

Multivariate data-modelling to separate light scattering from light absorbance: The Optimized Extended Multiplicative Signal Correction OEMSC

The importance of spectral preprocessing in data analysis for aquaphotonics

Harald Martens

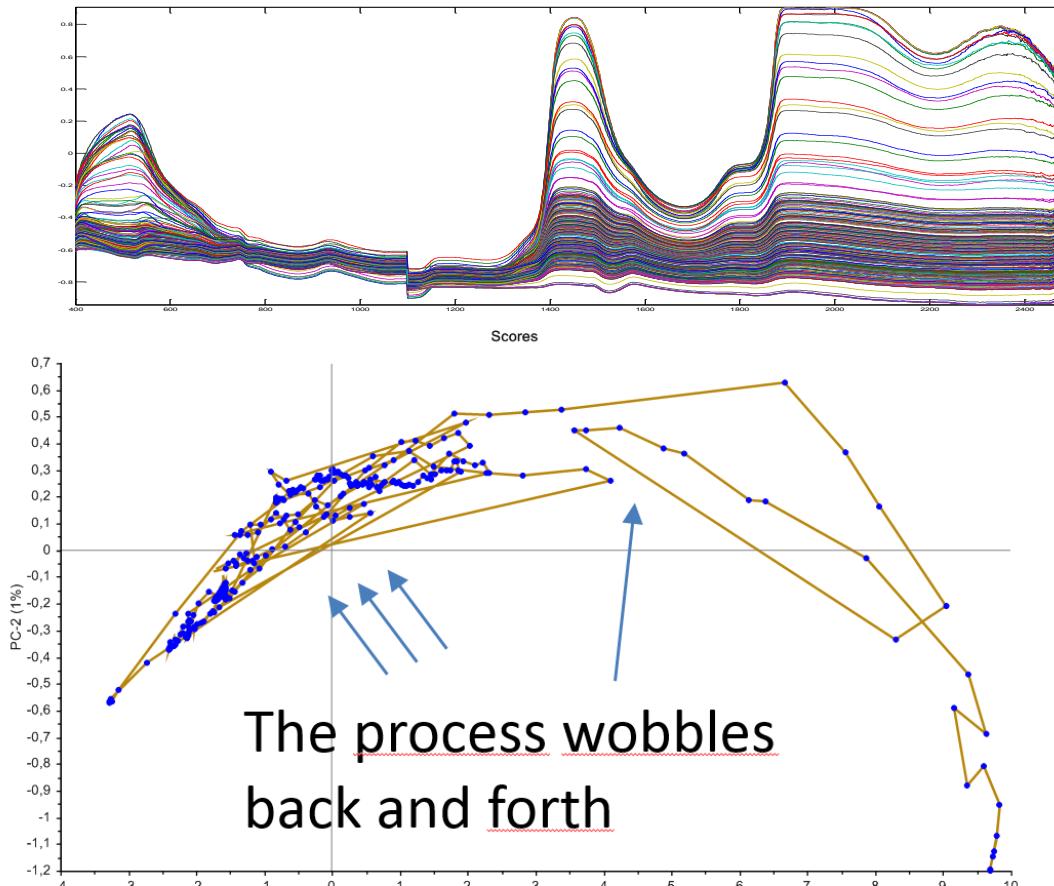
Dept. Engineering Cybernetics, Norwegian U. of Science and Technology,
Trondheim, Norway. Email: harald.martens@ntnu.no

When studying complex systems, vis/NIR spectroscopy is informative.

Example: Belusov-Zhabotinsky reaction.
Anna Zhyrova, Dalibor Stys, U. South Bohemia:



Many causes of variation: Light scattering, path length, nonlinear responses, organic absorbers, temperature effects, water concentration, water changes,...

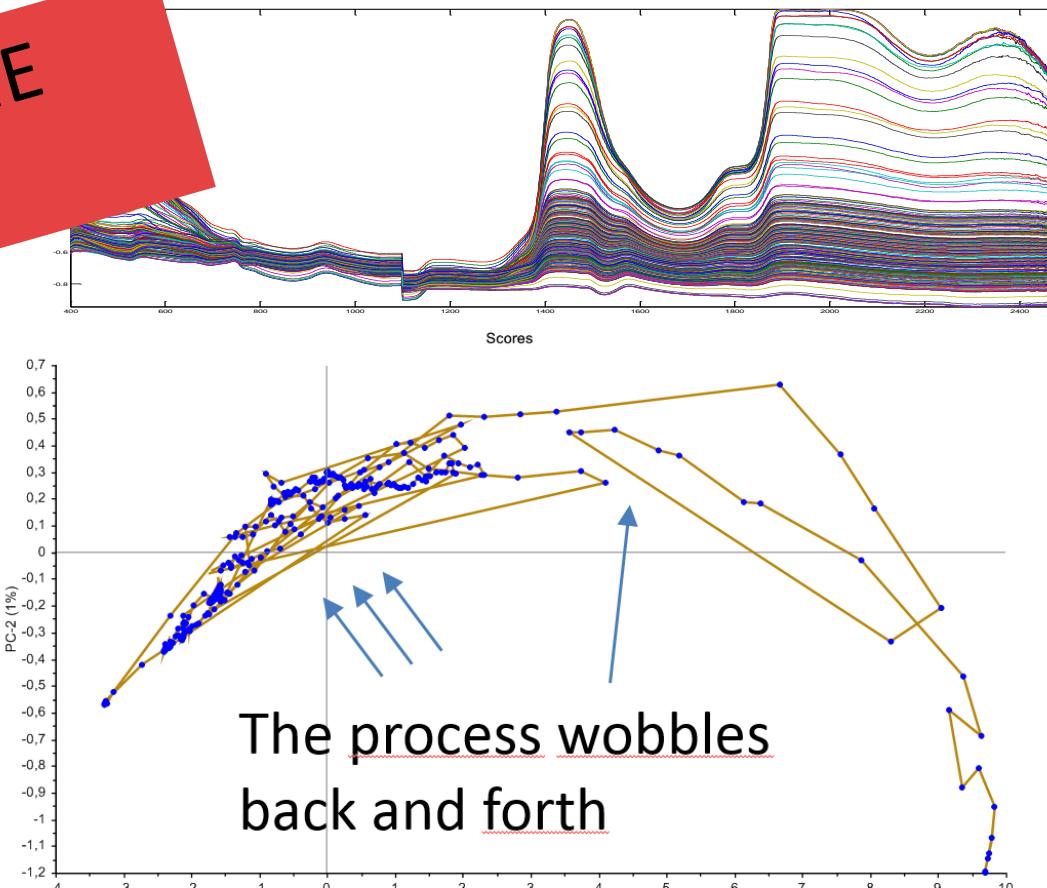


When studying complex systems, vis/NIR spectroscopy is informative.

