NIR based approach for counterfeit drugs' detection

Oxana Rodionova

Semenov Institute of Chemical Physics, RAS Moscow

Russian Chemometric Society

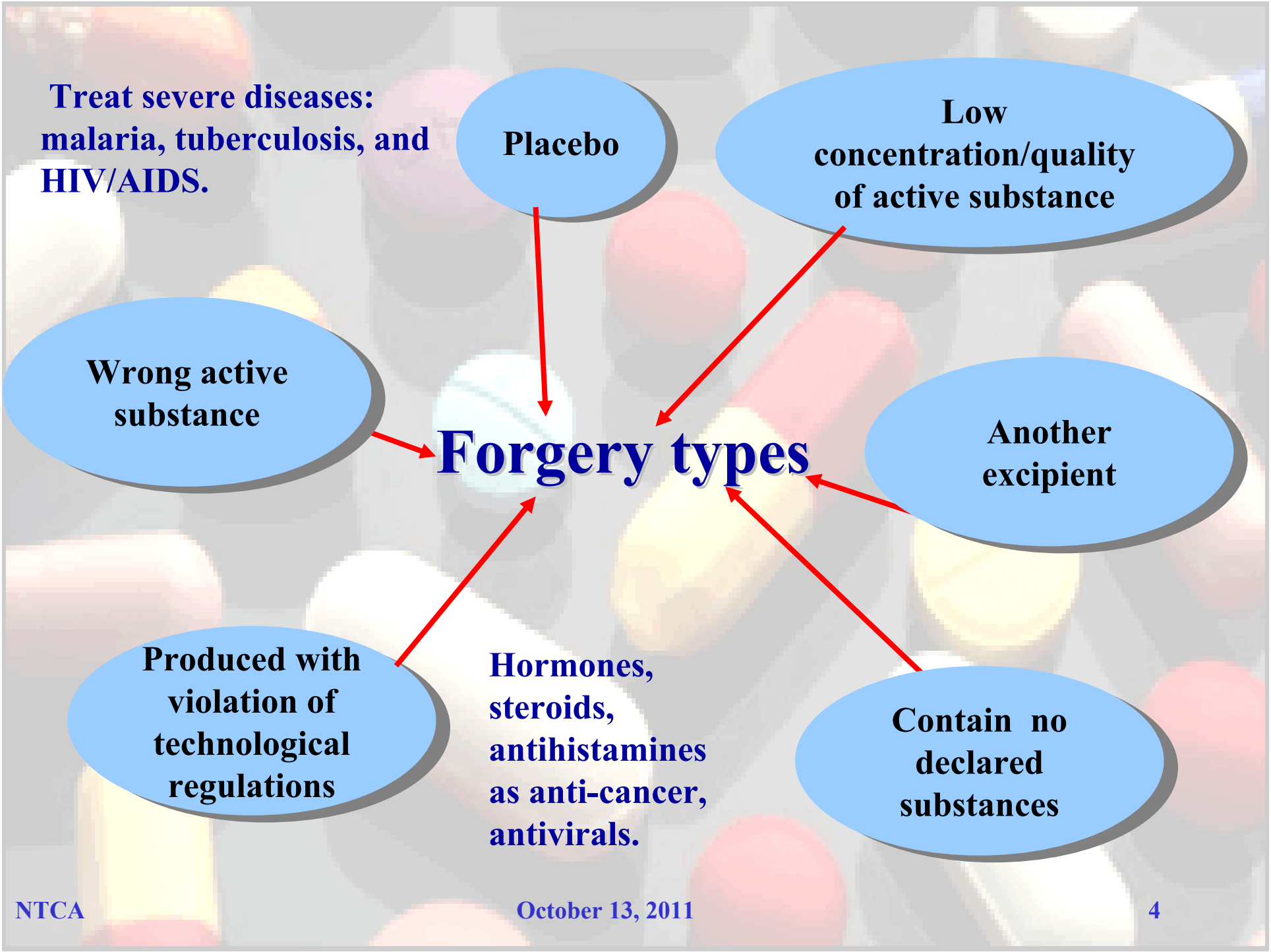
Outline

- What is counterfeit drug?
- Forgeries of different 'quality'
- Main steps of NIR-based approach
- Common Problems
- Conclusions

What is counterfeit drug?

- A counterfeit medicine is one which is **deliberately and fraudulently mislabeled** with respect to identity and/or source.
- Counterfeiting can apply to both **branded and generic products**
- Counterfeit products may include products with the **correct** ingredients or with the **wrong** ingredients, **without** active ingredients, with **insufficient** active ingredient or with fake packaging

WHO Counterfeit Drugs: Guidelines for the Development of Measures to Combat Counterfeit Drugs, WHO, Geneva, 1999.


Treat severe diseases: malaria, tuberculosis, and HIV/AIDS.
Placebo
Low concentration/quality of active substance
Wrong active substance
Forgery types
Another excipient
Produced with violation of technological regulations
Hormones, steroids, antihistamines as anti-cancer, antivirals.
Contain no declared substances

Traditional methods

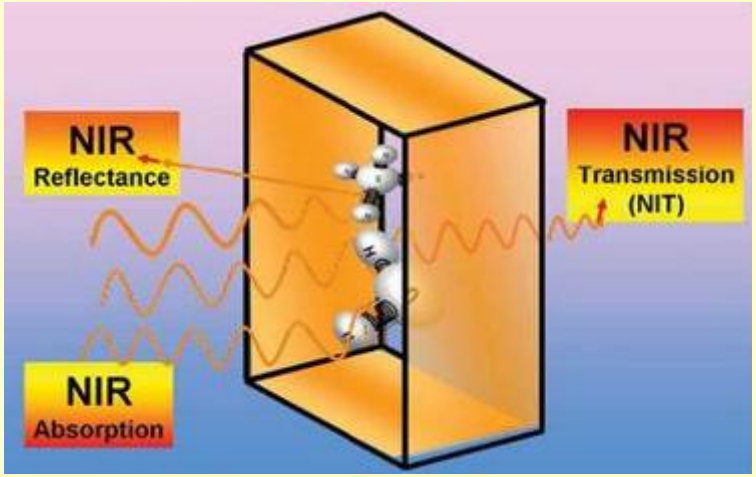
Methods described in pharmacopoeia

UV-Vis, mid-IR, GC, HLPC, melting points, viscosity, etc.

Rapid analysis

- simplified disintegration test
- simple qualitative reactions
- thin layer chromatography (TLC).

NIR spectroscopy (12500-4000 cm^{-1} or 800-2500 nm)



NIR spectra are much more complex than relatively easier for interpretation mid-IR spectra.

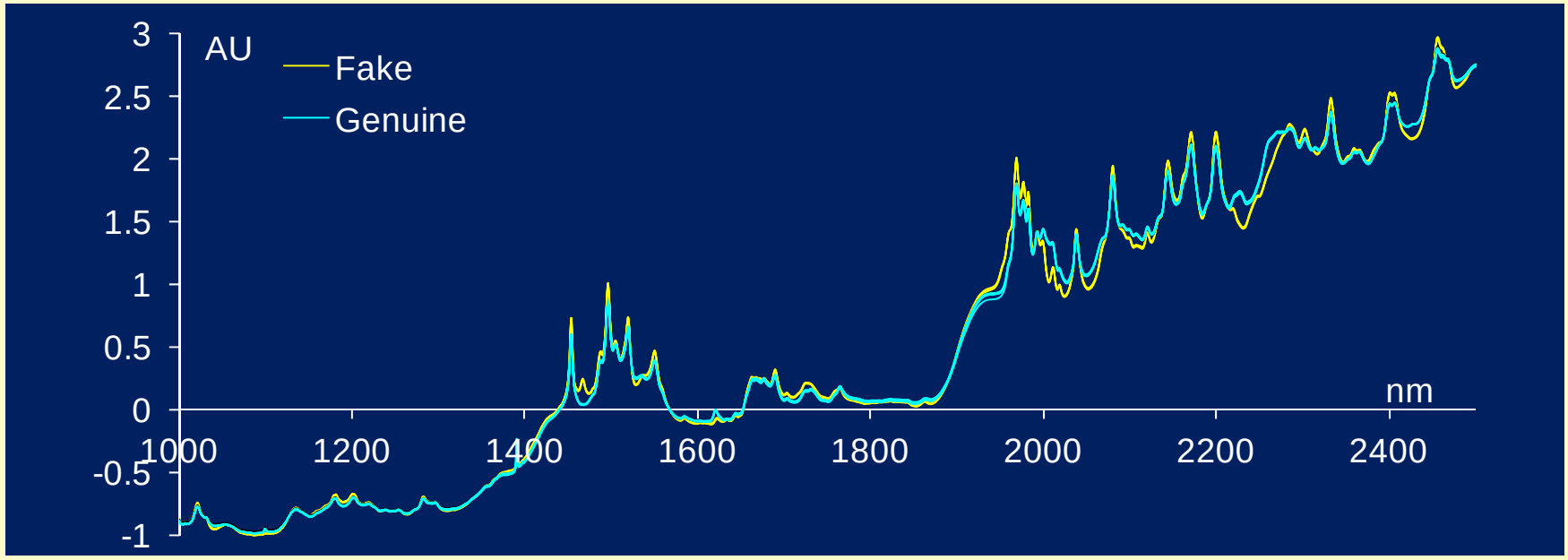
1. spectrum acquisition is fast compared to other analytical techniques

2. Minimal or no sample preparation

3. Carry information regarding not only chemical but also about physical phenomena

NIR spectra complexity ≠ Low-informative

NIR spectroscopy + **Multivariate data analysis (chemometric methods)** = **NIR based approach**



NIR Spectrometry is included in European Pharmacopoeia since 1997

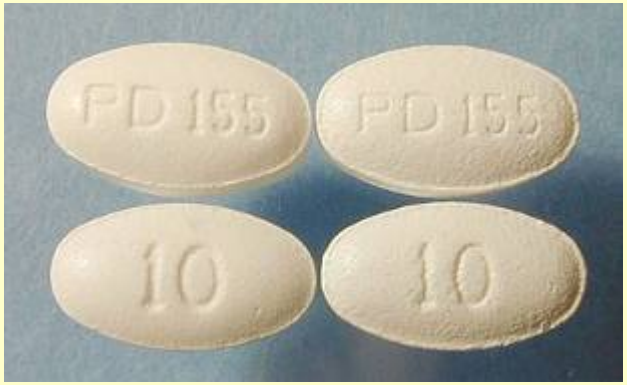
NIR-based approach

Measurements	NIR (800 – 2 500nm) reflectance spectra with integrating sphere
Data pre-processing	SNV/MSC + column-wise data centering
Variable reduction	PCA
Supervised pattern recognition	SIMCA+ special critical limits

