Separating Physics from Chemistry in Diffuse Spectroscopy:

Light scattering and light absorbance
separated by
Extended Multiplicative Signal Correction (EMSC)
and variations thereof

Harald Martens
Harald.Martens@matforsk.no

NIR 2003, Cordoba Spain April 2003

Separating Physics from Chemistry in Diffuse Spectroscopy:

Light scattering and light absorbance
separated by
Extended Multiplicative Signal Correction (EMSC)
and variations thereof

Harald Martens
Harald.Martens@matforsk.no

NIR 2003, Cordoba Spain April 2003

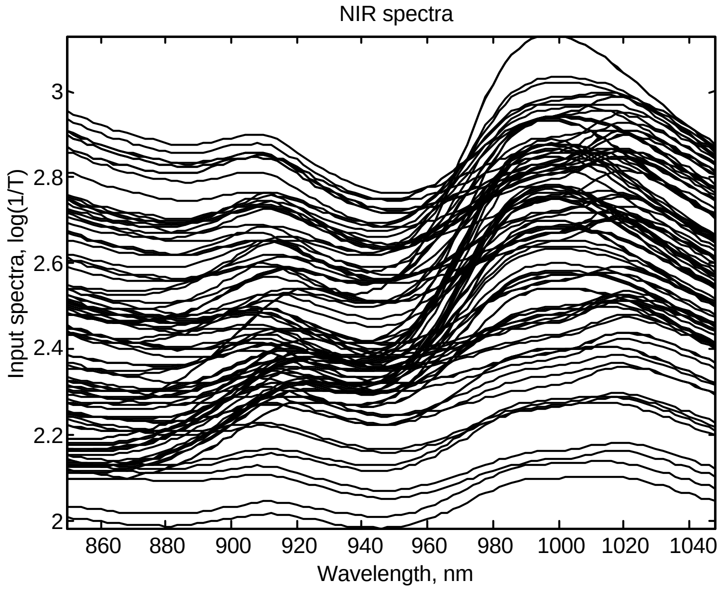
Acknowledgements:

Jesper Pram Nielsen, Michael Bom Froest,
Morten Beck Rye, Xuxin Lai and Achim Kohler

100 NIT spectra of mixtures of two powder types: wheat gluten and wheat starch.

Five different mixtures (0,25,50,75 and 100% gluten)

20 replicates of each (powder sampling, cuvette filling, sample packing, different sample holders, spectral parallels)



Martens, H., Pram Nielsen, J. and Balling Engelsen, S:

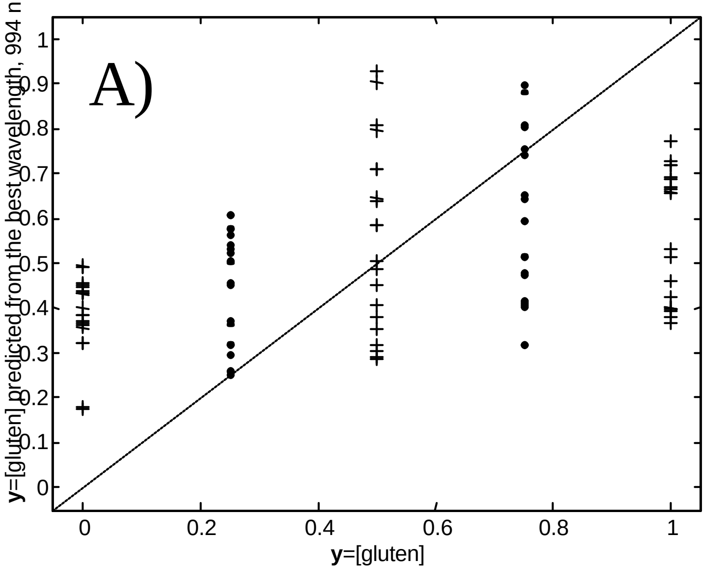
Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction.

Application to Near-Infrared Transmission Analysis of Powder Mixtures . Anal. Chem. 2003; **75** (3) pp 394 – 404.

Calibration for **y**=gluten fraction for **X** NIR log(1/T)

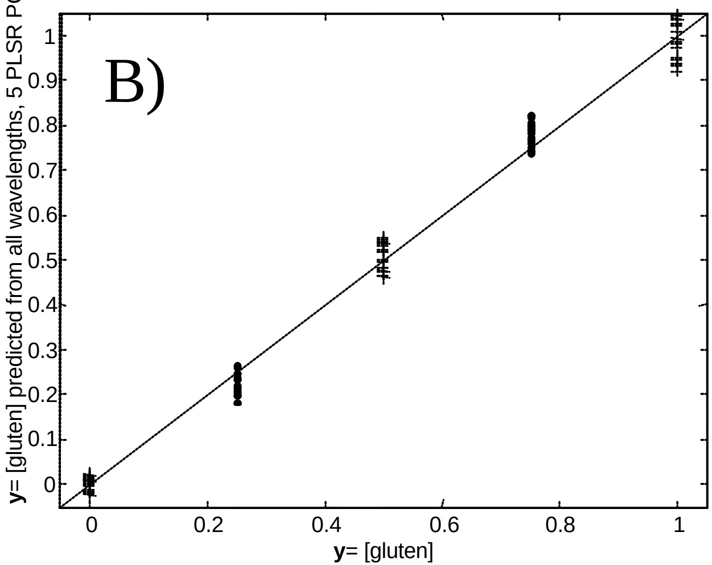
Best univariate calibration

(994 nm):

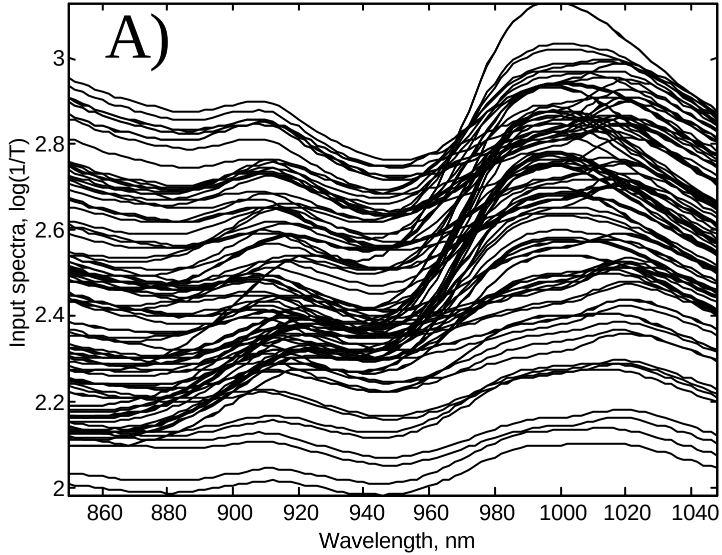


Best multivariate calibration

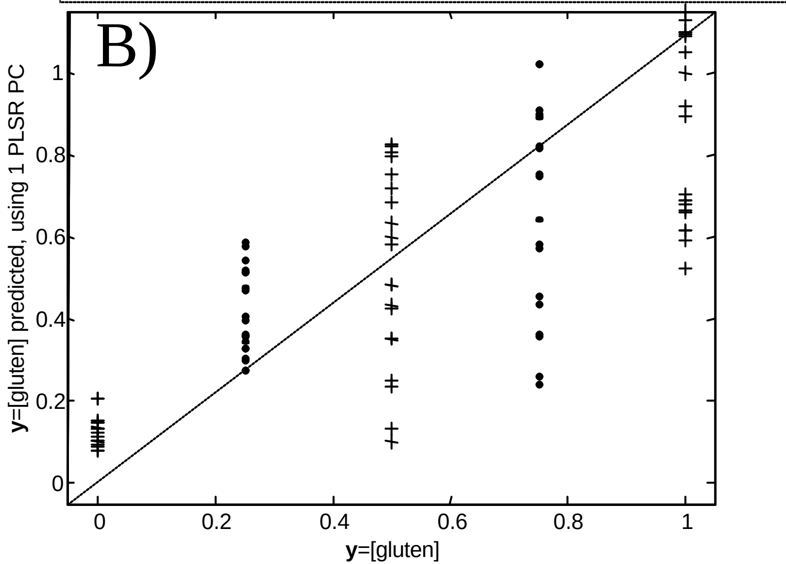
(via 5 PLSR PCs):

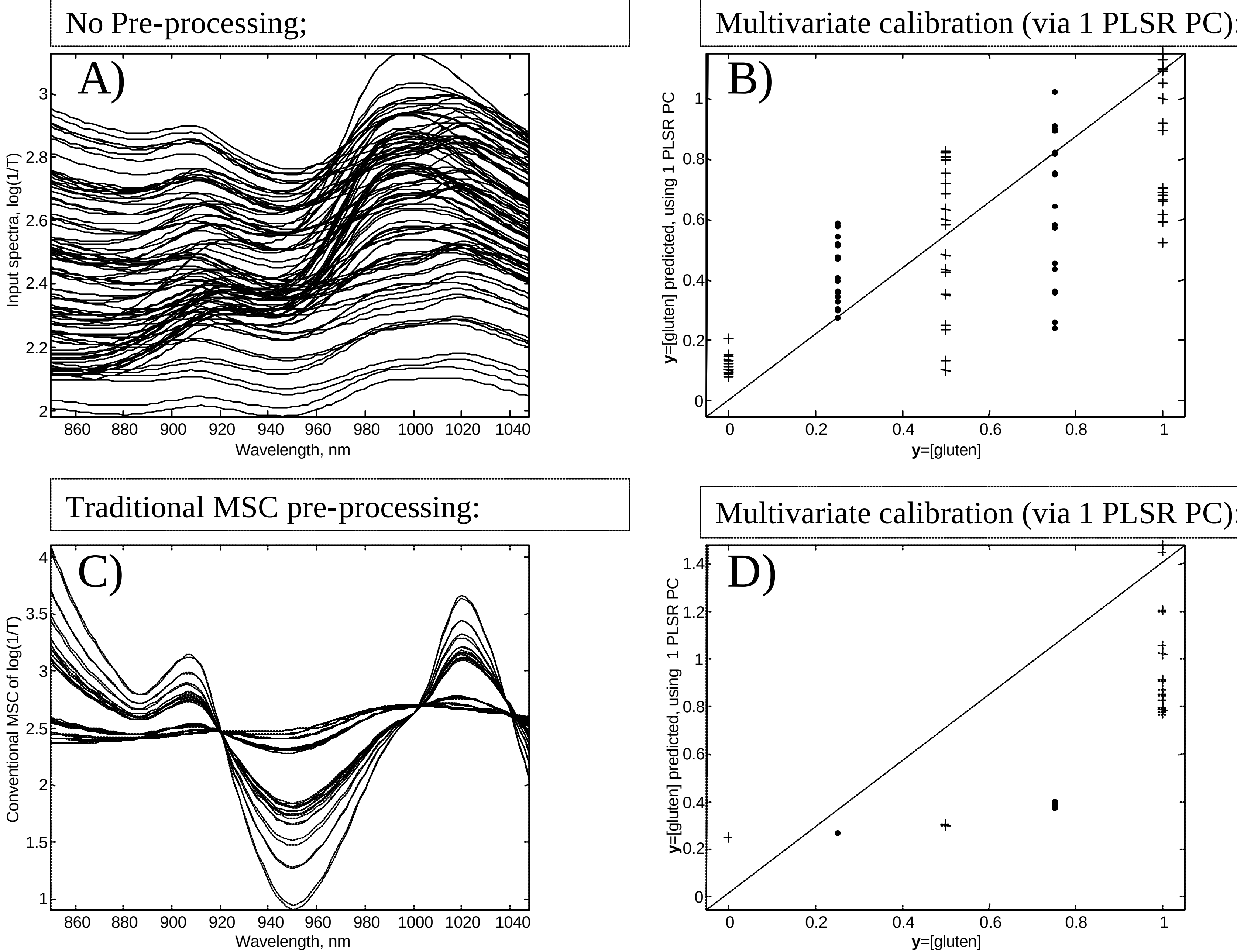


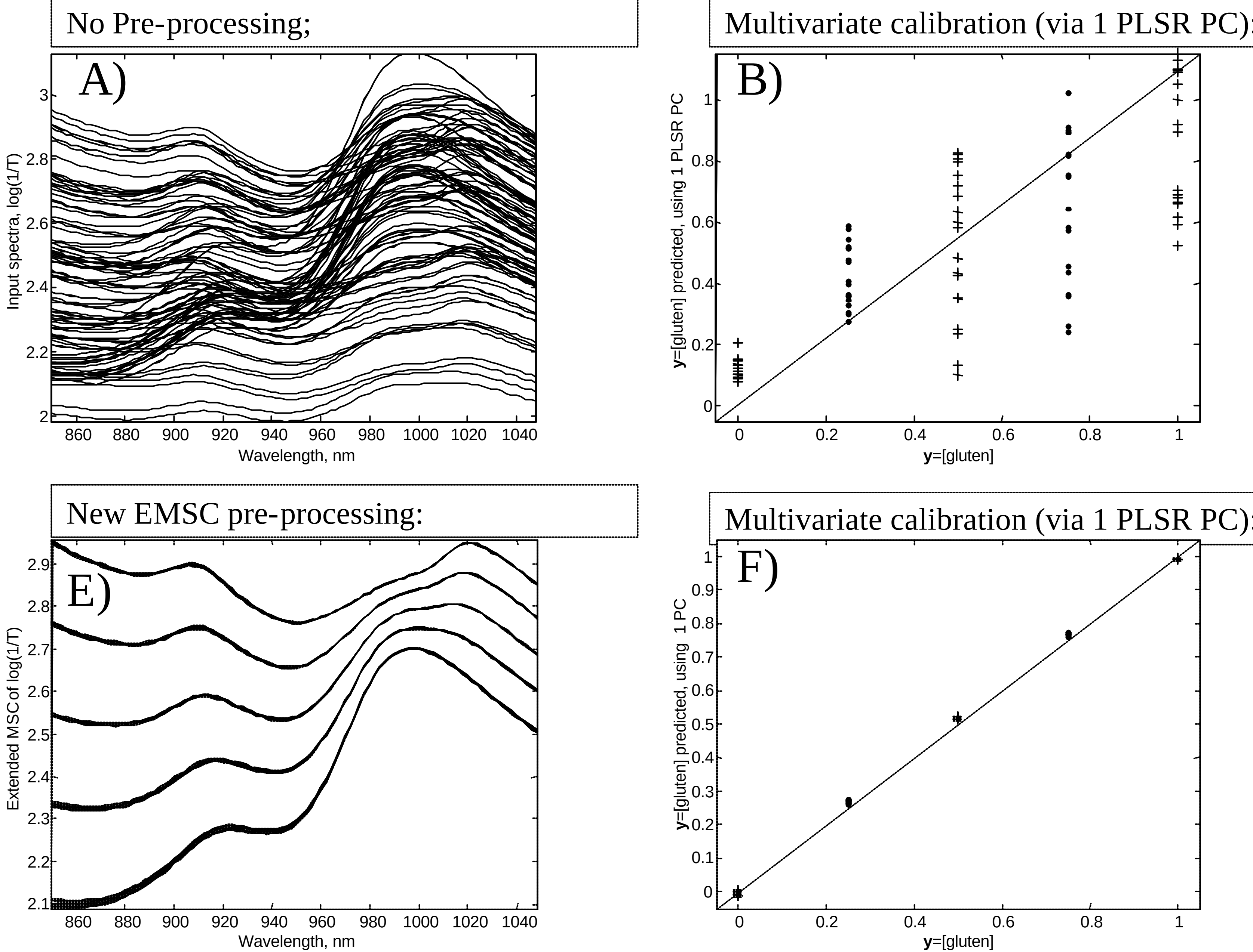
No Pre-processing;



Multivariate calibration (via 1 PLSR PC):




No Pre-processing;
A)
Input spectra, log(1/T)
Wavelength, nm
Multivariate calibration (via 1 PLSR PC):
B)
y=[gluten] predicted, using 1 PLSR PC
y=[gluten]
Traditional MSC pre-processing:
C)
Conventional MSC of log(1/T)
Wavelength, nm
Multivariate calibration (via 1 PLSR PC):
D)
y=[gluten] predicted, using 1 PLSR PC
y=[gluten]



No Pre-processing;
A)
Input spectra, log(1/T)
Wavelength, nm
Multivariate calibration (via 1 PLSR PC):
B)
y=[gluten] predicted, using 1 PLSR PC
y=[gluten]
New EMSC pre-processing:
E)
Extended MSC of log(1/T)
Wavelength, nm
Multivariate calibration (via 1 PLSR PC):
F)
y=[gluten] predicted, using 1 PC
y=[gluten]

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal "chemical" model $\mathbf{z}_{i,chem}$: Beer's Law(component spectra)

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal "chemical" model $\mathbf{z}_{i,chem}$: Beer's Law(component spectra)

* Ideal "physical" model: f($\mathbf{z}_{i,chem}$, pathlength/scattering, etc.)

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal “chemical” model $\mathbf{z}_{i,chem}$: Beer’s Law(component spectra)

* Ideal “physical” model: f($\mathbf{z}_{i,chem}$, pathlength/scattering, etc.)

* Linear approximation model for each input spectrum:

$$\mathbf{z}_i \approx f(\mathbf{z}_{i,chem}, \text{ scattering spectra })$$

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal "chemical" model $\mathbf{z}_{i,chem}$: Beer's Law(component spectra)

* Ideal "physical" model: f($\mathbf{z}_{i,chem}$, pathlength/scattering, etc.)

* Linear approximation model for each input spectrum:

$$\mathbf{z}_i \approx f(\mathbf{z}_{i,chem}, \text{ scattering spectra })$$

* Estimate parameters in model f( ).

* Correct spectra by EMSC:

$$\mathbf{z}_{i,chem} \approx \mathbf{z}_{i,corrected} = f^{-1}(\mathbf{z}_i, \text{ component spectra, scattering spectra})$$

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal "chemical" model $\mathbf{z}_{i,chem}$: Beer's Law(component spectra)

* Ideal "physical" model: f($\mathbf{z}_{i,chem}$, pathlength/scattering, etc.)

* Linear approximation model for each input spectrum:

$$\mathbf{z}_i \approx f(\mathbf{z}_{i,chem}, \text{scattering spectra})$$

* Estimate parameters in model f( ).

* Correct spectra by EMSC:

$$\mathbf{z}_{i,chem} \approx \mathbf{z}_{i,corrected} = f^{-1}(\mathbf{z}_i, \text{component spectra, scattering spectra})$$

Improve the model

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal "chemical" model $\mathbf{z}_{i,chem}$: Beer's Law(component spectra)

* Ideal "physical" model: f($\mathbf{z}_{i,chem}$, pathlength/scattering, etc.)

* Linear approximation model for each input spectrum:

$$\mathbf{z}_i \approx f(\mathbf{z}_{i,chem}, \text{ scattering spectra })$$

* Estimate parameters in model f( ).

* Correct spectra by EMSC:

$$\mathbf{z}_{i,chem} \approx \mathbf{z}_{i,corrected} = f^{-1}(\mathbf{z}_i, \text{ component spectra, scattering spectra})$$

Improve the model, the correction

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal "chemical" model $\mathbf{z}_{i,chem}$: Beer's Law(component spectra)

* Ideal "physical" model: f($\mathbf{z}_{i,chem}$, pathlength/scattering, etc.)

* Linear approximation model for each input spectrum:

$$\mathbf{z}_i \approx f(\mathbf{z}_{i,chem}, \text{ scattering spectra })$$

* Estimate parameters in model f( ).

* Correct spectra by EMSC:

$$\mathbf{z}_{i,chem} \approx \mathbf{z}_{i,corrected} = f^{-1}(\mathbf{z}_i, \text{ component spectra, scattering spectra})$$

Improve the model, the correction, the parameter estimation

EMSC theory:

Each input spectrum $\mathbf{z}_i$ is modelled and corrected:

* Ideal "chemical" model $\mathbf{z}_{i,chem}$: Beer's Law(component spectra)

* Ideal "physical" model: f($\mathbf{z}_{i,chem}$, pathlength/scattering, etc.)

* Linear approximation model for each input spectrum:

$$\mathbf{z}_i \approx f(\mathbf{z}_{i,chem}, \text{scattering spectra})$$

* Estimate parameters in model f( ).

* Correct spectra by EMSC:

$$\mathbf{z}_{i,chem} \approx \mathbf{z}_{i,corrected} = f^{-1}(\mathbf{z}_i, \text{component spectra, scattering spectra})$$

Improve the model, the correction, the parameter estimation, the component spectra

Ideal chemical model: Additive model (Beer's law):

$$\mathbf{z}_{i,\mathrm{chem}} = c_{i1}\mathbf{k}_1' + \ldots + c_{ij}\mathbf{k}_j' + \ldots + c_{iJ}\mathbf{k}_J'$$

$$= \mathbf{c}_i\mathbf{K}'$$

where c_{ij} =conc., $\mathbf{k}_j$ = absorptivity of constituent j

Ideal chemical model: Additive model (Beer's law):

$$\mathbf{z}_{i,chem} = c_{i1}\mathbf{k}_1{}' + \ldots + c_{ij}\mathbf{k}_j{}' + \ldots + c_{iJ}\mathbf{k}_J{}'$$
$$= \mathbf{c}_i\mathbf{K}'$$

where c_{ij} =conc., $\mathbf{k}_j$ = absorptivity of constituent j

Expressed as deviations around a reference spectrum:

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathbf{c}_i\mathbf{K}'$$

Ideal chemical model: Additive model (Beer's law):

$$\mathbf{z}_{i,\text{chem}} = c_{i1}\mathbf{k}_1' + \ldots + c_{ij}\mathbf{k}_j' + \ldots + c_{iJ}\mathbf{k}_J'$$
$$= \mathbf{c}_i\mathbf{K}'$$

where c_{ij} =conc., $\mathbf{k}_j$ = absorptivity of constituent j

Expressed as deviations around a reference spectrum:

$$\mathbf{z}_{i,\text{chem}} = \mathbf{m}' + \mathbf{c}_i\mathbf{K}'$$

The EMSC model of physical interferants:

$$\mathbf{z}_i \approx a_i\mathbf{1} + b_i\mathbf{z}_{i,\text{chem}} + d_i \mathrm{l} + e_i \mathrm{l}^2$$

Ideal chemical model: Additive model (Beer's law):

$$\mathbf{z}_{i,chem} = c_{i1}\mathbf{k}_1{}' + \ldots + c_{ij}\mathbf{k}_j{}' + \ldots + c_{iJ}\mathbf{k}_J{}'$$
$$= \mathbf{c}_i\mathbf{K}'$$

where c_{ij} =conc., $\mathbf{k}_j$ = absorptivity of constituent j

Expressed as deviations around a reference spectrum:

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathbf{c}_i\mathbf{K}'$$

The EMSC model of physical interferants:

$$\mathbf{z}_i \approx a_i\mathbf{1} + b_i\mathbf{z}_{i,chem} + d_i l + e_i l^2$$

Linear model for estimation:

$$\mathbf{z}_i = a_i\mathbf{1} + b_i\mathbf{m}' + b_i\mathbf{c}_i\mathbf{K}' + d_i l + e_i l^2 + \mathrm{e}_i$$

Ideal chemical model: Additive model (Beer's law):

$$\mathbf{z}_{i,chem} = c_{i1}\mathbf{k}_1{}' + \ldots + c_{ij}\mathbf{k}_j{}' + \ldots + c_{iJ}\mathbf{k}_J{}'$$

$$= \mathbf{c}_i\mathbf{K}'$$

where c_{ij} =conc., $\mathbf{k}_j$ = absorptivity of constituent j

Expressed as deviations around a reference spectrum:

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathbf{c}_i\mathbf{K}'$$

The EMSC model of physical interferants:

$$\mathbf{z}_i \approx a_i\mathbf{1} + b_i\mathbf{z}_{i,chem} + d_i\,\mathrm{l} + e_i\,\mathrm{l}^2$$

Linear model for estimation:

$$\mathbf{z}_i = a_i\mathbf{1} + b_i\mathbf{m}' + b_i\mathbf{c}_i\mathbf{K}' + d_i\,\mathrm{l} + e_i\,\mathrm{l}^2 + \mathrm{e}_i$$

Ideal chemical model: Additive model (Beer's law):

$$\mathbf{z}_{i,chem} = c_{i1}\mathbf{k}_1{}' + \ldots + c_{ij}\mathbf{k}_j{}' + \ldots + c_{iJ}\mathbf{k}_J{}'$$
$$= \mathbf{c}_i\mathbf{K}'$$

where c_{ij} =conc., $\mathbf{k}_j$ = absorptivity of constituent j

Expressed as deviations around a reference spectrum:

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathbf{c}_i\mathbf{K}'$$

The EMSC model of physical interferants:

$$\mathbf{z}_i \approx a_i\mathbf{1} + b_i\mathbf{z}_{i,chem} + d_i\mathrm{l} + e_i\mathrm{l}^2$$

Linear model for estimation:

$$\mathbf{z}_i = a_i\mathbf{1} + b_i\mathbf{m}' + b_i\mathbf{c}_i\mathbf{K}' + d_i\mathrm{l} + e_i\mathrm{l}^2 + \mathrm{e}_i$$
$$\mathbf{z}_i = a_i\mathbf{1} + b_i\mathbf{m}' + \mathbf{h}_i\mathbf{K}' + d_i\mathrm{l} + e_i\mathrm{l}^2 + \mathrm{e}_i$$

Ideal chemical model: Additive model (Beer's law):

$$\mathbf{z}_{i,chem} = c_{i1}\mathbf{k}_1' + \ldots + c_{ij}\mathbf{k}_j' + \ldots + c_{iJ}\mathbf{k}_J'$$
$$= \mathbf{c}_i\mathbf{K}'$$

where c_{ij} =conc., $\mathbf{k}_j$ = absorptivity of constituent j

Expressed as deviations around a reference spectrum:

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathbf{c}_i\mathbf{K}'$$

The EMSC model of physical interferants:

$$\mathbf{z}_i \approx a_i\mathbf{1} + b_i\mathbf{z}_{i,chem} + d_i\lambda + e_i\lambda^2$$

Linear model for estimation:

$$\mathbf{z}_i = a_i\mathbf{1} + b_i\mathbf{m}' + b_i\mathbf{c}_i\mathbf{K}' + d_i\lambda + e_i\lambda^2 + \mathbf{e}_i$$
$$\mathbf{z}_i = a_i\mathbf{1} + b_i\mathbf{m}' + \mathbf{h}_i\mathbf{K}' + d_i\lambda + e_i\lambda^2 + \mathbf{e}_i$$

The EMSC correction:

$$\mathbf{z}_{i,corrected} = (\mathbf{z}_i - a_i - d_i\lambda - e_i\lambda^2)/b_i$$

Deviations around a reference spectrum **m**
(e.g. the mean spectrum):

$$\mathbf{z}_{i,chem} = \mathbf{m}' + c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + ... + c_{iJ}\mathbf{k}_J'$$

where concentrations c_{ij} represent deviations around the unknown concentrations in the reference

Only two constituents

$$\mathbf{z}_{i,chem} = c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2'$$

Deviations around a reference spectrum $\mathbf{m}$

(e.g. the mean spectrum):

$$\mathbf{z}_{i,chem} = \mathbf{m}' + c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + \ldots + c_{iJ}\mathbf{k}_J'$$

where concentrations c_{ij} represent deviations around the unknown concentrations in the reference

Only two constituents

$$\mathbf{z}_{i,chem} \quad = \quad c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' \qquad\qquad = \mathbf{m}' + c_i\mathbf{K}'$$

where $\mathbf{K} = \mathbf{k}_1 - \mathbf{k}_2$ difference spectrum

Estimation of model parameters by weighted least squares:

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + \mathbf{h}_i\mathbf{K}' + d_i\,\mathrm{l} + e_i\,\mathrm{l}^2 + \mathrm{e}_i$$

Estimation of model parameters by weighted least squares:

$$\mathbf{z}_i = a_i \mathbf{1}' + b_i \mathbf{m}' + \mathbf{h}_i \mathbf{K}' + d_i \mathrm{l} + e_i \mathrm{l}^2 + \mathrm{e}_i$$

$$\mathbf{M} = [\mathbf{1}'; \mathbf{m}'; \mathbf{K}'; \mathrm{l}; \mathrm{l}^2]$$

$$\mathbf{p}_i = [a_i, b_i, \mathbf{h}_i, d_i, e_i]$$

$$\mathbf{z}_i = \mathbf{p}_i \mathbf{M} + \mathrm{e}_i$$

Estimation of model parameters by weighted least squares:

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + \mathbf{h}_i\mathbf{K}' + d_i l + e_i l^2 + \mathrm{e}_i$$

$$\mathbf{M} = [\mathbf{1}';\ \mathbf{m}';\ \mathbf{K}';\ l\ ;\ l^2]$$

$$\mathbf{p}_i = [a_i,\ b_i,\ \mathbf{h}_i,\ d_i,\ e_i]$$

$$\mathbf{z}_i = \mathbf{p}_i\,\mathbf{M} + \mathrm{e}_i$$

$$\mathbf{p}_i = \mathbf{z}_i\mathbf{V}\mathbf{M}'(\mathbf{M}\mathbf{V}\mathbf{M}')^{-1}$$

Estimation of model parameters by weighted least squares:

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + \mathbf{h}_i\mathbf{K}' + d_i l + e_i l^2 + \mathrm{e}_i$$

$$\mathbf{M}=[\mathbf{1}';\ \mathbf{m}';\ \mathbf{K}';\ l\ ;\ l^{2}]$$

$$\mathbf{p}_i = [a_i,\ b_i,\ \mathbf{h}_i,\ d_i,\ e_i]$$

$$\mathbf{z}_i = \mathbf{p}_i\,\mathbf{M} + \mathrm{e}_i$$

V = weight matrix

$$\mathbf{p}_i = \mathbf{z}_i\mathbf{V}\mathbf{M}'(\mathbf{M}\mathbf{V}\mathbf{M}')^{-1}$$

Estimation of model parameters by weighted least squares:

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + \mathbf{h}_i\mathbf{K}' + d_i\,\mathrm{l} + e_i\,\mathrm{l}^2 + \mathrm{e}_i$$

$$\mathbf{M} = [\mathbf{1}';\ \mathbf{m}';\ \mathbf{K}';\ \mathrm{l}\ ;\ \mathrm{l}^2]$$

$$\mathbf{p}_i = [a_i,\ b_i,\ \mathbf{h}_i,\ d_i,\ e_i]$$

$$\mathbf{z}_i = \mathbf{p}_i\,\mathbf{M} + \mathrm{e}_i$$

$$\mathbf{p}_i = \mathbf{z}_i\mathbf{V}\mathbf{M}'(\mathbf{M}\mathbf{V}\mathbf{M}')^{-1}$$

MVM´ full rank?

Estimation of model parameters by weighted least squares:

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + \mathbf{h}_i\mathbf{K}' + d_i\,\mathrm{l} + e_i\,\mathrm{l}^2 + \mathrm{e}_i$$

$$\mathbf{M}=[\mathbf{1}';\, \mathbf{m}';\, \mathbf{K}';\, \mathrm{l}\,;\, \mathrm{l}^2]$$

$$\mathbf{p}_i = [a_i, b_i, \mathbf{h}_i, d_i, e_i]$$

$$\mathbf{z}_i = \mathbf{p}_i\,\mathbf{M} + \mathrm{e}_i$$

$$\mathbf{p}_i = \mathbf{z}_i\mathbf{V}\mathbf{M}'(\mathbf{M}\mathbf{V}\mathbf{M}')^{-1}$$

$$\mathrm{e}_i = \mathbf{z}_i - \mathbf{p}_i\,\mathbf{M}$$

EMSC:

$$\mathbf{z}_{i,chem} = \mathbf{m}' + c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + \ldots + c_{iJ}\mathbf{k}_J'$$

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{z}_{i,chem} + e_i$$

<u>**MSC:**</u> *Ignore* $c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + \ldots + c_{iJ}\mathbf{k}_J$

$$\mathbf{z}_{i,chem} = \mathbf{m}' + d_i'$$

EMSC: $\mathbf{z}_{i,chem} = \mathbf{m}' + c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + \ldots + c_{iJ}\mathbf{k}_J'$

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{z}_{i,chem} + \mathrm{e}_i$$

<u>MSC:</u> *Ignore* $c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + \ldots + c_{iJ}\mathbf{k}_J$

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathrm{d}_i'$$

F*urther simplification:* $\mathrm{d}_i \approx \mathbf{0}$

$$\mathbf{z}_{i,chem} \approx \mathbf{m}'$$

EMSC: $\mathbf{z}_{i,chem} = \mathbf{m}' + c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + ... + c_{iJ}\mathbf{k}_J'$

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{z}_{i,chem} + \mathrm{e}_i$$

<u>**MSC:**</u> *Ignore* $c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + ... + c_{iJ}\mathbf{k}_J$

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathrm{d}_i'$$

F*urther simplification:* $\mathrm{d}_i \approx \mathbf{0}$

$$\mathbf{z}_{i,chem} \approx \mathbf{m}'$$

MSC model:

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + \mathrm{e}_i$$

EMSC:

$$\mathbf{z}_{i,chem} = \mathbf{m}' + c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + \ldots + c_{iJ}\mathbf{k}_J'$$

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{z}_{i,chem} + \mathrm{e}_i$$

<u>**MSC:**</u> *Ignore* $c_{i1}\mathbf{k}_1' + c_{i2}\mathbf{k}_2' + \ldots + c_{iJ}\mathbf{k}_J$

$$\mathbf{z}_{i,chem} = \mathbf{m}' + \mathrm{d}_i'$$

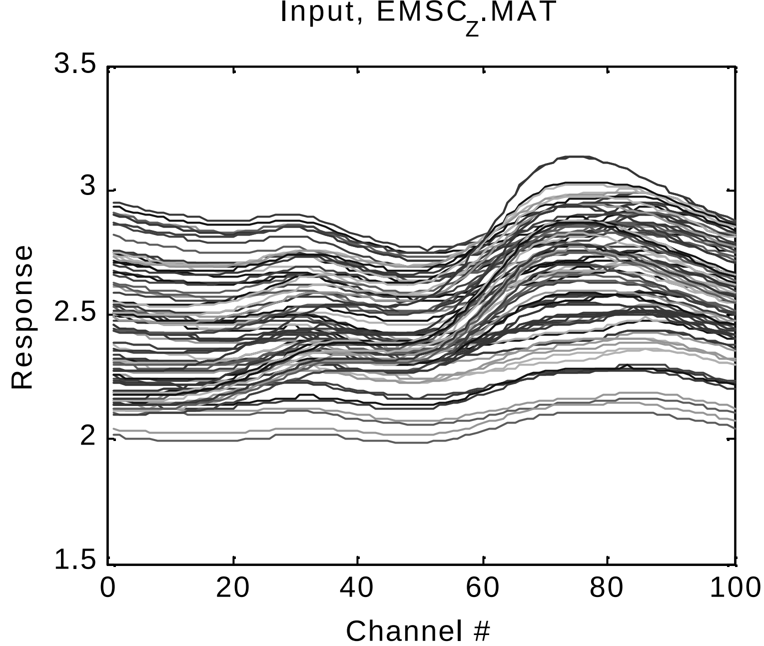
F*urther simplification:* $\mathrm{d}_i \approx \mathbf{0}$

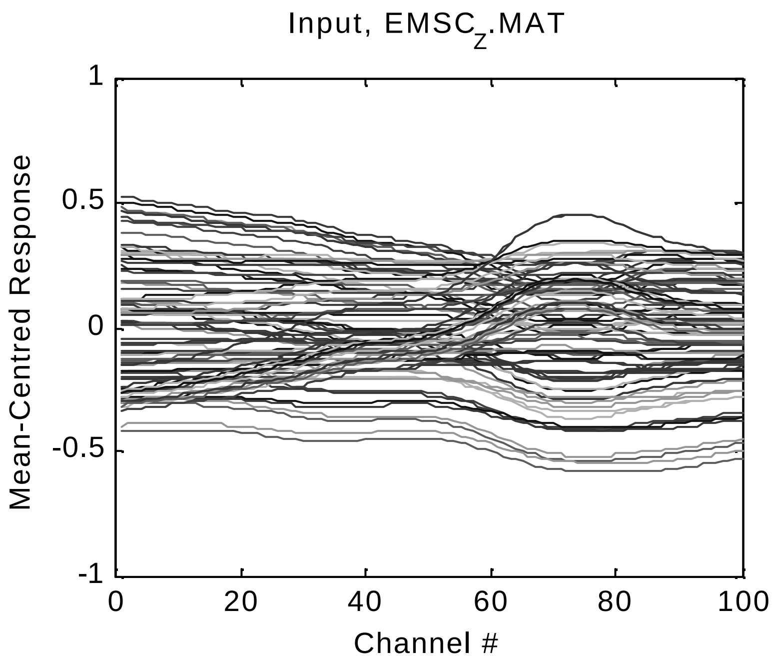
$$\mathbf{z}_{i,chem} \approx \mathbf{m}'$$

MSC model:

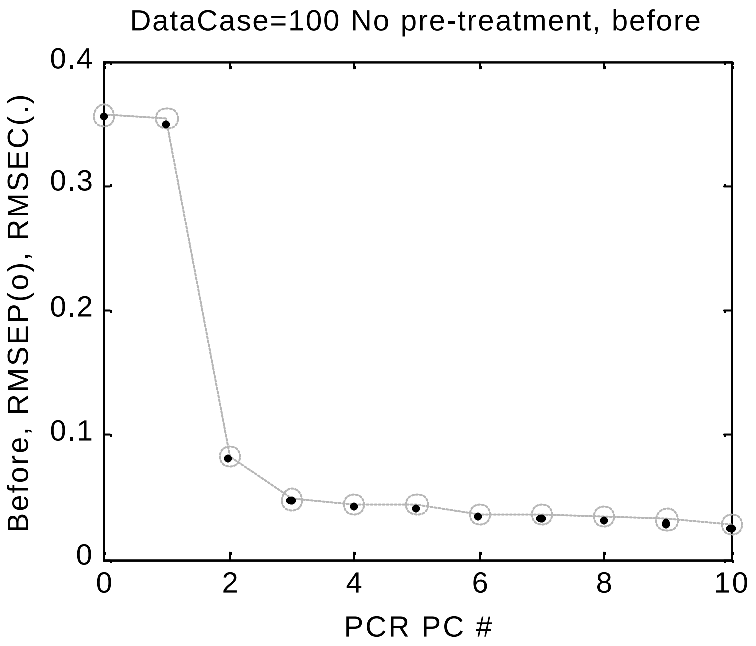
$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + \mathrm{e}_i$$

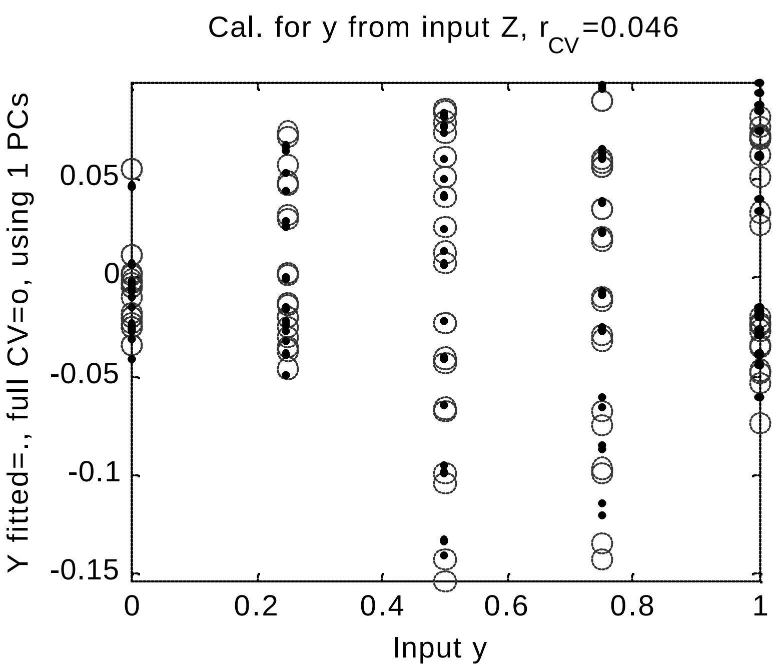
$$\mathbf{z}_{i,corrected} = (\mathbf{z}_i - a_i)/b_i$$





DataCase=100, No pre-treatment


DataCase=100 No pre-treatment, before
Before, RMSEP(o), RMSEC(.)
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
PCR PC #


Cal. for y from input Z, r_{CV}=0.046
Y fitted=., full CV=o, using 1 PCs
0.05
0
-0.05
-0.1
-0.15
0
0.2
0.4
0.6
0.8
1
Input y

MSC: Conventional modelling of offset a_i (pathlength or scaling) and slope b_i (baseline):

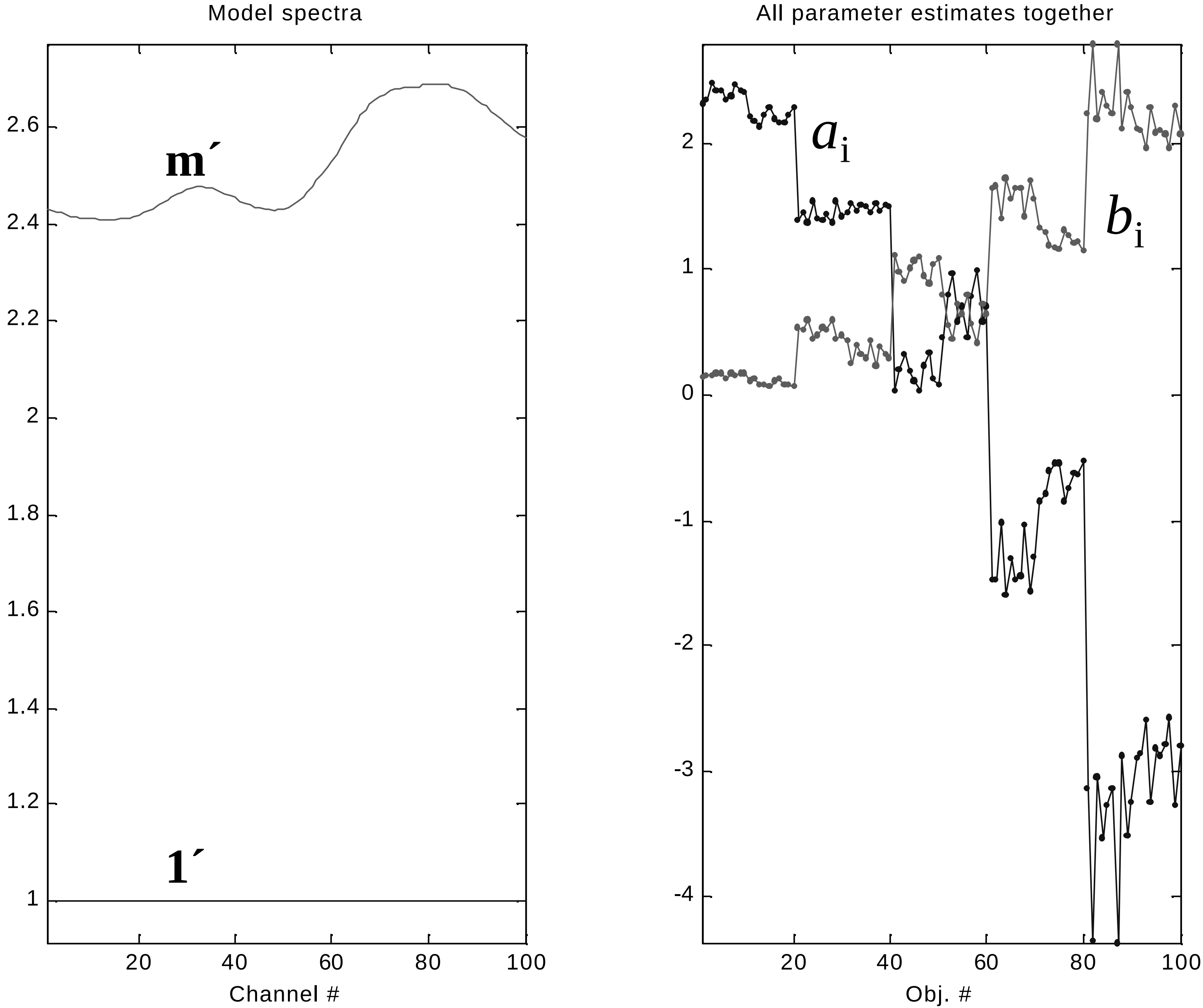
Model:

$$\mathbf{z}_i = a_i \mathbf{1}' + b_i \mathbf{m}' + e_i$$

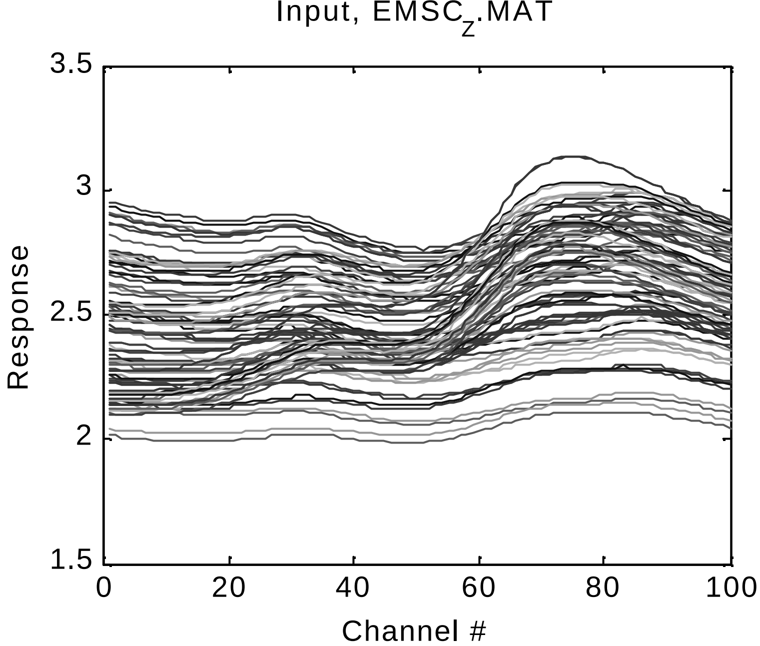
Estimate a_i and b_i from $\mathbf{z}_i$, $\mathbf{1}$ and $\mathbf{m}$

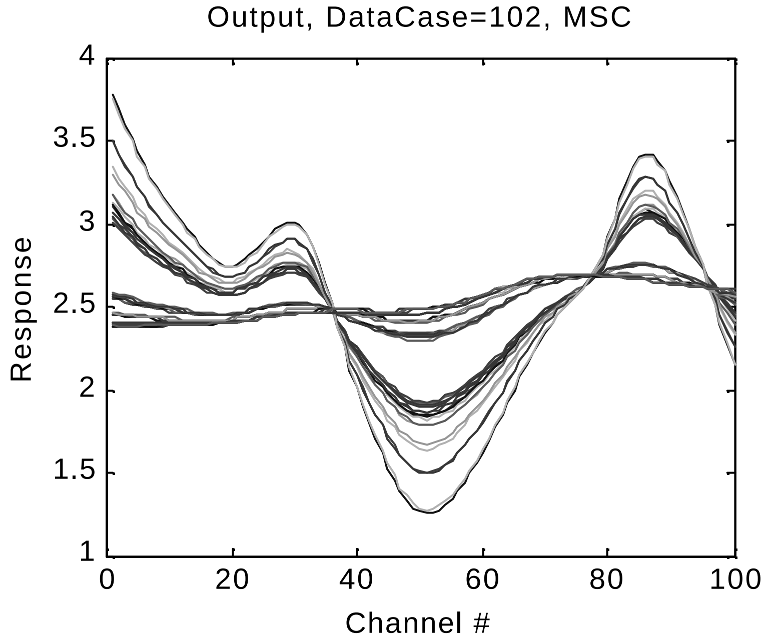
MSC correction:

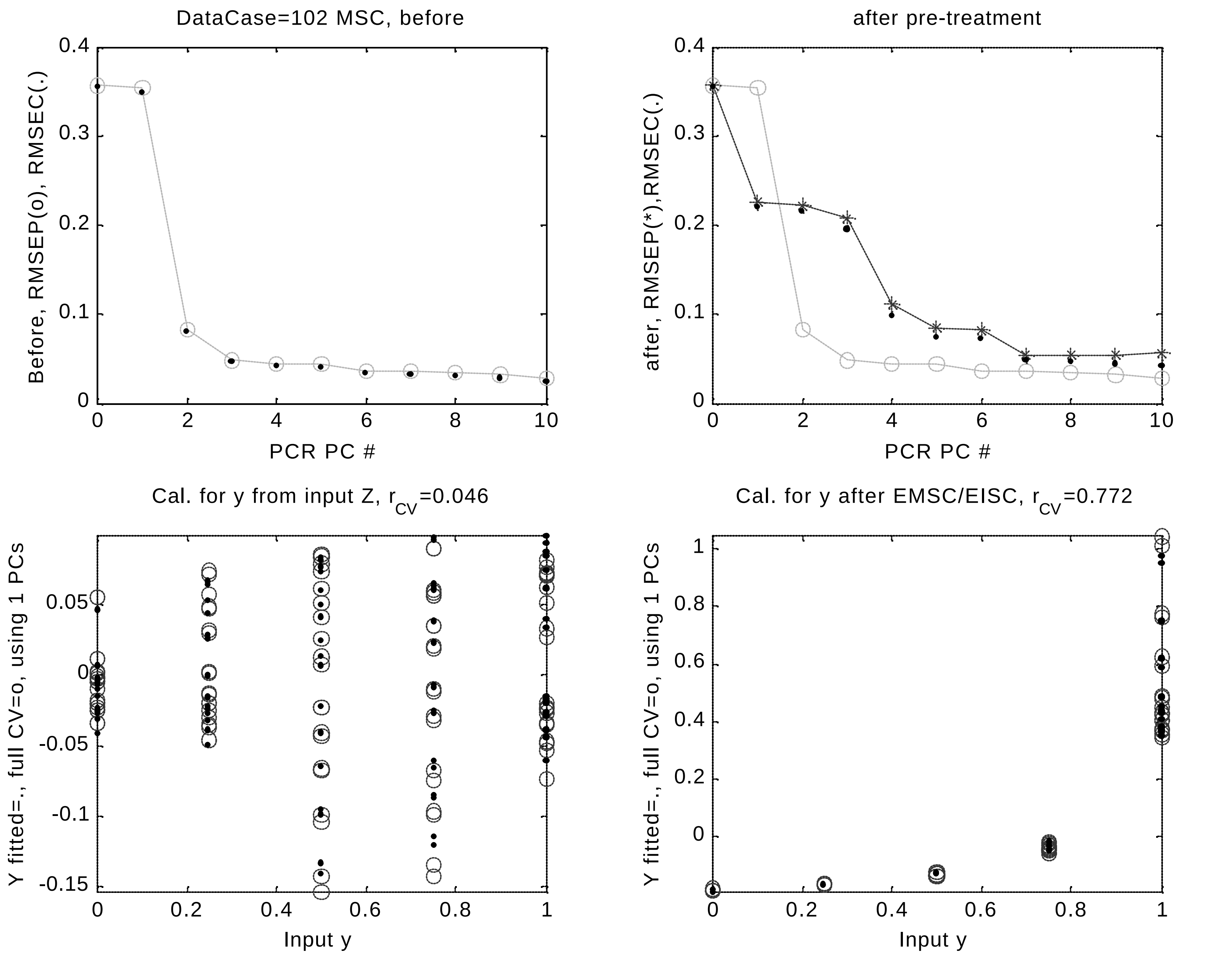
$$\mathbf{z}_{i,\text{corrected}} = (\mathbf{z}_i - a_i)/b_i$$



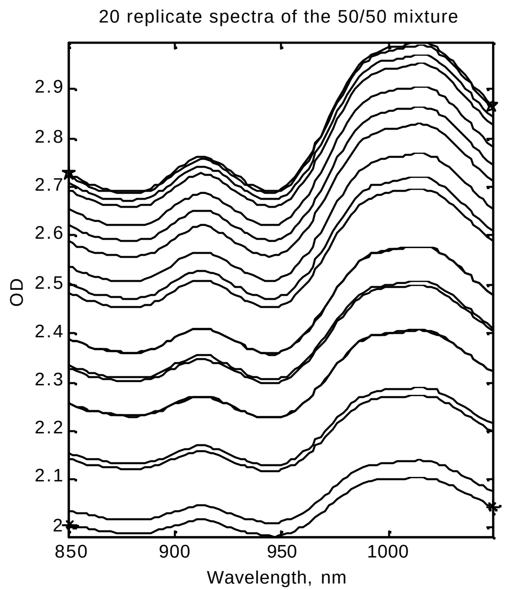
Model spectra
m´
1´
Channel #
All parameter estimates together
a_i
b_i
Obj. #


Input, EMSC$_Z$.MAT
3.5
3
2.5
2
1.5
Response
0
20
40
60
80
100
Channel #

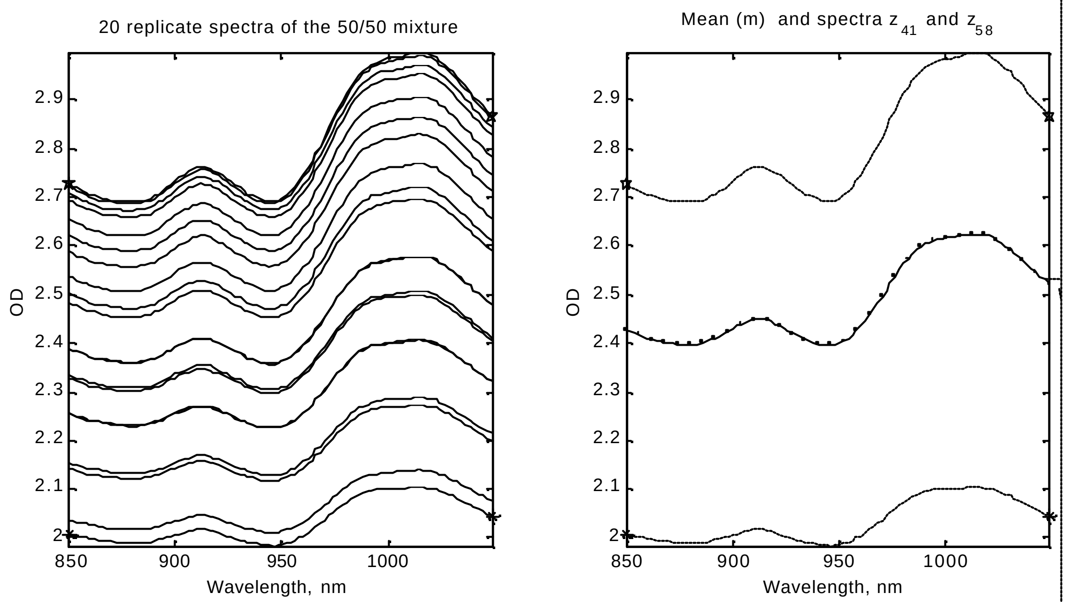

Output, DataCase=102, MSC
4
3.5
3
2.5
2
1.5
1
Response
0
20
40
60
80
100
Channel #

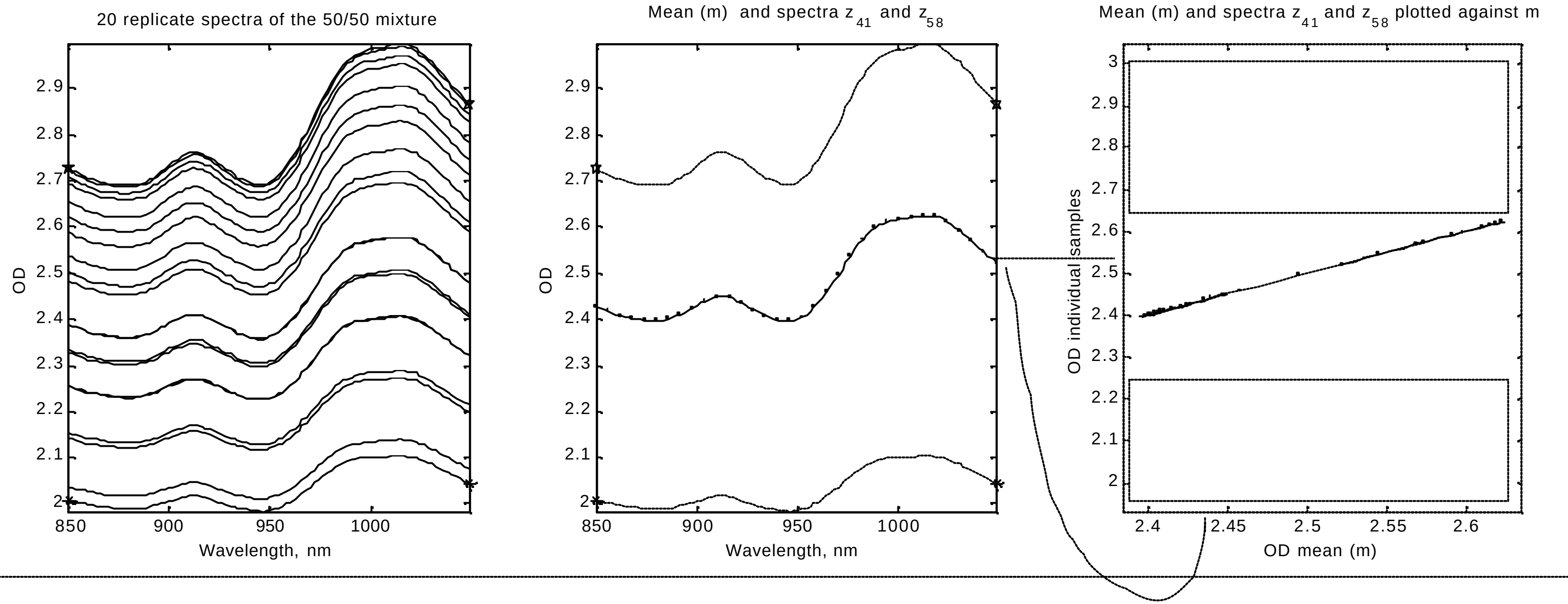


DataCase=102 MSC, before
Before, RMSEP(o), RMSEC(.)
PCR PC #
after pre-treatment
after, RMSEP(*),RMSEC(.)
PCR PC #
Cal. for y from input Z, r_{CV}=0.046
Y fitted=., full CV=o, using 1 PCs
Input y
Cal. for y after EMSC/EISC, r_{CV}=0.772
Y fitted=., full CV=o, using 1 PCs
Input y

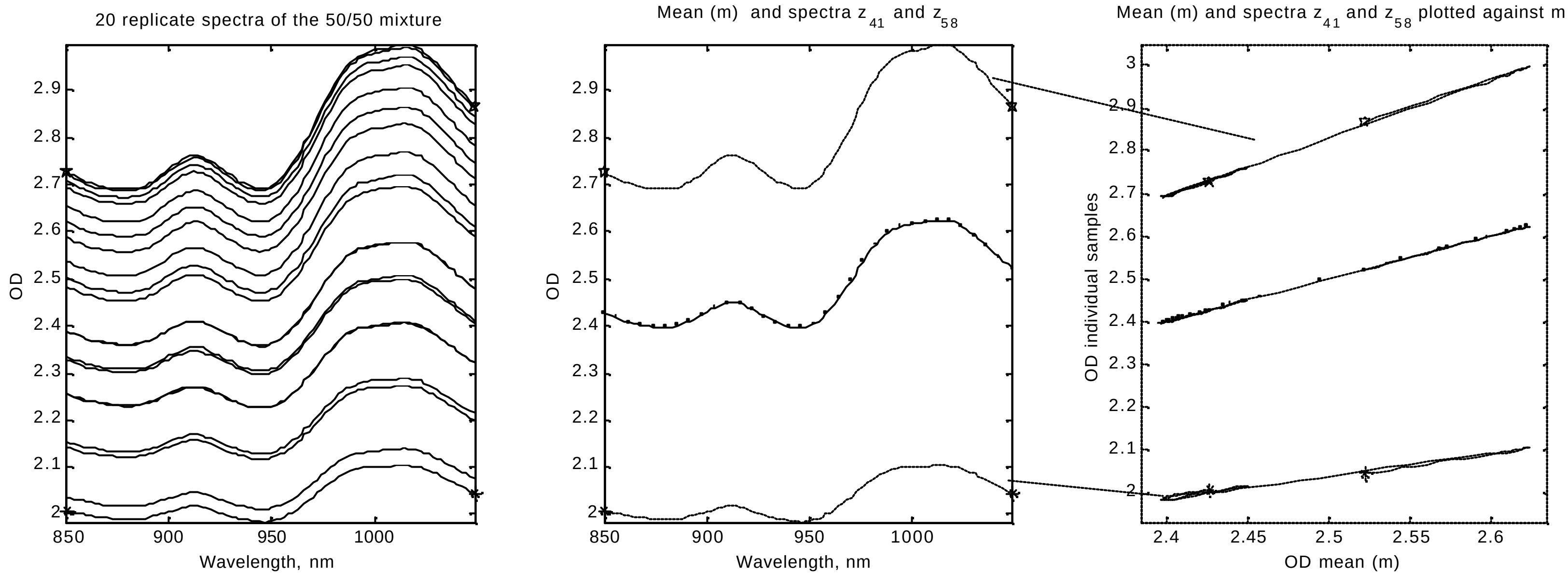


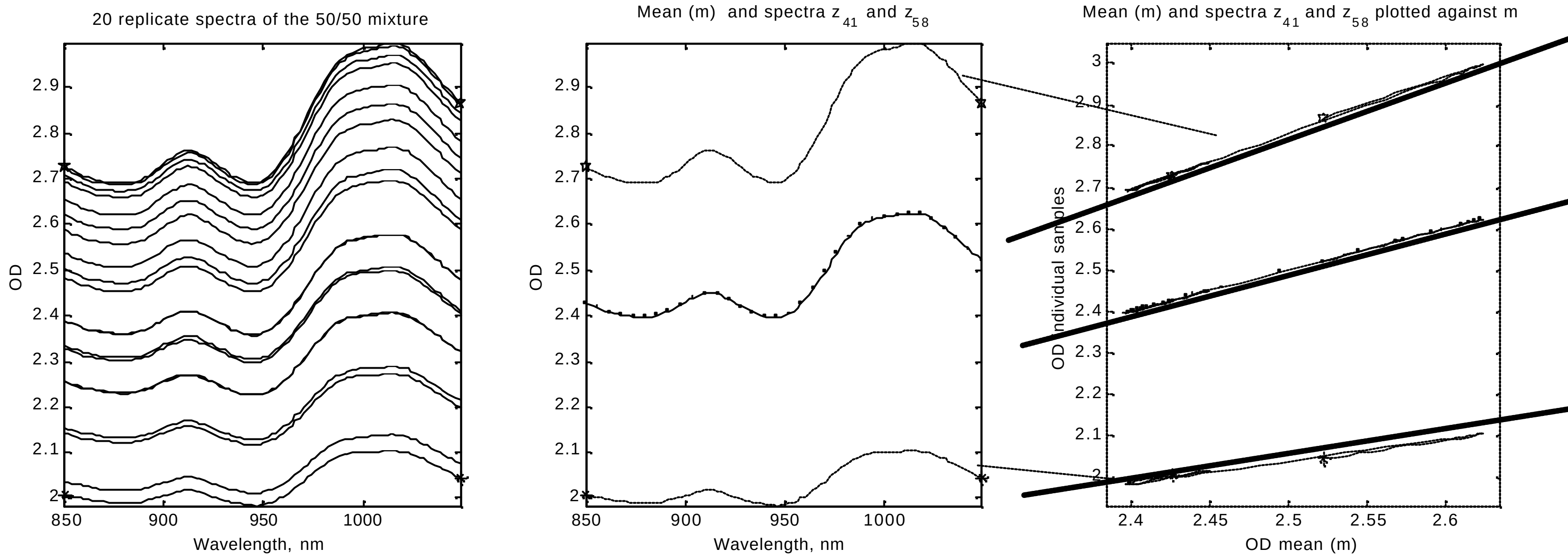
20 samples of identical chemical composition (50/50 gluten/starch), but different path length, powder packing etc)


20 replicate spectra of the 50/50 mixture
OD
Wavelength, nm
Mean (m) and spectra z_{41} and z_{58}
OD
Wavelength, nm

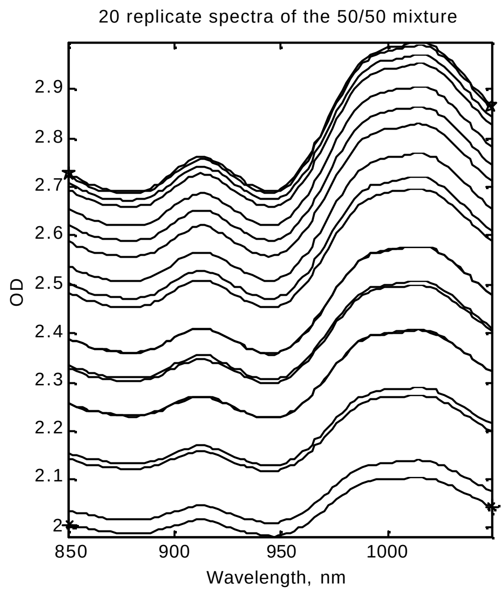


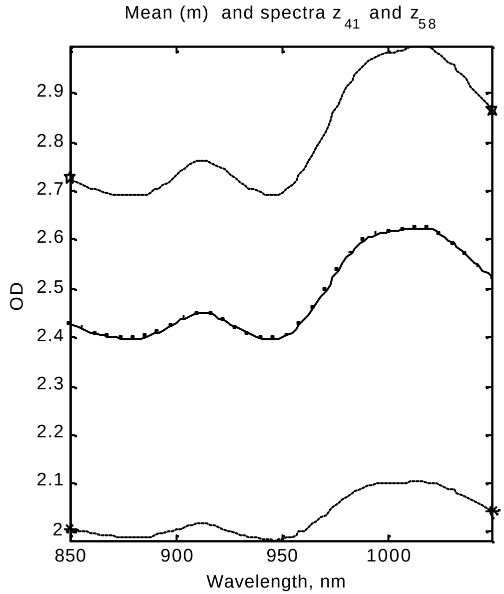
20 replicate spectra of the 50/50 mixture
OD
Wavelength, nm
Mean (m) and spectra z_{41} and z_{58}
OD
Wavelength, nm
Mean (m) and spectra z_{41} and z_{58} plotted against m
OD individual samples
OD mean (m)

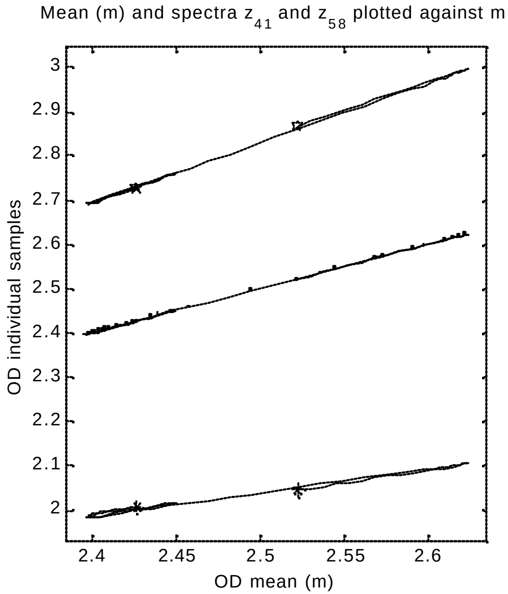

20 replicate spectra of the 50/50 mixture
Mean (m) and spectra z_{41} and z_{58}
Mean (m) and spectra z_{41} and z_{58} plotted against m
OD
Wavelength, nm
OD individual samples
OD mean (m)

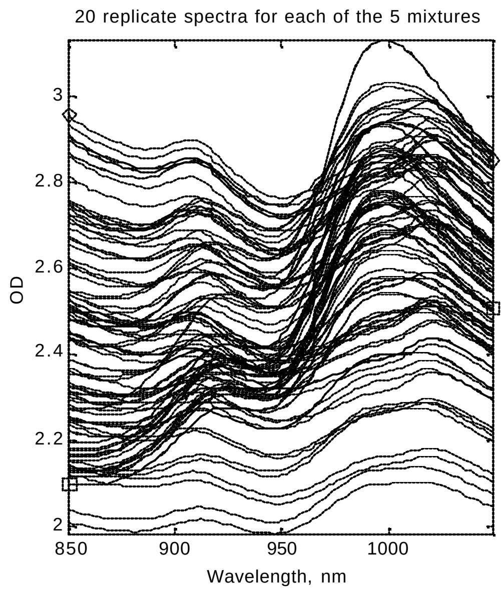


20 replicate spectra of the 50/50 mixture
OD
Wavelength, nm
Mean (m) and spectra z_{41} and z_{58}
OD
Wavelength, nm
Mean (m) and spectra z_{41} and z_{58} plotted against m
OD individual samples
OD mean (m)