Example: Belusov-Zhabotinsky reaction.
Anna Zhyrova, Dalibor Stys, U. South Bohemia:



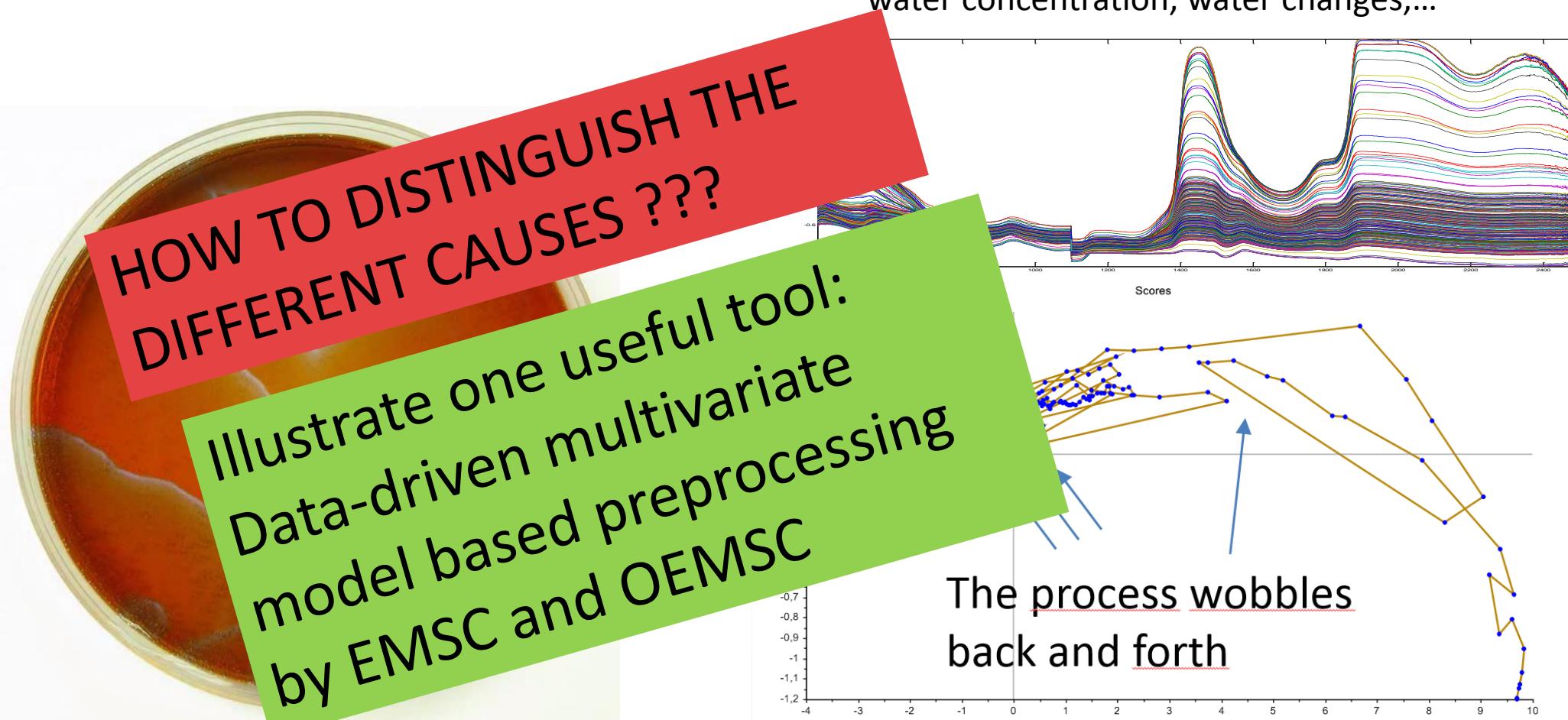
Many causes of variation: Light scattering, path length, nonlinear responses, organic absorbers, temperature effects, water concentration, water changes,...



When studying complex systems, vis/NIR spectroscopy is informative.

Example: Belusov-Zhabotinsky reaction.
Anna Zhyrova, Dalibor Stys, U. South Bohemia:

Many causes of variation: Light scattering, path length, nonlinear responses, organic absorbers, temperature effects, water concentration, water changes,...



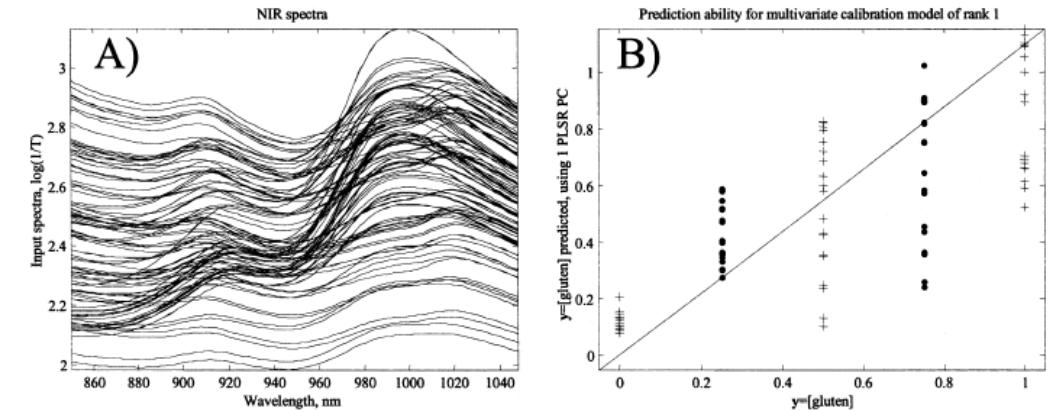
NIT of «difficult samples» :

Mixtures of wheat protein and wheat starch powders

(5 different ratios: 0,25,50,75,100% protein)

Different sample thickness, different sample packing

2 technical replicates



Go to Multivariate calibration?

