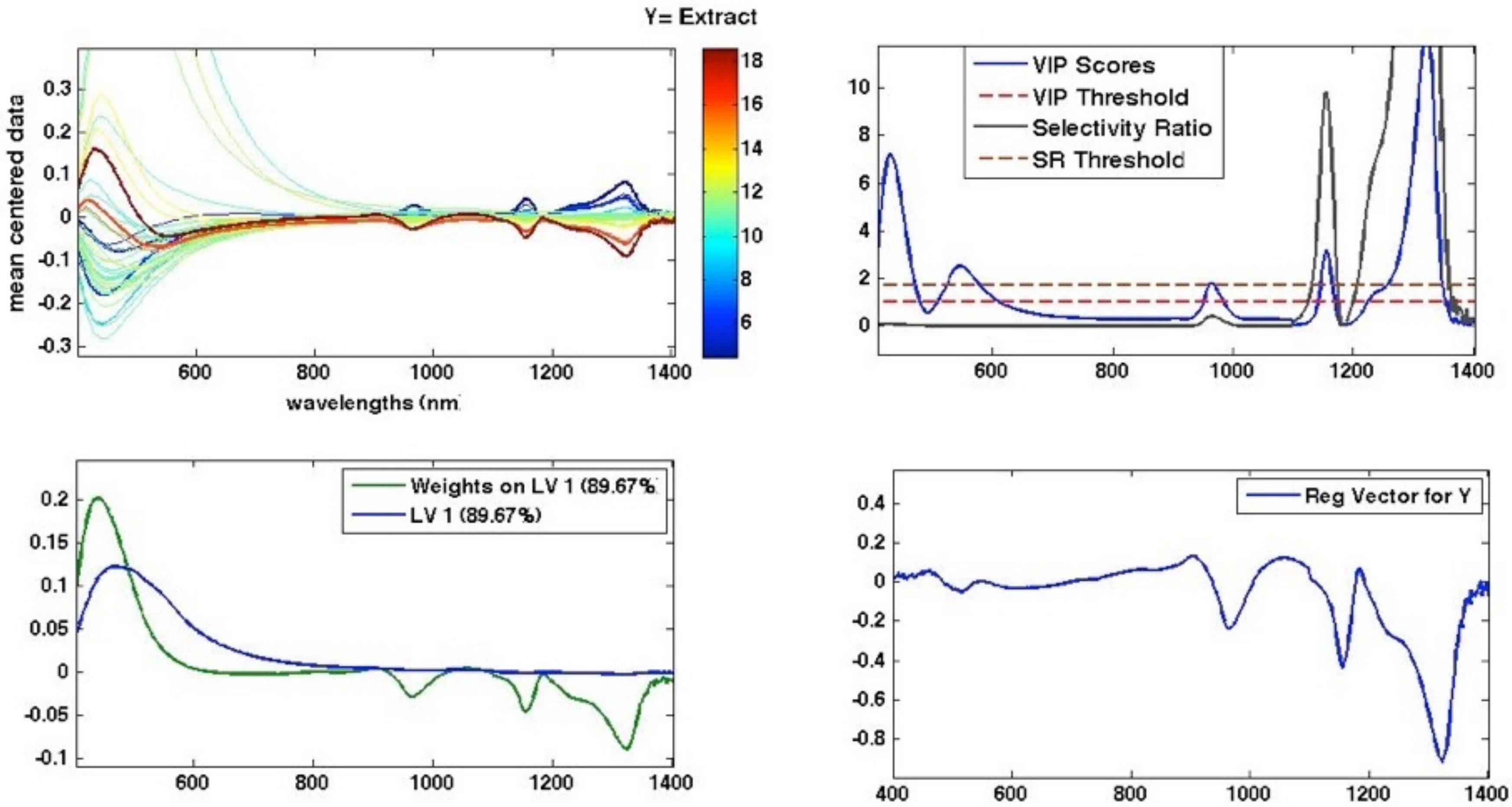
$$SR = var(\widehat{X})/var(X-\widehat{X})$$

SR express for each x-var the ratio among the variance explained by the target component and the residuals variance. The higher the more relevant

[1] Wold S, Johansson E, Cocchi M., 1993. PLS- Partial Least Squares Projections to Latent Structures, in: 3DQSAR in Drug Design. H. Kubinyi Ed., Leiden, Holland. [2] Chong *et al. Chemom. Int. Lab. Syst. 78 (2005)) 103-112.*
[2] T. Rajalahti, R. Arneberg, F.S. Berven, K.M. Myhr, R.J. Ulvik, O.M. Kvalheim, *Chemom. Int. Lab. Syst.* . 95 (2009) 35–48.

which are most important variables?

- Interpreting Variable importance



1. Model

- **Fit**: R^2, % Explained Variance of **Y (as well, for each y-variable)** [R2YCUM, R2VY]
- **Predictive capability**: **RMSECV / RMSEP** **C**ross-**V**alidation or monitoring set

(Use also to choose the number of significant PLS components) [*Q2YCUM..]*

- **Contribution to the PLS model**: % Explained Variance of **X (as well for each x-variable)**
- **Validation: RMSEP** test set (truly independent)

2. Objects (samples, systems)

- **in X-space:** scores plots $\mathbf{t}_1$, $\mathbf{t}_2$, …
- **in Y-space:** scores plots $\mathbf{u}_1$, $\mathbf{u}_2$, …
- **inner relation U/T**: scores plots $\mathbf{t}_1$, $\mathbf{u}_1$, ; $\mathbf{t}_2$, $\mathbf{u}_2$, ; …

Check for outliers/trends

- **distance from PC model of X:** Plot X-residuals **E**
- **distance from PC model of Y:** Plot Y- residuals **F**
- **check randomness:** Plot Y-residuals vs Y, vs Order of spectra aquisition,..
- **Leverage:** $T(T^TT)^{-1}T^T$ (X-block)/ $U(U^TU)^{-1}U^T$ (Y-block) how influential objects are in determining X or Y models

3. Variables

- **Correlation among X and Y**: PLS weights, $\mathbf{w}_1$, $\mathbf{w}_2$, ..; regression cfs $\mathbf{B}_{PLS}$
- **Trends- Correlation among X variables**: X- Loadings plot **p**1, **p**2; …
- **Trends –Correlation among Y variables**: Y- Loadings plot **q**1, **q**2; …

PLS ALGORITHMS a bit

More in references

PLS steps

1. First PLS-component is calculated as the latent variable which has MAXIMUM COVARIANCE between the scores and modeled property y (or **Y scores)**.
Note that the criterion ''covariance'' is a compromise between maximum correlation coefficient (OLS) and maximum variance (PCA).

2. Next, the information (variance) of this component is removed from the **X**. This process is called PEELING or DEFLATION giving residual matrix **X**res (depending on algorithm **Y** can be deflated as well).
Actually it is a projection of the x-space on to a (hyper-)plane that is orthogonal to the direction of the found component.

3. From the residual matrix, the next PLS component is derived—again with maximum covariance between the scores and y (or **Y scores)**.

4. This procedure is continued until no improvement of modeling y is achieved. The number of PLS components defines the complexity of the model

In the standard versions of PLS, the scores of the PLS components are uncorrelated; the corresponding loading vectors, however, are in general not orthogonal.

Consideration about Algorithms

A complicating aspect of most PLS algorithms is the stepwise calculation of the components. After a component is computed, the residual matrices for X (and eventually Y) are determined.
The next PLS component is calculated from the residual matrices and therefore its parameters (scores, loadings, weights) do not relate to X but to the residual matrices. However, equations exist, that relate the PLS-x-loadings and PLS-x-scores to the original x-data, and that also provide the regression coefficients of the final model for the original x-data.

In the following slides the most used Algorithms:
NIPALS , SIMPLS are schematically reported

NIPALS

NIPALS

(1) initialize $\boldsymbol{u}_1$ for instance by the first column of $\boldsymbol{Y}$
(2) $\boldsymbol{w}_1 = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{u}_1/(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u}_1)$
(3) $\boldsymbol{w}_1 = \boldsymbol{w}_1/\|\boldsymbol{w}_1\|$
(4) $\boldsymbol{t}_1 = \boldsymbol{X}\boldsymbol{w}_1$
(5) $\boldsymbol{c}_1 = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{t}_1/(\boldsymbol{t}_1^{\mathrm{T}}\boldsymbol{t}_1)$
(6) $\boldsymbol{c}_1 = \boldsymbol{c}_1/\|\boldsymbol{c}_1\|$
(7) $\boldsymbol{u}_1^* = \boldsymbol{Y}\boldsymbol{c}_1$
(8) $\boldsymbol{u}_\Delta = \boldsymbol{u}_1^* - \boldsymbol{u}_1$
(9) $\Delta u = \boldsymbol{u}_\Delta^{\mathrm{T}}\boldsymbol{u}_\Delta$
(10) stop if $\Delta u < \varepsilon$ (with ε for instance set to 10^{-6}); otherwise $\boldsymbol{u}_1 = \boldsymbol{u}_1^*$ and go to step 2

iterate

at convergence

(11) $\boldsymbol{p}_1 = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{t}_1/(\boldsymbol{t}_1^{\mathrm{T}}\boldsymbol{t}_1)$
(12) $\boldsymbol{q}_1 = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{u}_1/(\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u}_1)$
(13) $d_1 = \boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{t}_1/(\boldsymbol{t}_1^{\mathrm{T}}\boldsymbol{t}_1)$
(14) $\boldsymbol{X}_1 = \boldsymbol{X} - \boldsymbol{t}_1\boldsymbol{p}_1^{\mathrm{T}}$ and $\boldsymbol{Y}_1 = \boldsymbol{Y} - d_1\boldsymbol{t}_1\boldsymbol{c}_1^{\mathrm{T}}$ **Go to next component calculation**

Re-expressing for prediction $\boldsymbol{B} = \boldsymbol{W}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{W})^{-1}\boldsymbol{C}^{\mathrm{T}}$, $\Rightarrow$ $\widehat{\mathbf{Y}} = \mathbf{X}\mathbf{B}_{\mathrm{PLS}}$

SIMPLS

INPUT $n \times p$ matrix $\mathbf{X}$,
$n \times m$ matrix $\mathbf{Y}$,
number of factors A.