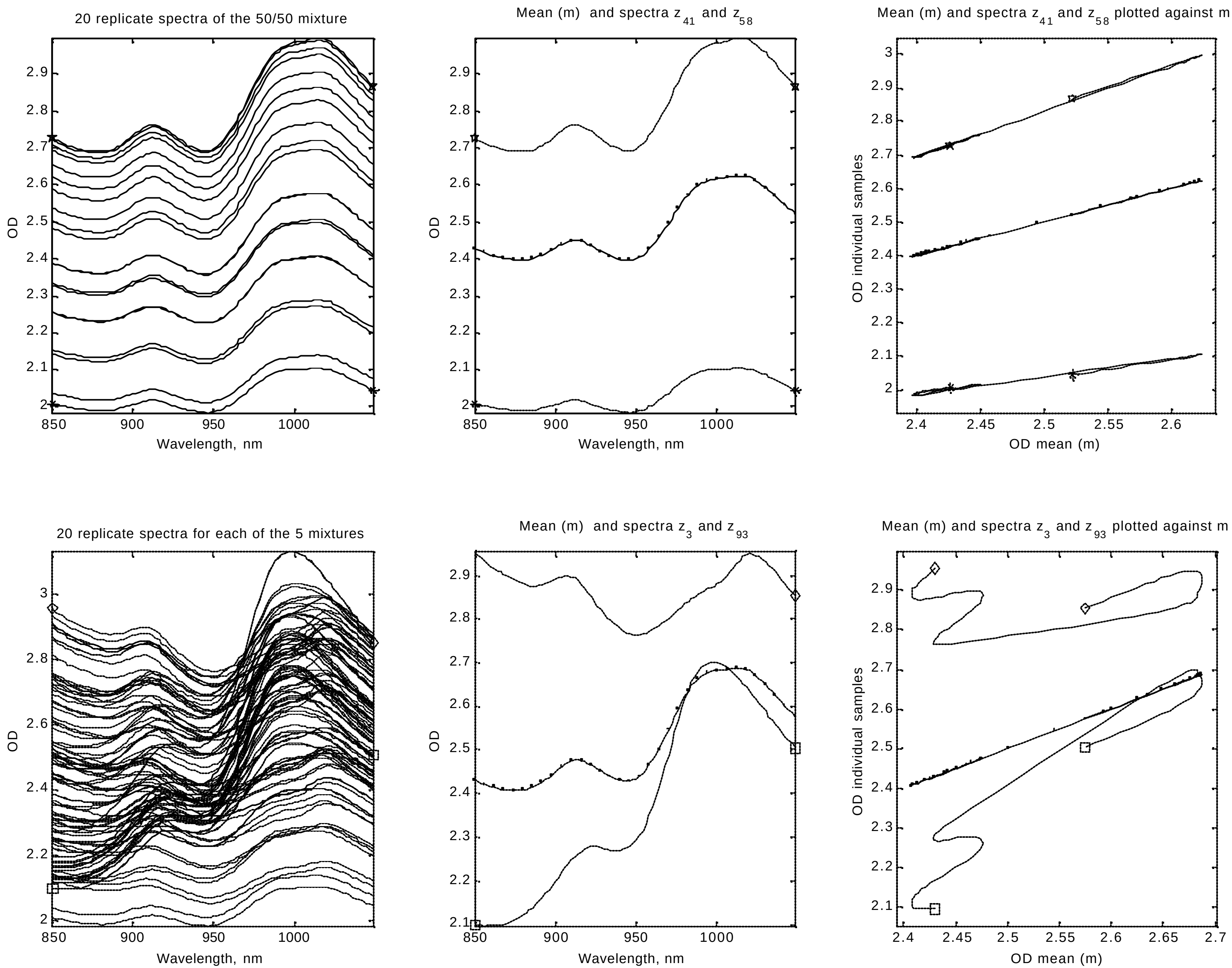


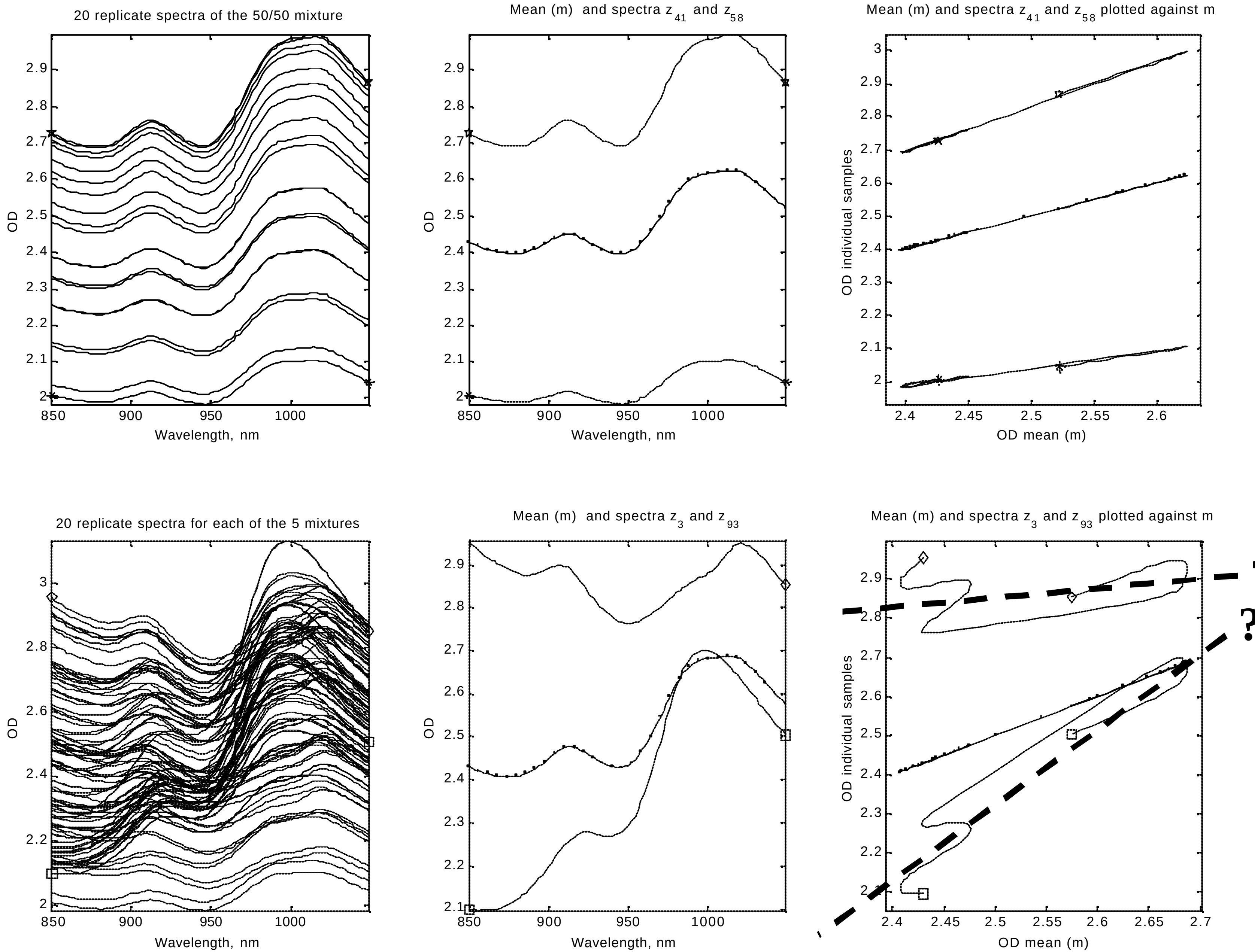






All 5 chemical mixtures


20 replicate spectra of the 50/50 mixture
OD
Wavelength, nm
Mean (m) and spectra z_{41} and z_{58}
OD
Wavelength, nm
Mean (m) and spectra z_{41} and z_{58} plotted against m
OD individual samples
OD mean (m)
20 replicate spectra for each of the 5 mixtures
OD
Wavelength, nm
Mean (m) and spectra z_3 and z_{93}
OD
Wavelength, nm
Mean (m) and spectra z_3 and z_{93} plotted against m
OD individual samples
OD mean (m)


20 replicate spectra of the 50/50 mixture
Mean (m) and spectra z41 and z58
Mean (m) and spectra z41 and z58 plotted against m
OD
Wavelength, nm
OD individual samples
OD mean (m)
20 replicate spectra for each of the 5 mixtures
Mean (m) and spectra z3 and z93
Mean (m) and spectra z3 and z93 plotted against m
?

EMSC Default: Physical model, but no chemical model parameters

Ideal chemical model: Like MSC:

$\mathbf{z}_{\mathrm{i,chem}} = \mathbf{m}' + {}_{\mathrm{d_i}}{}'$

The EMSC model of physical interferants:

$\mathbf{z}_{\mathrm{i}} \approx a_{\mathrm{i}}\mathbf{1}' + b_{\mathrm{i}}\mathbf{z}_{\mathrm{i,chem}} + d_{\mathrm{i}}\,\mathrm{l} + e_{\mathrm{i}}\,\mathrm{l}^{2}$

EMSC: $\mathbf{z}_{\mathrm{i}} = a_{\mathrm{i}}\mathbf{1}' + b_{\mathrm{i}}\mathbf{m}' + h_{\mathrm{i}}\mathbf{k}' + d_{\mathrm{i}}\,\mathrm{l} + e_{\mathrm{i}}\,\mathrm{l}^{2} + \mathrm{e}_{\mathrm{i}}$

EMSC correction: $\mathbf{z}_{\mathrm{i,corrected}} = (\mathbf{z}_{\mathrm{i}} - a_{\mathrm{i}} - d_{\mathrm{i}}\,\mathrm{l} - e_{\mathrm{i}}\,\mathrm{l}^{2})/b_{\mathrm{i}}$

EMSC Toolbox for Matlab:

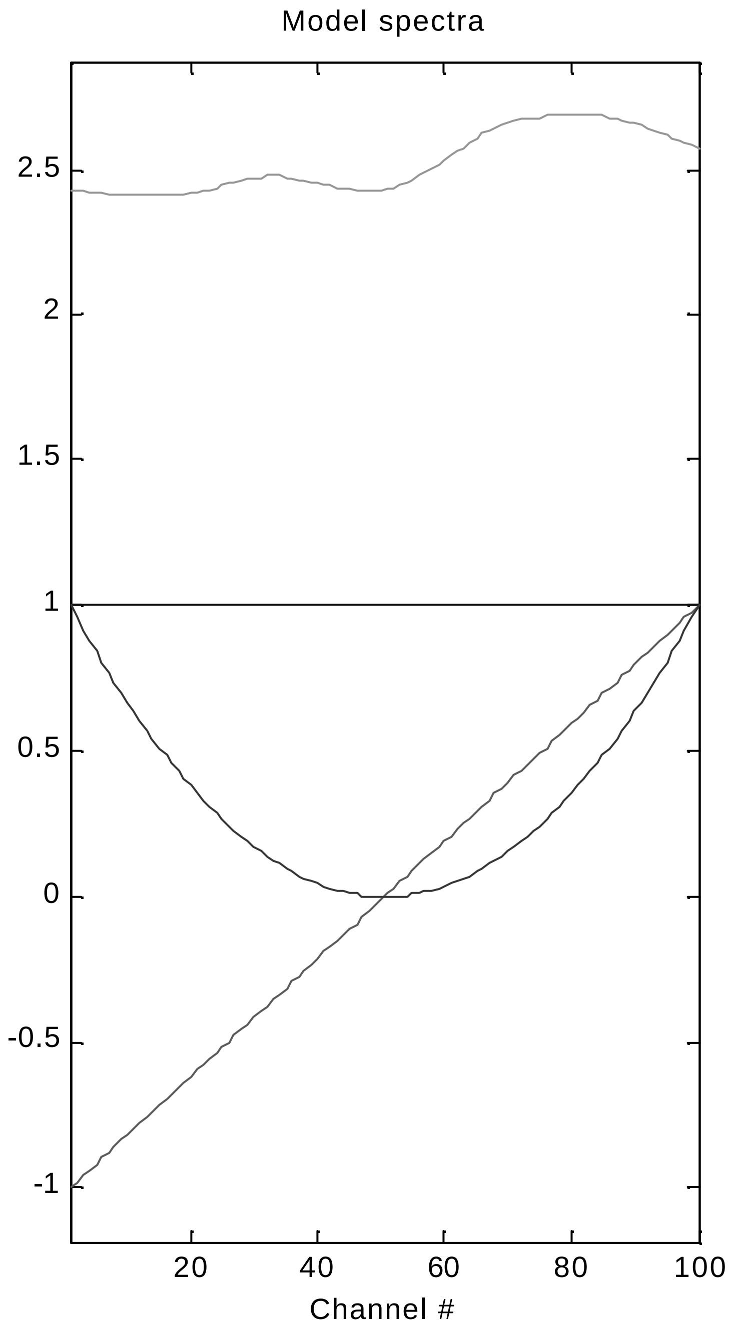
http://www.models.kvl.dk/source/EMSCtoolbox/index.asp

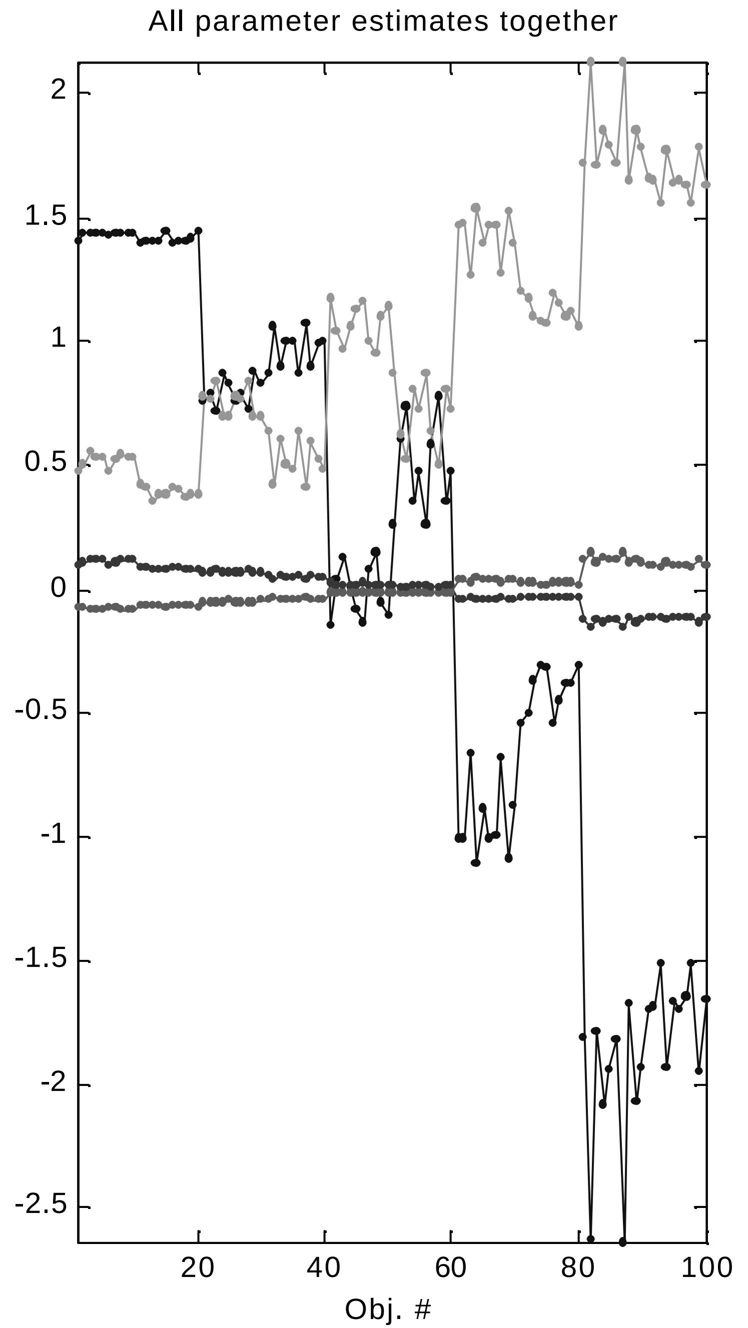
Pre-defined default EMSC in the software:

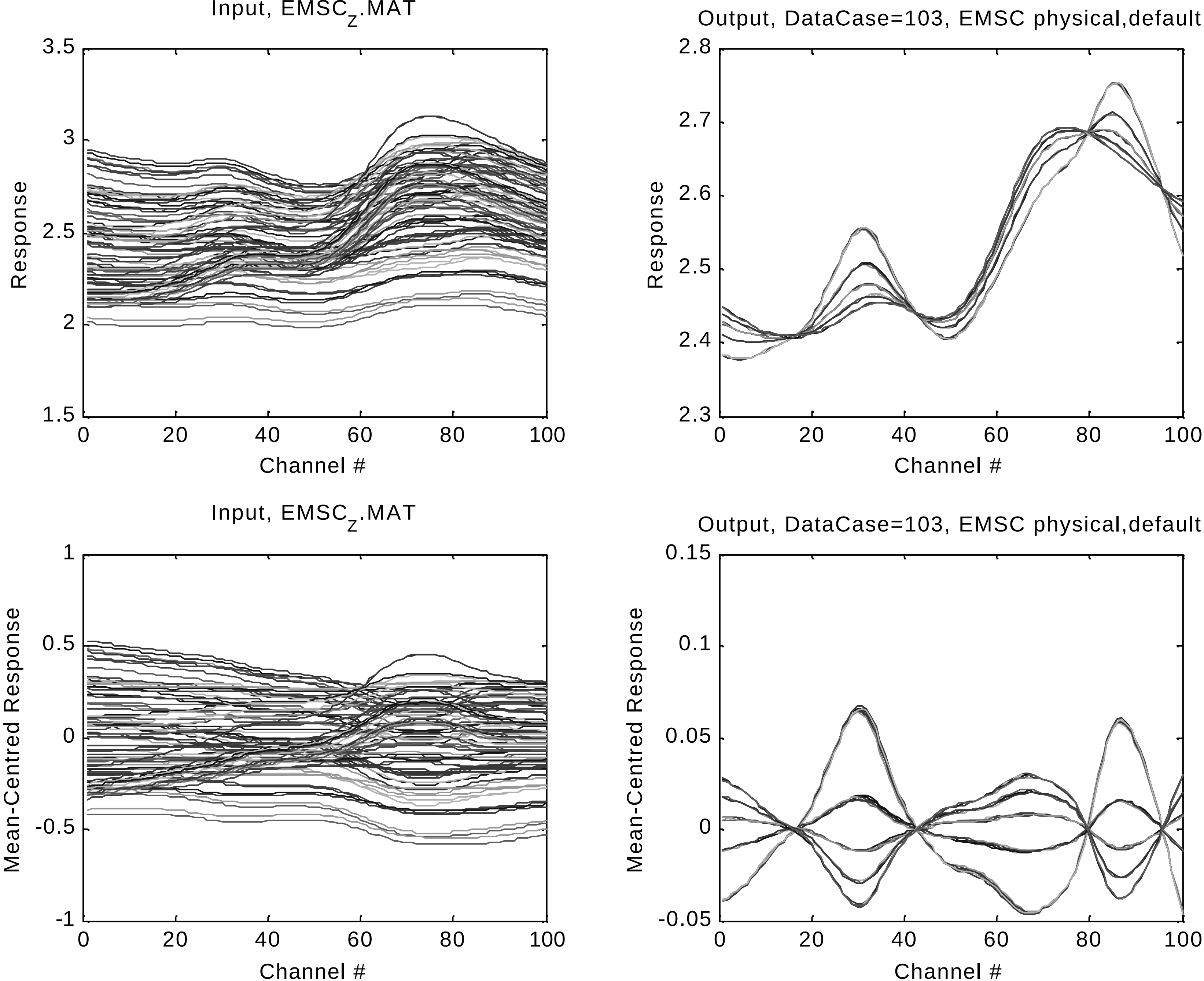
```
…
elseif DataCase==(103)
       DataCaseName='EMSC physical,default'
elseif …
```

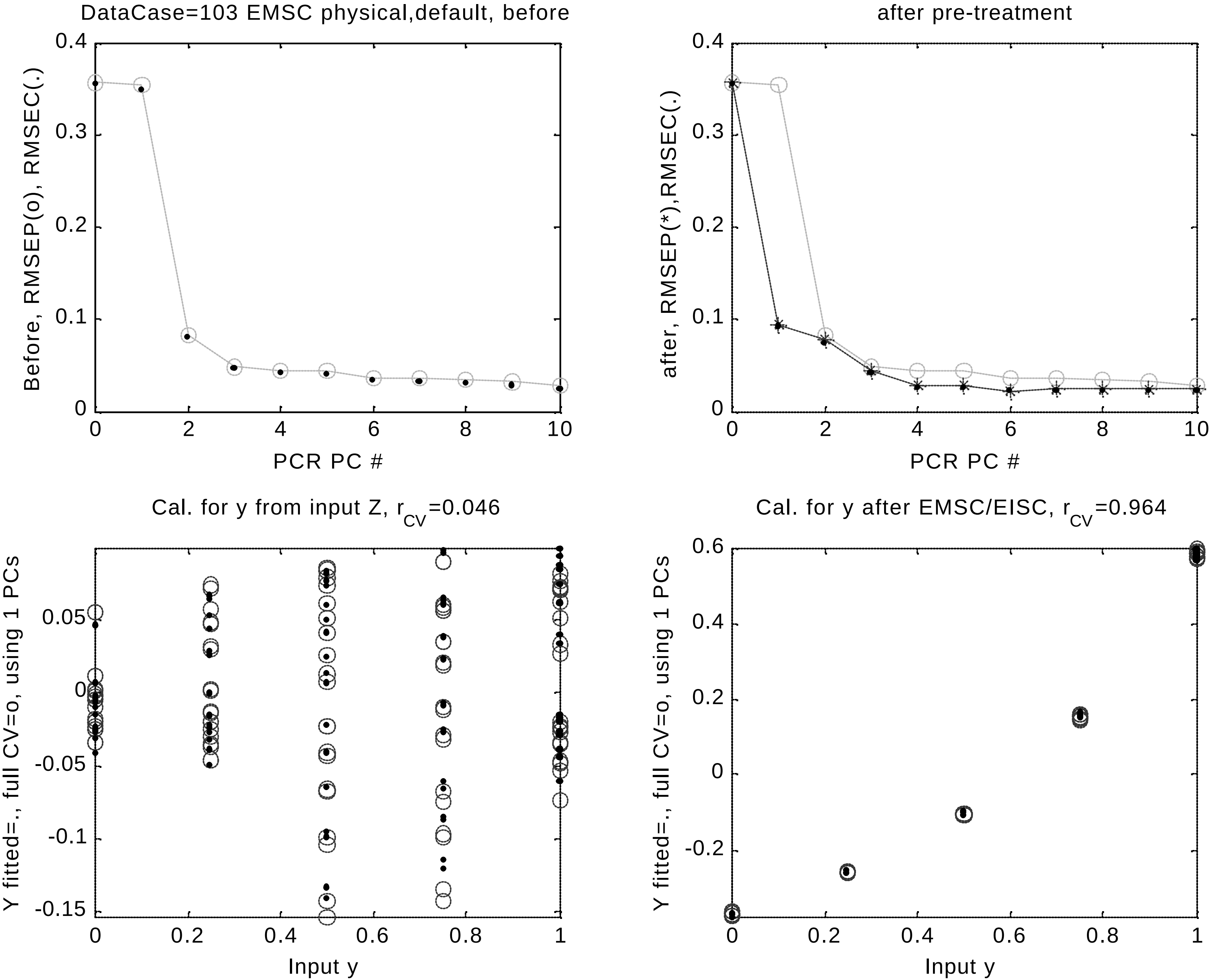
Running the program:

DataCase=? 103


Model spectra
2.5
2
1.5
1
0.5
0
-0.5
-1
20
40
60
80
100
Channel #


All parameter estimates together
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
20
40
60
80
100
Obj. #


Input, EMSC$_Z$.MAT
Output, DataCase=103, EMSC physical,default
Response
Channel #
Input, EMSC$_Z$.MAT
Output, DataCase=103, EMSC physical,default
Mean-Centred Response
Channel #



DataCase=103 EMSC physical,default, before
Before, RMSEP(o), RMSEC(.)
PCR PC #
after pre-treatment
after, RMSEP(*),RMSEC(.)
PCR PC #
Cal. for y from input Z, r_{CV}=0.046
Y fitted=., full CV=o, using 1 PCs
Input y
Cal. for y after EMSC/EISC, r_{CV}=0.964
Y fitted=., full CV=o, using 1 PCs
Input y

EMSC: Default physical model, no chemical model parameters, but **optimize the Reference spectrum m**

Instead of just using the mean spectrum:

$$\mathbf{m} = \bar{\mathbf{z}}'$$

Make bilinear model around the mean spectrum:

$$\mathbf{m} = t_{m,0}\bar{\mathbf{z}}' + t_{m,1}\mathbf{p}_1' + t_{m,2}\mathbf{p}_2' + t_{m,3}\mathbf{p}_3'$$

Estimate unknown parameters t_0, t_1, t_2, t_3 by SIMPLEX Opt., minimizing RMSEP(Y).