Bad news

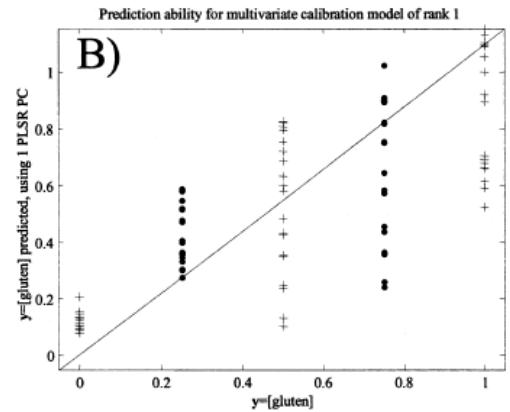
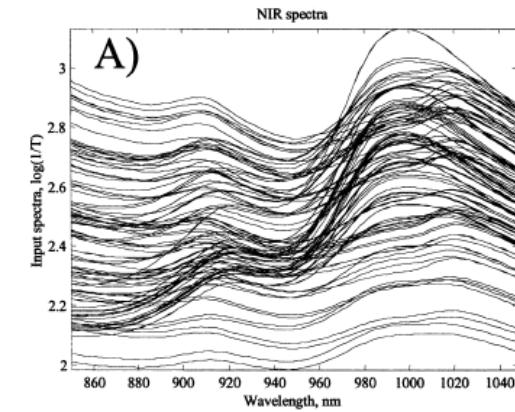
NIT of «difficult samples» :

Mixtures of wheat protein and wheat starch powders

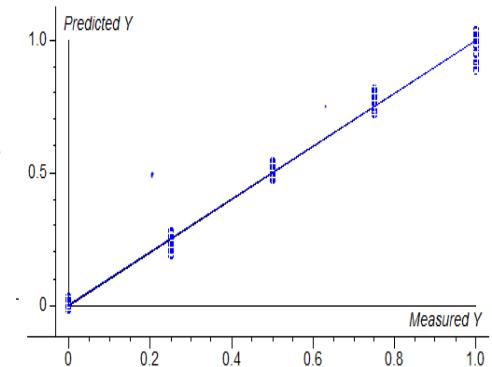
(5 different ratios: 0,25,50,75,100% protein)

Different sample thickness, different sample packing

2 technical replicates



PLS regression:
Imperfect. Needs 5 PCs.
Should need only 1 PC



Bad news

NIT of «difficult samples» :

Mixtures of wheat protein and wheat starch powders

(5 different ratios: 0,25,50,75,100% protein)

Different sample thickness, different sample packing

2 technical replicates

MSC:

Martens, H., Jensen, S.Å. and Geladi, P. (1983)

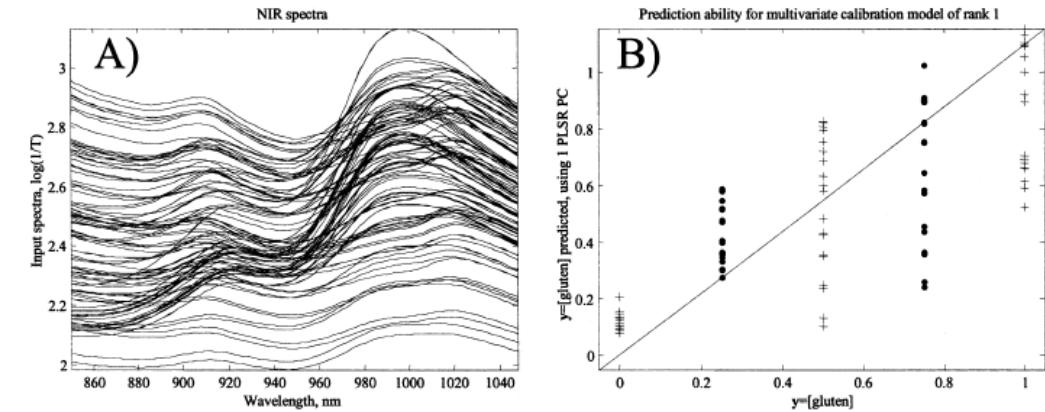
Multivariate linearity transformation for near-infrared reflectance spectrometry.

Proc. Nordic Symp. on Applied Statistics, (O.H.J. Christie, ed.) June 12.-14. 1983. Stokkand Forlag

Publ., Skagenkaien 12, N-4000 Stavanger, Norway, ISBN 82-90496-02-8, 208-234.

Geladi, P., MacDougall, D. and Martens, H. (1985): *Linearization and scatter-correction for near-infrared reflectance of meat.*

Applied Spectroscopy, 39, 3, 491-500.



The solution, of course: MSC or SNV!

NIT of «difficult samples» :

Mixtures of wheat protein and wheat starch powders

(5 different ratios: 0,25,50,75,100% protein)

Different sample thickness, different sample packing

2 technical replicates

MSC:

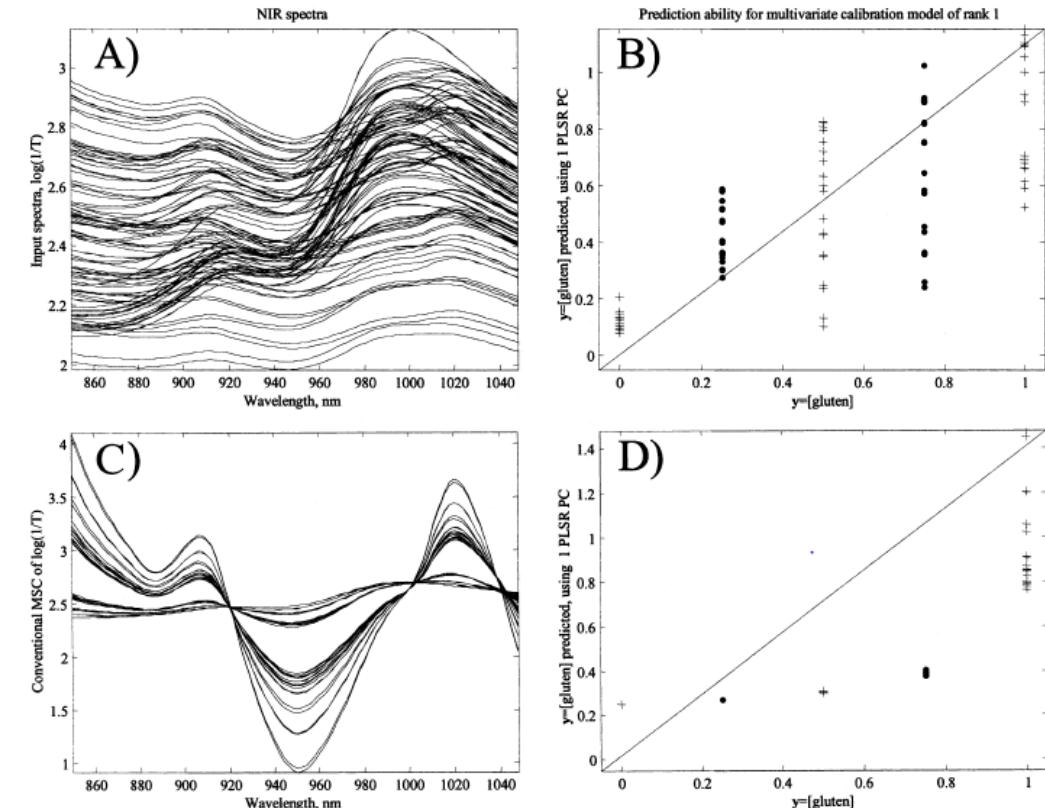
Martens, H., Jensen, S.Å. and Geladi, P. (1983)

Multivariate linearity transformation for near-infrared reflectance spectrometry.