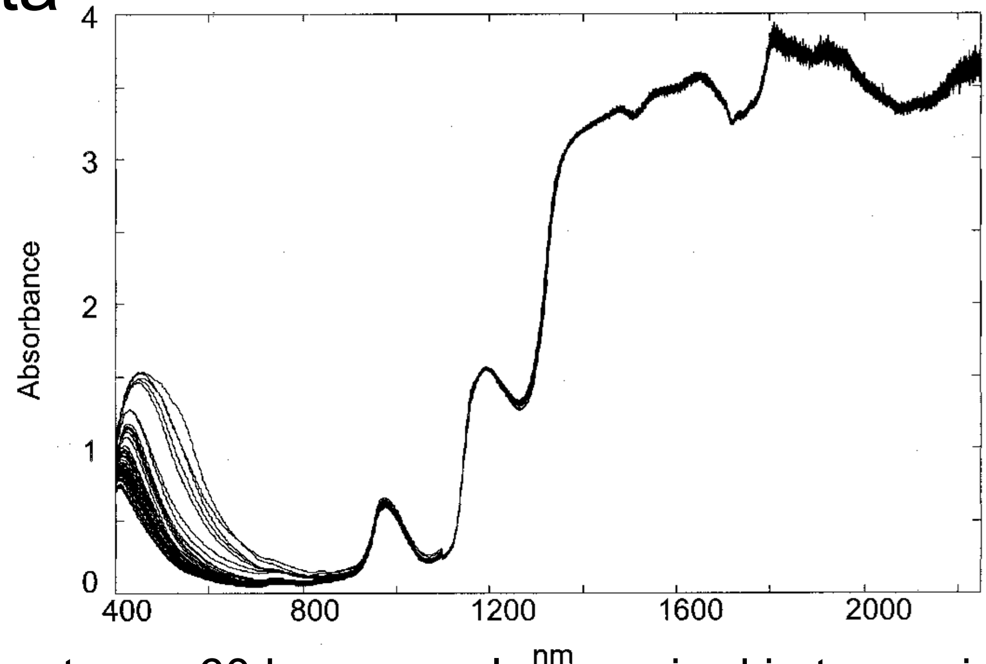
SIMPLS

$\mathbf{Y}_0 = \mathbf{Y} - \mathrm{MEAN}(\mathbf{Y})$	center $\mathbf{Y}$
$\mathbf{S} = \mathbf{X}'^* \mathbf{Y}_0$	cross-product
For $a = 1, \ldots, A$	per dimension
q = dominant eigenvector of $\mathbf{S}'^*\mathbf{S}$	Y block factor weights
$r = \mathbf{S}^* q$	X block factor weights
$t = \mathbf{X}^* r$	X block factor scores
$t = t - \mathrm{MEAN}(t)$	center scores
$normt = \mathrm{SQRT}(t'^* t)$	compute norm
$t = t/normt$	normalize scores
$r = r/normt$	adapt weights accordingly
$p = \mathbf{X}'^* t$	X block factor loadings
$q = \mathbf{Y}_0'^* t$	Y block factor loadings
$u = \mathbf{Y}_0^* q$	Y block factor scores
$v = p$	initialize orthogonal loadings
if $a > 1$ then	
$v = v - \mathbf{V}^*(\mathbf{V}'^* p)$	make $v \perp$ previous loadings
$u = u - \mathbf{T}^*(\mathbf{T}'^* u)$	make $u \perp$ previous t' values
end	
$v = v/\mathrm{SQRT}(v'^* v)$	normalize orthogonal loadings
$\mathbf{S} = \mathbf{S} - v^*(v'^* \mathbf{S})$	deflate $\mathbf{S}$ with respect to current loadings
Store r, t, p, q, u, and v into into $\mathbf{R}$, $\mathbf{T}$, $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{U}$, and $\mathbf{V}$, respectively.	
End	

$\mathbf{B} = \mathbf{R}^* \mathbf{Q}'$	regression coefficients
$h = \mathrm{DIAG}(\mathbf{T}^* \mathbf{T}') + 1/n$	leverages of objects
$\mathbf{varX} = \mathrm{DIAG}(\mathbf{P}'^* \mathbf{P})/(n-1)$	variance explained for X variables
$\mathbf{varY} = \mathrm{DIAG}(\mathbf{Q}'^* \mathbf{Q})/(n-1)$	variance explained for Y variables

$\Rightarrow \hat{\mathbf{Y}} = \mathbf{X}\mathbf{B}_{PLS}$

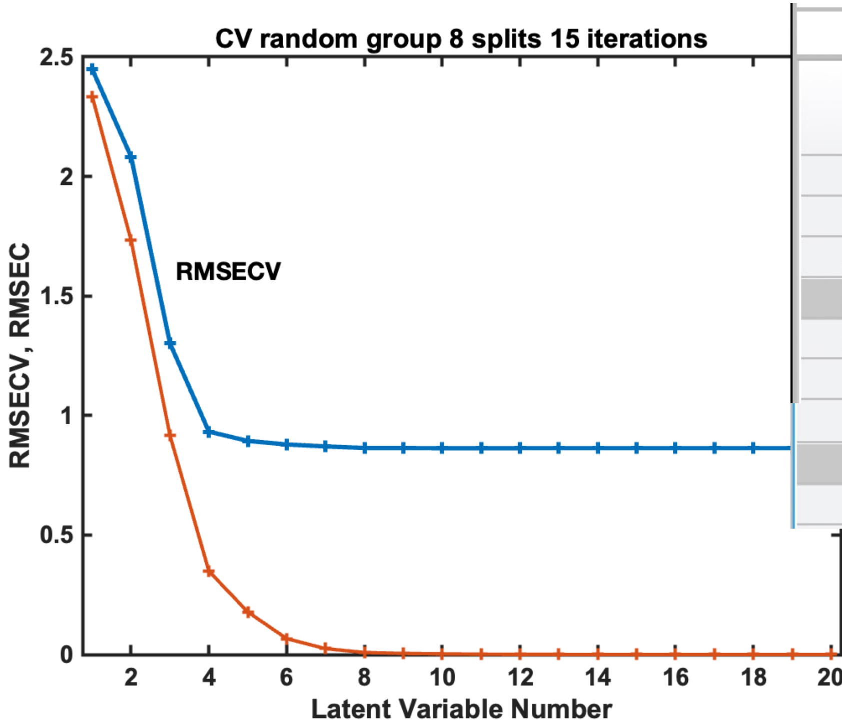
- Beer data



Vis-NIR spectra on 60 beer samples acquired in transmission mode (transformed in absorbance)
40 calibration; 20 validation. Samples

Want to calibrate the “extract” concentration which is indicating the substrate potential for the yeast to ferment alcohol and serving as a taxation parameter.

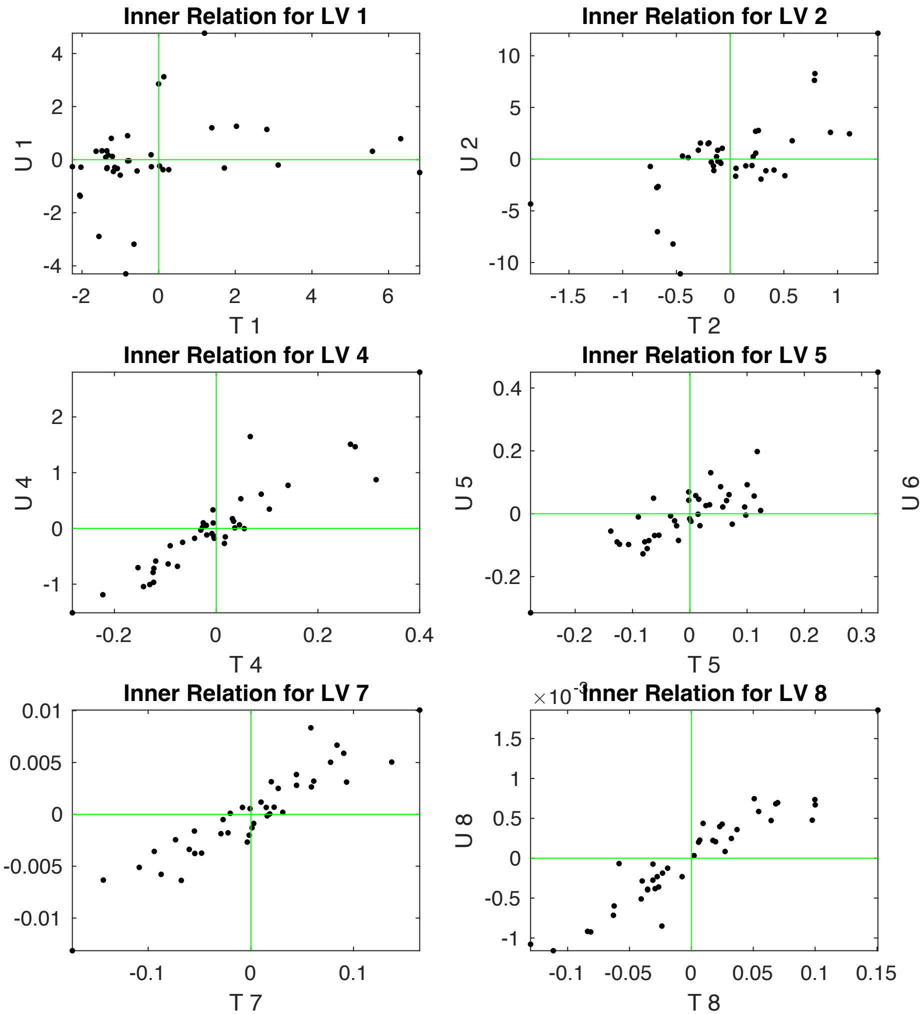
- Beer data



Percent Variance Captured by Model (* = suggested)

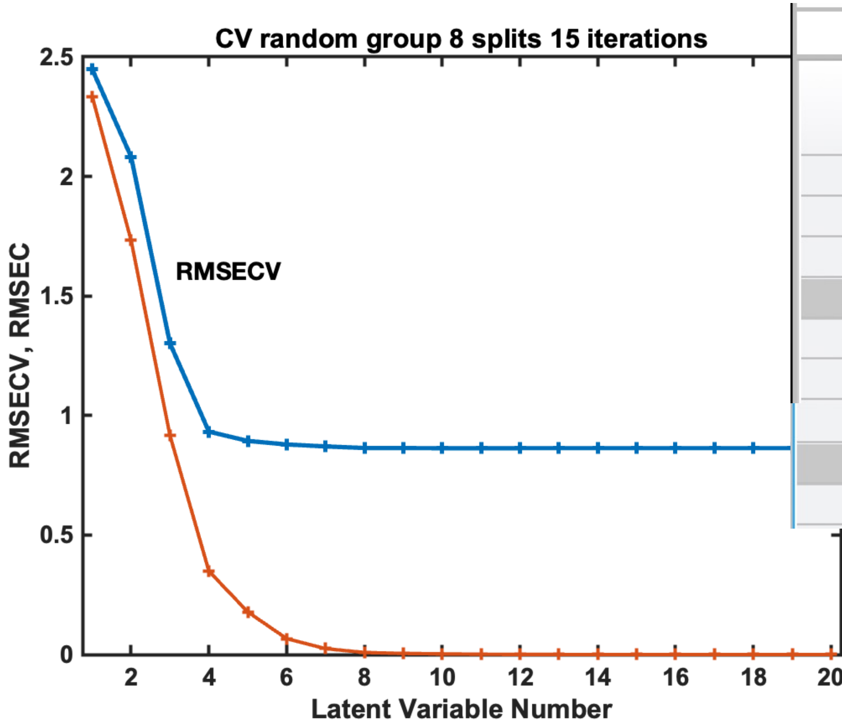
	X-Block LV	X-BLock Cumulative	Y-Block LV	y-Block Cumulative	RMSECV
1	89.67	89.67	5.94	5.94	2.4439
2	6.15	95.81	42.27	48.20	2.0971
3	0.98	96.79	37.32	85.52	1.288
4	0.35	97.14	12.38	97.90	0.92057
5	0.18	97.32	1.55	99.45	0.86862
6	0.15	97.47	0.47	99.92	0.85681
7	0.09	97.57	0.07	99.99	0.84714
8	0.07	97.63	0.01	100.00	0.8398
9	0.12	97.75	0.00	100.00	0.83929

• Beer data



going for 8 LVs ?