Forgeries of different 'quality'

1. Easily detected without instruments

Dietary supplements



2. Easily detected by a regular pharmacopeia test as well as by the NIR-based approach

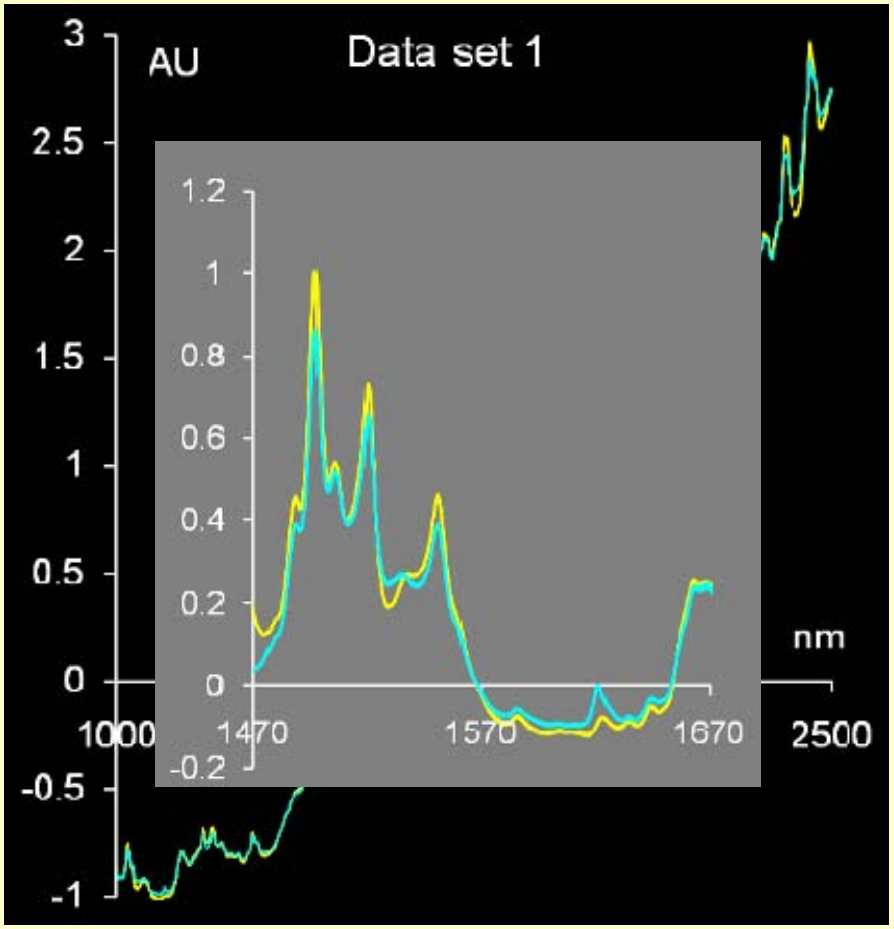
3. Easily detected by NIR-based approach but are not detected by the pharmacopeia tests

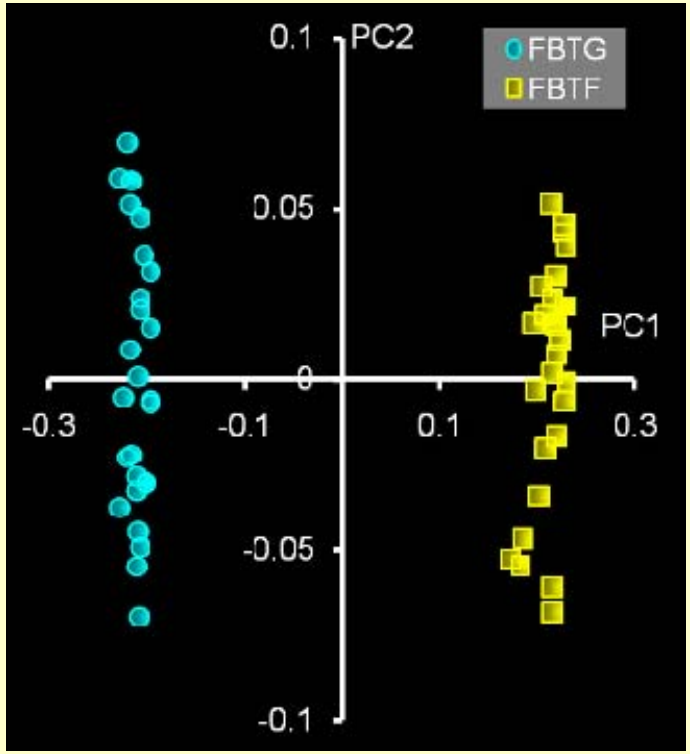
4. Various intricate cases for NIR approach and are not detected by the pharmacopeia tests

Data Set Overview

N	Description	Genuine batches	Forgery batches	Total
1	Complex antibacterial drug (2 active substances)	5 (5)	5 (5)	50
2	Antispasmodic drug	1 (10)	1 (9)	19
3	Antibiotic drug	17 (5-10)	2 (5)	109
4	Digestive enzyme (Manufacture 1)	4 (5)	1 (10)	30
5	Digestive enzyme (Manufacture 2)	11 (5)	4 (5)	75
6	Sildenafil	3 (5)	-	15

Complex antibacterial drug (2 active substances)