Proc. Nordic Symp. on Applied Statistics, (O.H.J. Christie, ed.) June 12.-14. 1983. Stokkand Forlag
Publ., Skagenkaien 12, N-4000 Stavanger, Norway, ISBN 82-90496-02-8, 208-234.

Geladi, P., MacDougall, D. and Martens, H. (1985): *Linearization and scatter-correction for near-infrared reflectance of meat.*

Applied Spectroscopy, 39, 3, 491-500.



Problem: MSC regression model could not estimate the offset and slope properly. Confused by large chemical variations.

Really bad news

NIT of «difficult samples» :

Mixtures of wheat protein and wheat starch powders

(5 different ratios: 0,25,50,75,100% protein)

Different sample thickness, different sample packing

2 technical replicates

MSC:

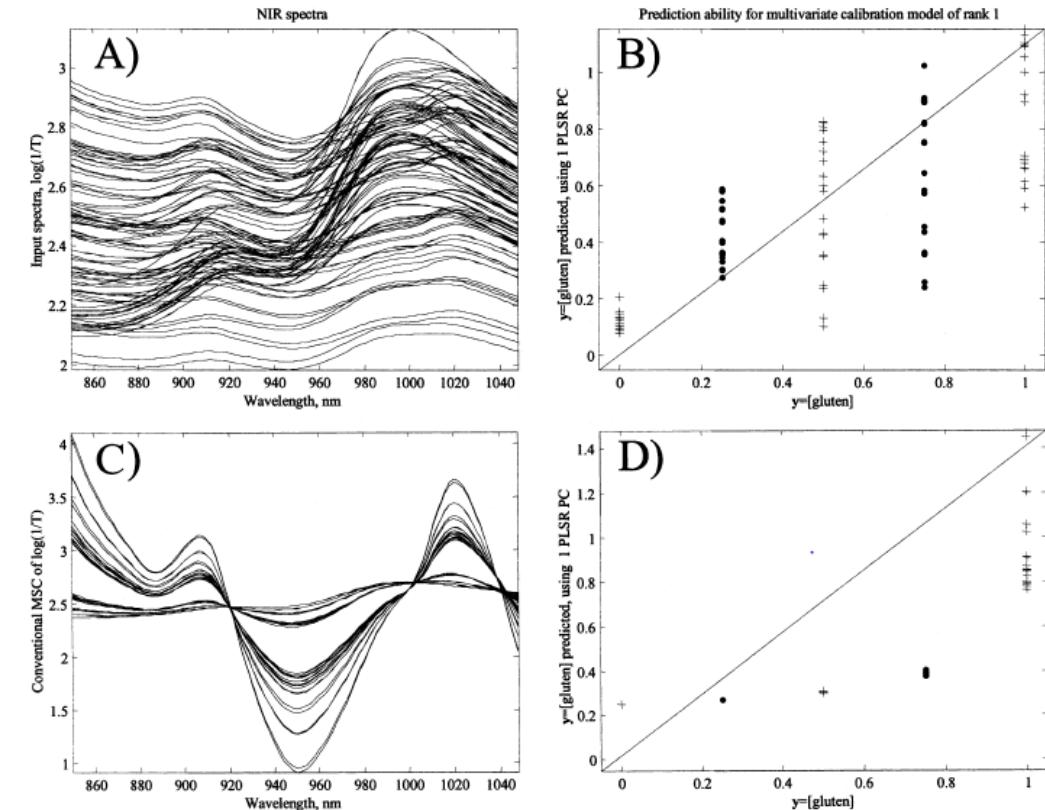
Martens, H., Jensen, S.Å. and Geladi, P. (1983)

Multivariate linearity transformation for near-infrared reflectance spectrometry.

Proc. Nordic Symp. on Applied Statistics, (O.H.J. Christie, ed.) June 12.-14. 1983. Stokkand Forlag
Publ., Skagenkaien 12, N-4000 Stavanger, Norway, ISBN 82-90496-02-8, 208-234.

Geladi, P., MacDougall, D. and Martens, H. (1985): *Linearization and scatter-correction for near-infrared reflectance of meat.*

Applied Spectroscopy, 39, 3, 491-500.



EMSC: H.Martens and E.Stark (1991)

Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. J.Pharmaceutical & Biomedical Analysis 9(8),625-635.

Martens, H., Pram Nielsen, J. and Balling Engelsen, S (2003)

Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. Anal. Chem. 75 (3) pp 394 – 404.

The solution, of course: EMSC with known analyte spectra !

NIT of «difficult samples» :

Mixtures of wheat protein and wheat starch powders

(5 different ratios: 0,25,50,75,100% protein)

Different sample thickness, different sample packing

2 technical replicates

MSC:

Martens, H., Jensen, S.Å. and Geladi, P. (1983)

Multivariate linearity transformation for near-infrared reflectance spectrometry.

Proc. Nordic Symp. on Applied Statistics, (O.H.J. Christie, ed.) June 12.-14. 1983. Stokkand Forlag
Publ., Skagenkaien 12, N-4000 Stavanger, Norway, ISBN 82-90496-02-8, 208-234.

Geladi, P., MacDougall, D. and Martens, H. (1985): *Linearization and scatter-correction for near-infrared reflectance of meat.*

Applied Spectroscopy, 39, 3, 491-500.

EMSC: H.Martens and E.Stark (1991)