- Beer data



Percent Variance Captured by Model (* = suggested)

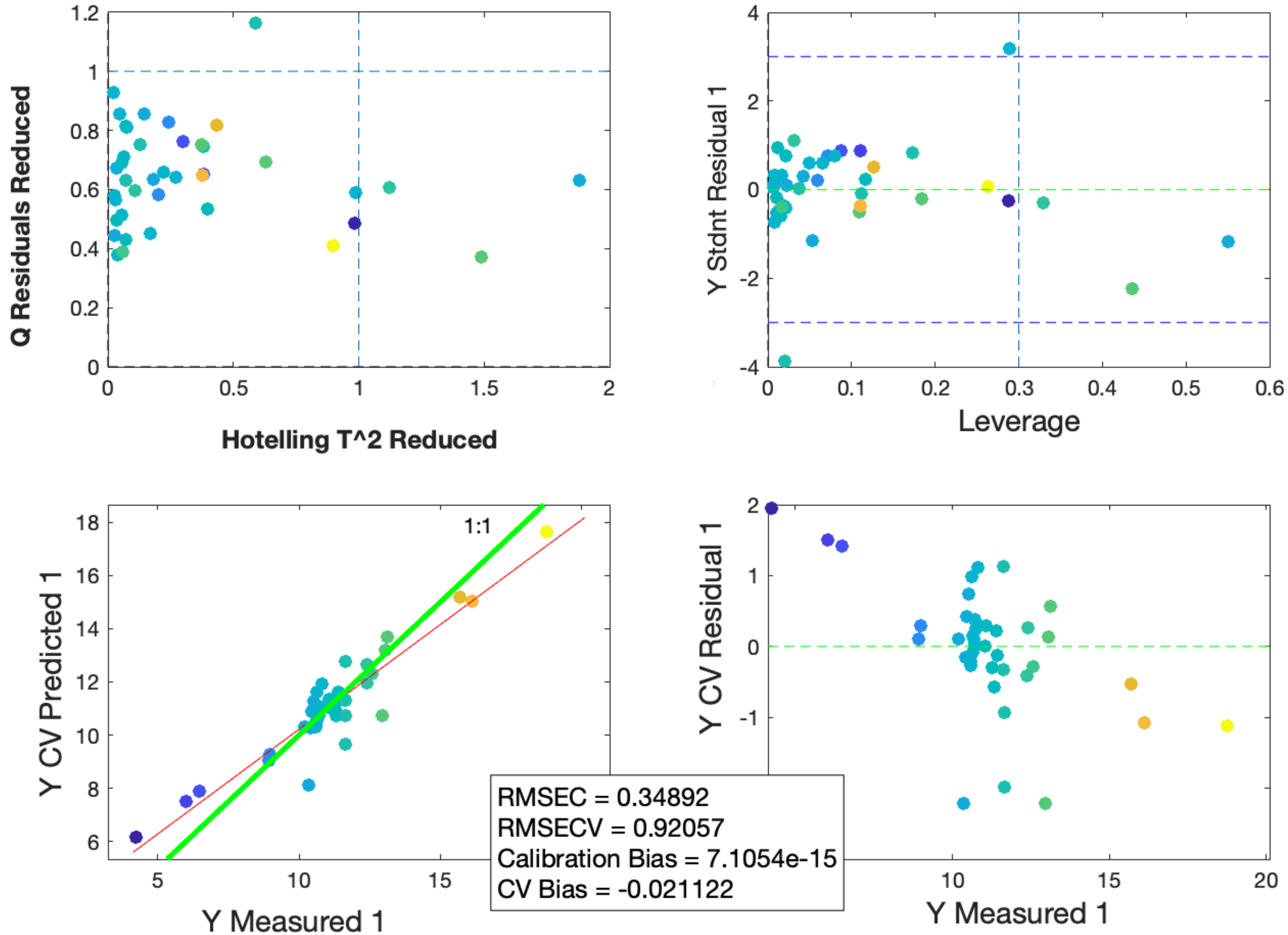
	X-Block LV	X-BLock Cumulative	Y-Block LV	y-Block Cumulative	RMSECV
1	89.67	89.67	5.94	5.94	2.4439
2	6.15	95.81	42.27	48.20	2.0971
3	0.98	96.79	37.32	85.52	1.288
4	0.35	97.14	12.38	97.90	0.92057
5	0.18	97.32	1.55	99.45	0.86862
6	0.15	97.47	0.47	99.92	0.85681
7	0.09	97.57	0.07	99.99	0.84714
8	0.07	97.63	0.01	100.00	0.8398
9	0.12	97.75	0.00	100.00	0.83929

going for 8 LVs ?

- Beer data

PLS 4LVs

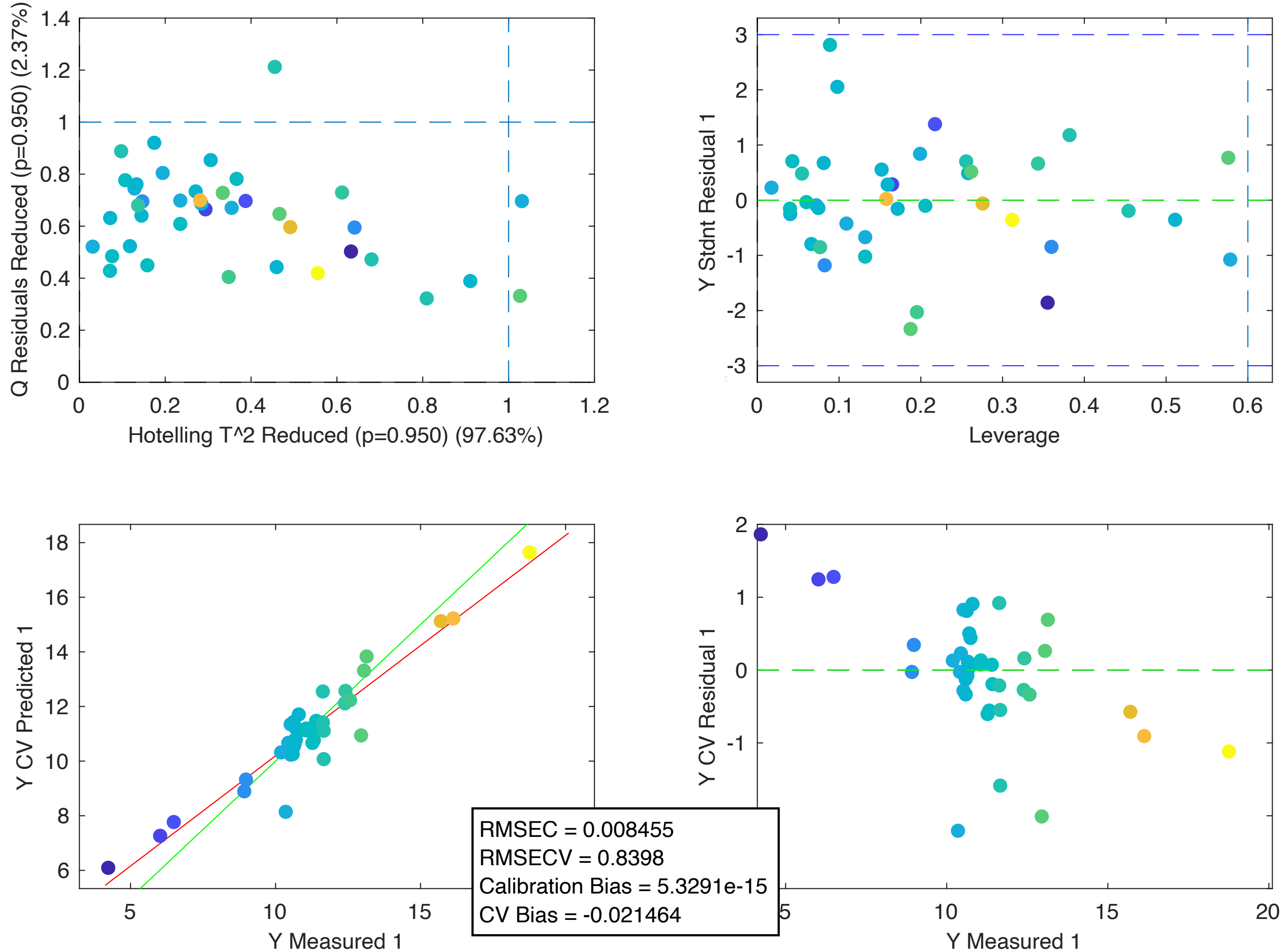
Colored by y-values



- Beer data

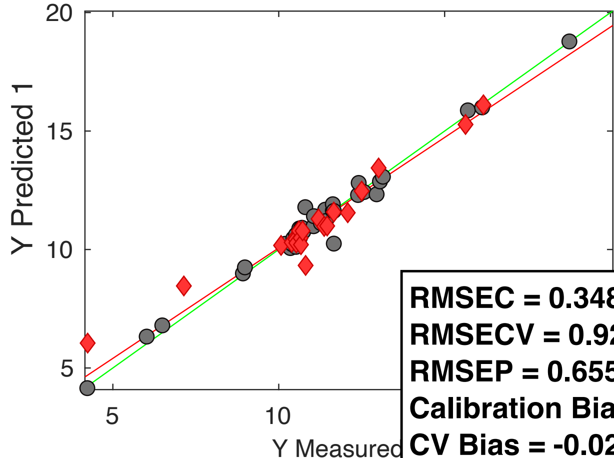
PLS 8LVs

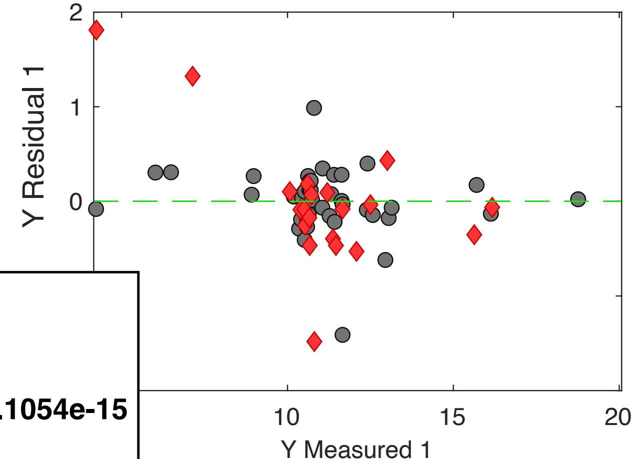
Colored by y-values



- Beer data

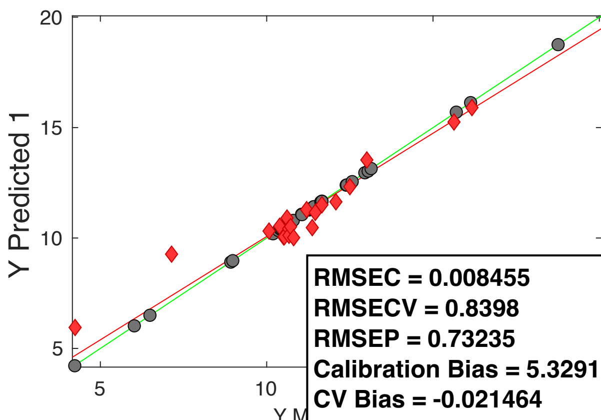
PLS 4 LVs

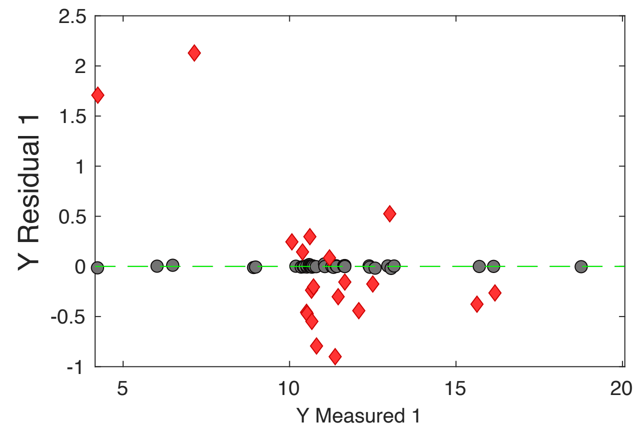




RMSEC = 0.34892
RMSECV = 0.92057
RMSEP = 0.65507
Calibration Bias = 7.1054e-15
CV Bias = -0.021122
Prediction Bias = -0.023083