(estimated leverage-corrected PCR,1 PC)

EMSC Toolbox for Matlab:

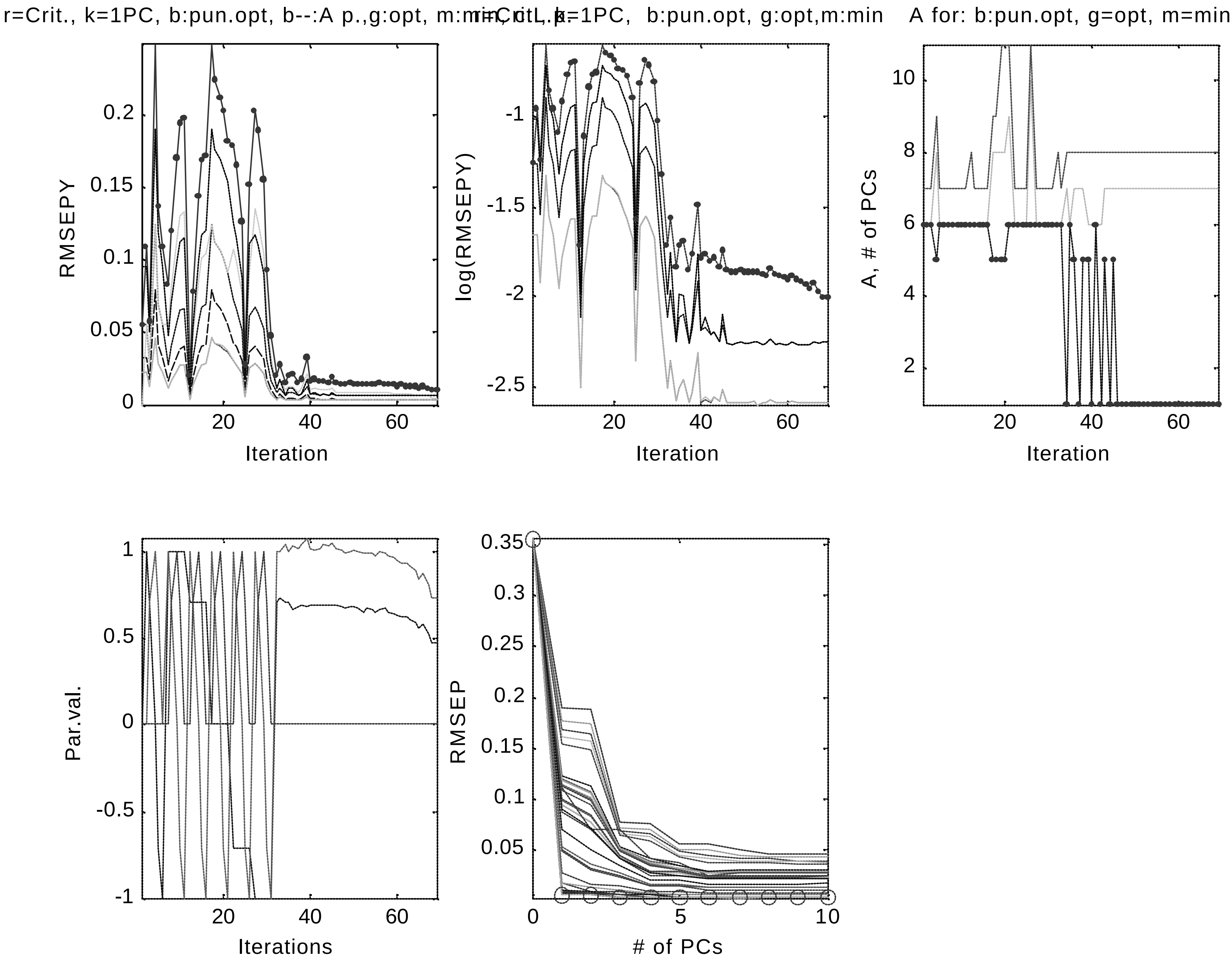
http://www.models.kvl.dk/source/EMSCtoolbox/index.asp

Pre-defined default EMSC in the software:

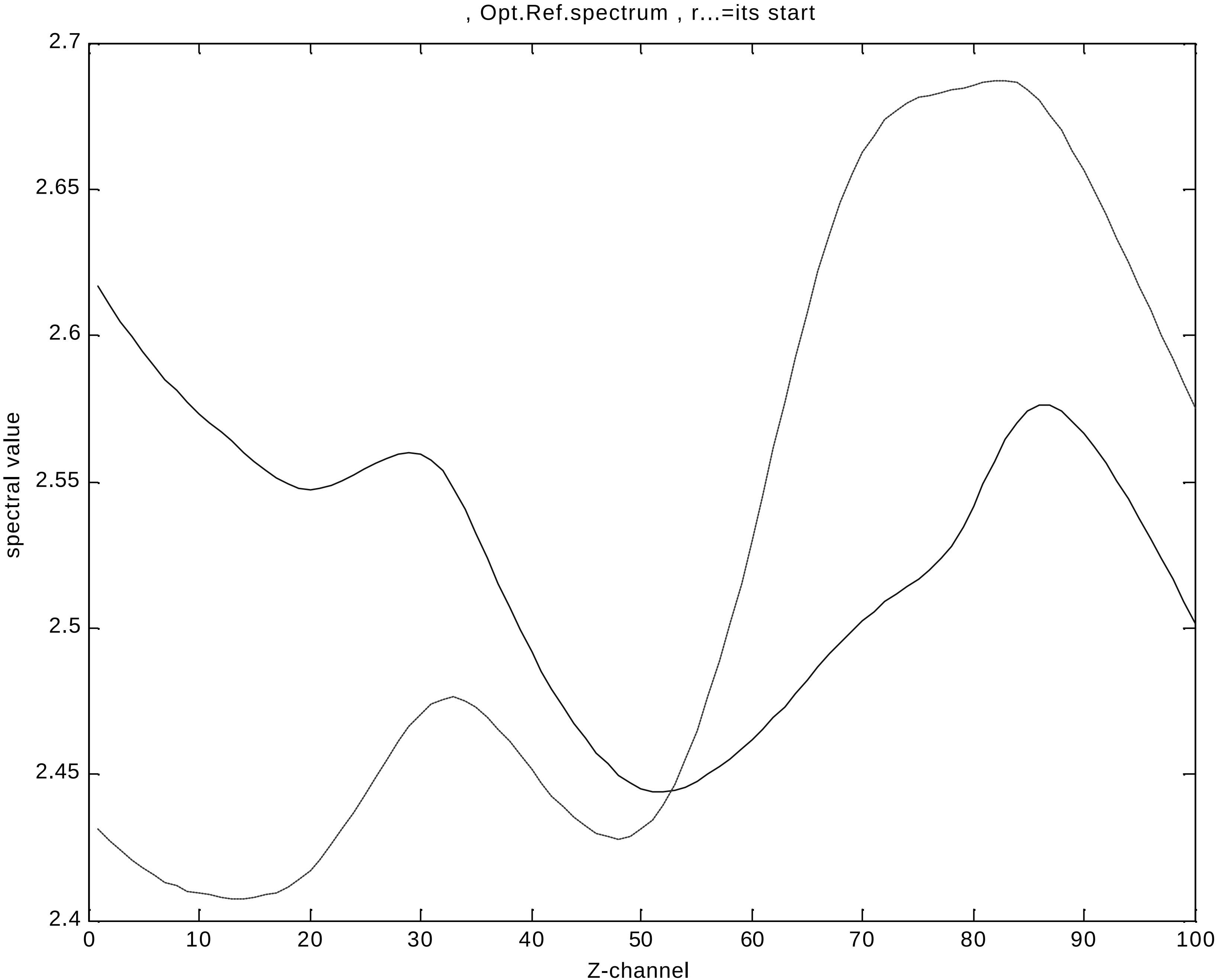
```
…
elseif DataCase==(103)
    DataCaseName=' EMSC, opt. the Ref.spectrum, starting from the mean spectrum'
    OptPar=1
    ASearchDim=3
elseif …
```

Running the program:

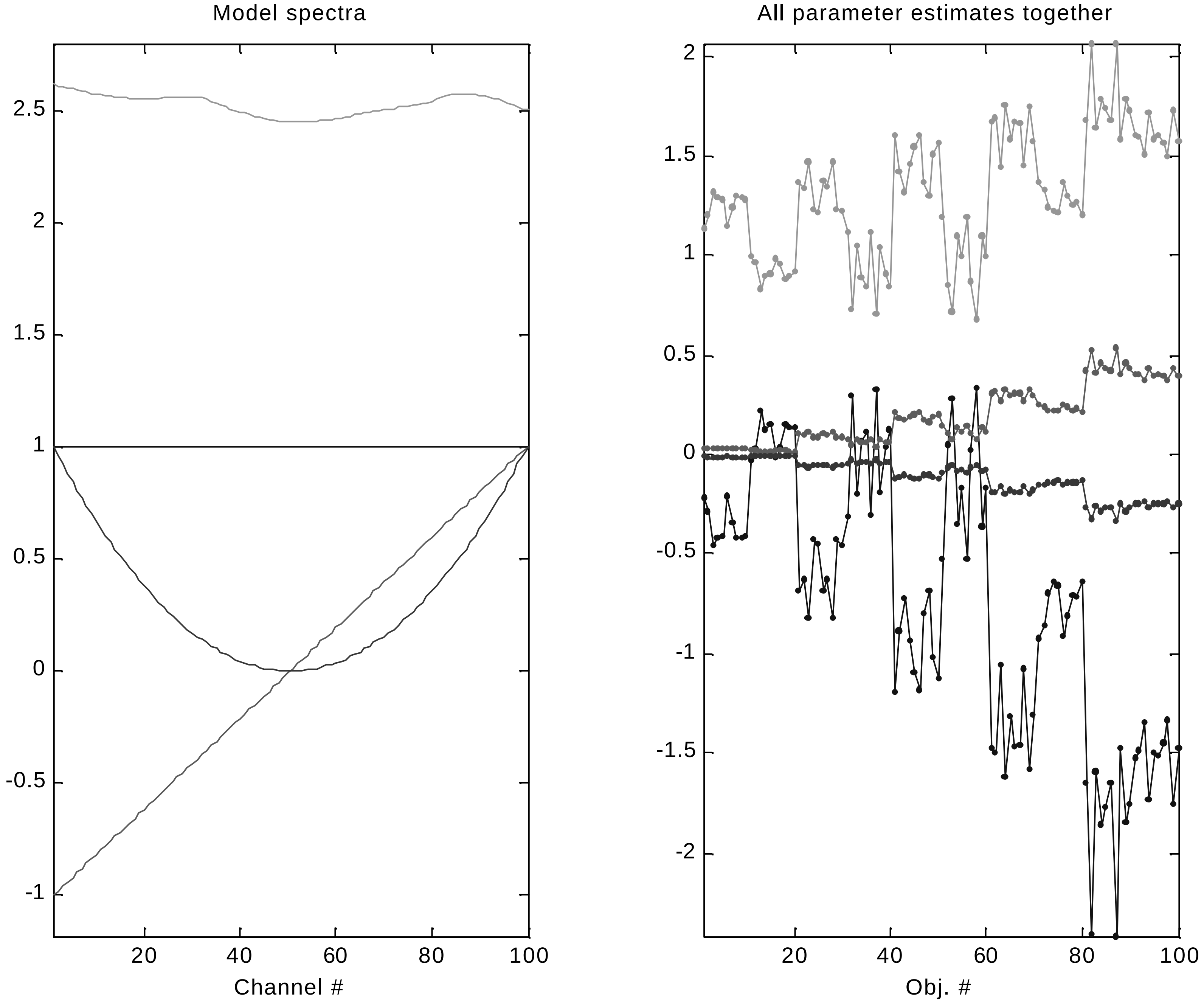
DataCase=? 151



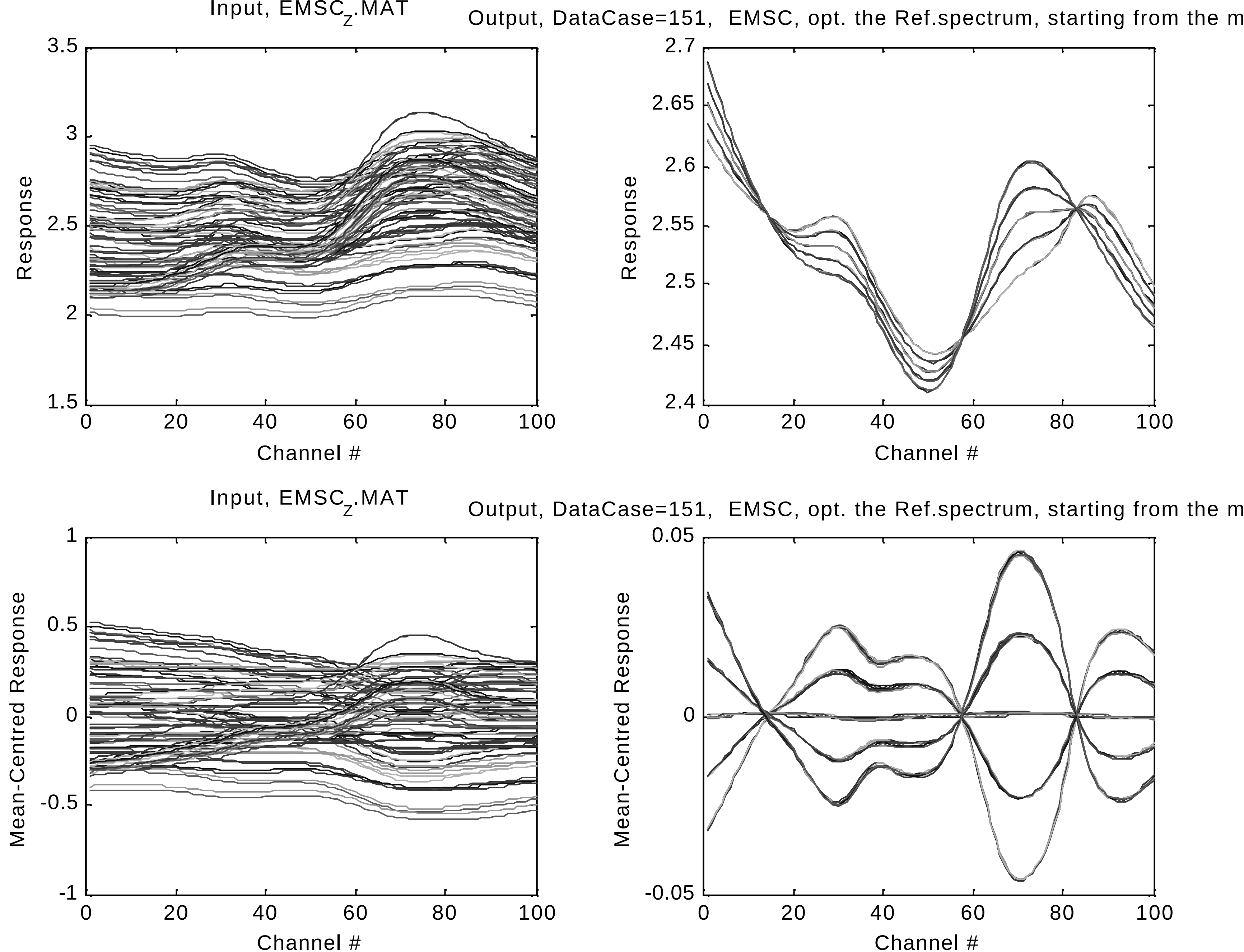
r=Crit., k=1PC, b:pun.opt, b--:A p.,g:opt, m:m
r=CritL, k=1PC, b:pun.opt, g:opt,m:min
A for: b:pun.opt, g=opt, m=min
RMSEPY
Iteration
log(RMSEPY)
Iteration
A, # of PCs
Iteration
Par.val.
Iterations
RMSEP
# of PCs



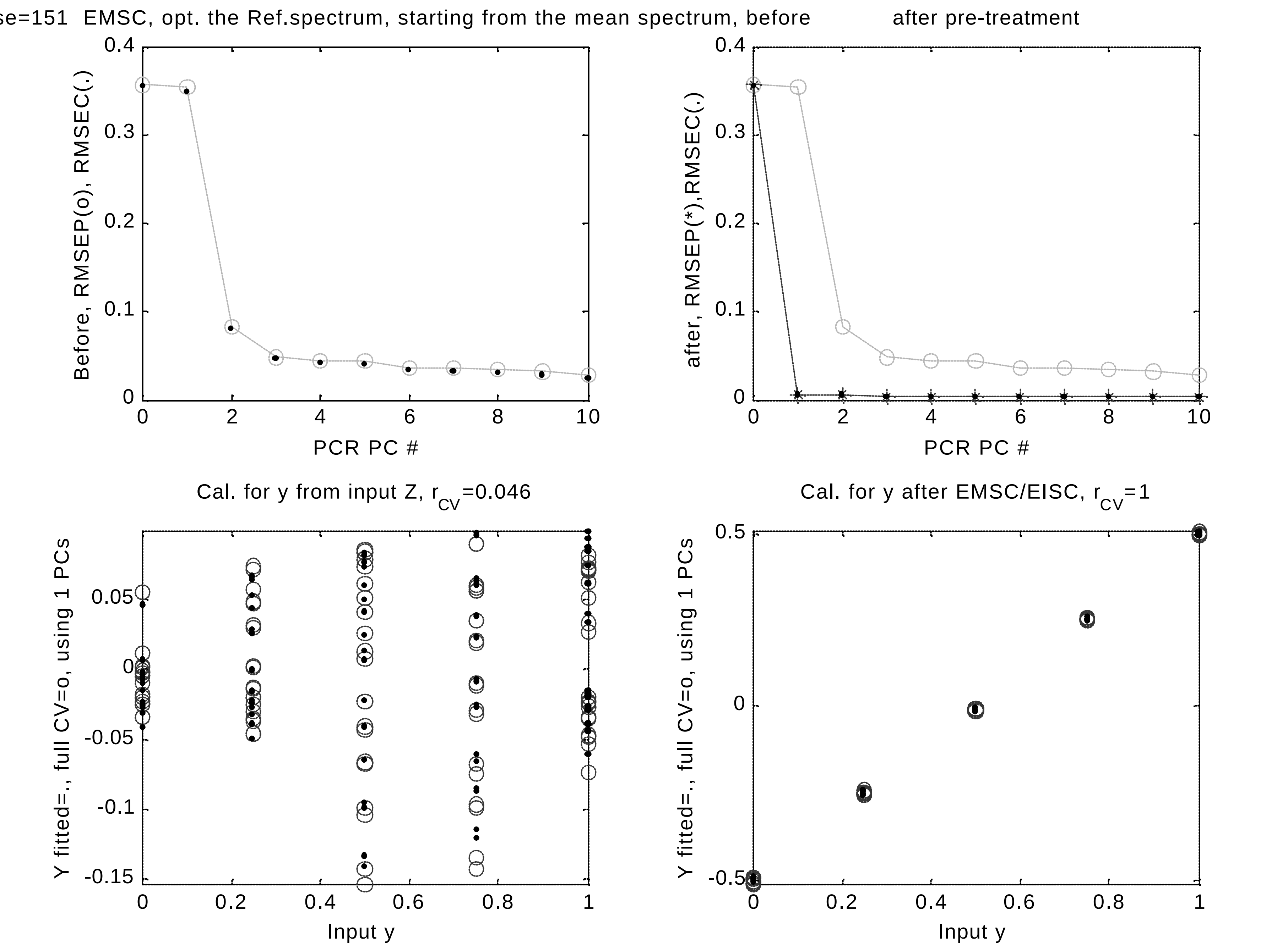
, Opt.Ref.spectrum , r...=its start
spectral value
Z-channel
2.7
2.65
2.6
2.55
2.5
2.45
2.4
0
10
20
30
40
50
60
70
80
90
100



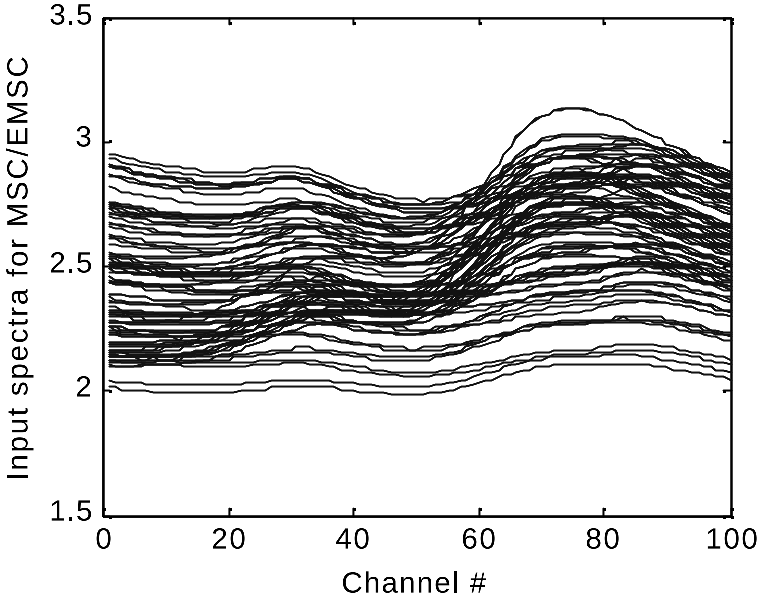
Model spectra
2.5
2
1.5
1
0.5
0
-0.5
-1
20
40
60
80
100
Channel #
All parameter estimates together
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
20
40
60
80
100
Obj. #

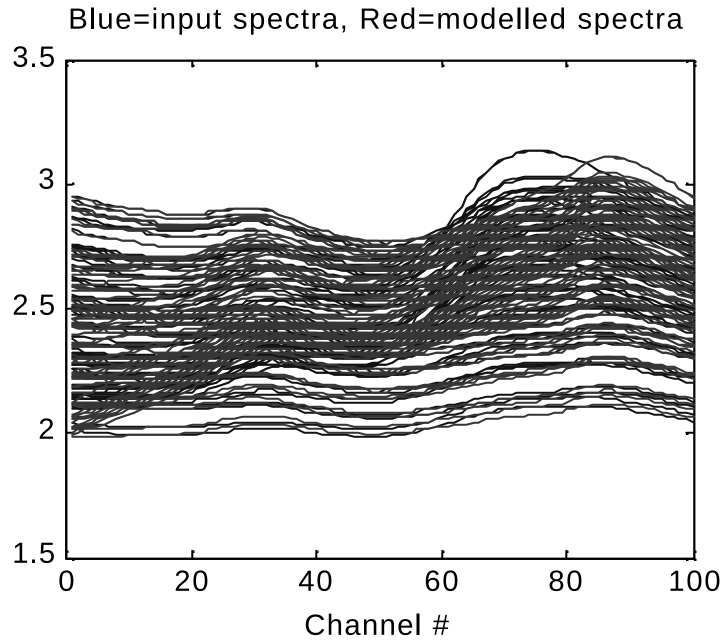


Input, $EMSC_Z$.MAT
Output, DataCase=151, EMSC, opt. the Ref.spectrum, starting from the m
Response
Channel #
Input, $EMSC_Z$.MAT
Output, DataCase=151, EMSC, opt. the Ref.spectrum, starting from the m
Mean-Centred Response
Channel #



se=151 EMSC, opt. the Ref.spectrum, starting from the mean spectrum, before
after pre-treatment
Before, RMSEP(o), RMSEC(.)
PCR PC #
after, RMSEP(*),RMSEC(.)
PCR PC #
Cal. for y from input Z, r$_{CV}$=0.046
Y fitted=., full CV=o, using 1 PCs
Input y
Cal. for y after EMSC/EISC, r$_{CV}$=1
Y fitted=., full CV=o, using 1 PCs
Input y

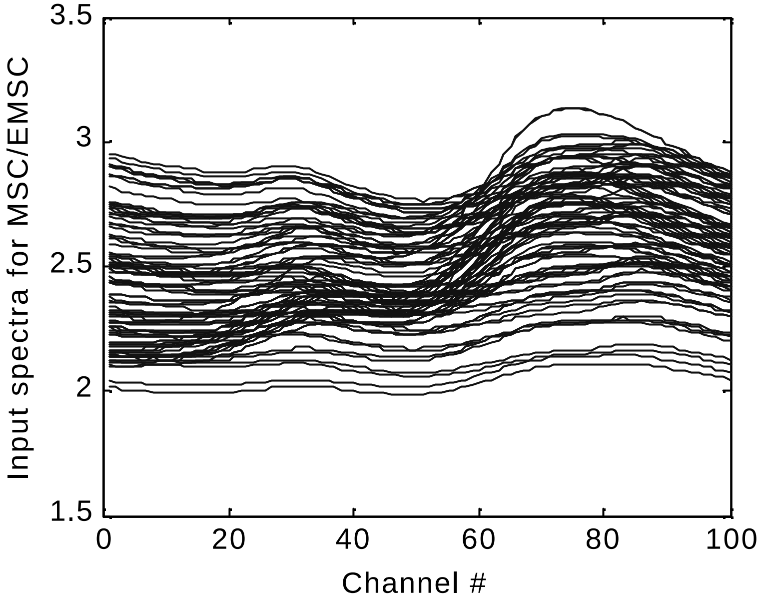


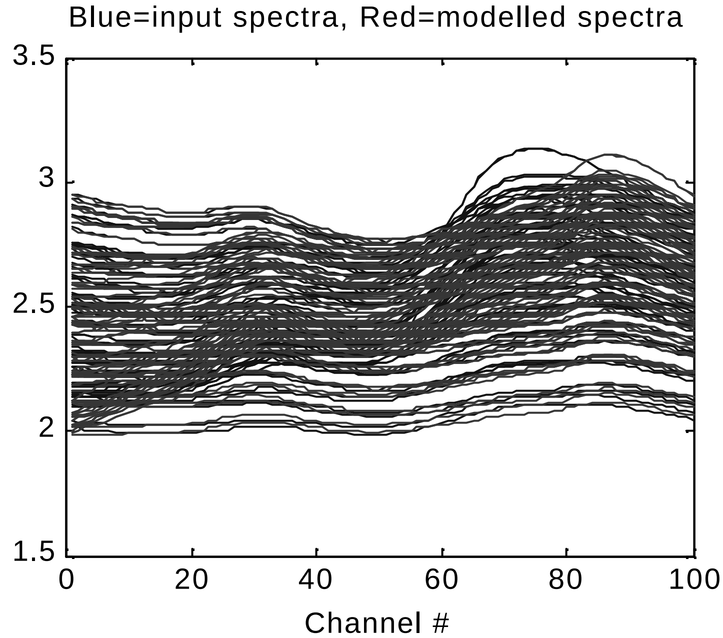


Input data: $\mathbf{z}_i$

EMSC reconstruction :

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + h_i\mathbf{k}' + d_i\lambda + e_i\lambda^2$$