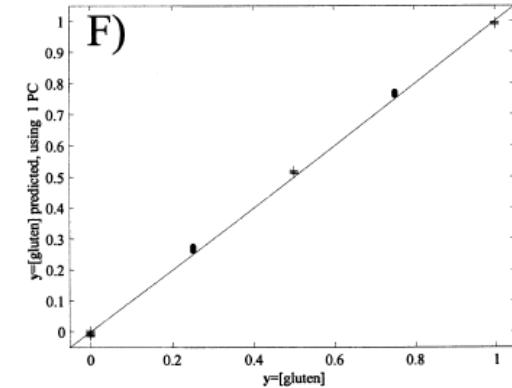
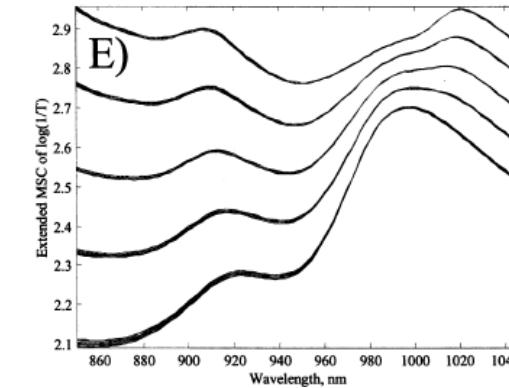
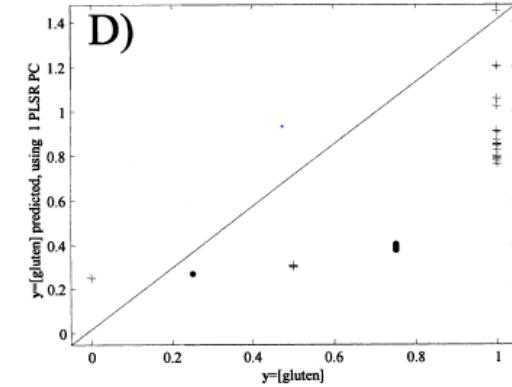
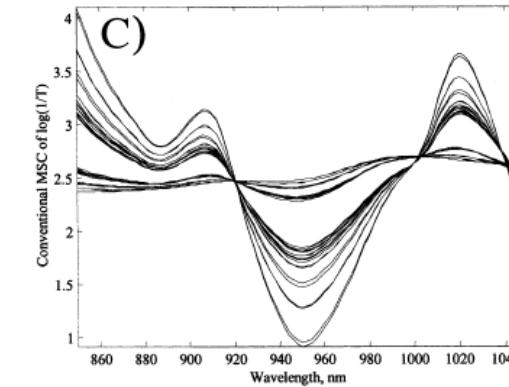
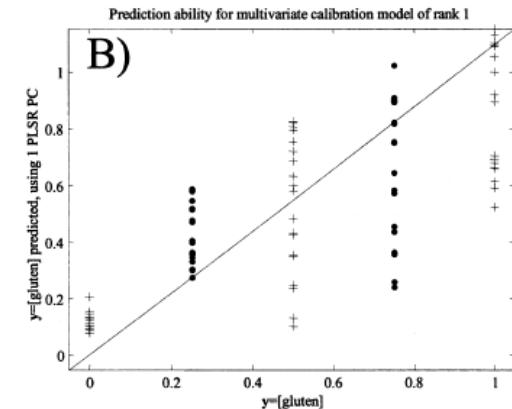
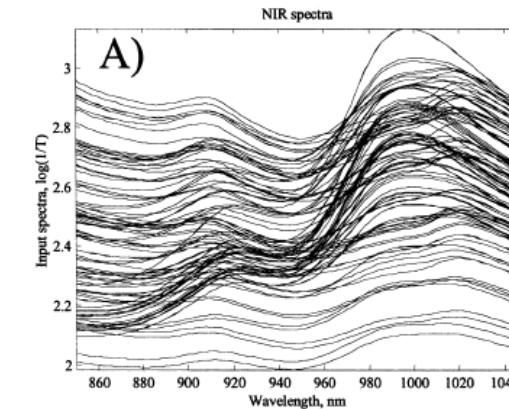
Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. J.Pharmaceutical & Biomedical Analysis 9(8),625-635.

Martens, H., Pram Nielsen, J. and Balling Engelsen, S (2003)

Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures.
Anal. Chem. 75 (3) pp 394 – 404.

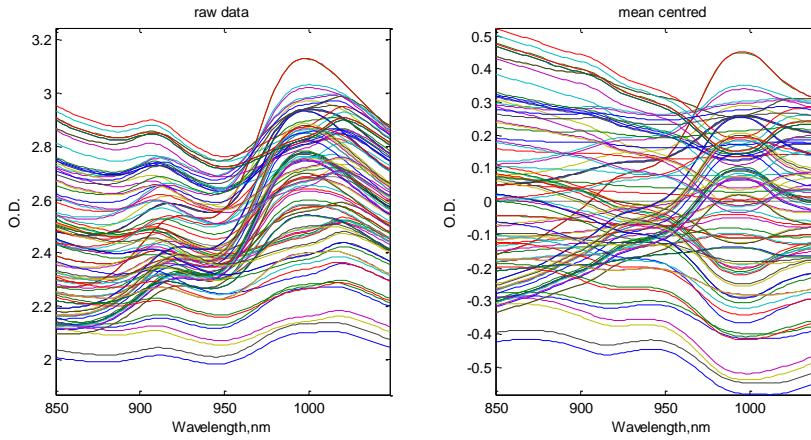
Using known constituent spectra in the EMSC model:

Difference spectrum between pure protein and pure starch



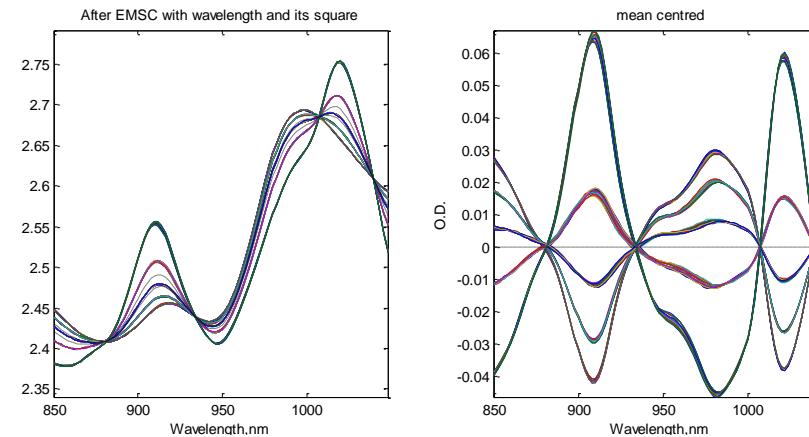
Good news, but needs extra info

Stabilize the preprocessing without any prior knowledge



**EMSC without known analyte spectra
(Morten Beck Rye):**

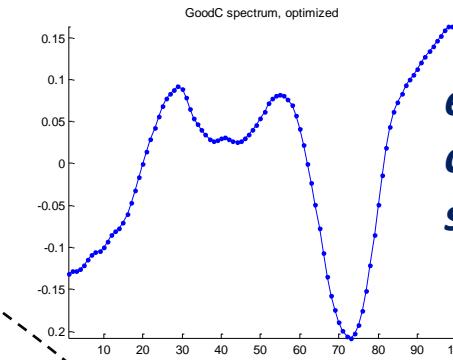
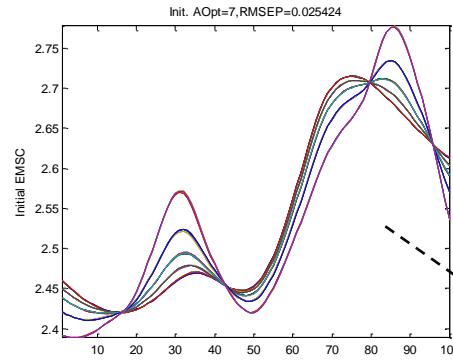
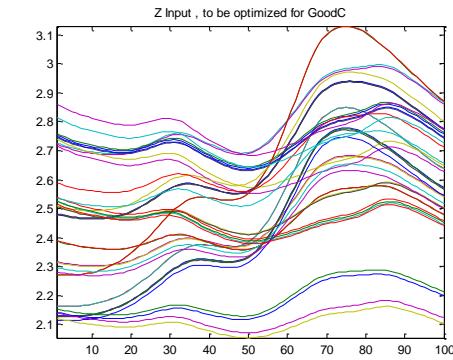
**No known constituent spectra used in the EMSC model.
Instead: Included known wavelength info:
(λ, λ^2)**