PLS 8 LVs

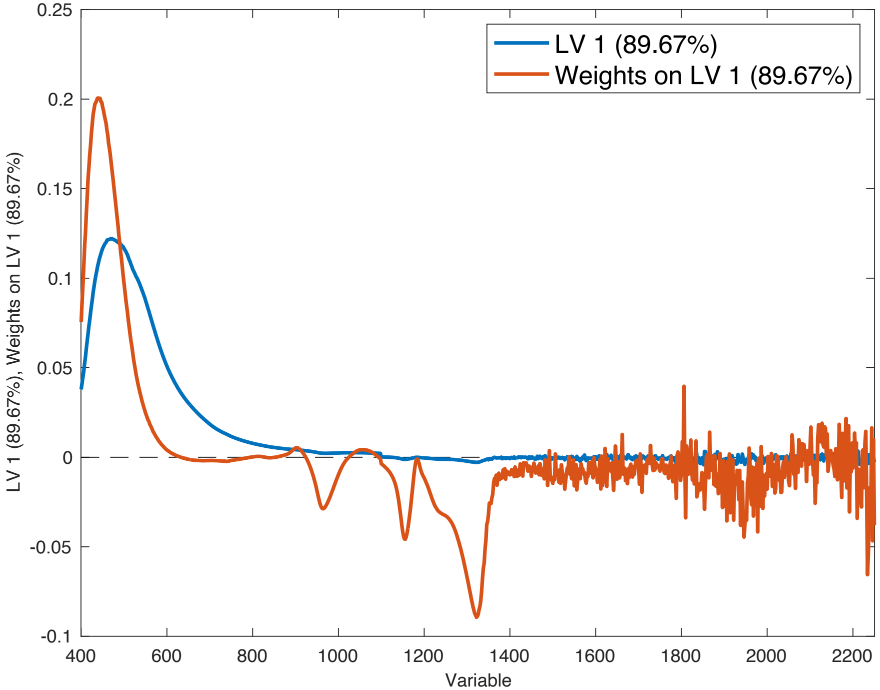


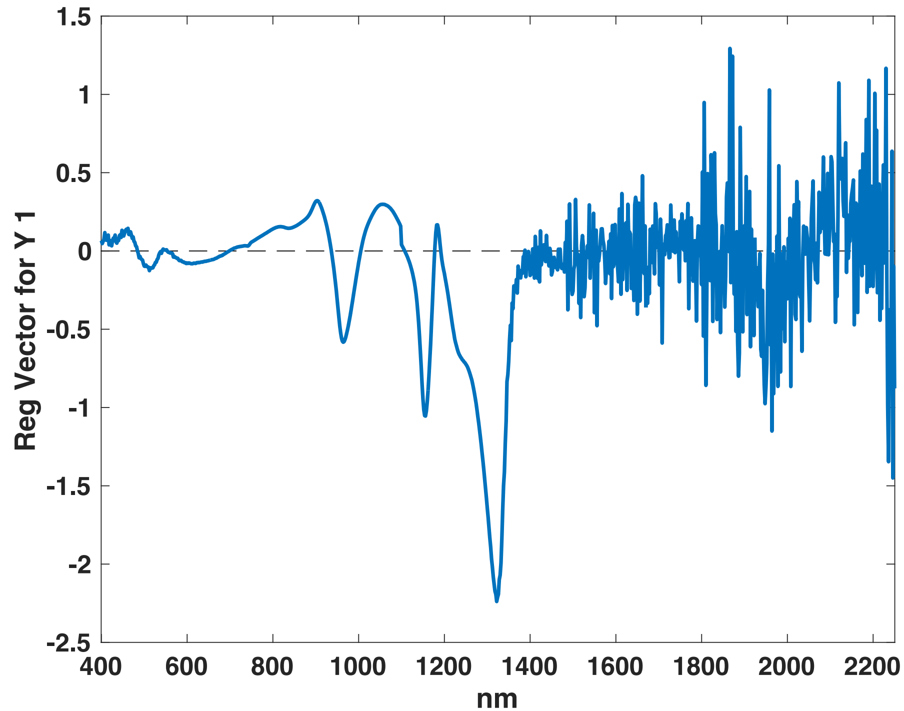


RMSEC = 0.008455
RMSECV = 0.8398
RMSEP = 0.73235
Calibration Bias = 5.3291e-15
CV Bias = -0.021464
Prediction Bias = -0.0098749