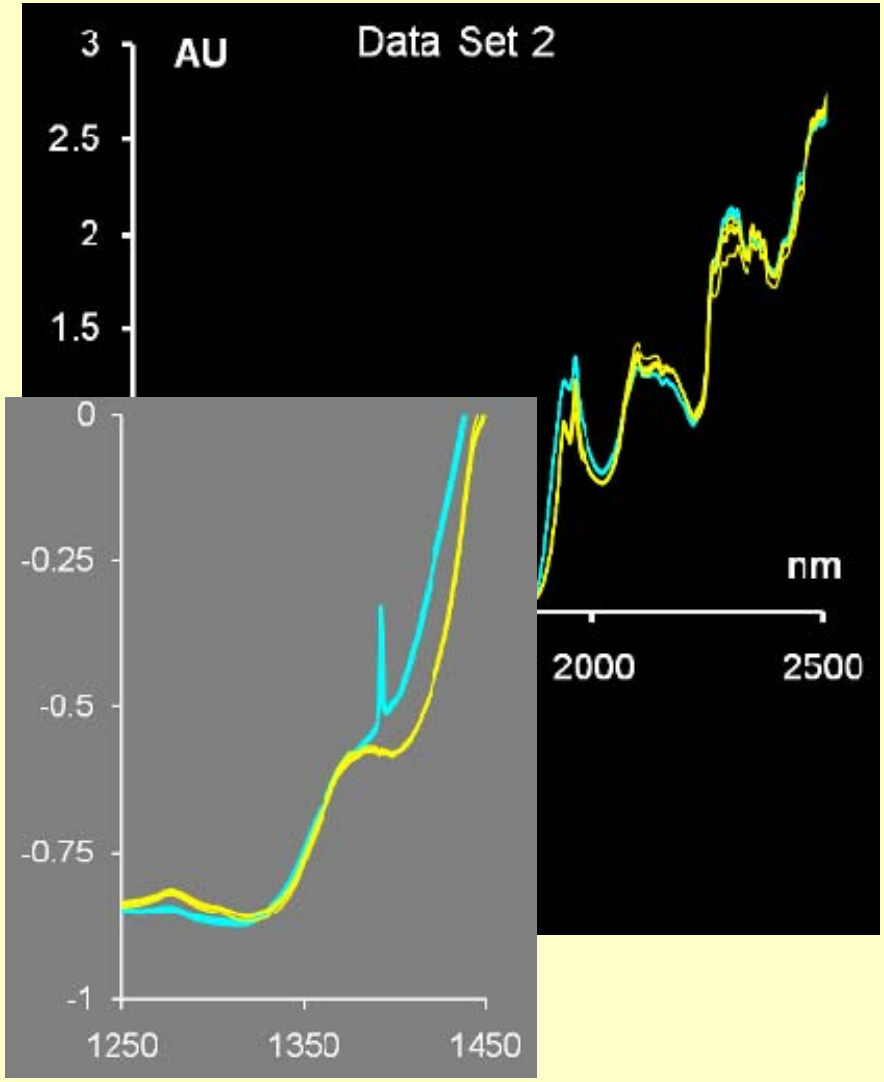


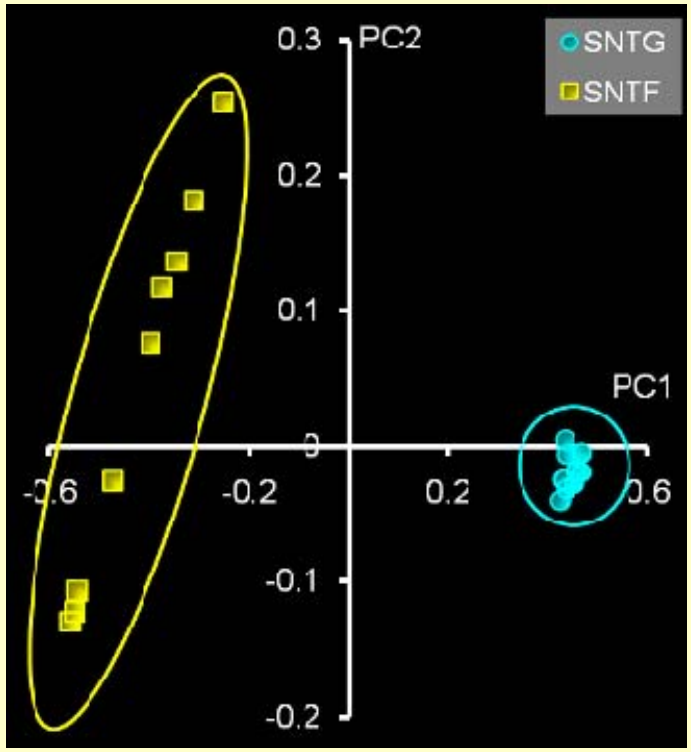


25 **original** tablets from 5 batches

25 **counterfeit** tablets from 5 batches

Antispasmodic drug

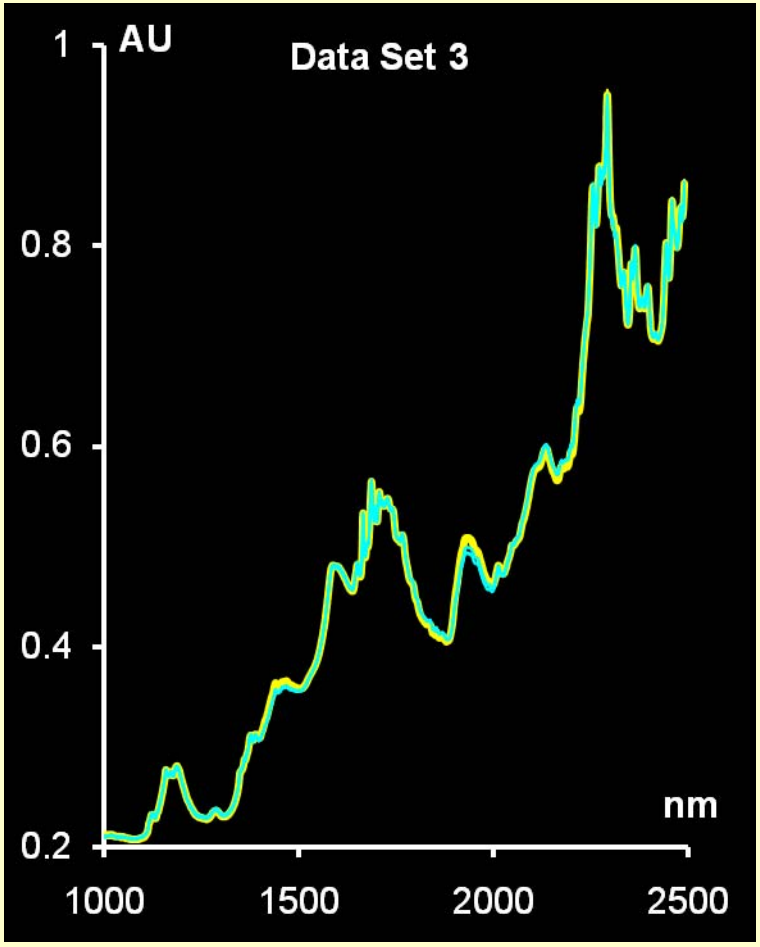


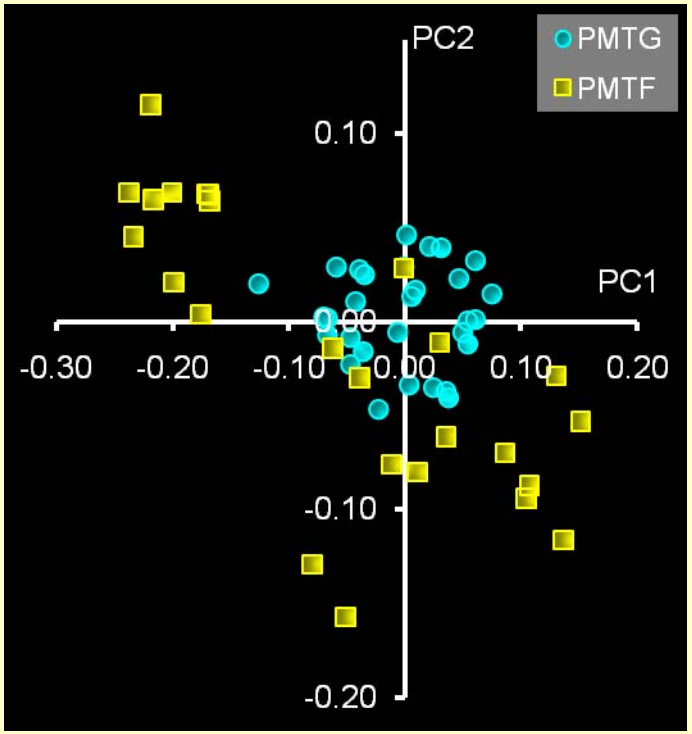


10 **original** pills

9 **counterfeit** pills

Antibiotic drug



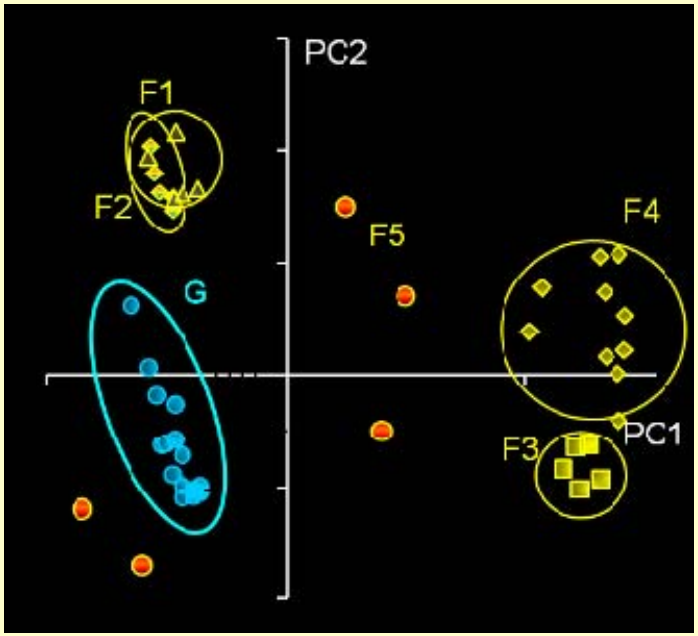


89 **original** tablets from 17 batches

20 **counterfeit** tablets from 2 batches

Main steps of NIR-based approach

PCA model for genuine samples



SIMCA classification with class limits for genuine samples

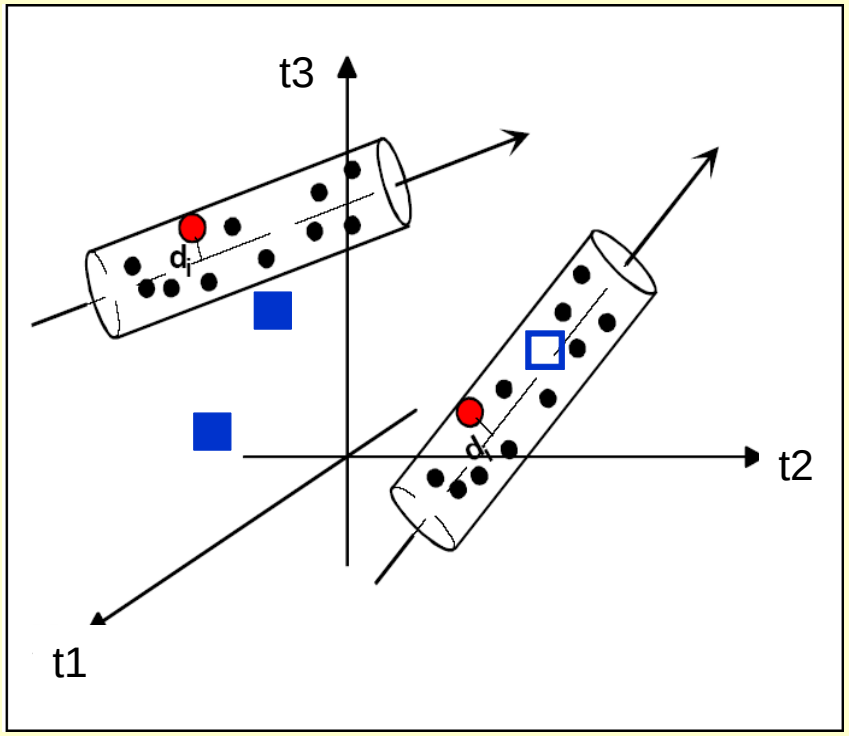
Representative calibration set

Pertinent external validation

Pre-processing

Proper variable selection

SIMCA (Soft Independent Modeling of Class Analogy)

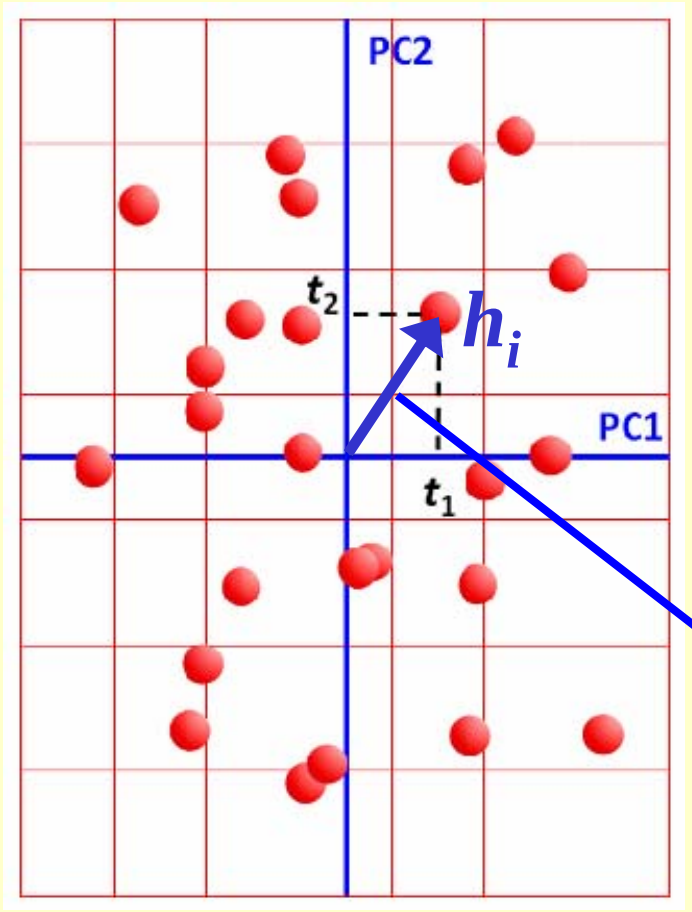


Independent PCA class - modeling

New object is compared with each class

S. Wold 1976

Score distance (SD), h_i

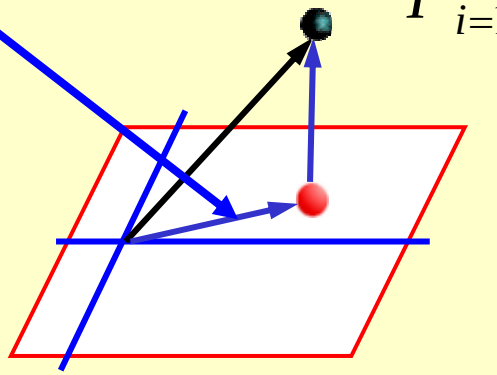


$$h_i = \mathbf{t}_i^{\mathrm{t}} (\mathbf{T}_A^{\mathrm{t}} \mathbf{T}_A)^{-1} \mathbf{t}_i = \sum_{a=1}^{A} \frac{t_{ia}^2}{\lambda_a}, \quad i = 1, \ldots, I$$