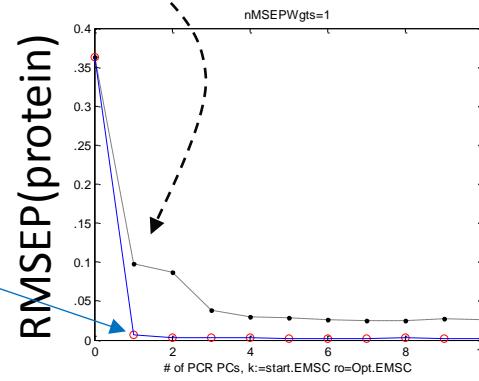
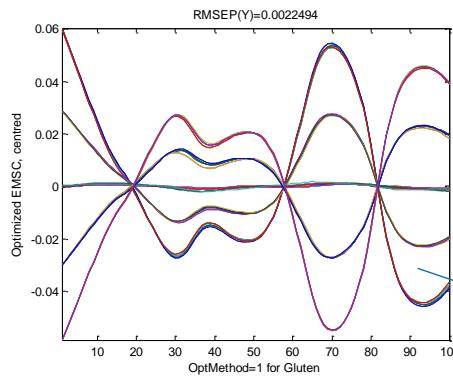
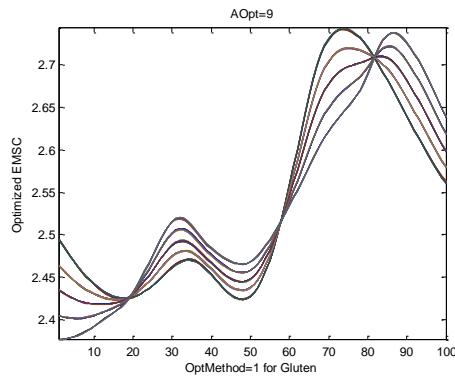
Good news, but nonlinear

Optimized EMSC (OEMSC)