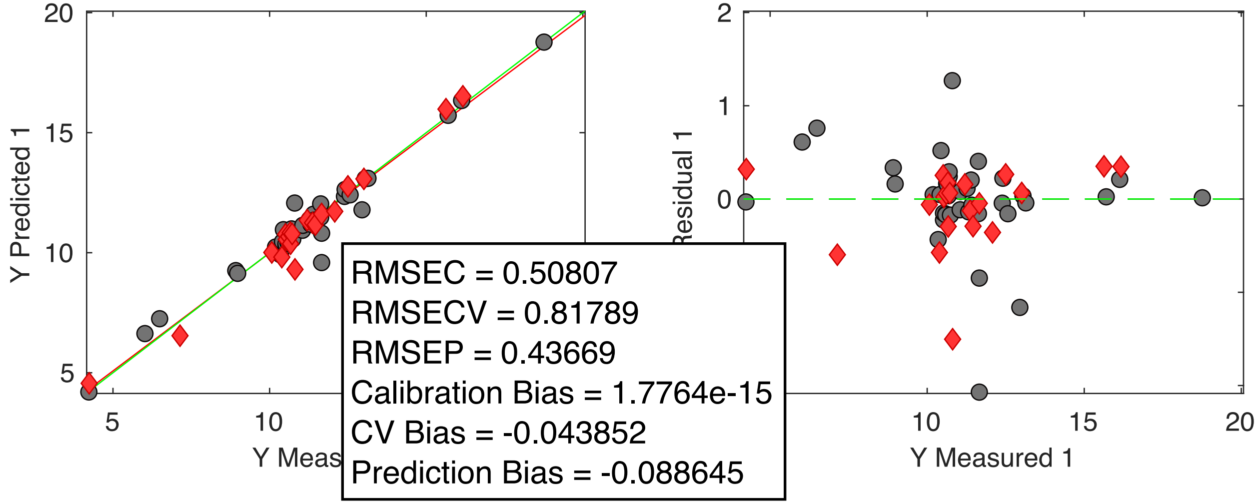
- Beer data

Regression cfs 4 LVs

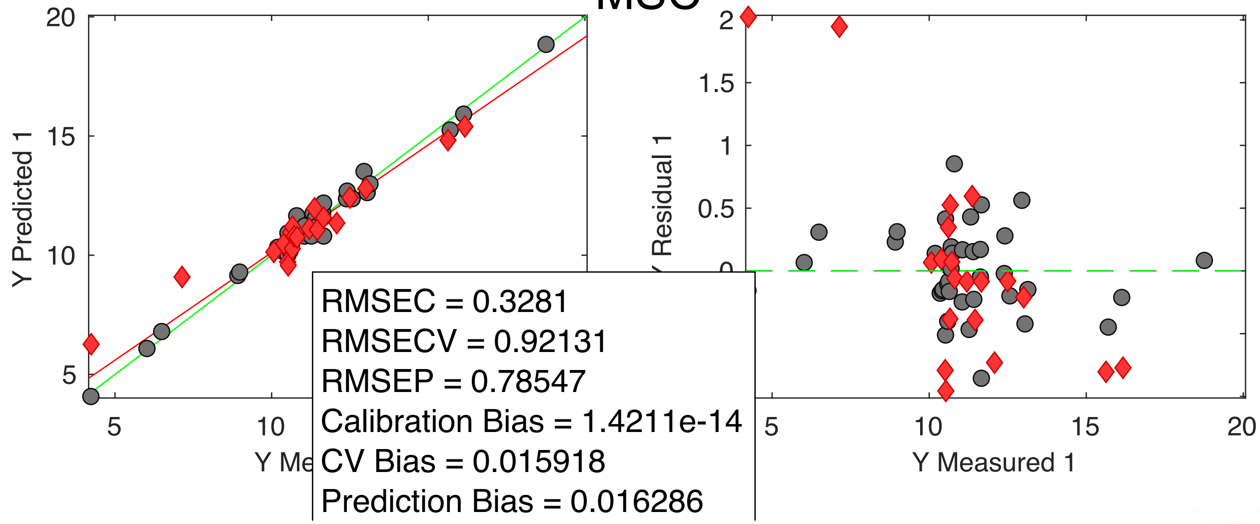




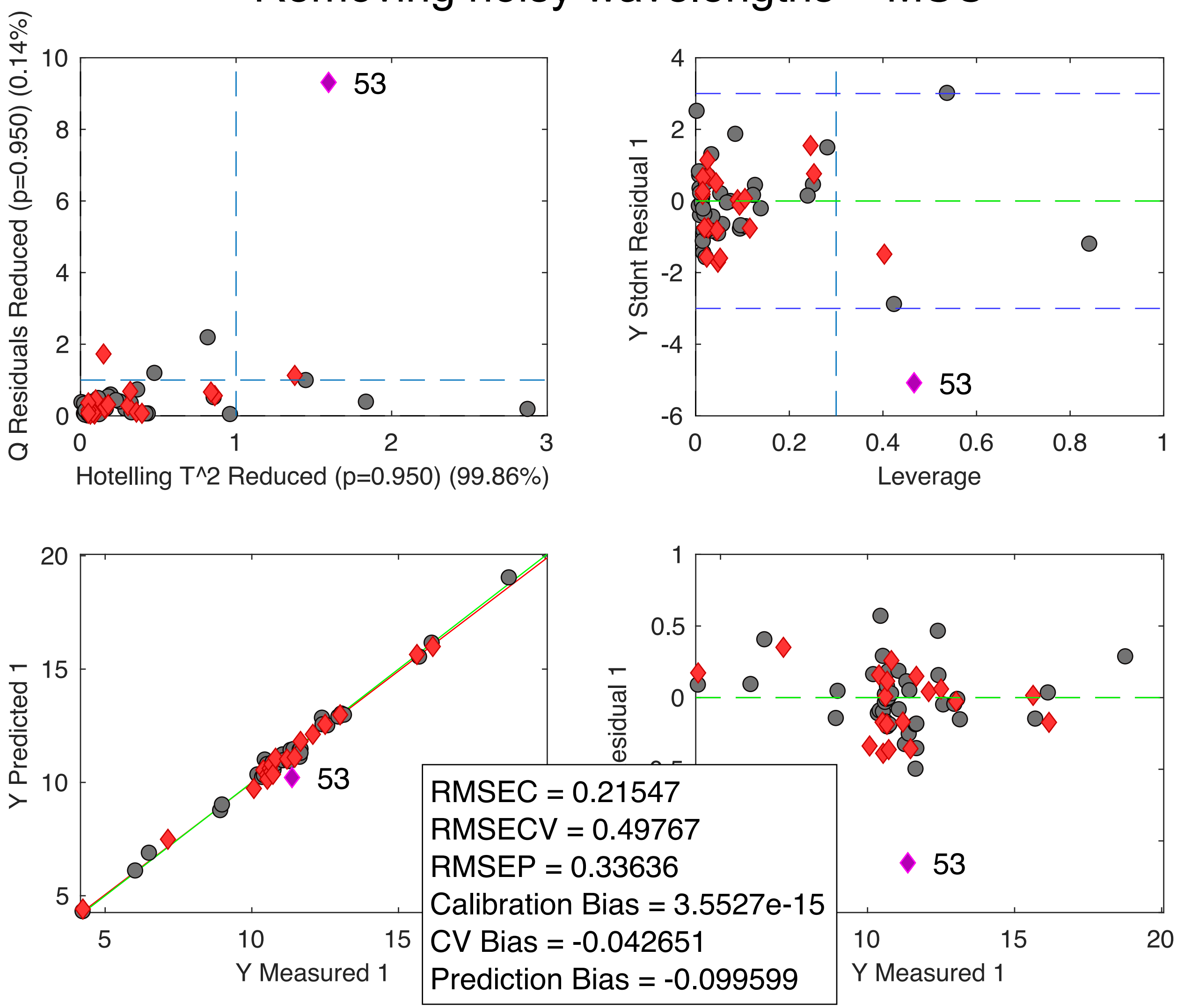
Removing noisy wavelengths



MSC

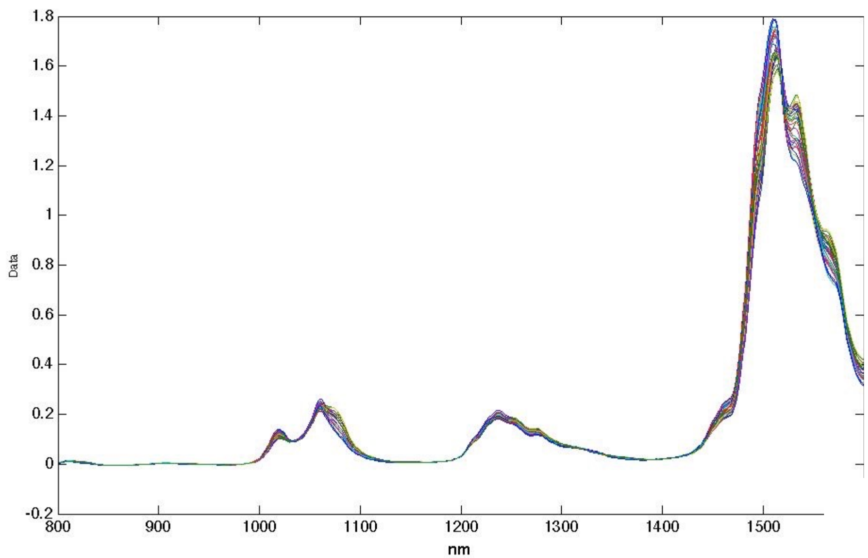


Removing noisy wavelengths + MSC

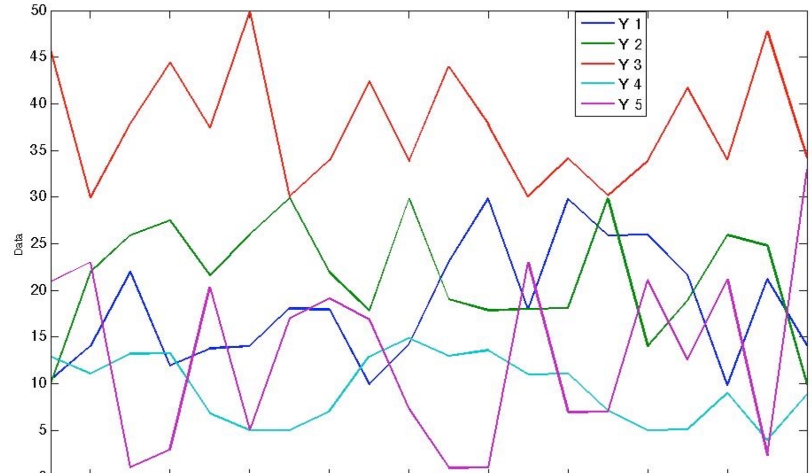


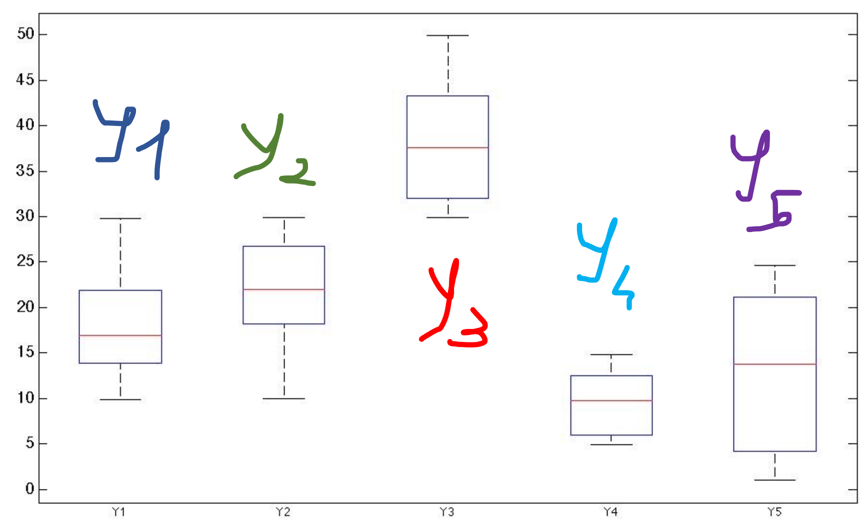
Calibration Example: 1. plot raw data

NIR gasoline data



Y 5 analytes

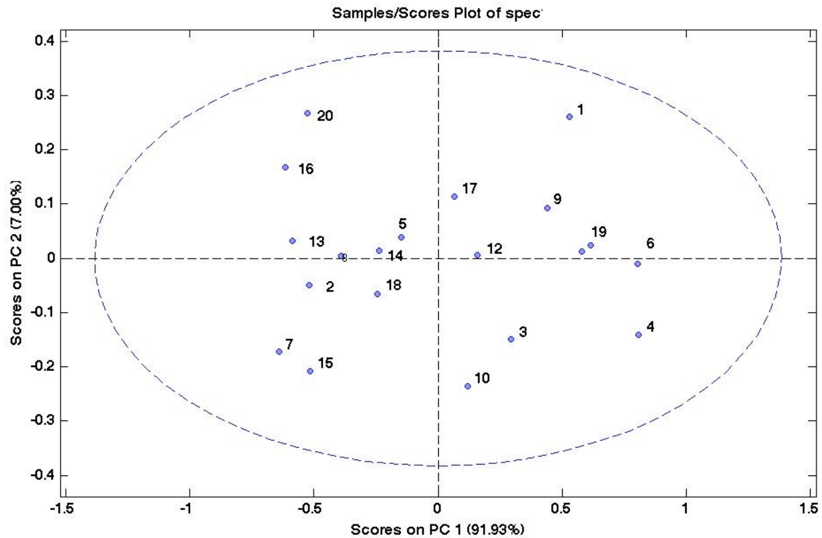


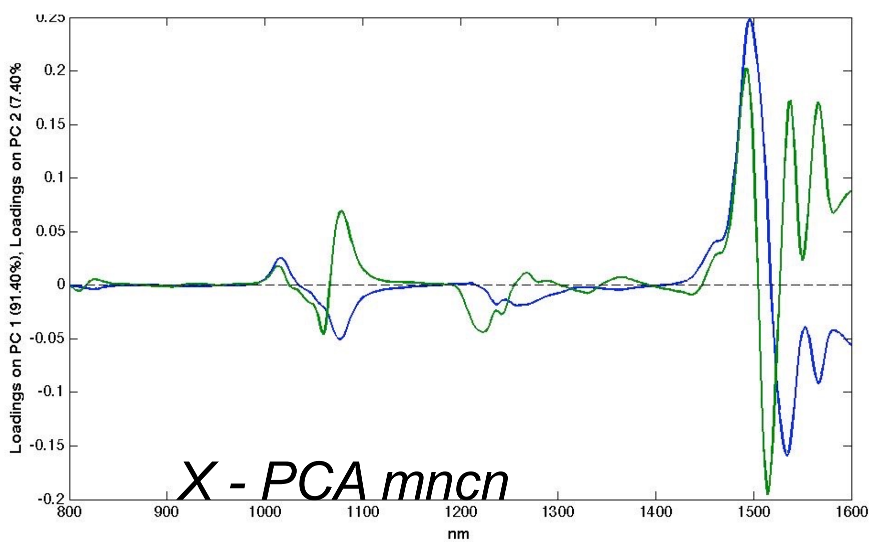


Calibration Example: 2. Explorative PCA

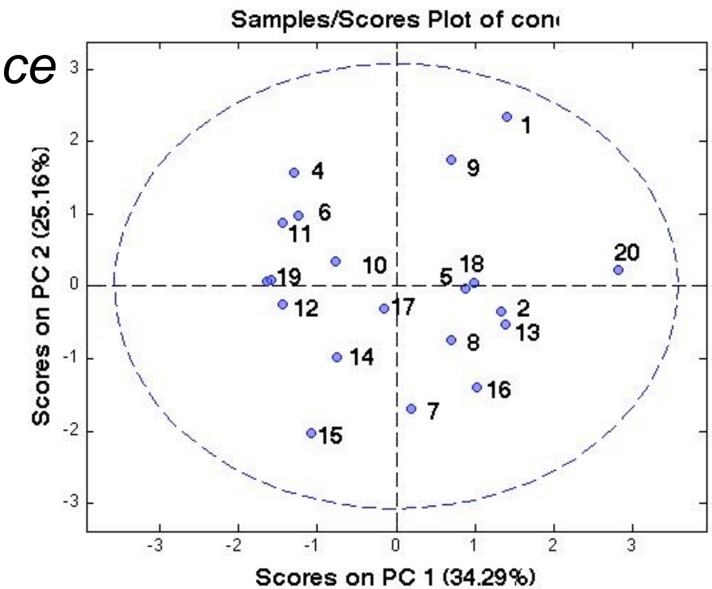
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