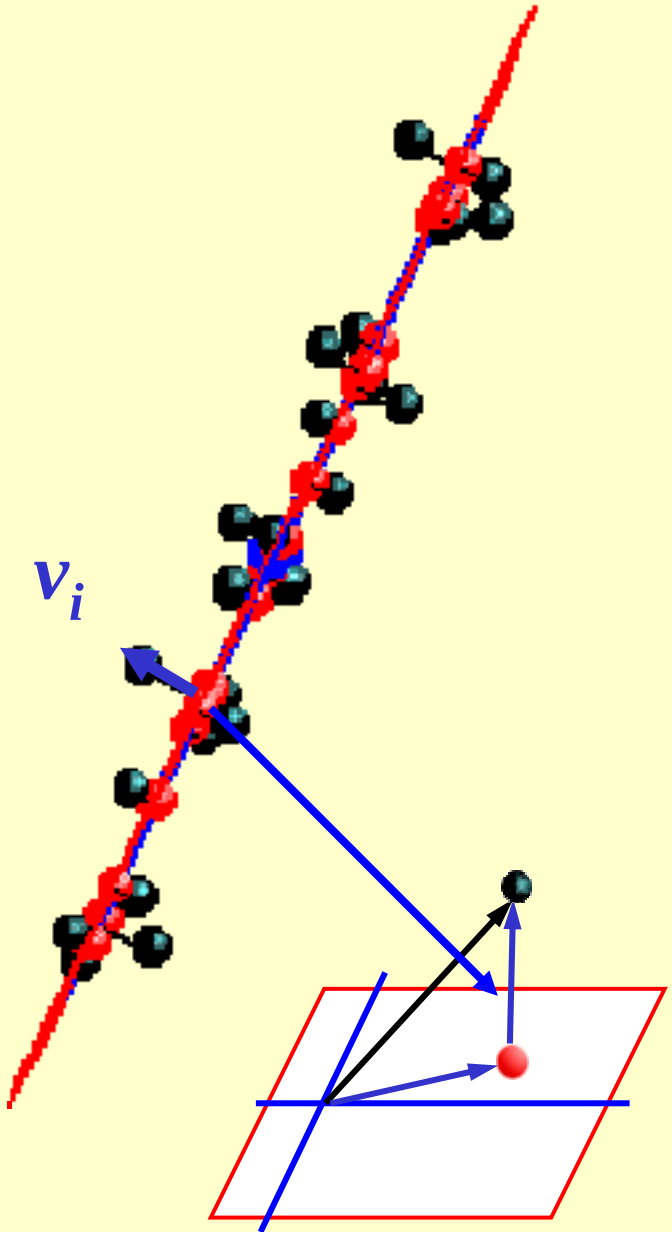
$$\textit{Leverage} = h_i + 1/I$$

$$\textit{Mahalanobis} = (h_i)^{1/2}$$

$$h_0 = \frac{1}{I} \sum_{i=1}^{I} h_i \equiv \frac{A}{I}$$



Orthogonal distance (OD), v_i



$$v_i = \sum_{j=1}^{J} e_{ij}^2 = \sum_{a=A+1}^{K} t_{ia}^2 = L_0 - \sum_{a=1}^{A} t_{ia}^2$$

Variance per sample $= v_i / J$

Q statistics $= v_i$

$$v_0 = \frac{1}{I}\sum_{i=1}^{I} v_i \equiv \frac{L_0}{I}\left(1 - R(A)\right)$$

Distribution of distances: DoF?

$$x = \begin{cases} = h/h_0 \\ = v/v_0 \end{cases} \qquad x_1, \ldots, x_I \sim \chi^2(N)/N \quad \Rightarrow \quad N = ?$$

Method of Moments

$$S^2 = \frac{1}{I}\sum_{i=1}^{I}(x_i - 1)^2$$

$$\hat{N} = \frac{2}{S^2}$$

Interquartile Approach

¼ ***IQR*** ¼

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(I\text{-}1)} \le x_{(I)}$$

$$\frac{\chi^{-2}(N, 3/4) - \chi^{-2}(N, 1/4)}{N} = IQR$$

Acceptance areas

Calculated by the PCA decomposition

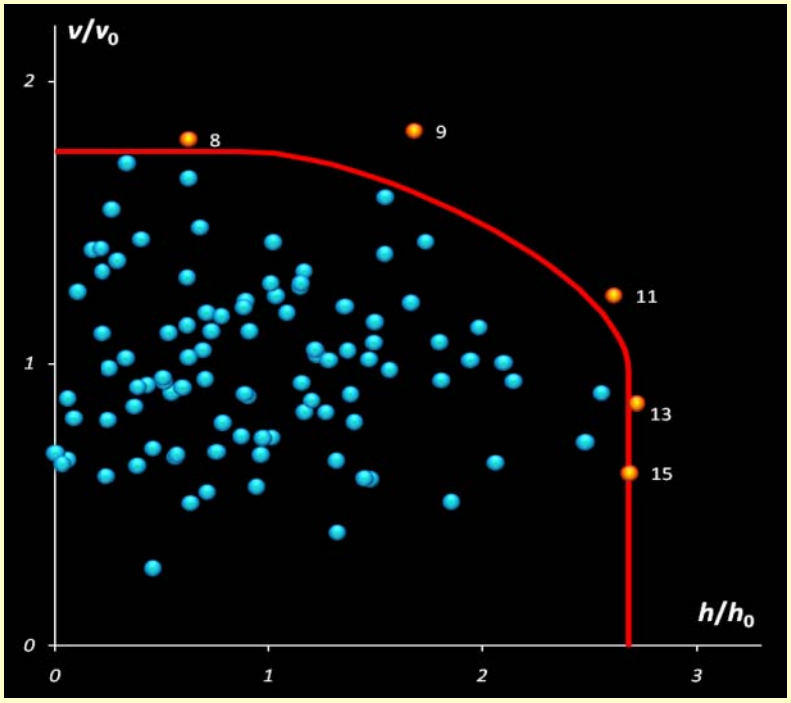
$$v/v_0 \sim \chi^2(N_v)/N_v$$
$$h/h_0 \sim \chi^2(N_h)/N_h$$

Estimated DoF

N_v, N_h

Set by a researcher

Type I Error = α



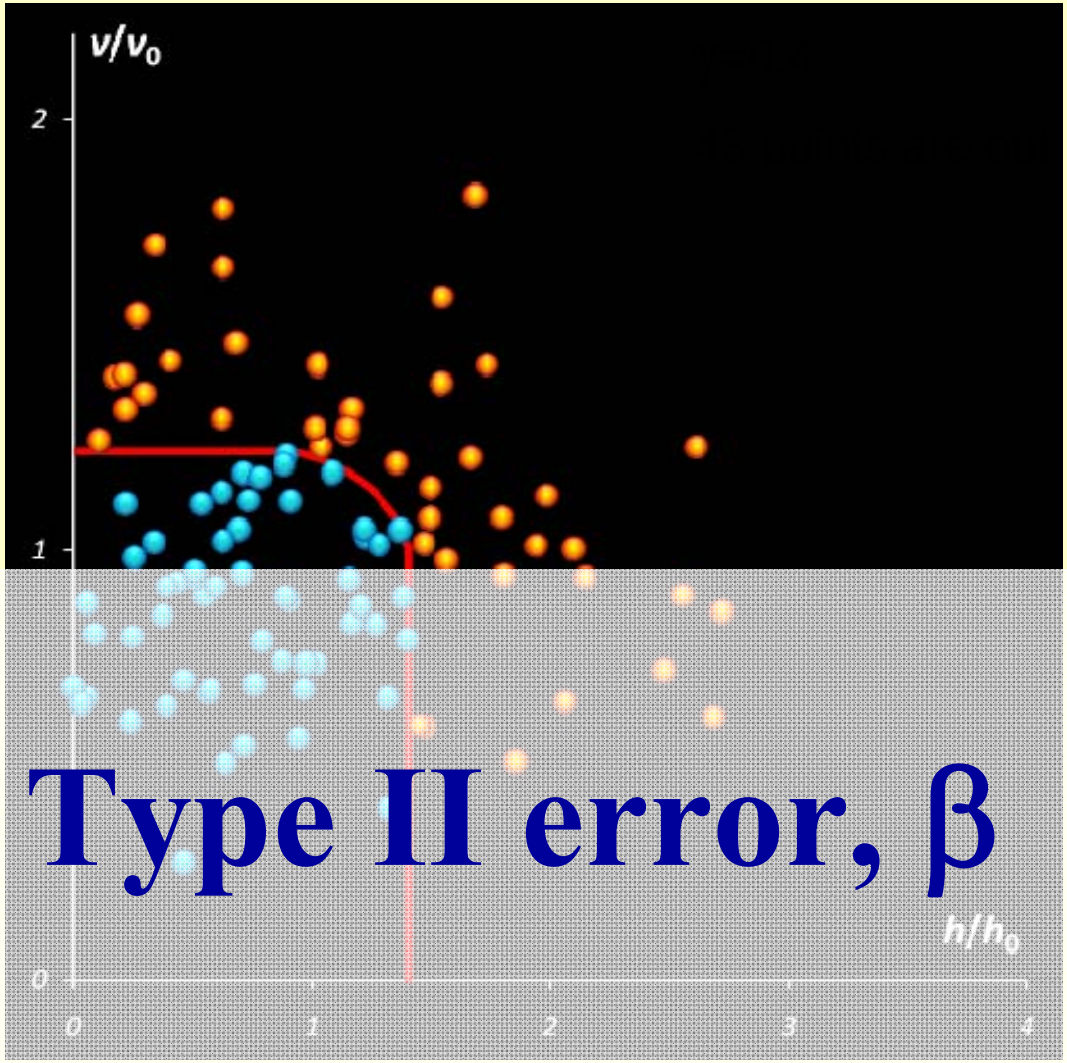
J. Chemometrics 2008; 22;
A. Pomerantsev