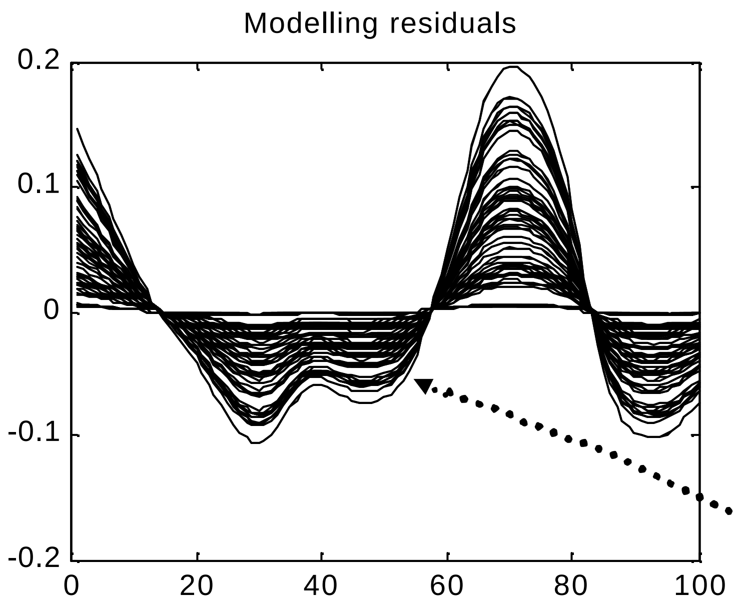




Input data: $\mathbf{z}_i$

EMSC reconstruction :

$$\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + h_i\mathbf{k}' + d_i l + e_i l^2$$



Large un-modelled residuals e_i in model

$$\mathbf{z}_i = \mathbf{z}_i + e_i$$

EMSC Default physical model,

+ a known "good" (analyte difference spectrum)

+ a known "bad" spectrum (water)

Extended ideal chemical model:

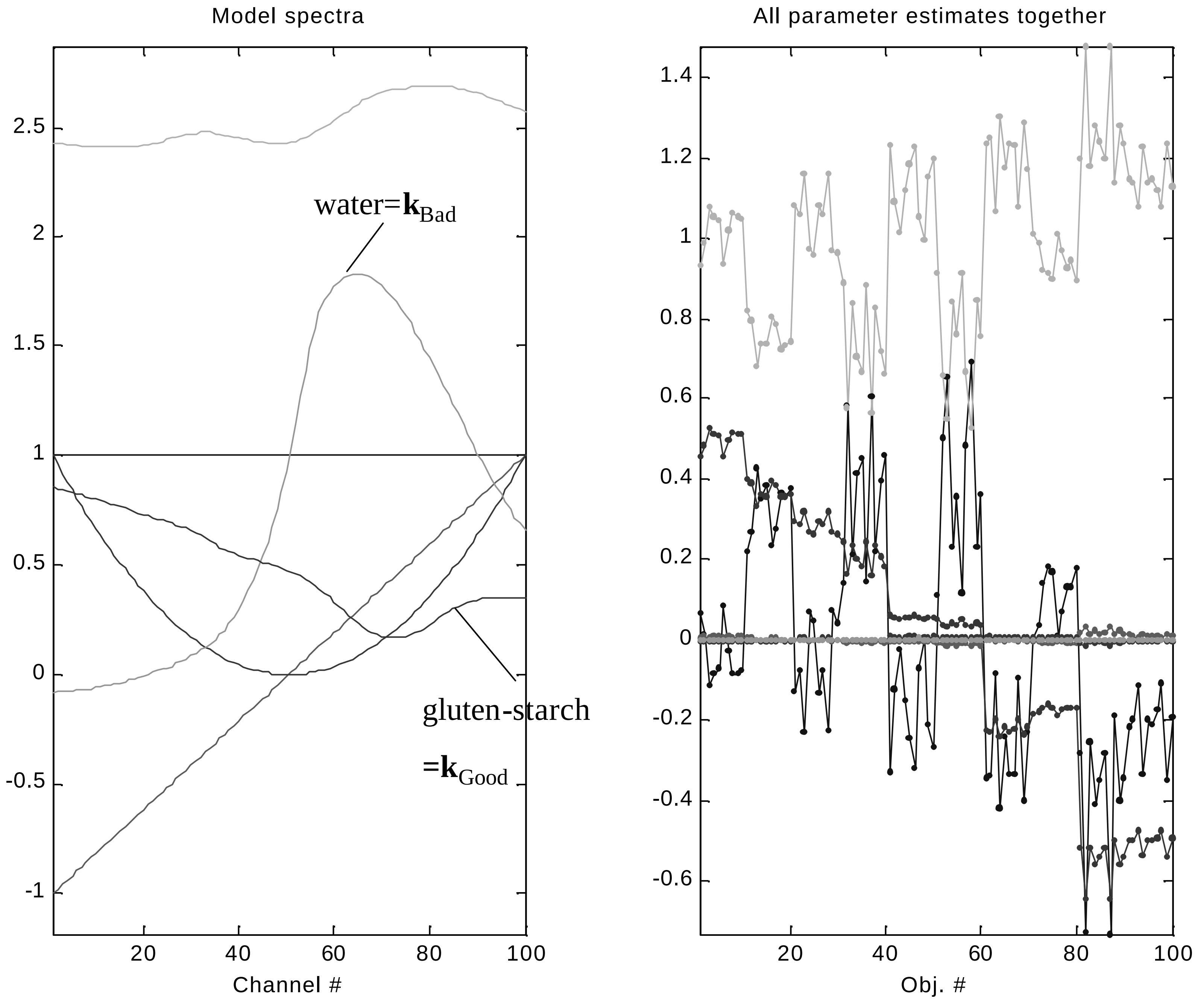
$$\mathbf{z}_{\mathrm{i,chem}} = \mathbf{m}' + c_{\mathrm{iGood}}\mathbf{K}_{\mathrm{Good}}' + c_{\mathrm{iBad}}\mathbf{K}_{\mathrm{Bad}}'$$

The EMSC model of physical interferants:

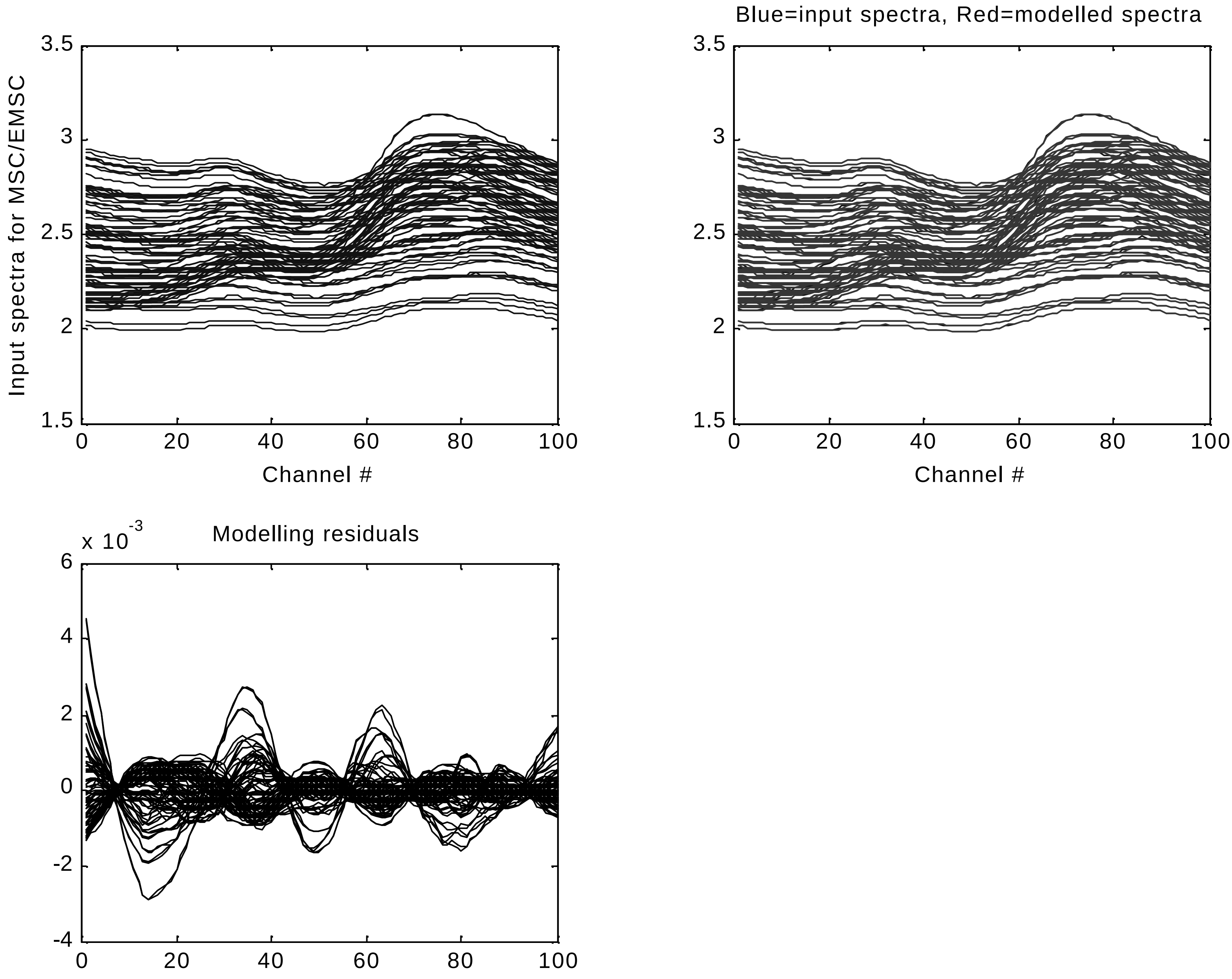
$$\mathbf{z}_{\mathrm{i}} \approx a_{\mathrm{i}}\mathbf{1}' + b_{\mathrm{i}}\mathbf{z}_{\mathrm{i,chem}} + d_{\mathrm{i}}\,\mathrm{l} + e_{\mathrm{i}}\,\mathrm{l}^{2}$$

$$\mathbf{z}_{\mathrm{i}} = a_{\mathrm{i}}\mathbf{1}' + b_{\mathrm{i}}\mathbf{m}' + h_{\mathrm{iGood}}\mathbf{k}_{\mathrm{Good}}' + h_{\mathrm{iBad}}\mathbf{k}_{\mathrm{Bad}}' + d_{\mathrm{i}}\,\mathrm{l} + e_{\mathrm{i}}\,\mathrm{l}^{2} + \mathrm{e}_{\mathrm{i}}$$

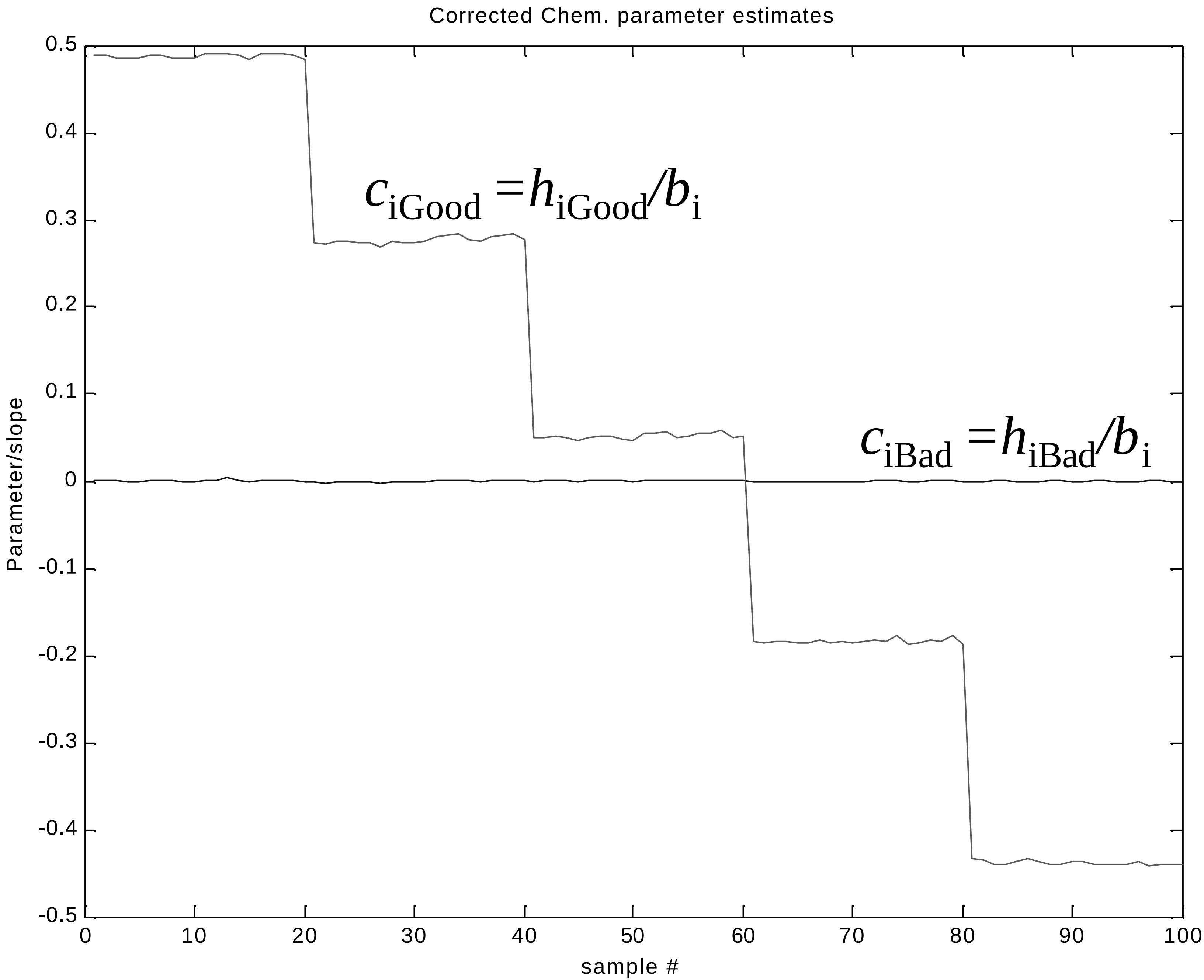
EMSC corr.: $\mathbf{z}_{\mathrm{i,corrected}} = (\mathbf{z}_{\mathrm{i}} - a_{\mathrm{i}} - h_{\mathrm{iBad}}\mathbf{k}_{\mathrm{Bad}}' - d_{\mathrm{i}}\,\mathrm{l} - e_{\mathrm{i}}\,\mathrm{l}^{2})/b_{\mathrm{i}}$

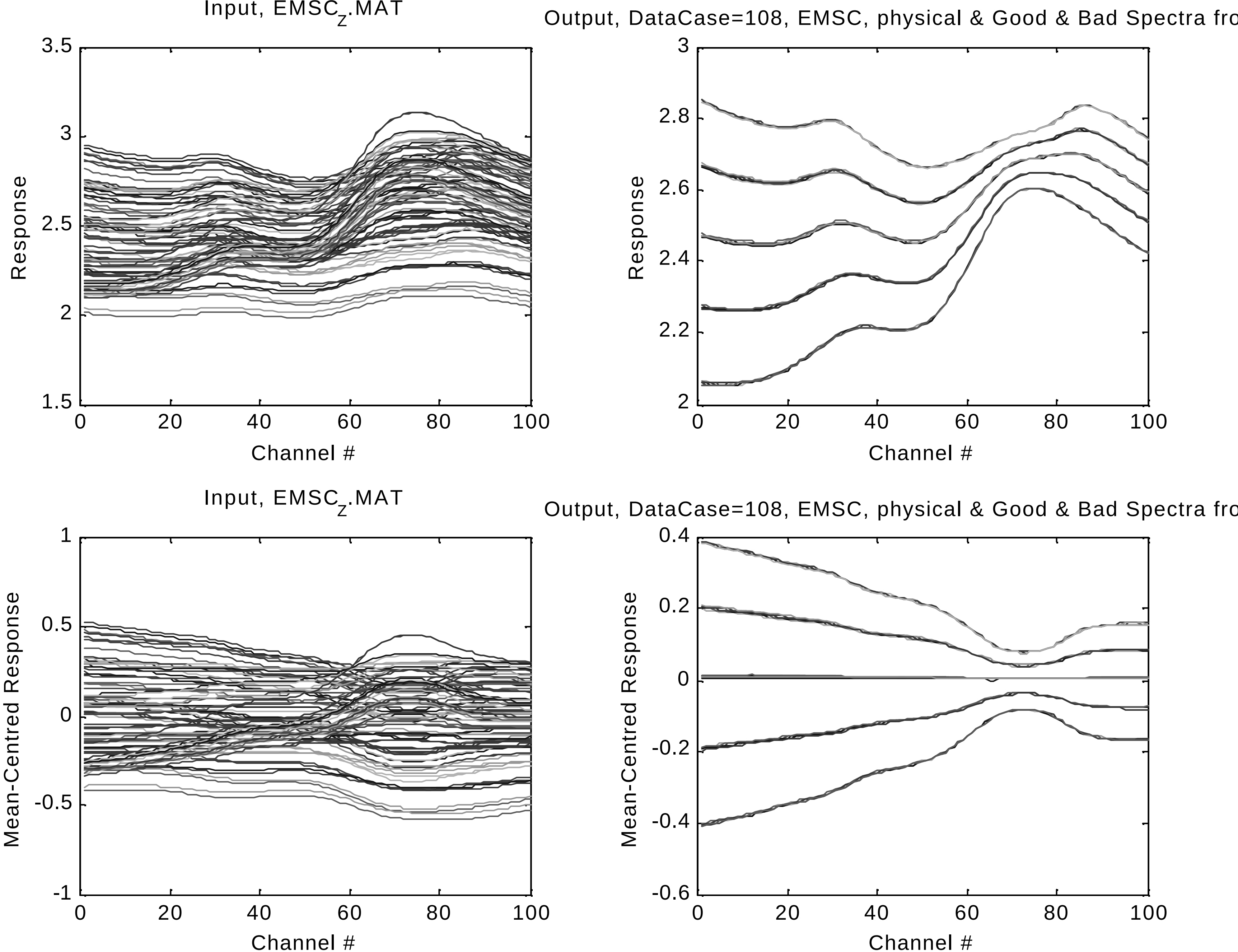


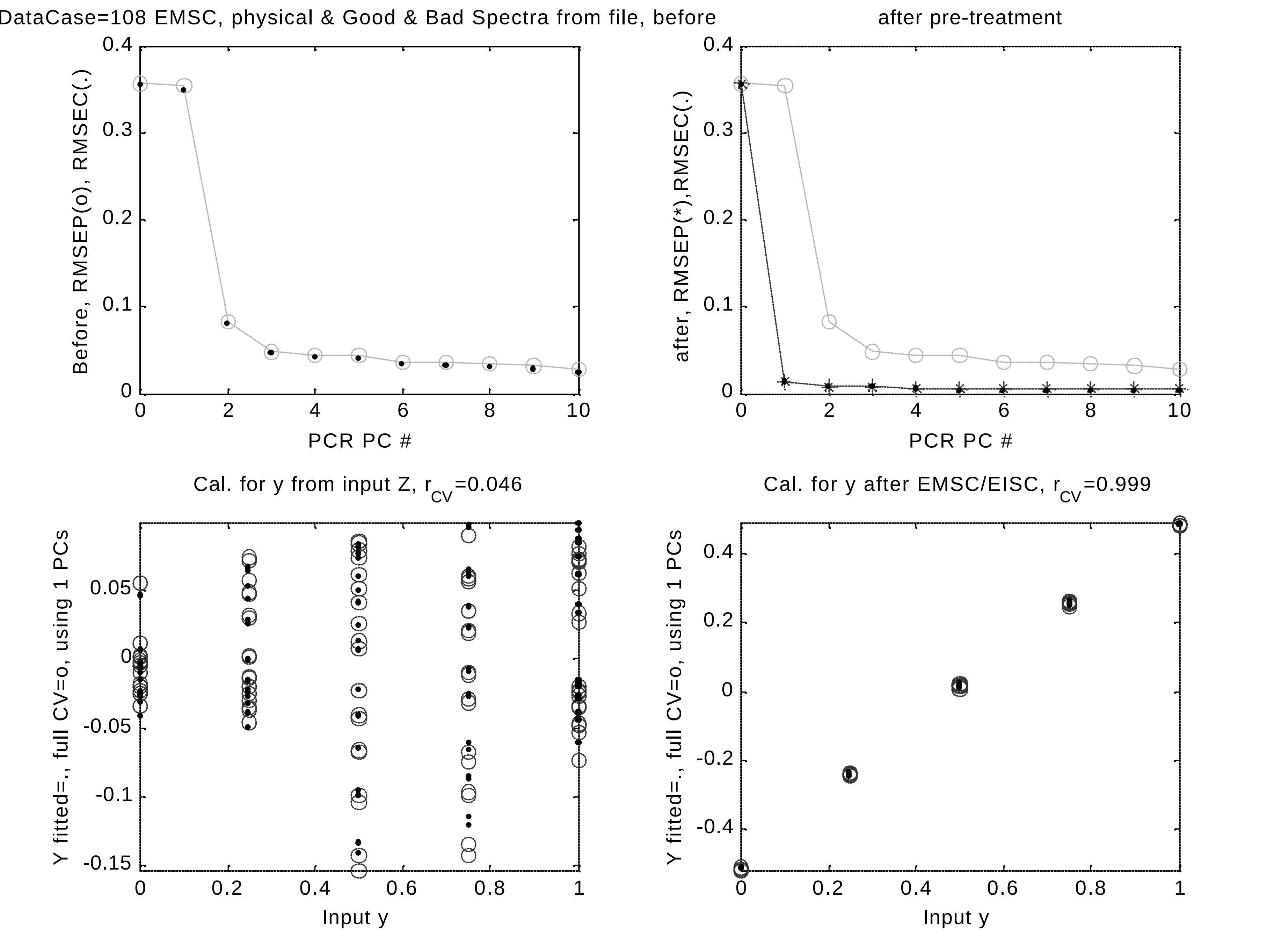
Model spectra
water=$\mathbf{k}_{Bad}$
gluten-starch
=$\mathbf{k}_{Good}$
Channel #
All parameter estimates together
Obj. #



Blue=input spectra, Red=modelled spectra
Input spectra for MSC/EMSC
Channel #
Channel #
Modelling residuals
x 10^-3
3.5
3
2.5
2
1.5
0
20
40
60
80
100
6
4
2
0
-2
-4


Corrected Chem. parameter estimates
$c_{\mathrm{iGood}} = h_{\mathrm{iGood}}/b_{\mathrm{i}}$
$c_{\mathrm{iBad}} = h_{\mathrm{iBad}}/b_{\mathrm{i}}$
Parameter/slope
sample #
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
0
10
20
30
40
50
60
70
80
90
100


Input, EMSC_Z.MAT
Output, DataCase=108, EMSC, physical & Good & Bad Spectra fro
Response
Channel #
Input, EMSC_Z.MAT
Output, DataCase=108, EMSC, physical & Good & Bad Spectra fro
Mean-Centred Response
Channel #



DataCase=108 EMSC, physical & Good & Bad Spectra from file, before
after pre-treatment
Before, RMSEP(o), RMSEC(.)
after, RMSEP(*),RMSEC(.)
PCR PC #
PCR PC #
Cal. for y from input Z, r_{CV}=0.046
Cal. for y after EMSC/EISC, r_{CV}=0.999
Y fitted=., full CV=o, using 1 PCs
Y fitted=., full CV=o, using 1 PCs
Input y
Input y

EMSC: Default physical model, no known good or bad spectra,

but **automatically find and optimize**

an unknown “good” (analyte) spectrum

EMSC: Default physical model, no known good or bad spectra,

but **automatically find and optimize**

an unknown “good” (analyte) spectrum

based on minimising RMSEP(Y)