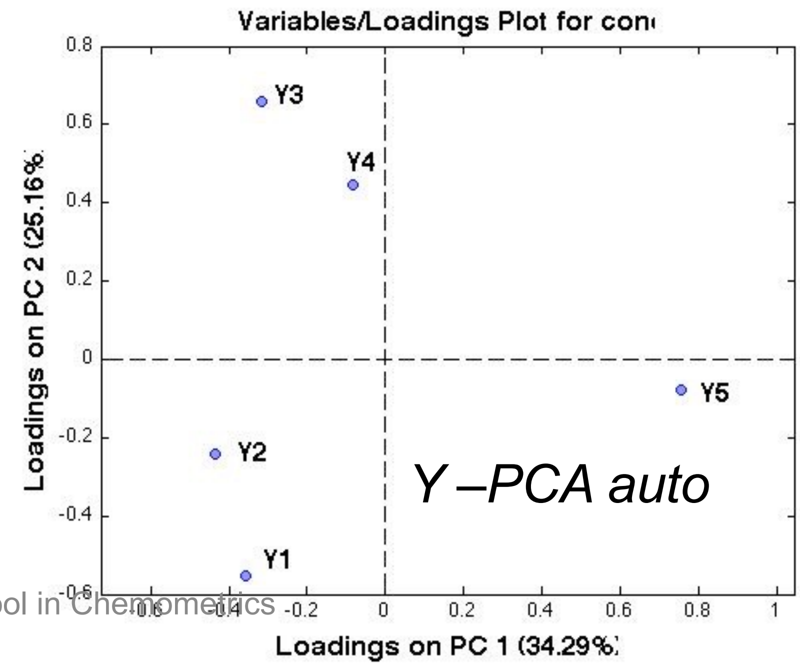
X - *space*



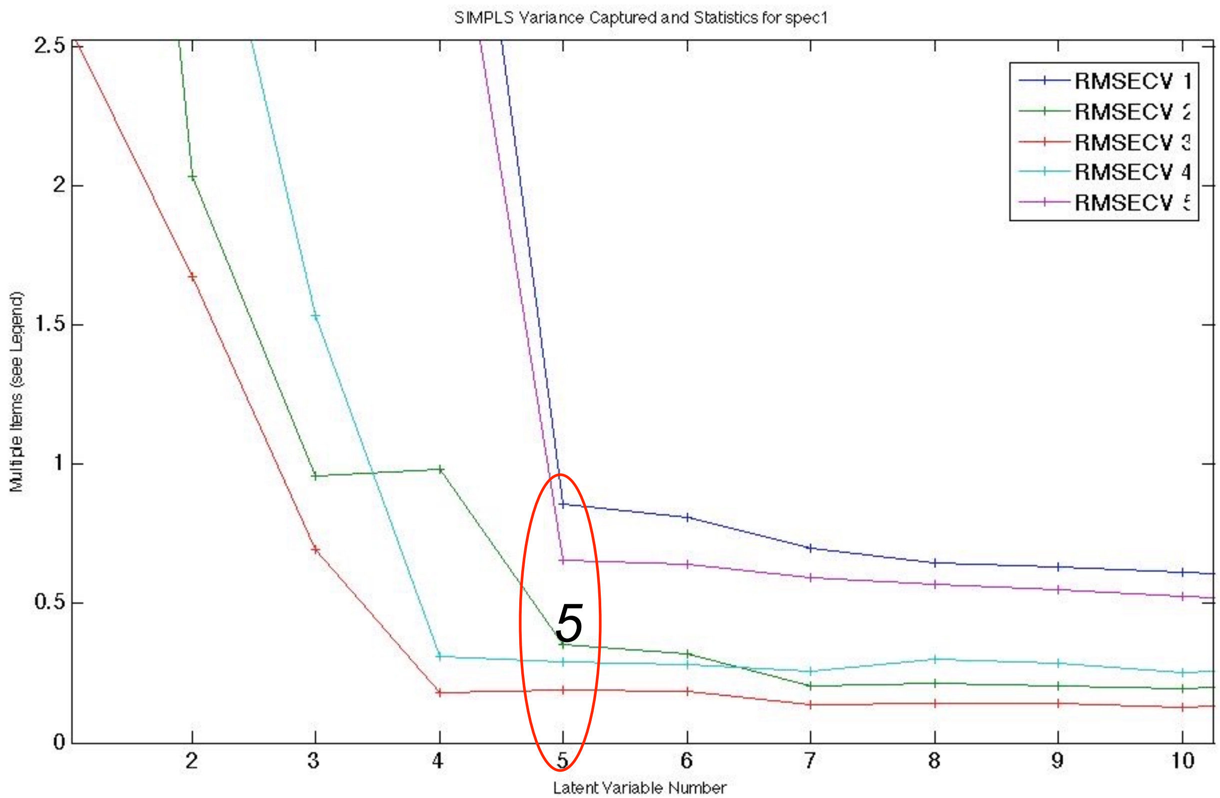


Y - *space*





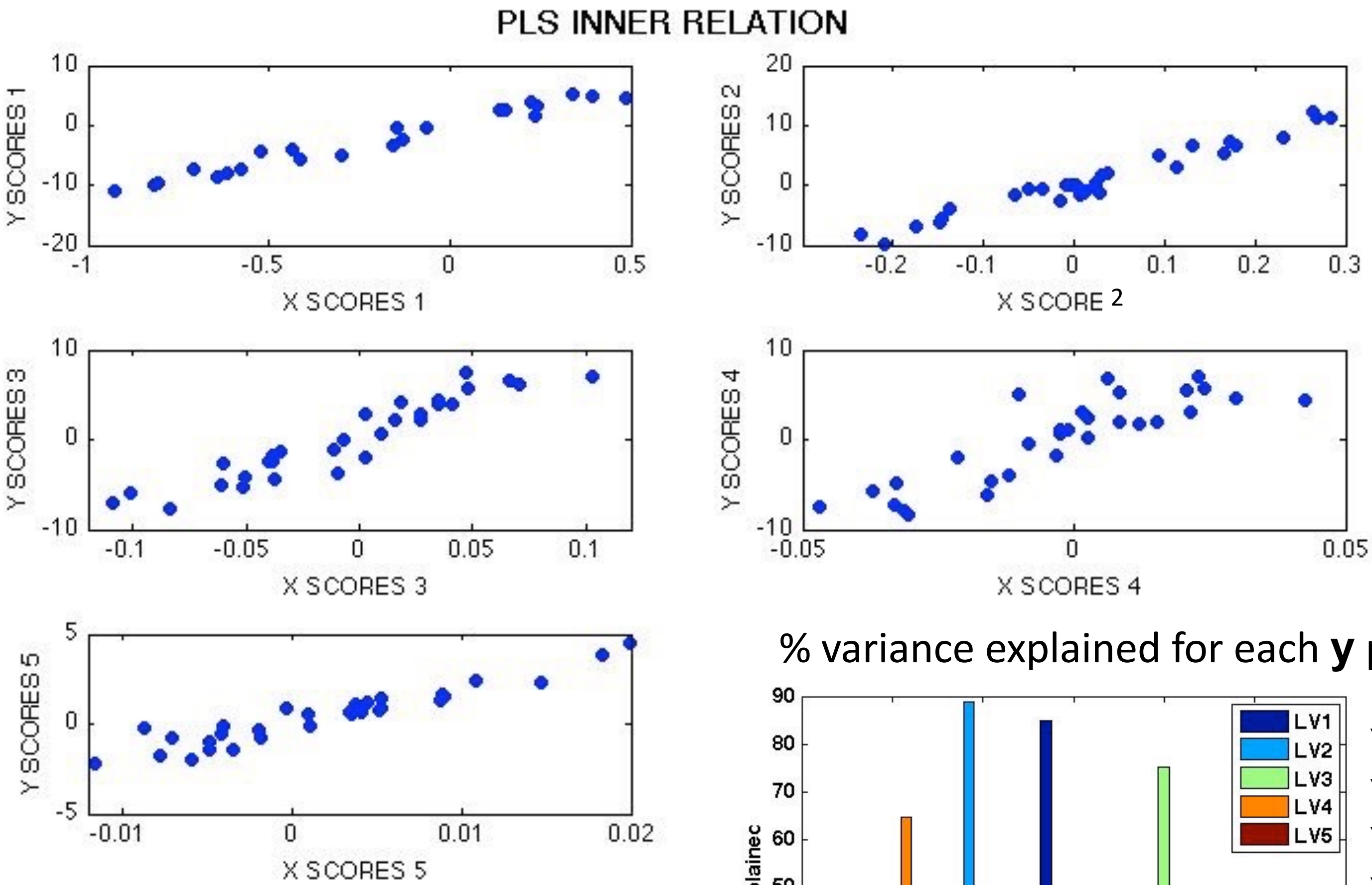
Calibration Example: 3. Fit & choose PLS model



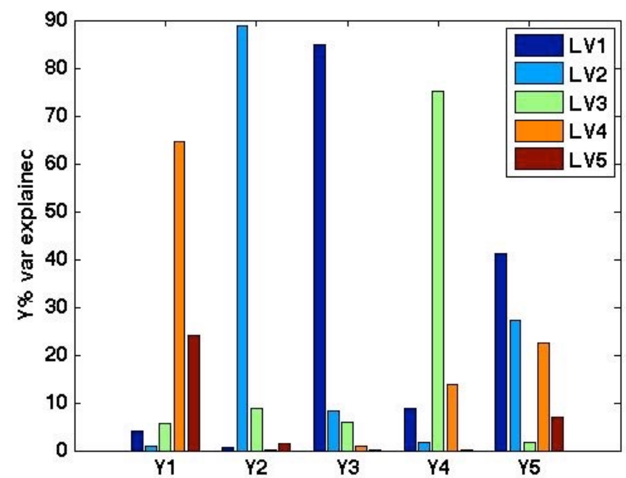
Choose 5LV
Cross validation
Venetian blinds 5 split