Acceptance areas for multivariate classification derived by projection methods

Type I error α. *I*=100

α=0.4

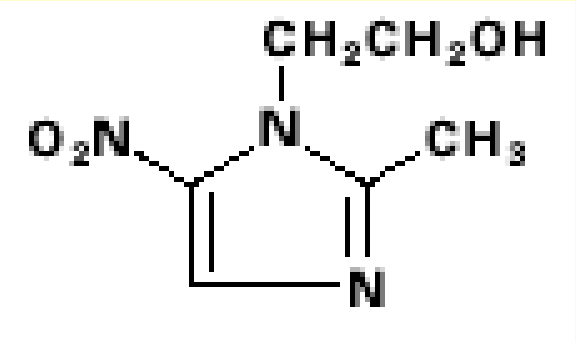
OUT
43 object

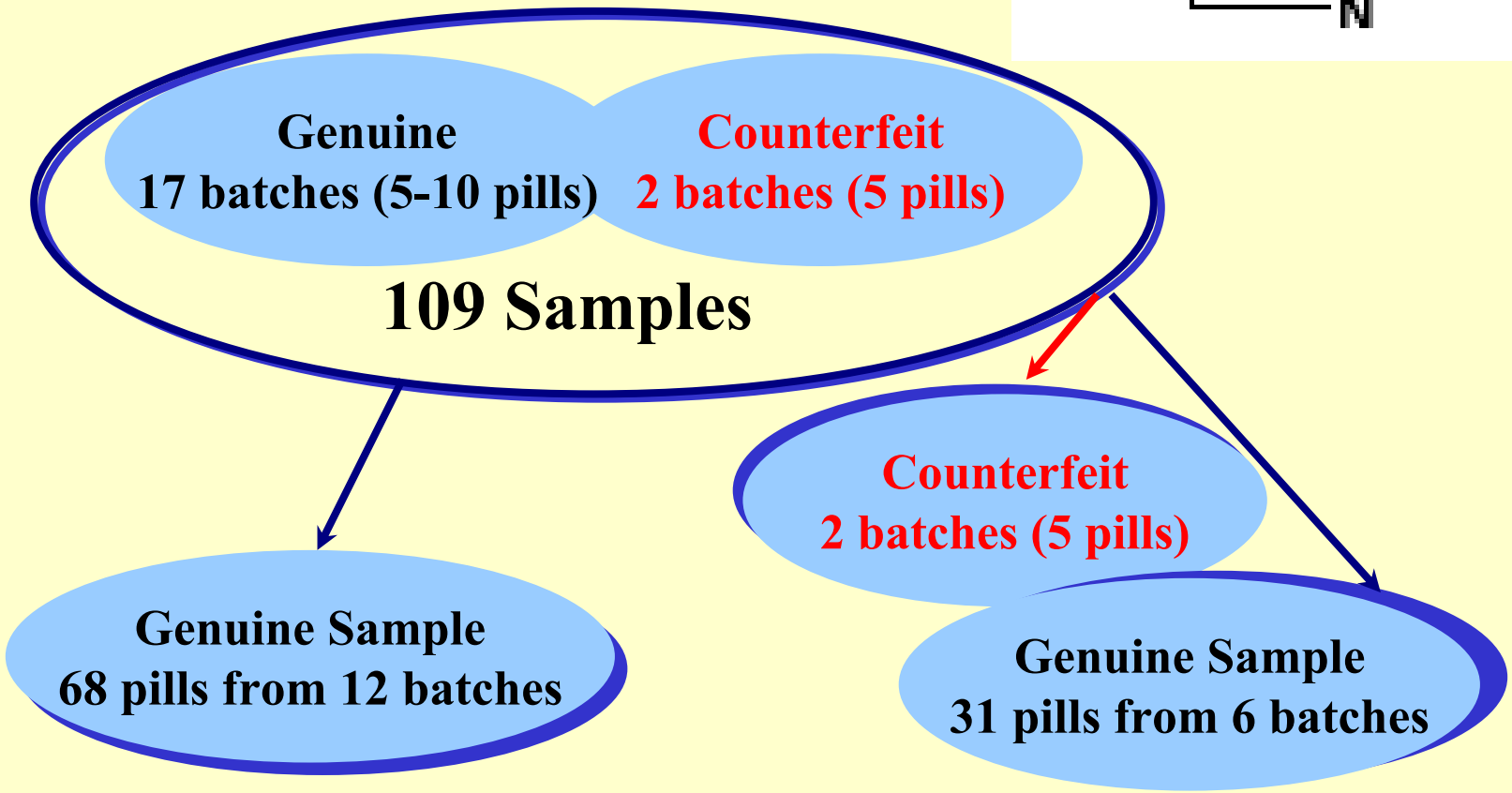


Type II error, β ?

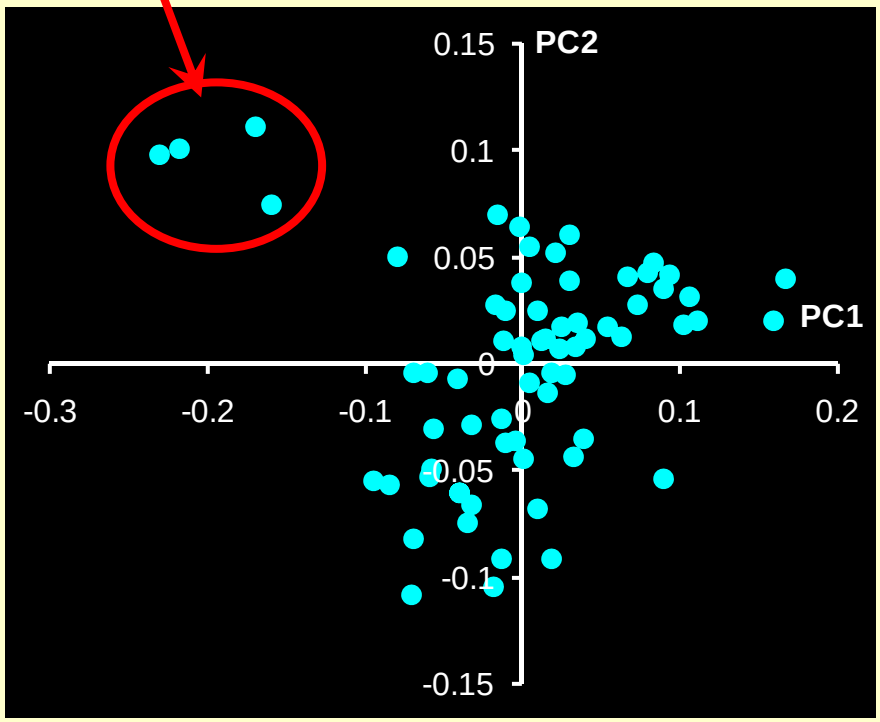
Step-by-step classification

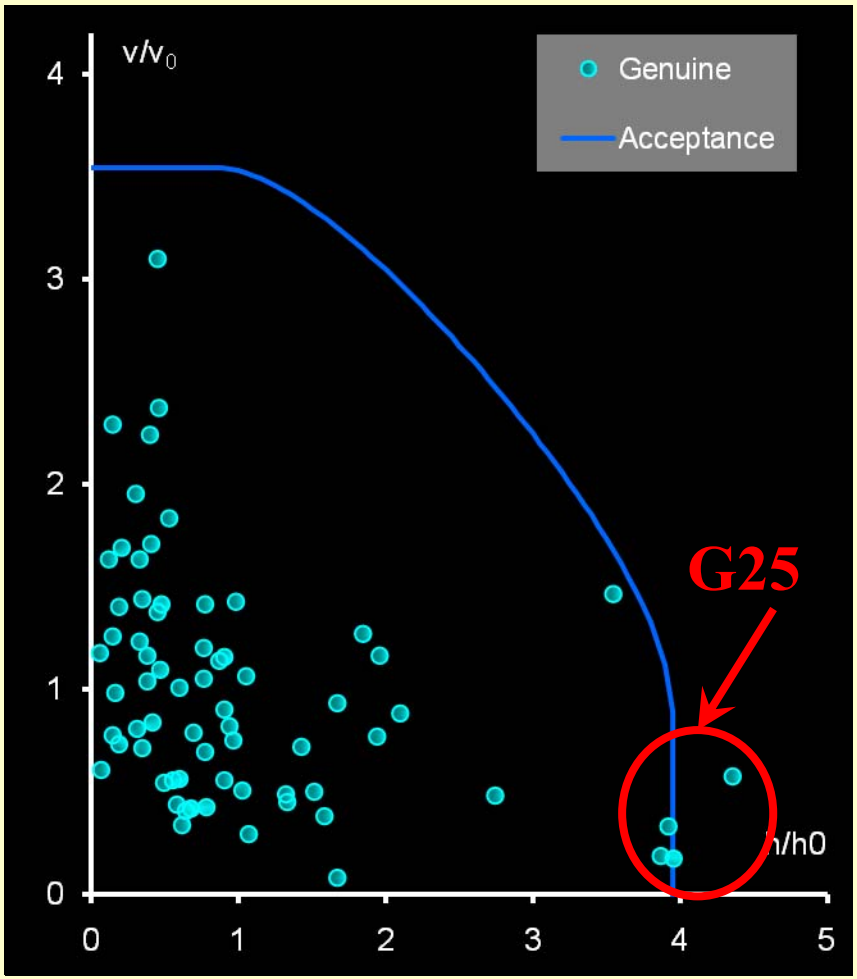
Synthetic Antiprotozoal and Antibacterial Agent,
1-(β-hydroxy-ethyl)-2-methyl-5-nitroimidazole,