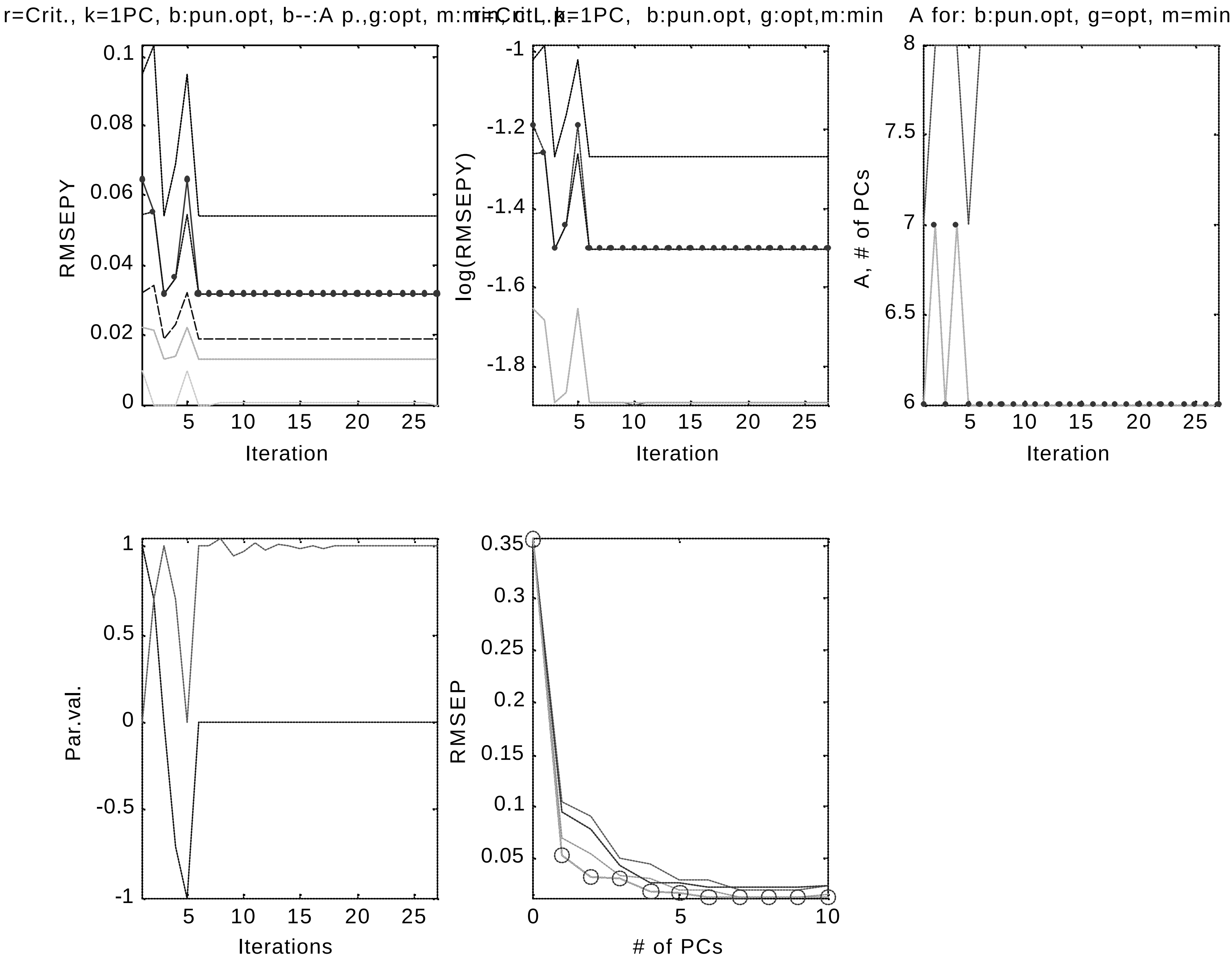
EMSC: Default physical model, no known good or bad spectra,

but **automatically find and optimize**

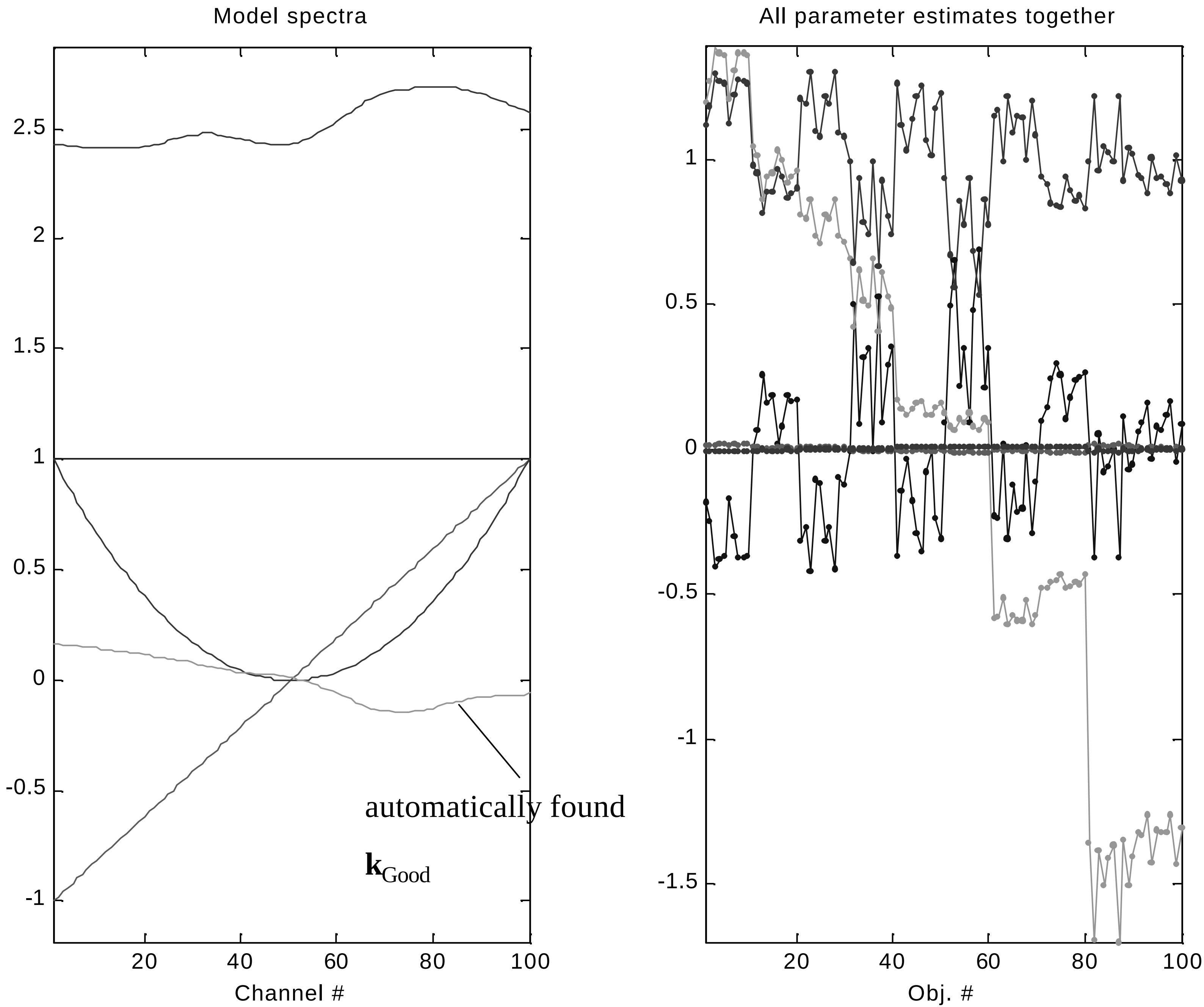
an unknown “good” (analyte) spectrum

based on minimising RMSEP(Y)

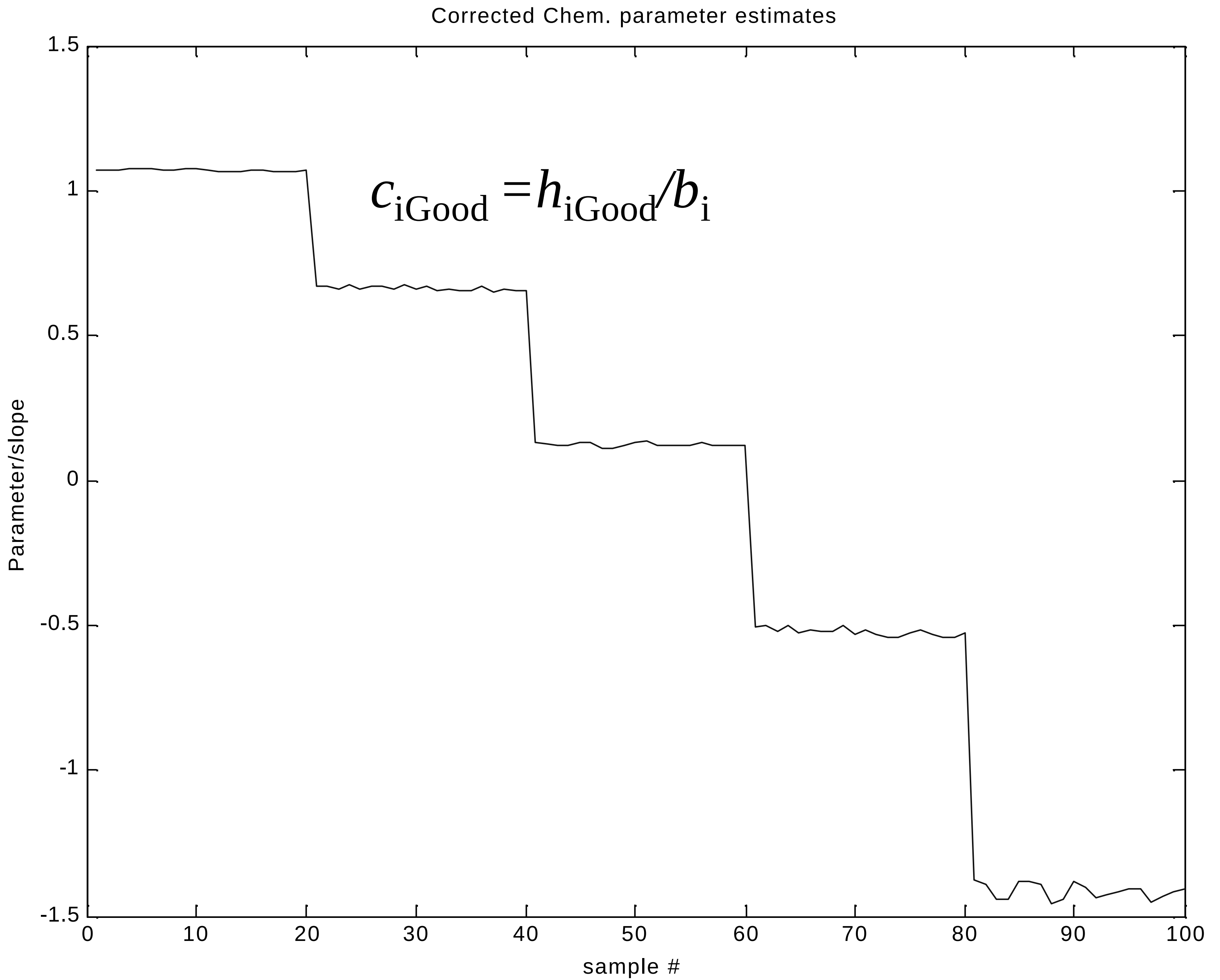
(estimated leverage-corrected PCR,1 PC)

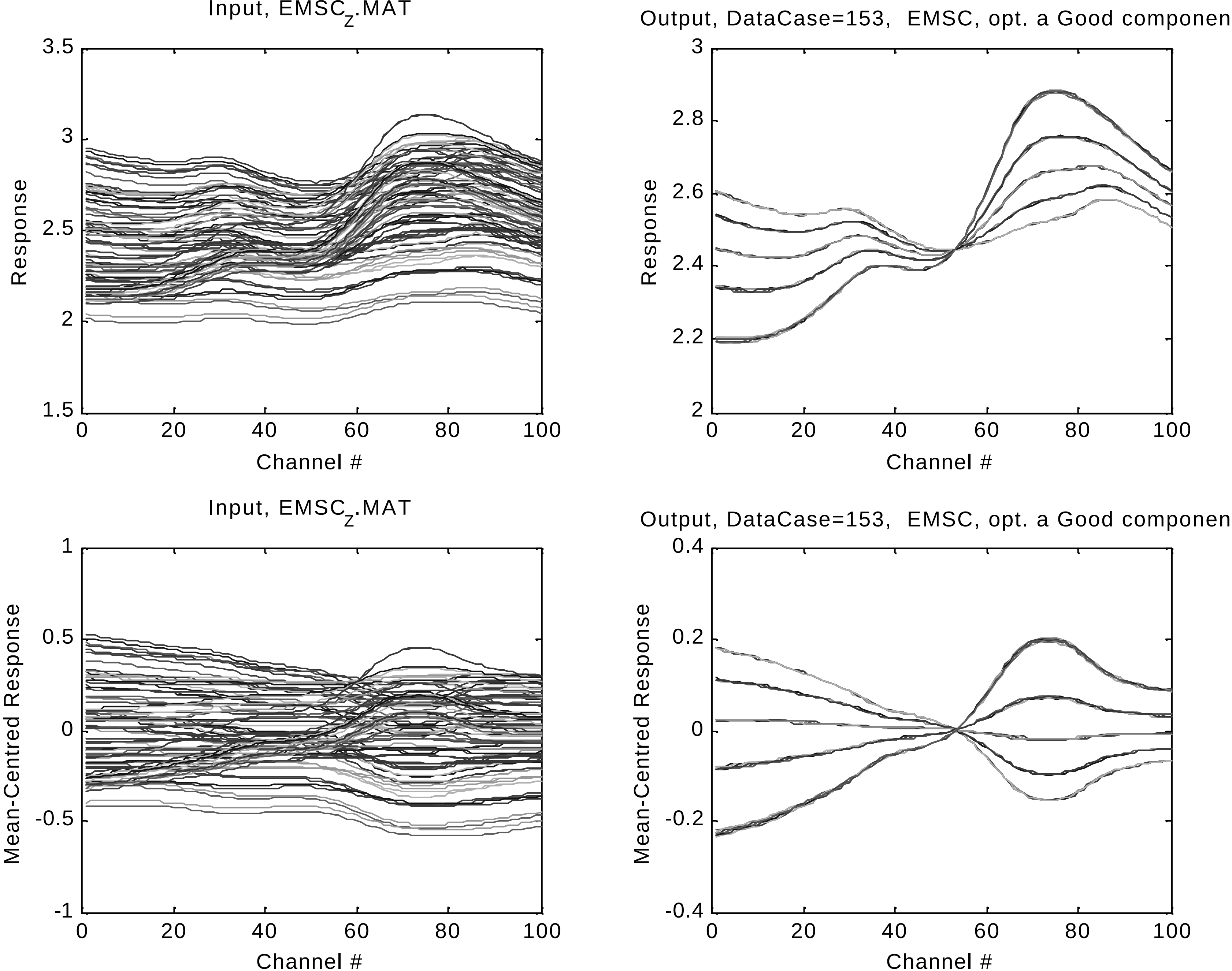


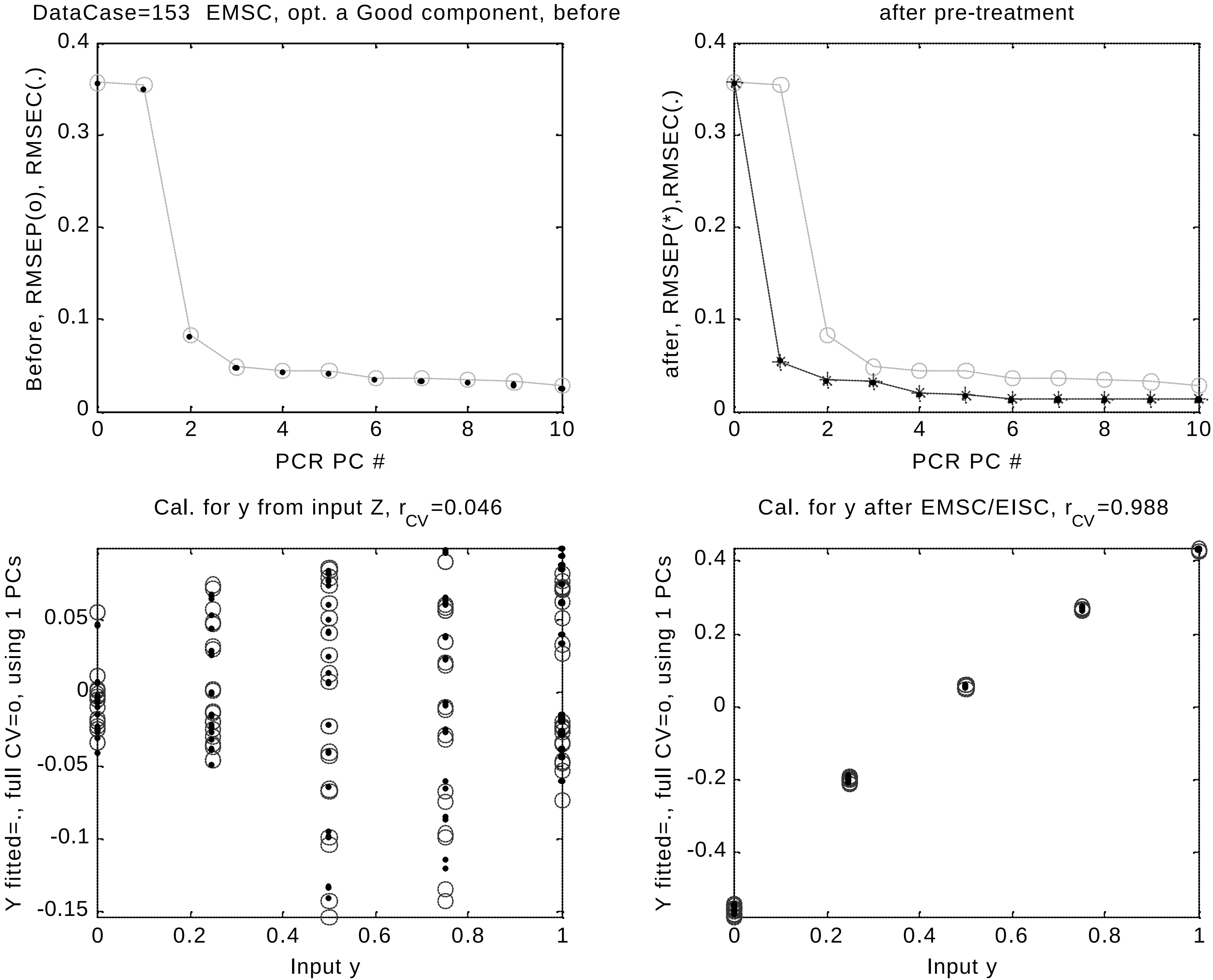
r=Crit., k=1PC, b:pun.opt, b--:A p.,g:opt, m:min
r=Crit.L, k=1PC, b:pun.opt, g:opt,m:min
A for: b:pun.opt, g=opt, m=min
RMSEPY
log(RMSEPY)
A, # of PCs
Iteration
Par.val.
Iterations
RMSEP
# of PCs



Model spectra
2.5
2
1.5
1
0.5
0
-0.5
-1
20
40
60
80
100
Channel #
automatically found
$\mathbf{k}_{\text{Good}}$
All parameter estimates together
1
0.5
0
-0.5
-1
-1.5
20
40
60
80
100
Obj. #


Corrected Chem. parameter estimates
$c_{\mathrm{iGood}} = h_{\mathrm{iGood}}/b_{\mathrm{i}}$
Parameter/slope
sample #
1.5
1
0.5
0
-0.5
-1
-1.5
0
10
20
30
40
50
60
70
80
90
100

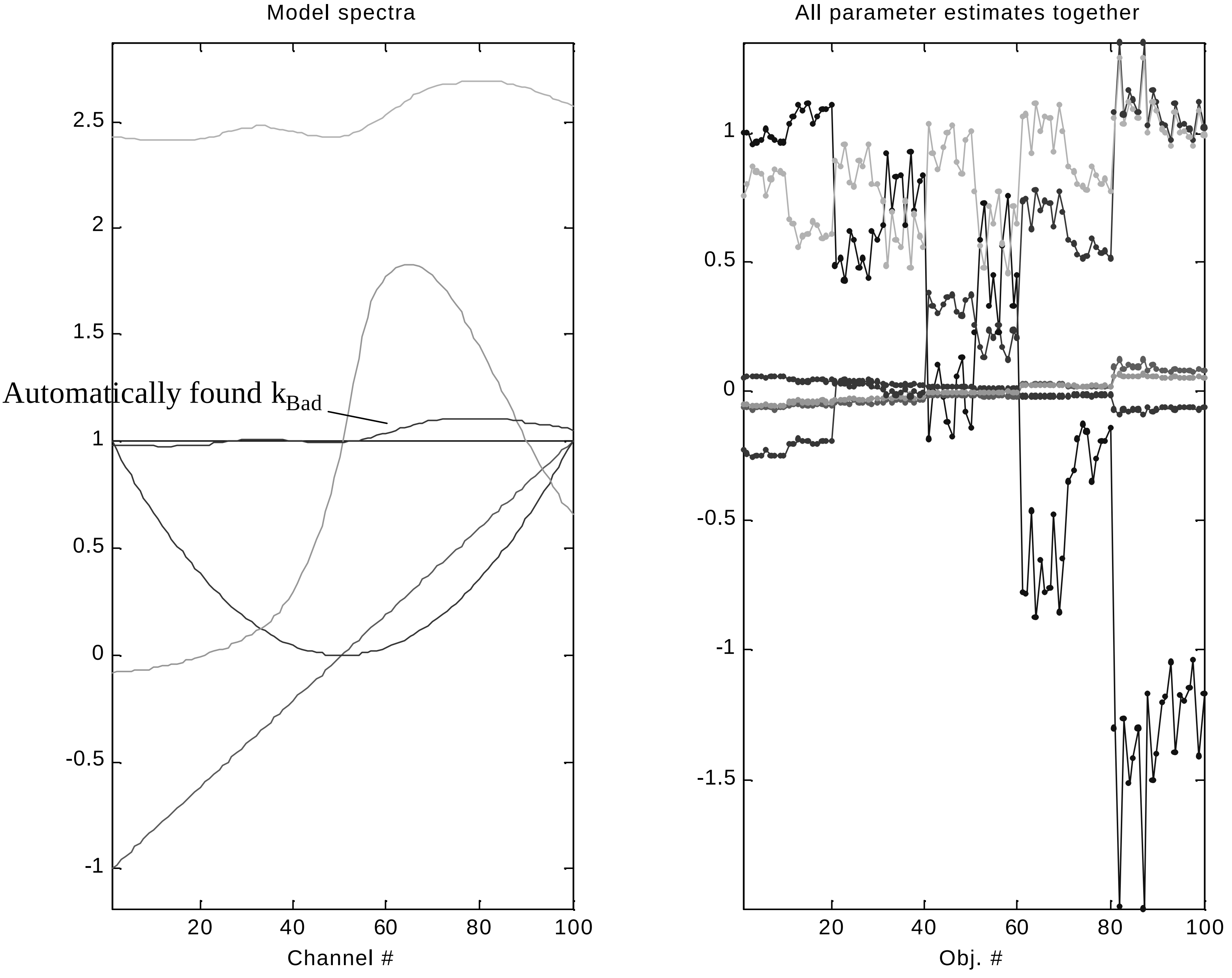

Input, EMSC$_Z$.MAT
Output, DataCase=153, EMSC, opt. a Good component
Response
Channel #
Input, EMSC$_Z$.MAT
Output, DataCase=153, EMSC, opt. a Good component
Mean-Centred Response
Channel #