Cal.
subset



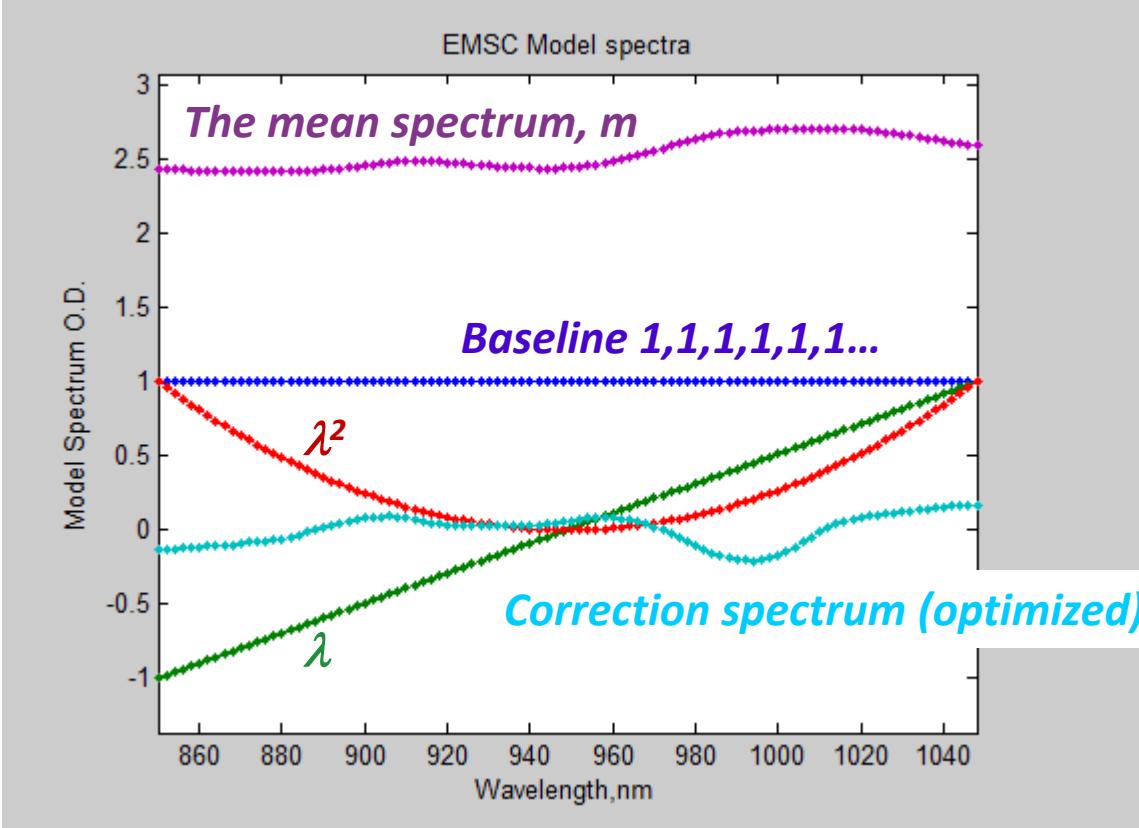
*estimated
correction
spectrum*



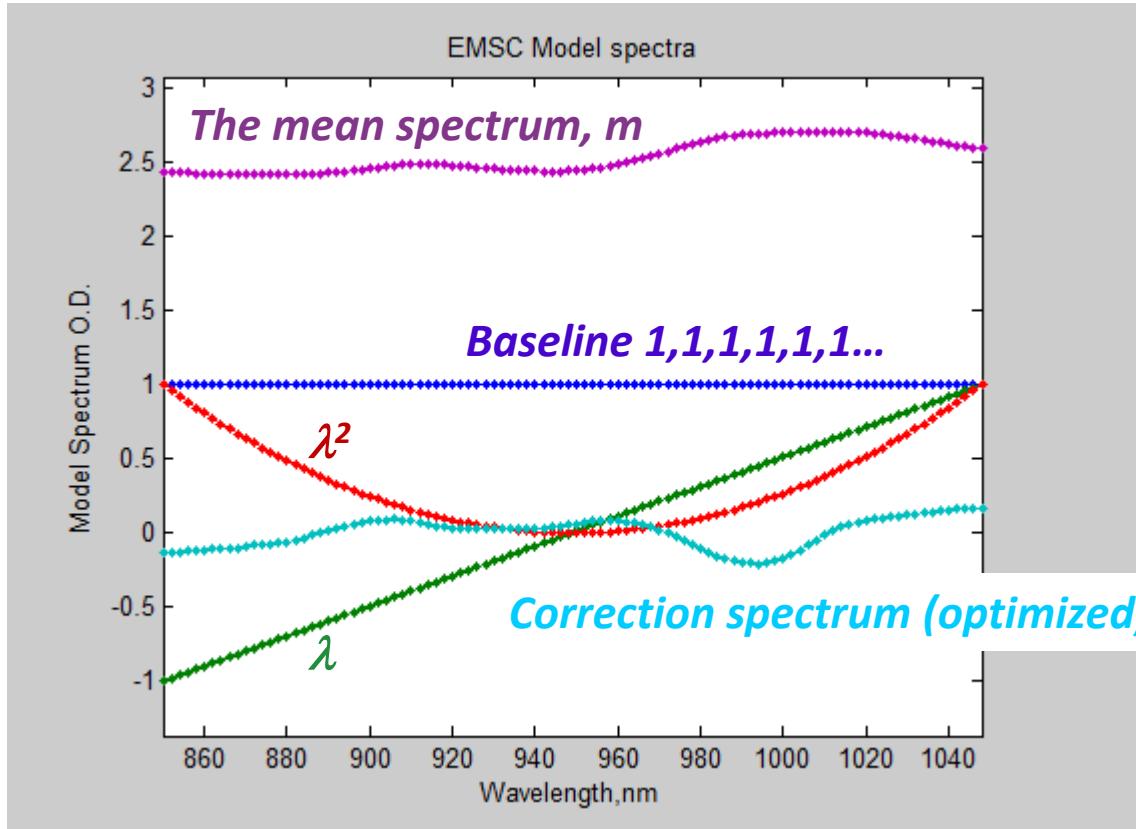
Estimate a correction spectrum in the EMSC model (in addition to λ, λ^2)
Simplex optimization of rmsep(protein)

Good news: linear

Optimize the EMSC model automatically



Optimize the EMSC model automatically



EMSC:

a) Estimation: Project input each sample spectrum on these model spectra.

b) Correction: *Subtract baseline level and λ^2 - effects. Divide by m-effects.*

OEMSC:

1) Define V = First e.g. 5 PCA PCs of the untreated sample spectra.

The correction spectrum will be a linear combination tV by optimizing $t(1 \times 5)$.

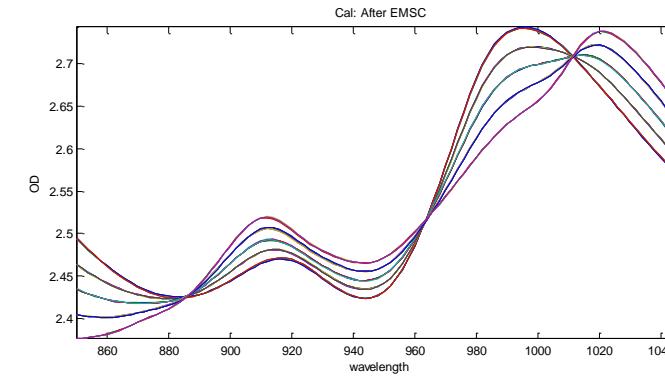
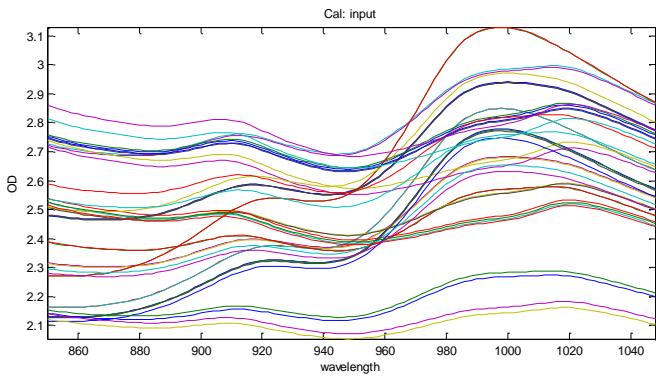
2) Find optimal correction spectrum.

Here: Guess initial t . Then Simplex optimization of t for a chosen criterion (e.g. RMSEP(y) of leverage-corrected PCR).

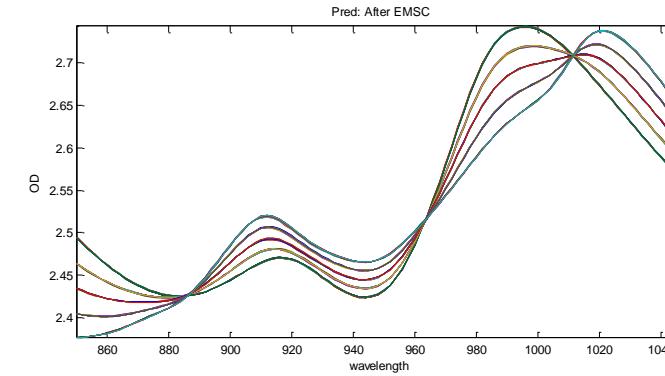
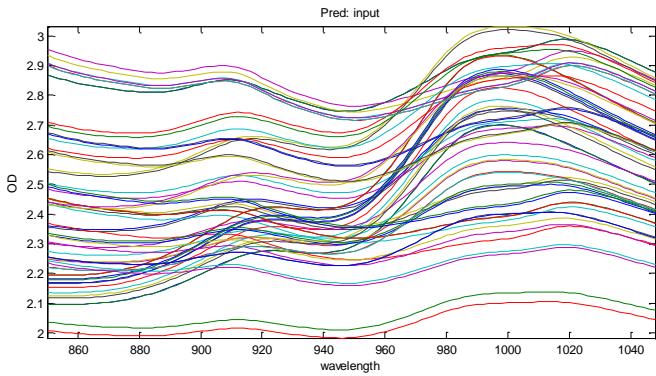
Many alternative opt.criteria possible!

OEMSC model overfitted ?