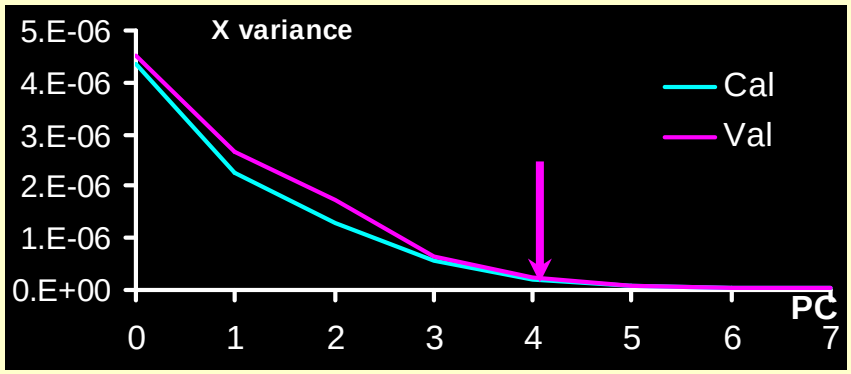




Initial PCA model



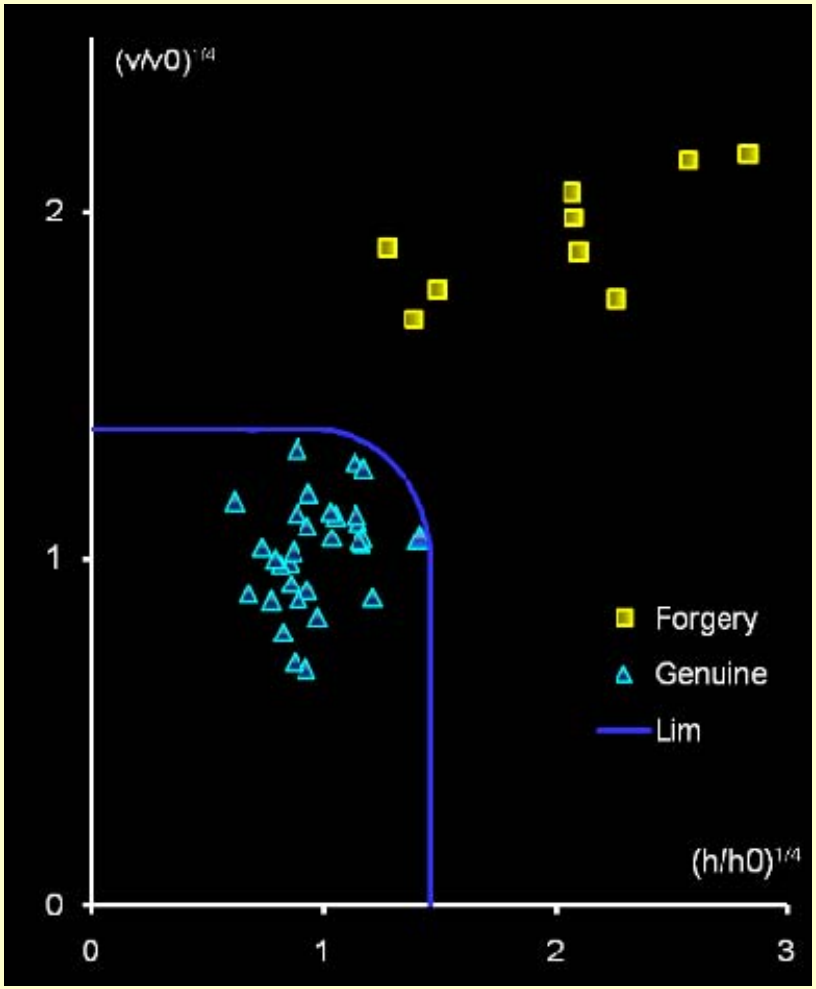




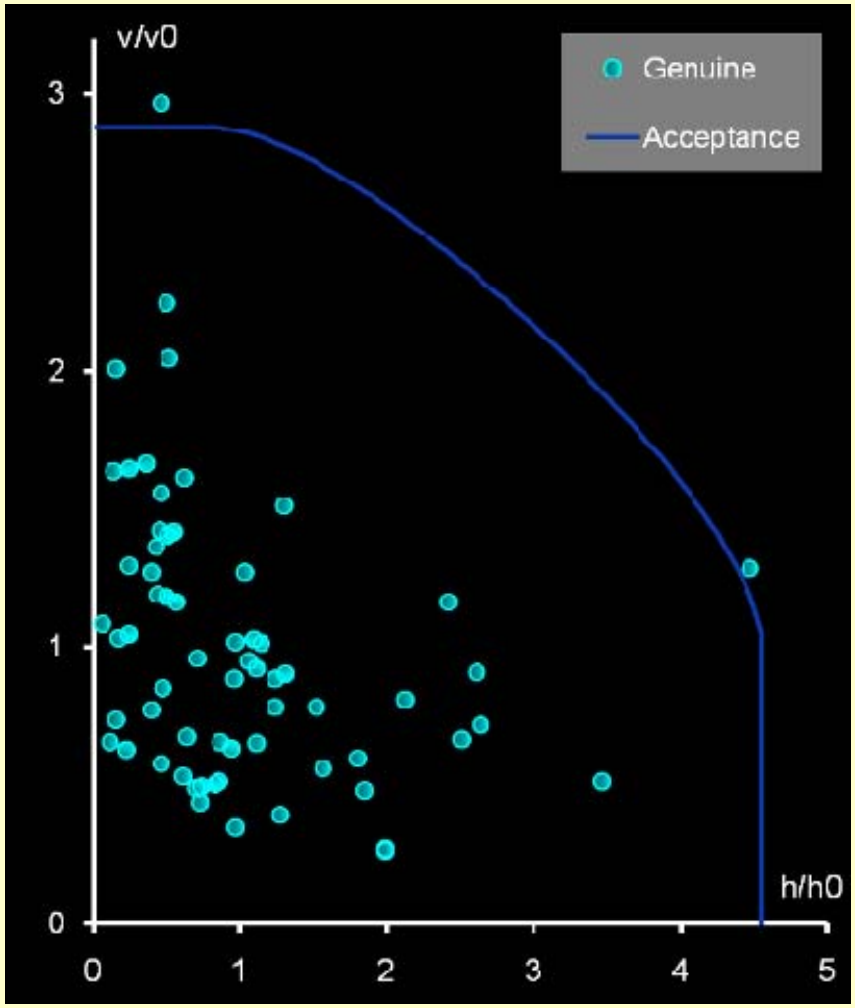
N=68, A=4

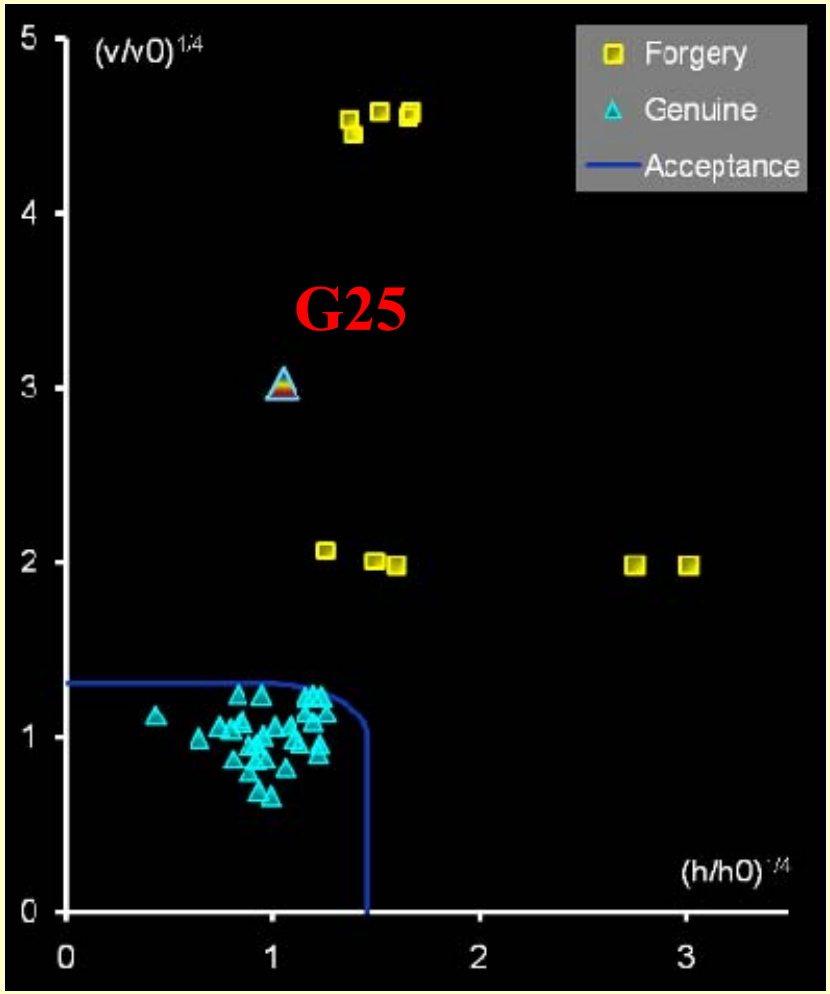
γ=0.01

Classification



Classification without batch G25



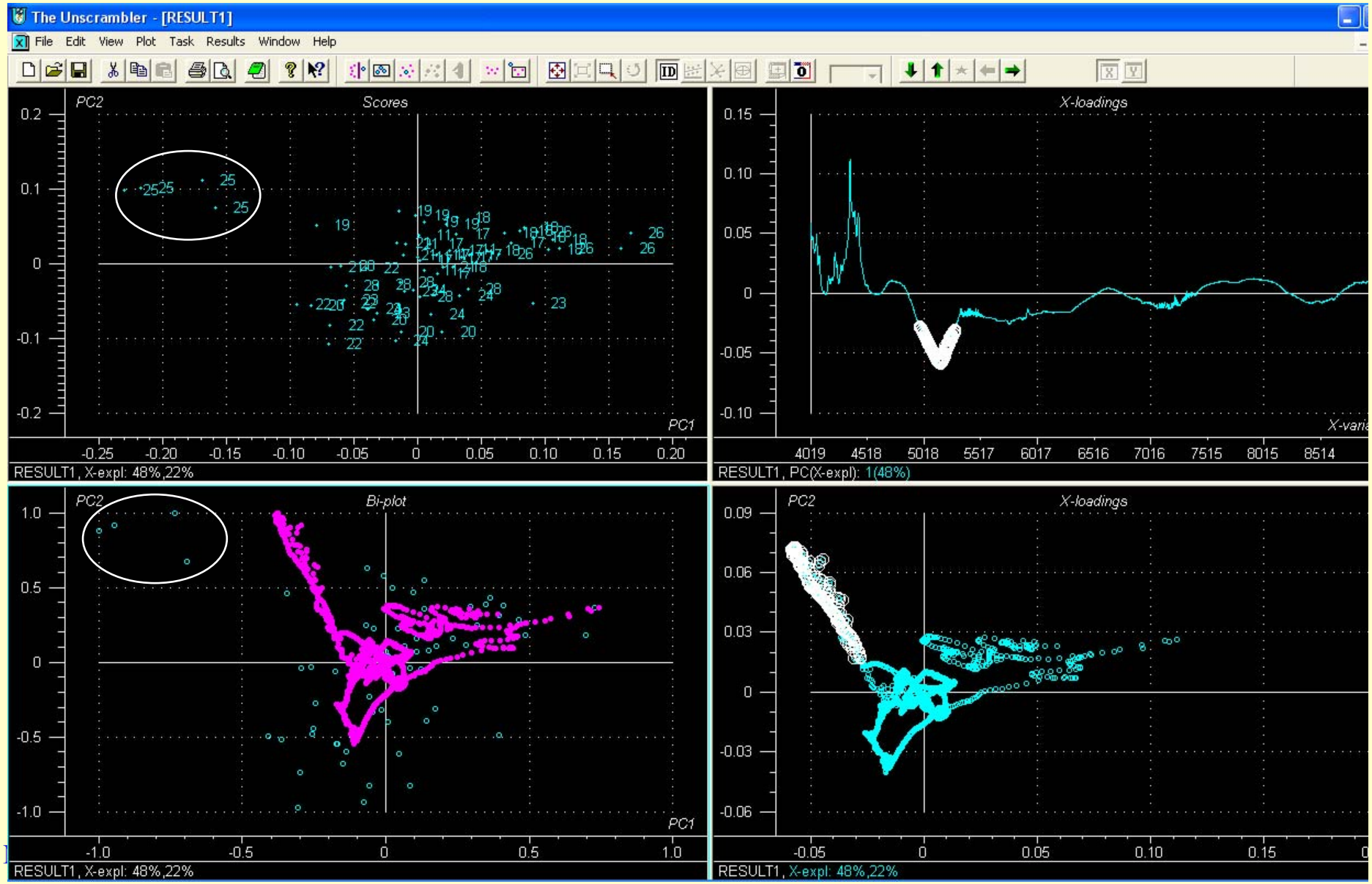


Calibration

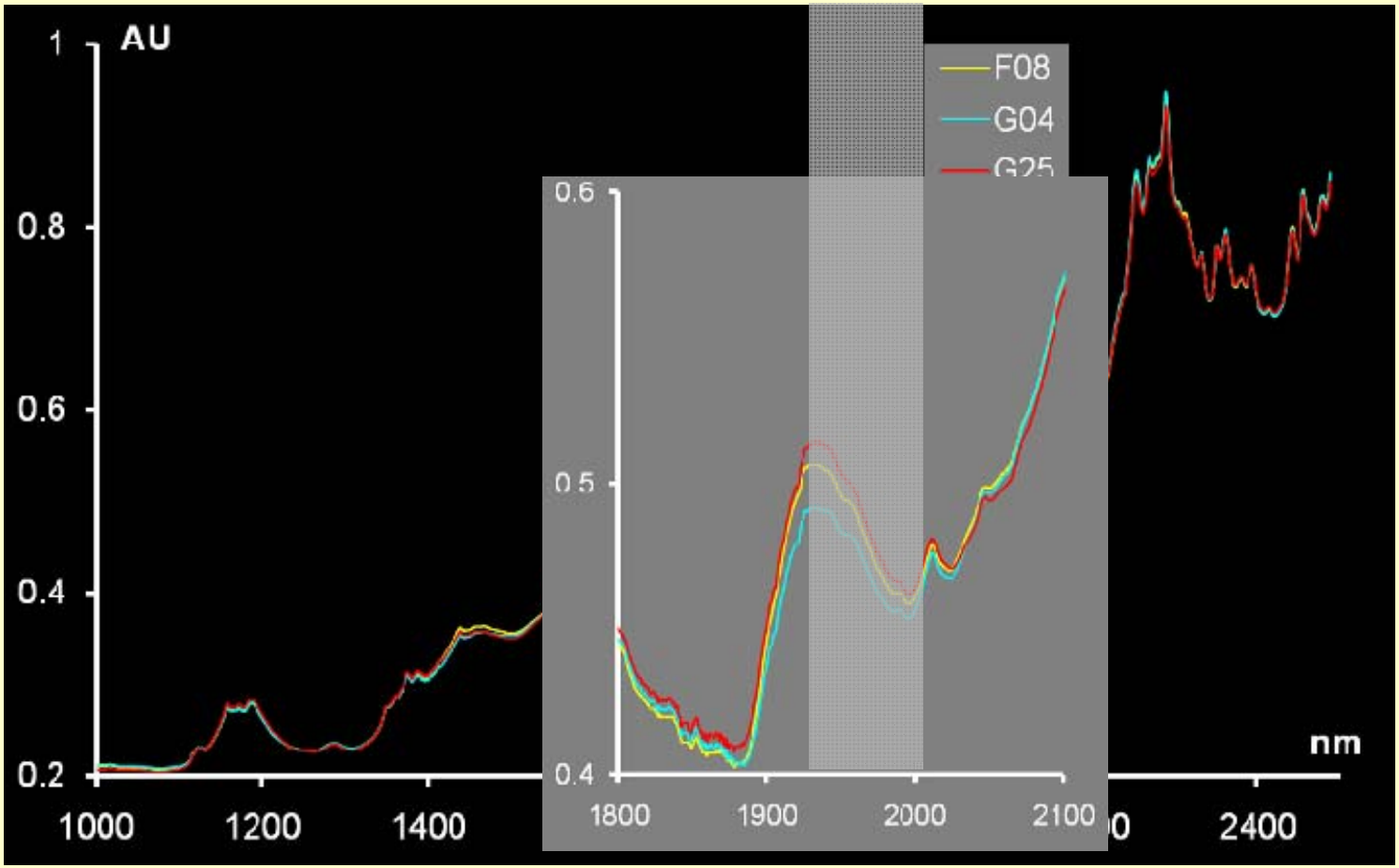
N=64, A=3

Test

Peculiarity of G25