% V (fit)
X 99.9
Y 99.6

Calibration Example: 3. inspect PLS model

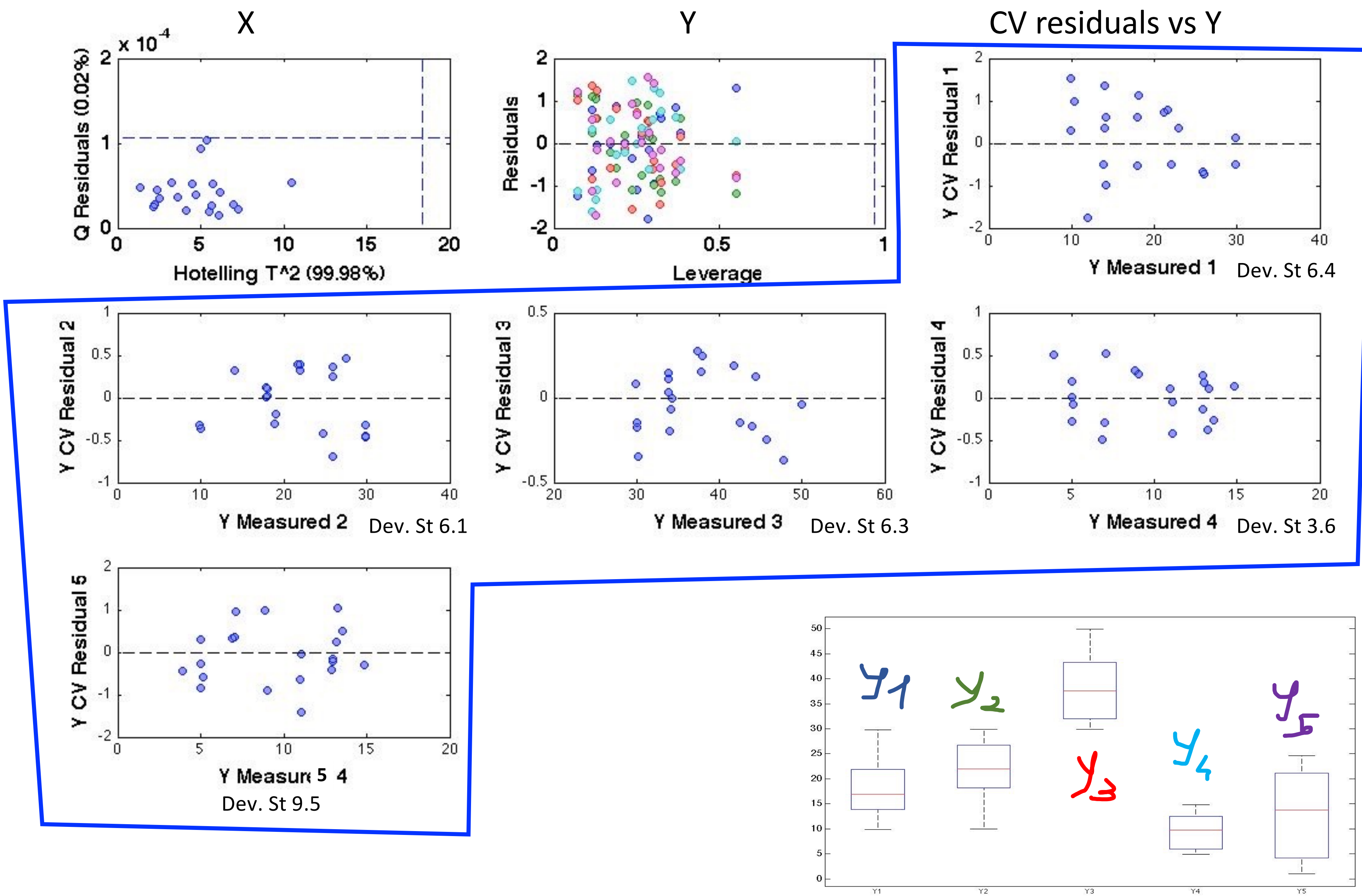


% variance explained for each **y** per LV



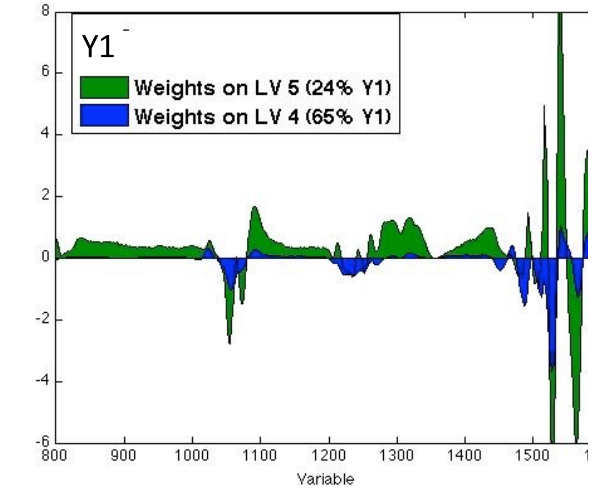
Y1 mainly LV4 (LV5)
Y2 mainly LV2
Y3 mainly LV1
Y4 mainly LV3
Y5 mainly LV1,2,4

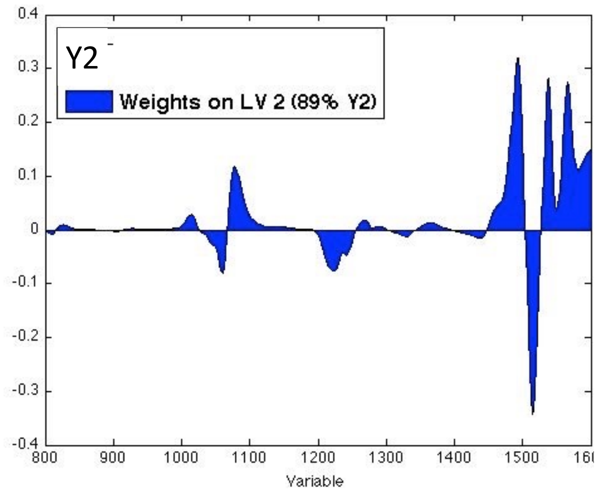
Calibration Example: 3. inspect PLS model

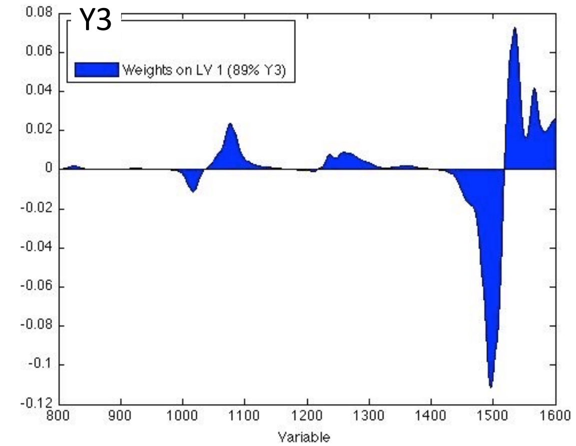


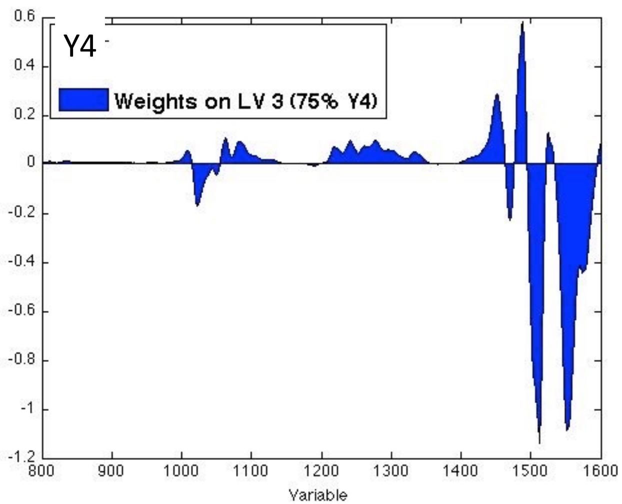
Calibration Example: 4. interpret PLS model