Cal.
subset

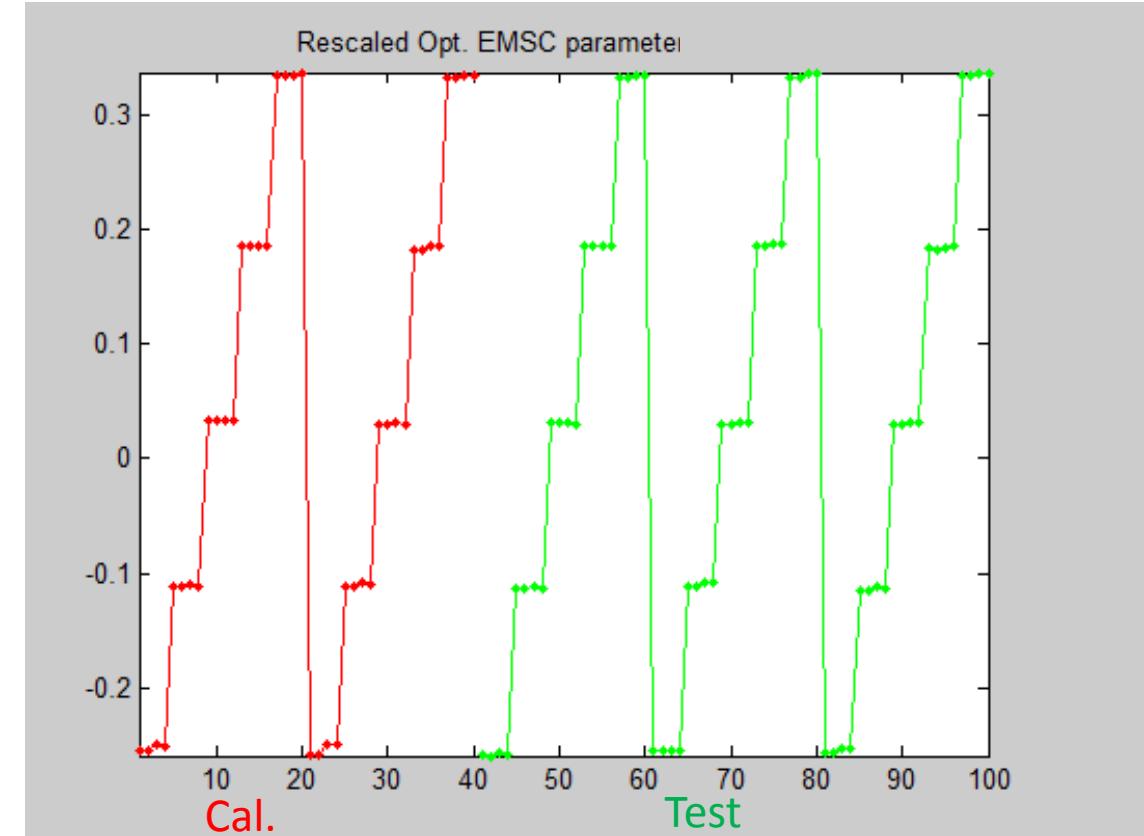
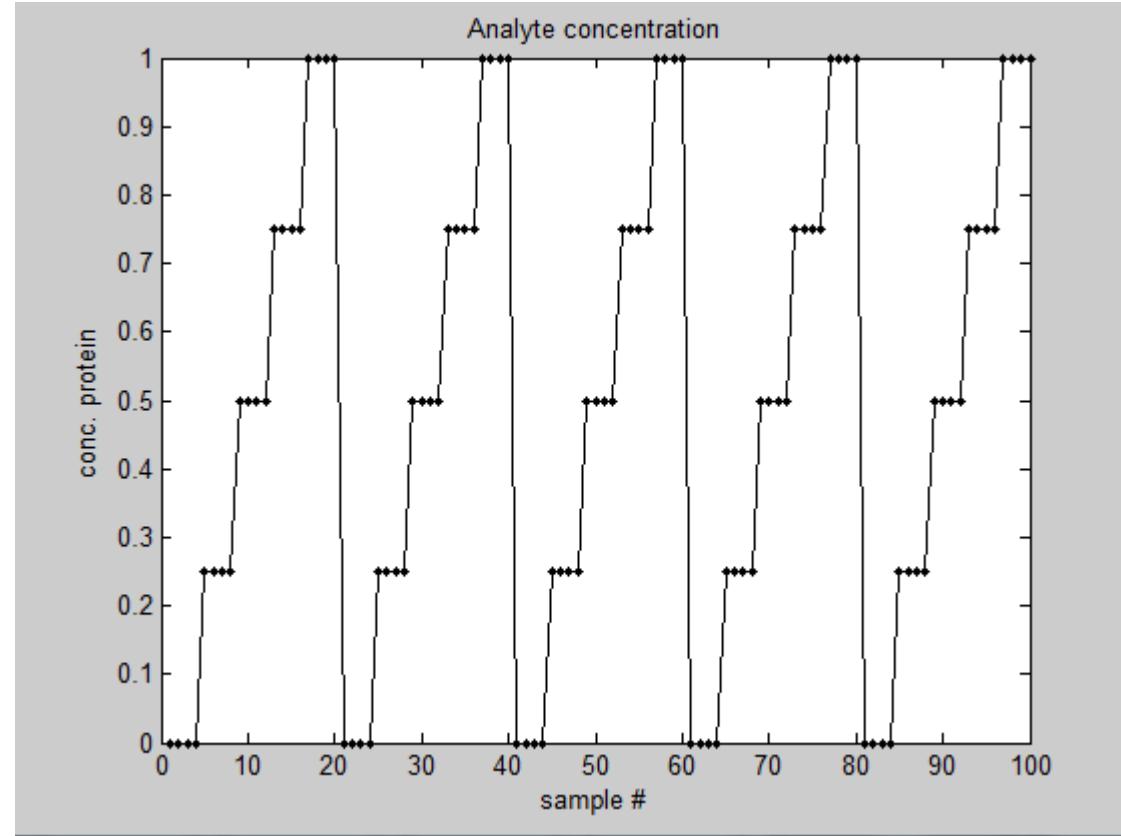


Test
subset



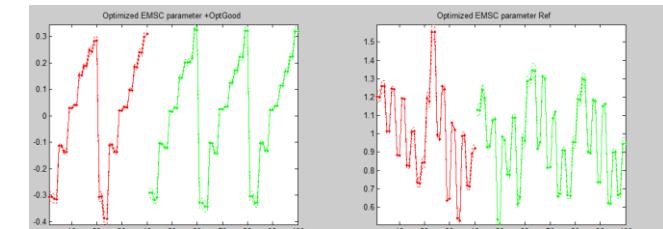
Not overfitted

Optimized EMSC parameter: proportional to analyte variable optimized for (protein)



Multivariate calibration unnecessary?

Conventional EMSC:



Does OEMSC also work for other data?

- Limited experience
- One example: protein in ground wheat from $\log(1/R)$
- Water structure: which criterion to optimize?

Software

- EMSC:
 - The Unscrambler (www.camo.com)
 - PLS Toolbox (www.eigenvector.com)
- OEMSC:
 - Matlab (preliminary) from harald.martens@ntnu.no