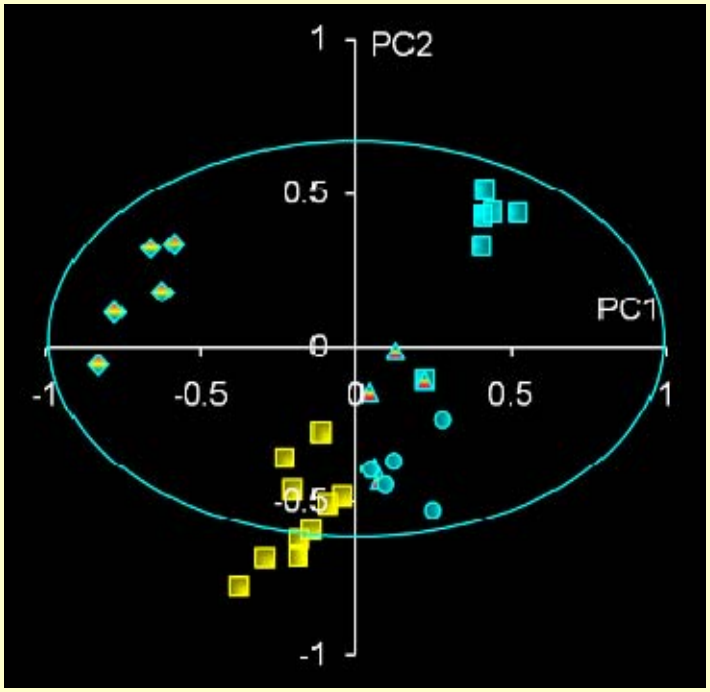
NIR spectra for G25



Common Problems

Variability of genuine drugs
Digestive enzyme

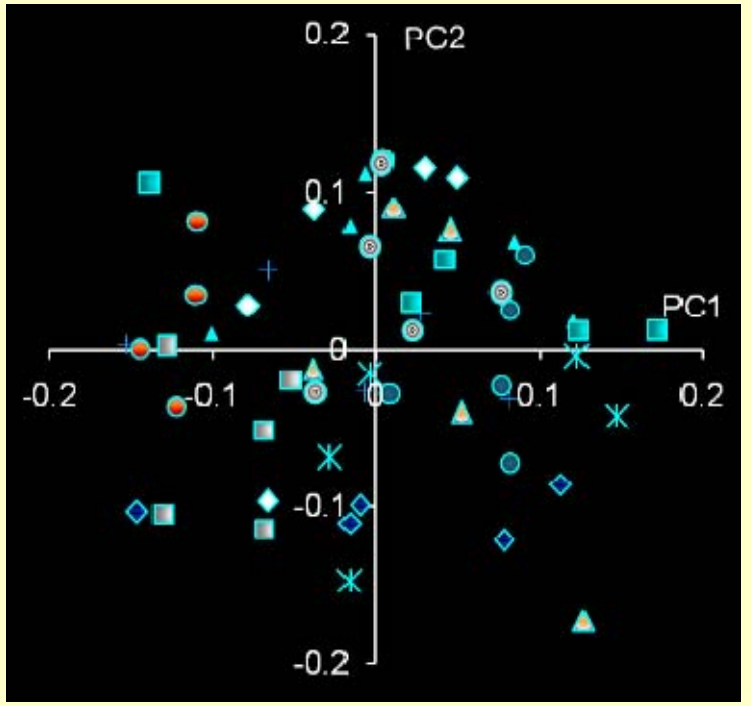
Manufacture N1
(Dataset 4)



20 **original** tablets from 4 batches

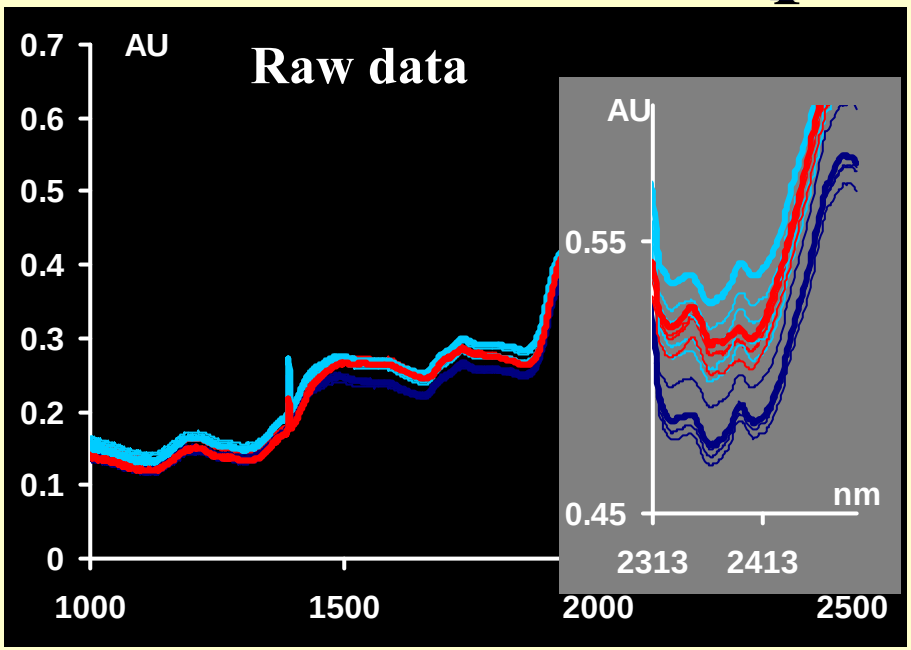
10 **counterfeit** tablets from 1 batch

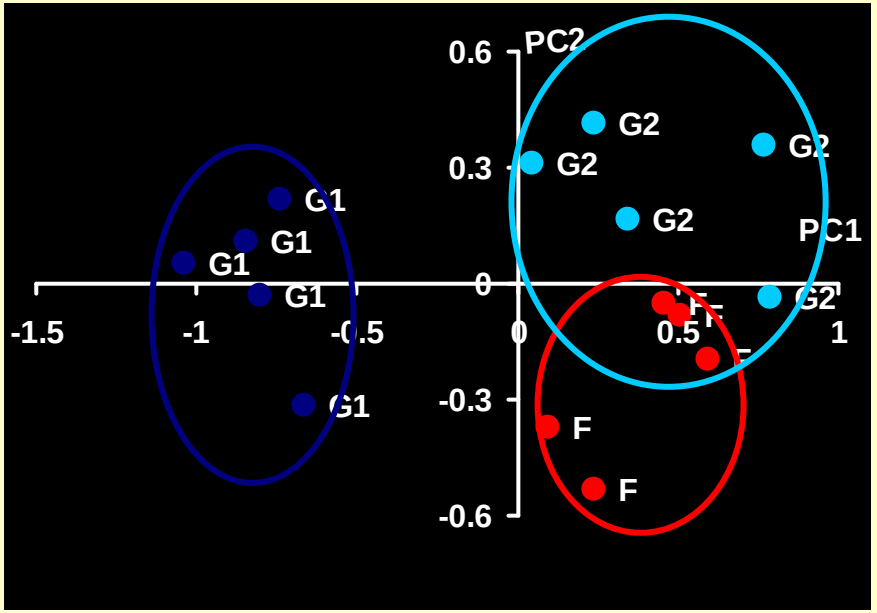
Manufacture N2
(Dataset 5)