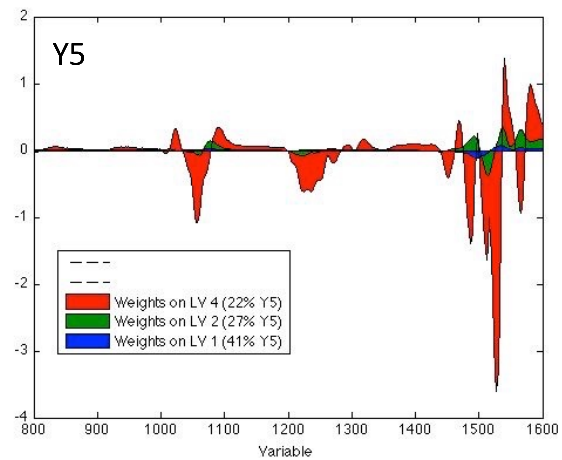
X-variables weights



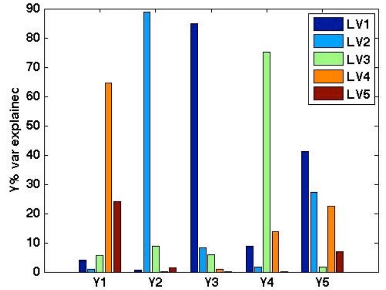






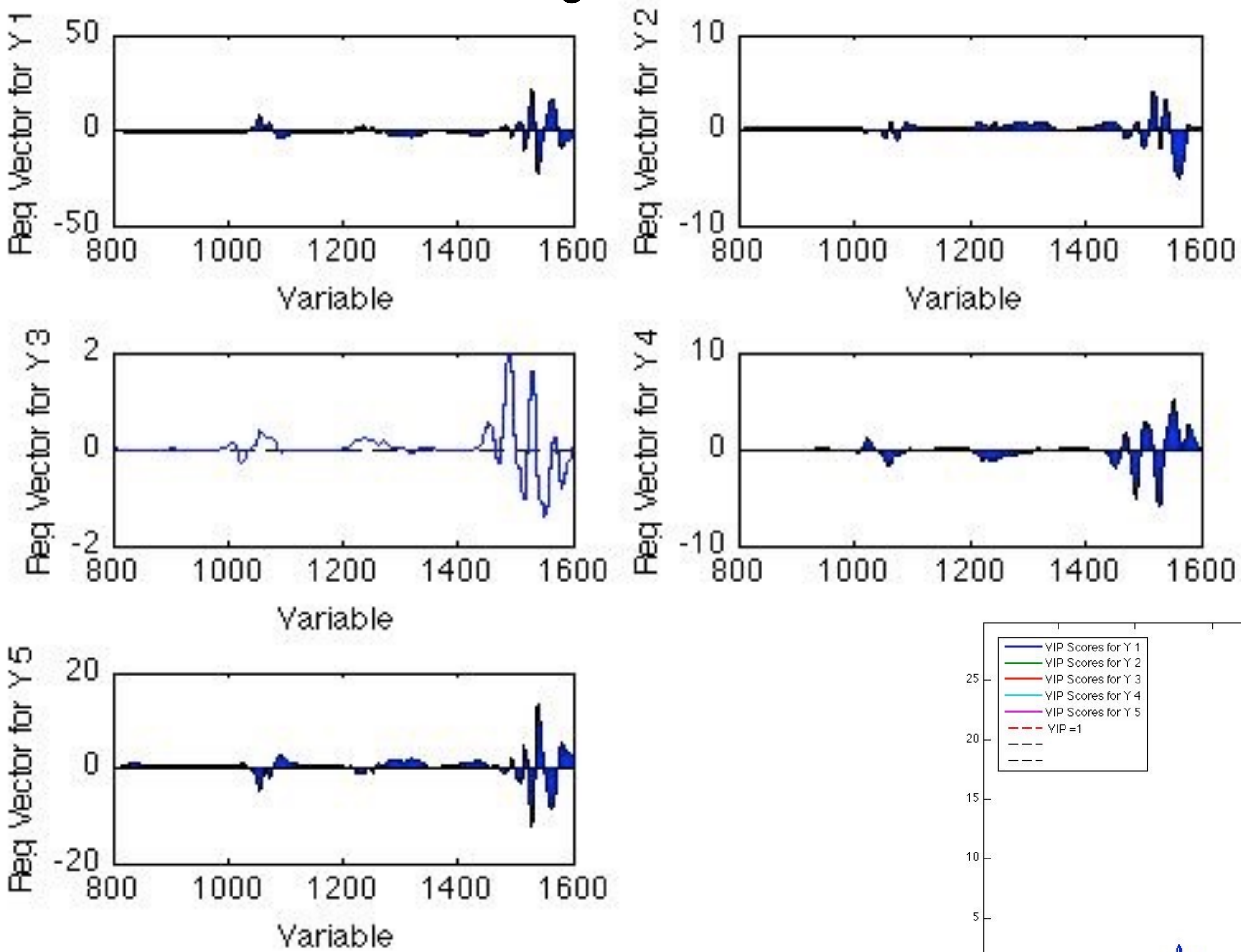


Y1 mainly LV4 (LV5)
Y2 mainly LV2
Y3 mainly LV1
Y4 mainly LV3
Y5 mainly LV1,2,4

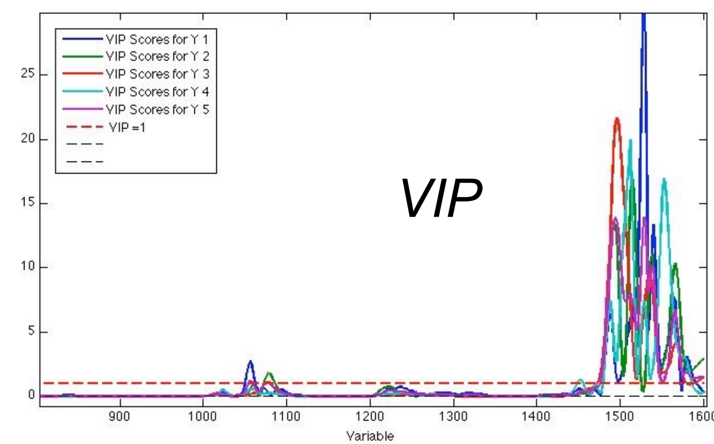


Calibration Example: 4. interprete PLS model

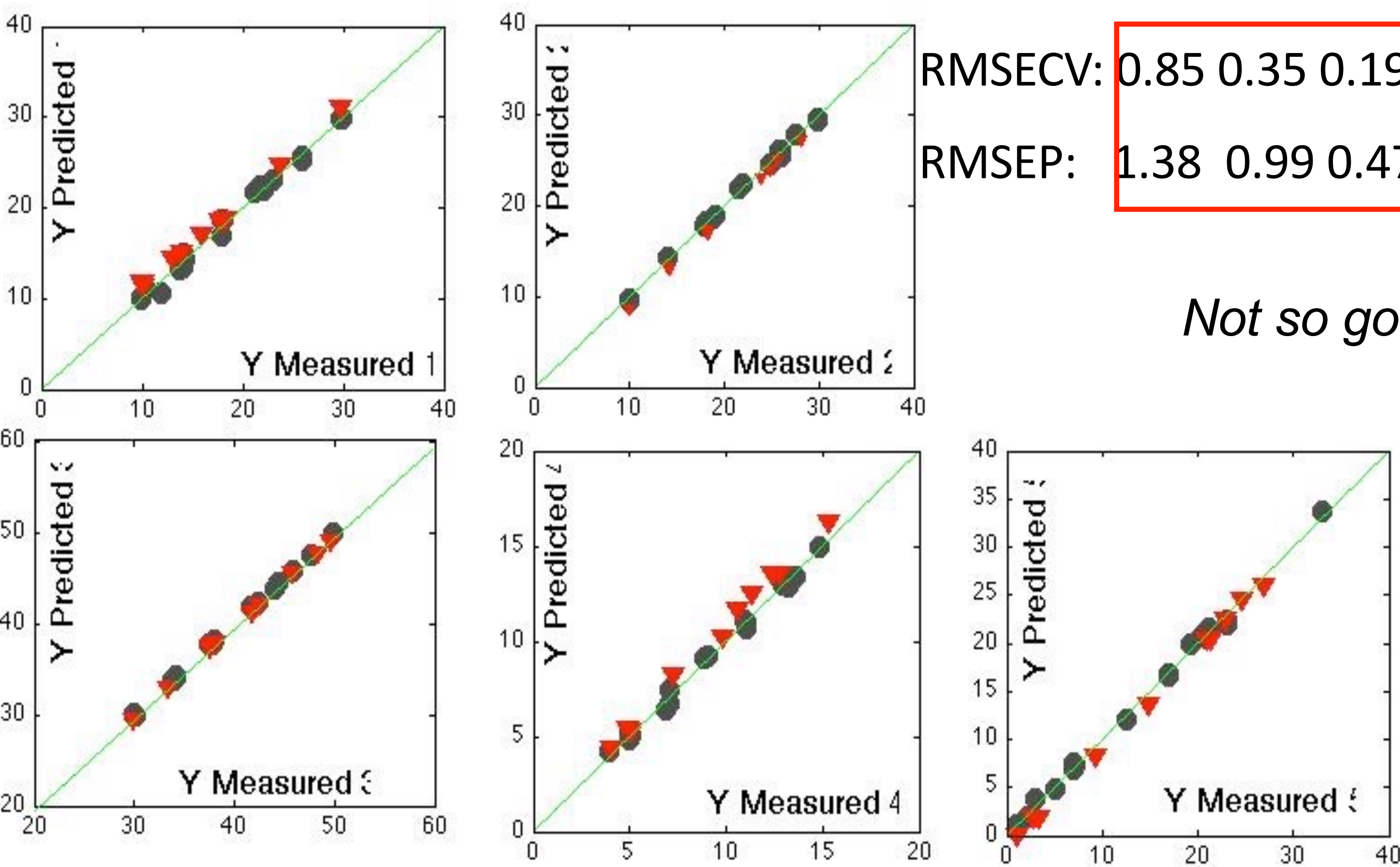
Regression cfs



VIP



Calibration Example: 5. validate PLS model

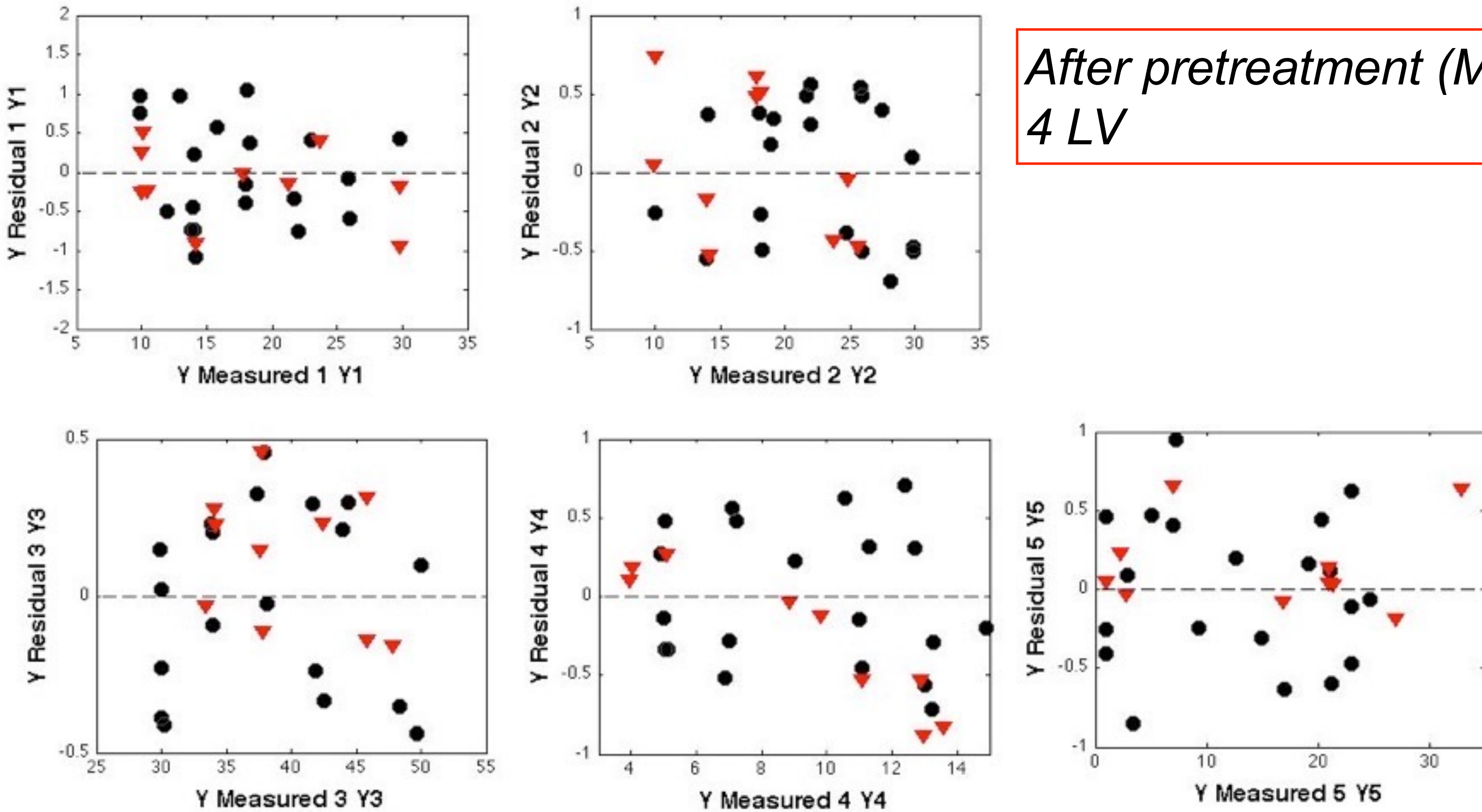


RMSECV: 0.85 0.35 0.19 0.29 0.65

RMSEP: 1.38 0.99 0.47 0.88 0.84

Not so good

Calibration Example: 5. validate PLS model



After pretreatment (MSC)
4 LV

RMSECV:	**1.05**	**0.62**	**0.38**	**0.43**	**0.85**
RMSEP:	**0.48**	**0.46**	**0.24**	**0.57**	**0.31**

- Plan for tomorrow:
  - PLS practical work (you) 2 data sets in chemflow
  1. Triglycerides 2. Apples

  - PLS for discrimination (PLS-DA) (me)

  - PLS practical work (you) 2 data sets in chemflow
  1. FeedMIR 2. FeedNIRmap

General workflow

aim	tools	result
General data examination	Plots, Histogrammes, PCA, Lab notebook, etc.	Error / outlier detection
Preprocessing data independent from the model	Interpolations, zone selection non linear transforms	Using theoretical knowledge e.g. Beer Lambert law
Test set selection	Kennard & Stones, Duplex y sorting, PCA	Dataset simulating the future using of the model
Preprocessing spectra	Linearization clutter removal	Useful subspace
Model calculation	PLS, Local, SIMCA, SVM, Sel + MLR, PLS-DA, Sel+LDA, …	Figures of merit (R^2, SEC), loadings, weights ; regression cfs, Model Interpretation
Model validation	(double / repeated) Cross Validation	Figures of merit (R^2, SECV) Model variation
Model test	Application preprocessing + model	Figures of merit (R^2, SEP, Biais)
Robustification	DOE. Standardisation Orthogonalisation, preprocessing	Figures of merit (R^2, SEP, Biais)