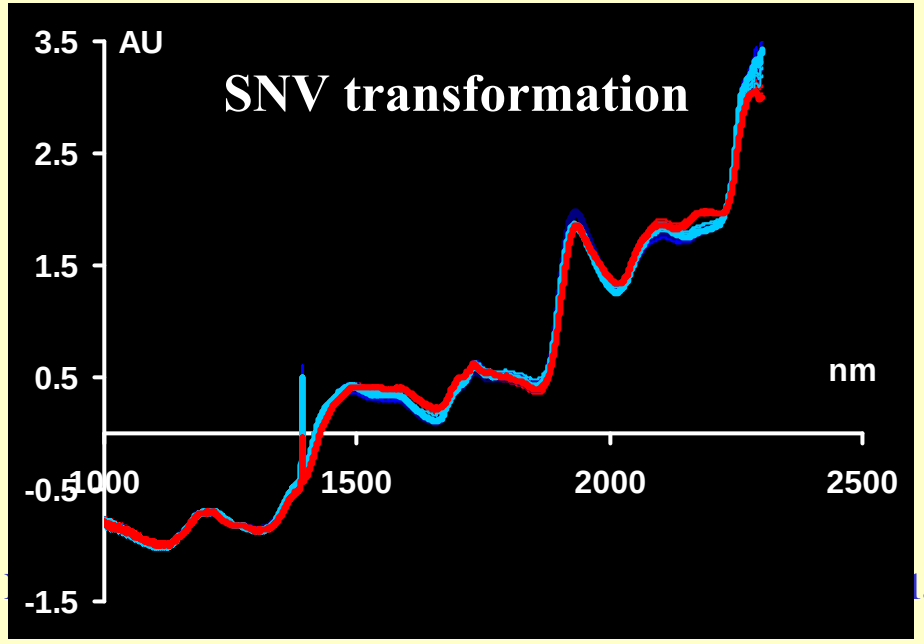


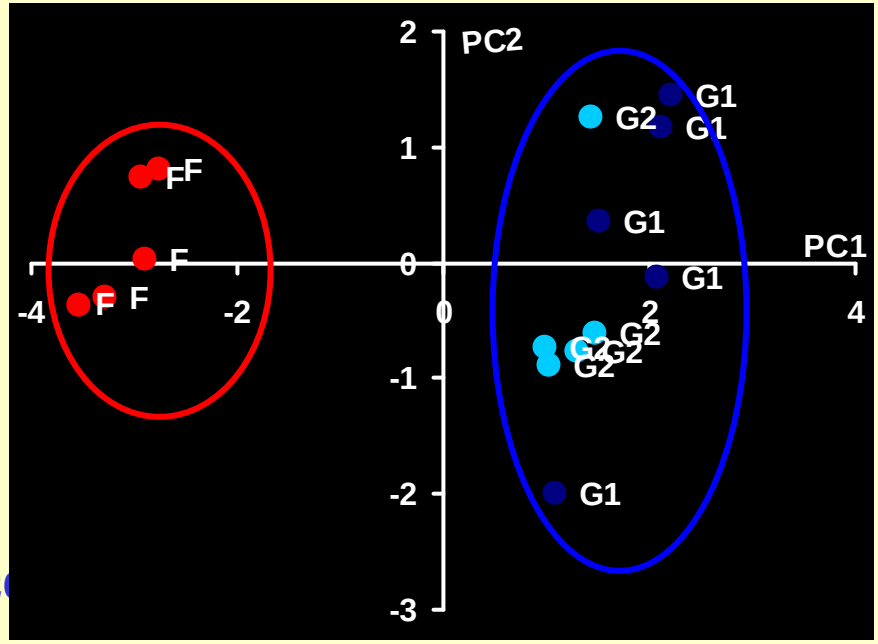
55 **original** tablets from 11 batches

Pre-processing

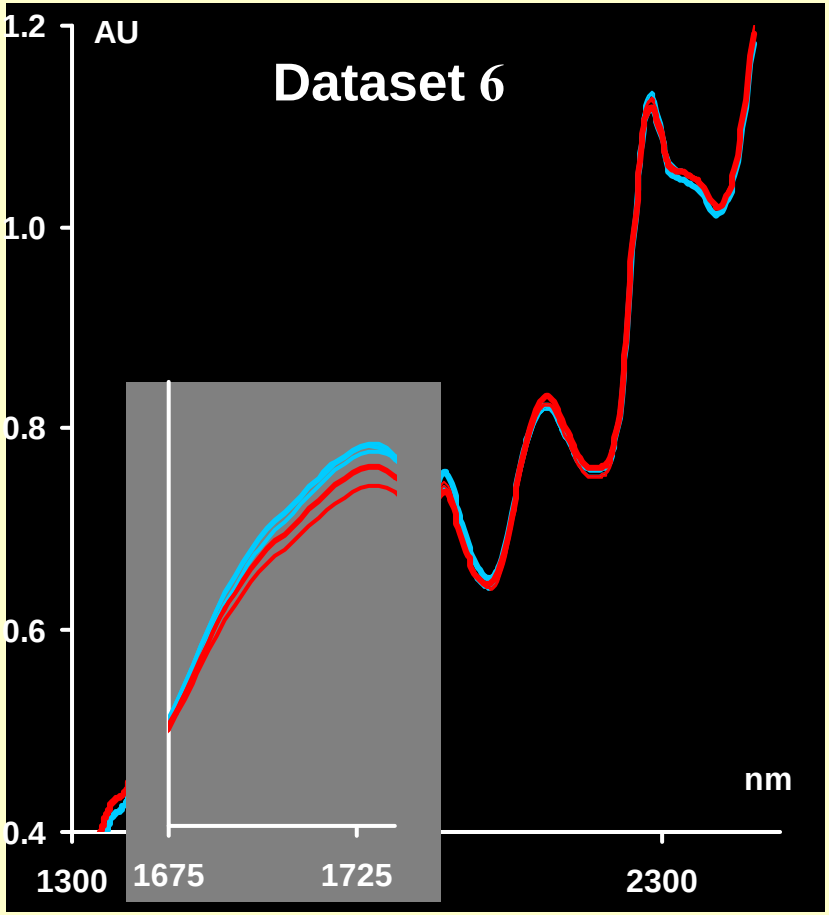








Influence of spectral region



Film-coated tablets of Sildenafil

15 **original** tablets from 3 batches

Conclusions

Variability in the genuine drug production should be fully investigated. Batch-to-batch variability should be studed.

The NIR spectra should be preprocessed before chemometric analysis.

The selection of a spectral region should be done for each type of medicine individually. The choice of the spectral region may essentially influence the final classification results.

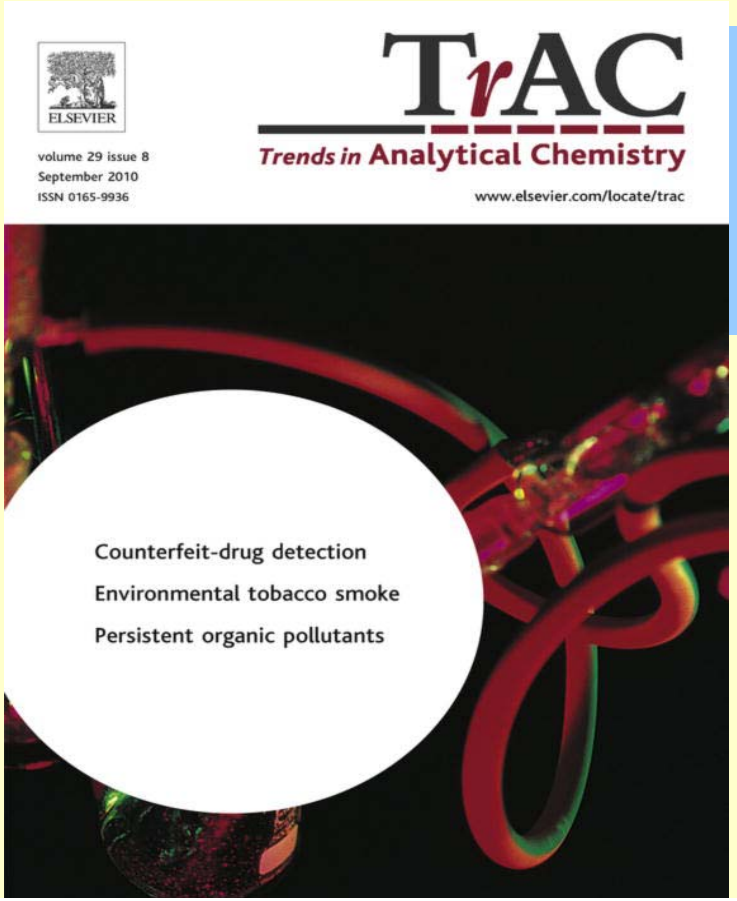
Conclusions

The model construction requires representative sample distribution between the calibration and the test sets

It is crucial not only to recognize forgeries but also to avoid misclassification of genuine samples. The application of reliable acceptance limits is of great importance

Methods based only on quantitative determination of API are insufficient. It is necessary to investigate a remedy as a whole object

Thank you for attention!



NIR based approach to counterfeit-drug detection

O.Rodionova, A. Pomerantsev

29 (8), 781-938 (2010)

Raw spectra for two datasets are located at
http://rcs.chph.ras.ru/data/