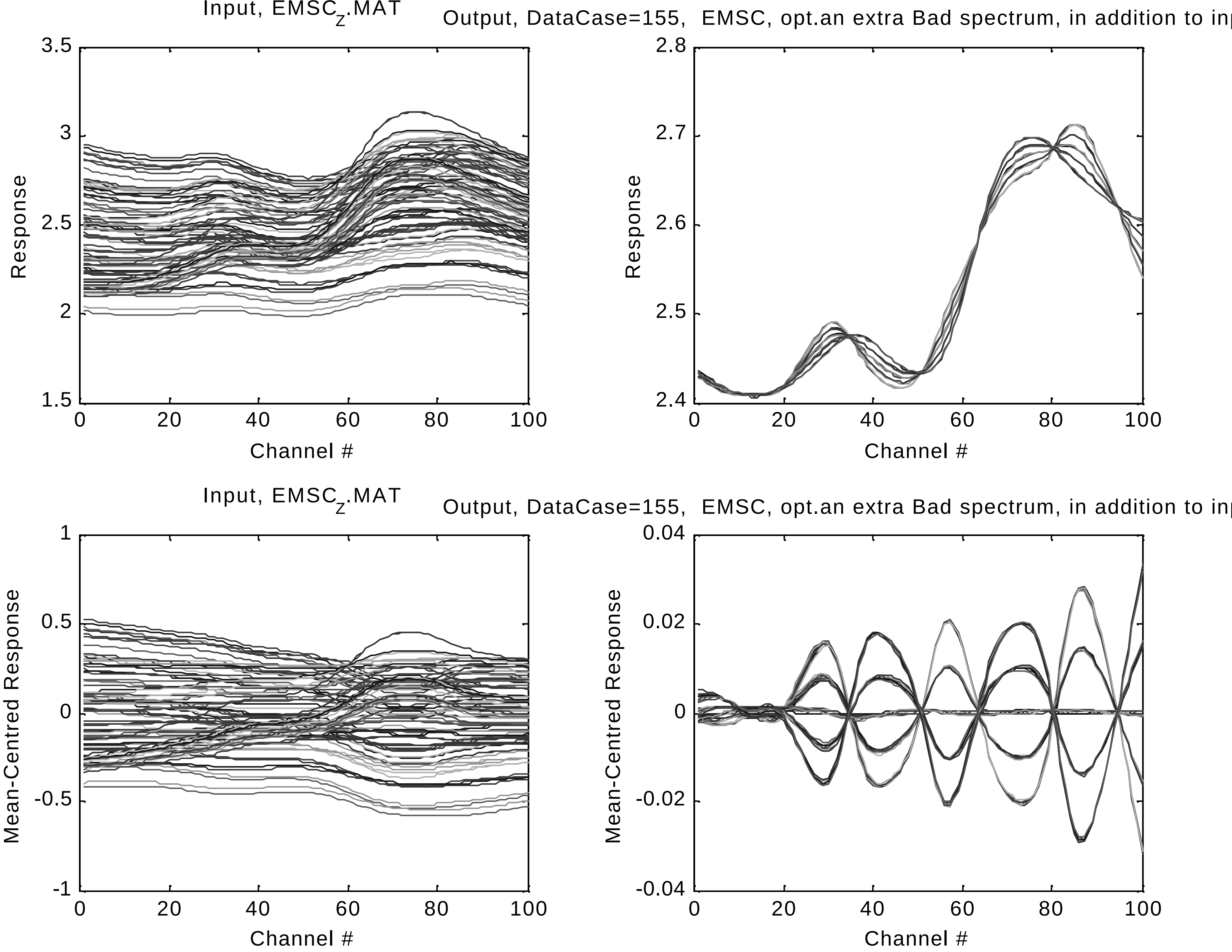


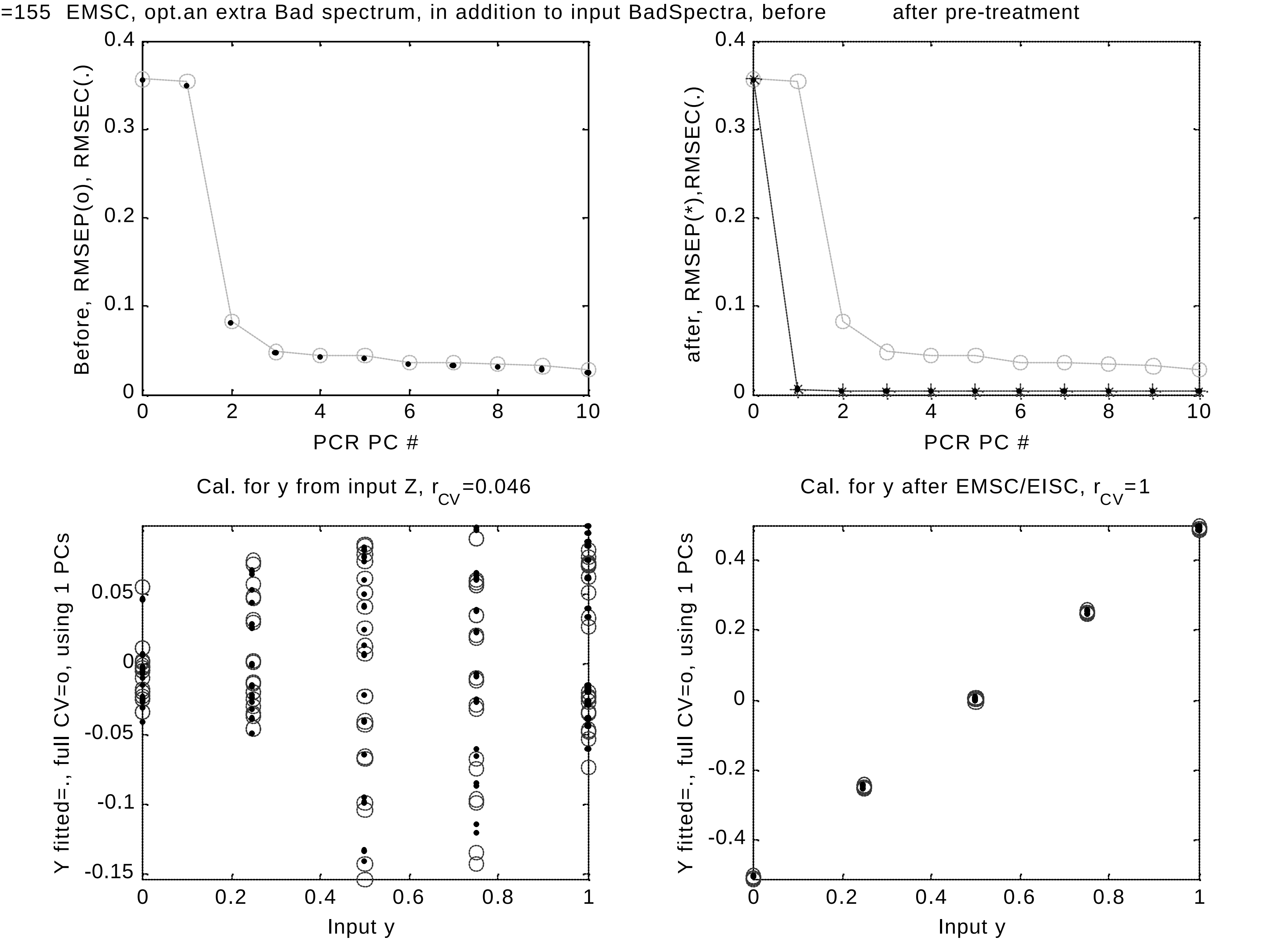
DataCase=153 EMSC, opt. a Good component, before
after pre-treatment
Before, RMSEP(o), RMSEC(.)
after, RMSEP(*),RMSEC(.)
PCR PC #
Cal. for y from input Z, r_{CV}=0.046
Cal. for y after EMSC/EISC, r_{CV}=0.988
Y fitted=., full CV=o, using 1 PCs
Input y

EMSC: Default physical model, input “bad” water spectrum

but **find and optimized another “bad” spectrum**

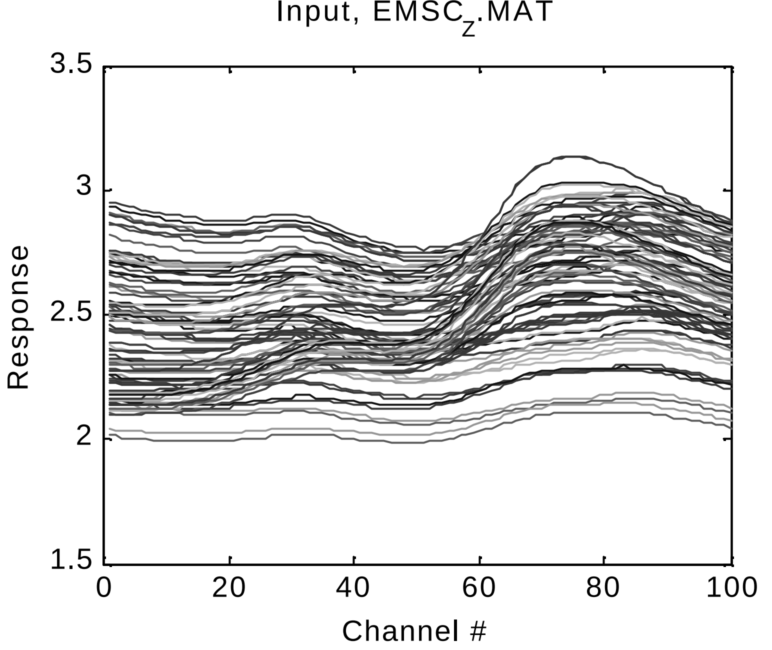

Model spectra
All parameter estimates together
Automatically found k_{Bad}
Channel #
Obj. #

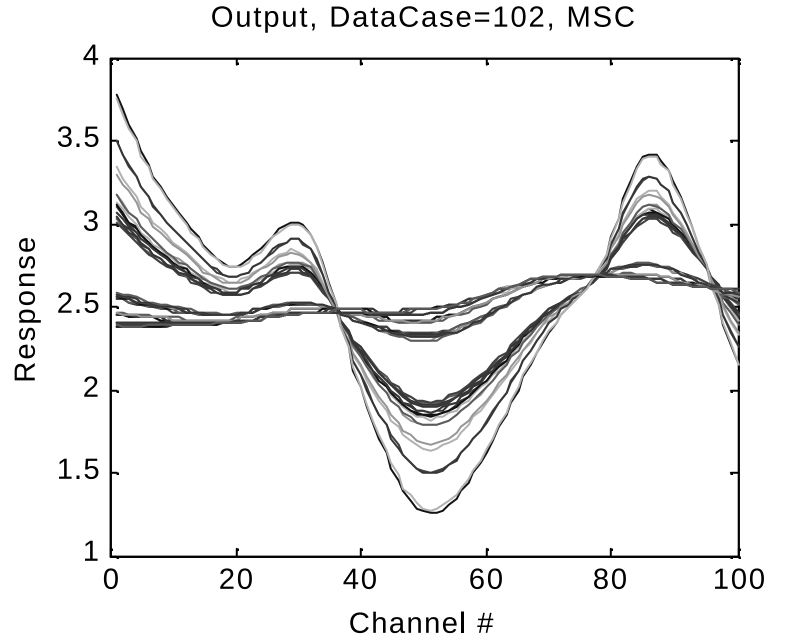

Input, EMSC$_Z$.MAT
Output, DataCase=155, EMSC, opt.an extra Bad spectrum, in addition to inp
Response
Channel #
Input, EMSC$_Z$.MAT
Output, DataCase=155, EMSC, opt.an extra Bad spectrum, in addition to inp
Mean-Centred Response
Channel #

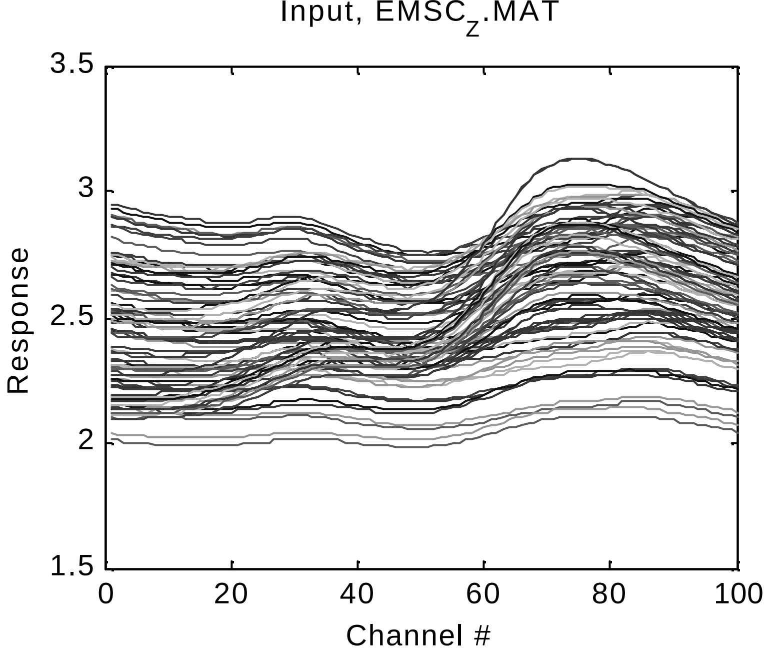


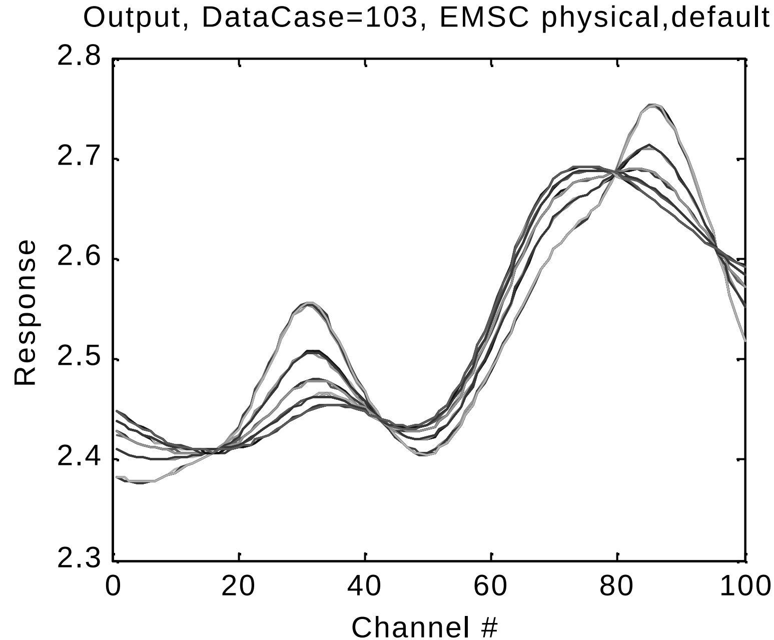
=155 EMSC, opt.an extra Bad spectrum, in addition to input BadSpectra, before
after pre-treatment
Before, RMSEP(o), RMSEC(.)
after, RMSEP(*),RMSEC(.)
PCR PC #
PCR PC #
Cal. for y from input Z, r_CV=0.046
Cal. for y after EMSC/EISC, r_CV=1
Y fitted=., full CV=o, using 1 PCs
Y fitted=., full CV=o, using 1 PCs
Input y
Input y

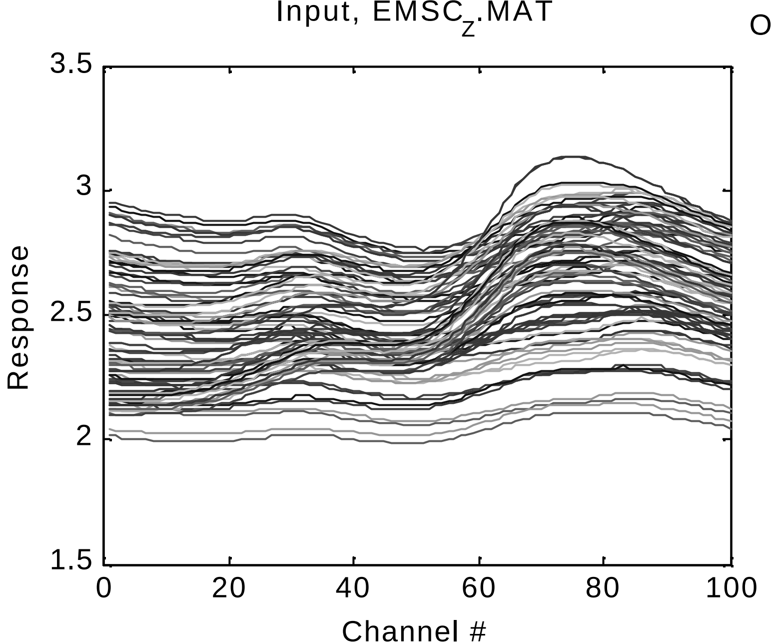
Comparisons of these models:


Input, $EMSC_Z$.MAT
Response
Channel #
3.5
3
2.5
2
1.5
0
20
40
60
80
100

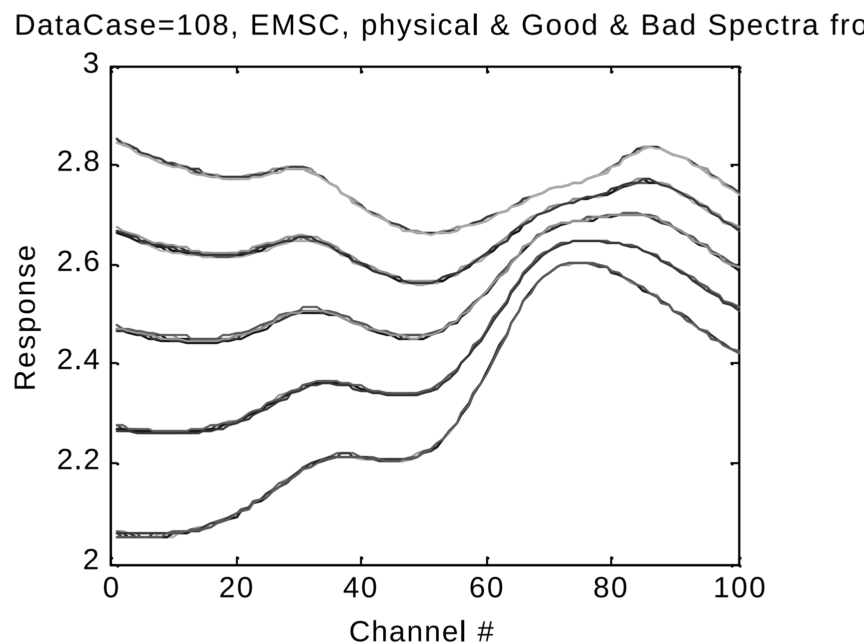

Output, DataCase=102, MSC
Response
Channel #
4
3.5
3
2.5
2
1.5
1
0
20
40
60
80
100


Input, $EMSC_Z$.MAT
3.5
3
2.5
2
1.5
Response
0
20
40
60
80
100
Channel #


Output, DataCase=103, EMSC physical,default
2.8
2.7
2.6
2.5
2.4
2.3
Response
0
20
40
60
80
100
Channel #

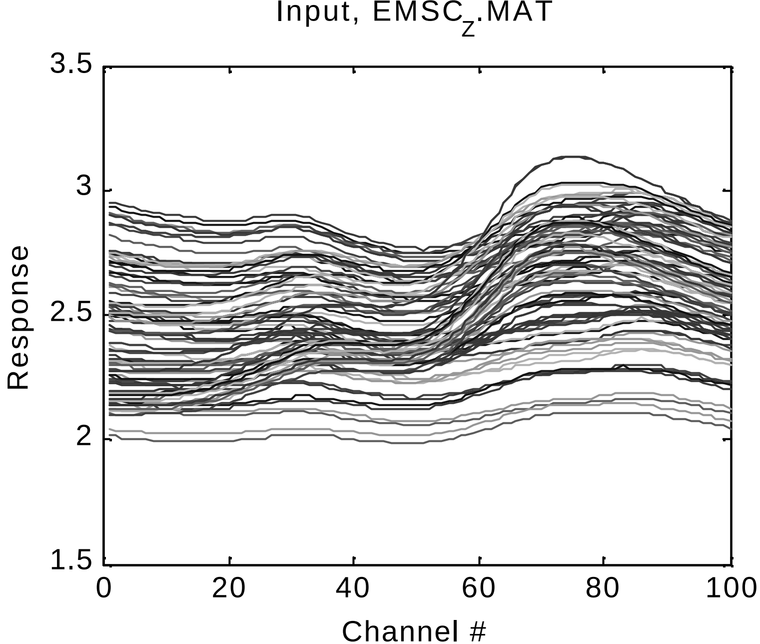


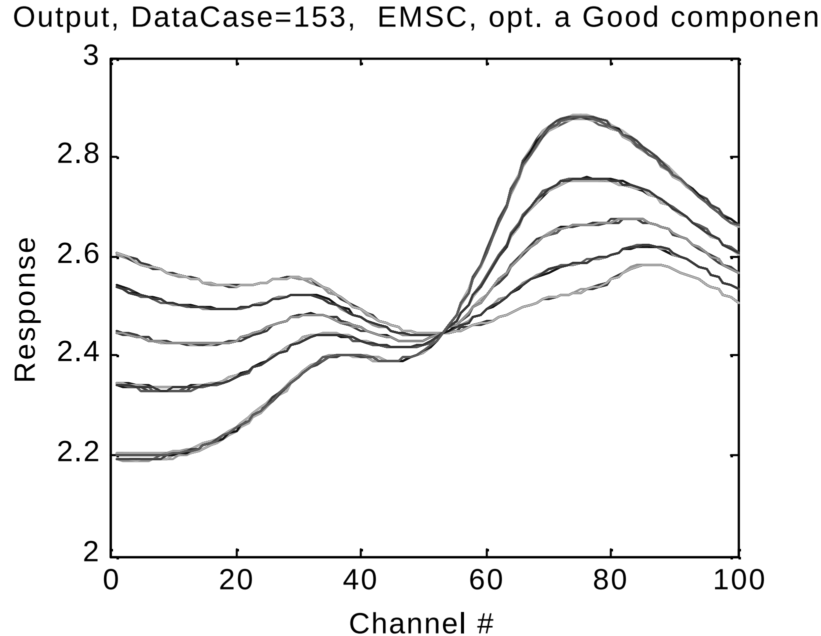
Input, EMSC$_Z$.MAT
3.5
3
2.5
2
1.5
Response
0
20
40
60
80
100
Channel #



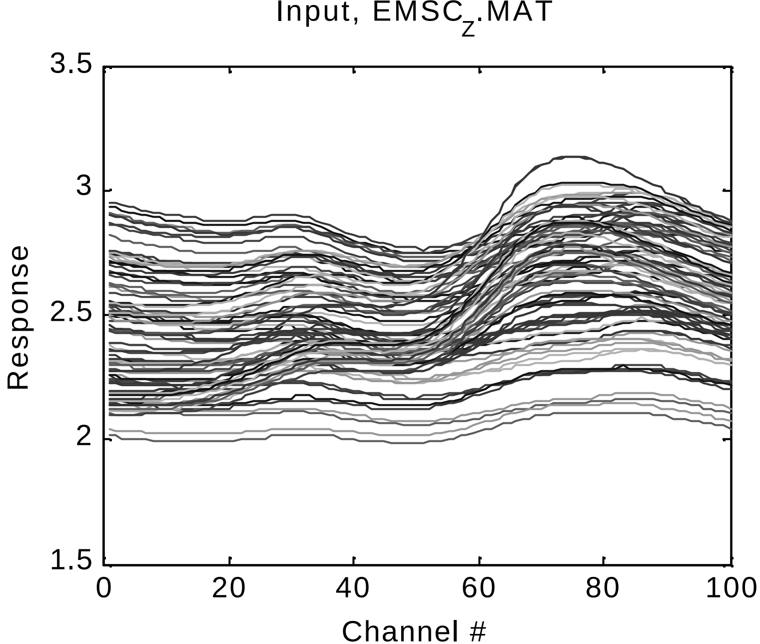
Output, DataCase=108, EMSC, physical & Good & Bad Spectra fro
3
2.8
2.6
2.4
2.2
2
Response
0
20
40
60
80
100
Channel #

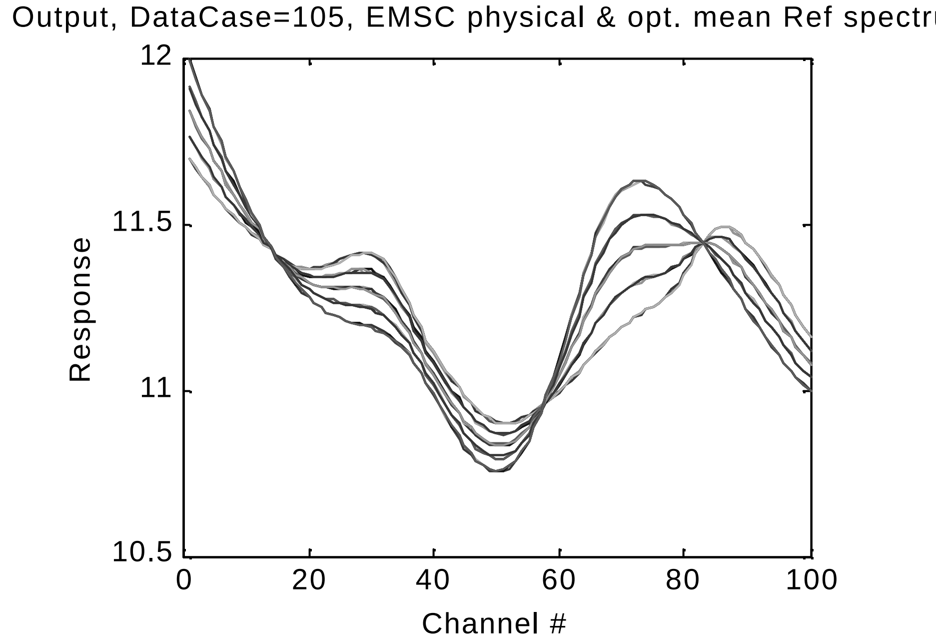
ra fro

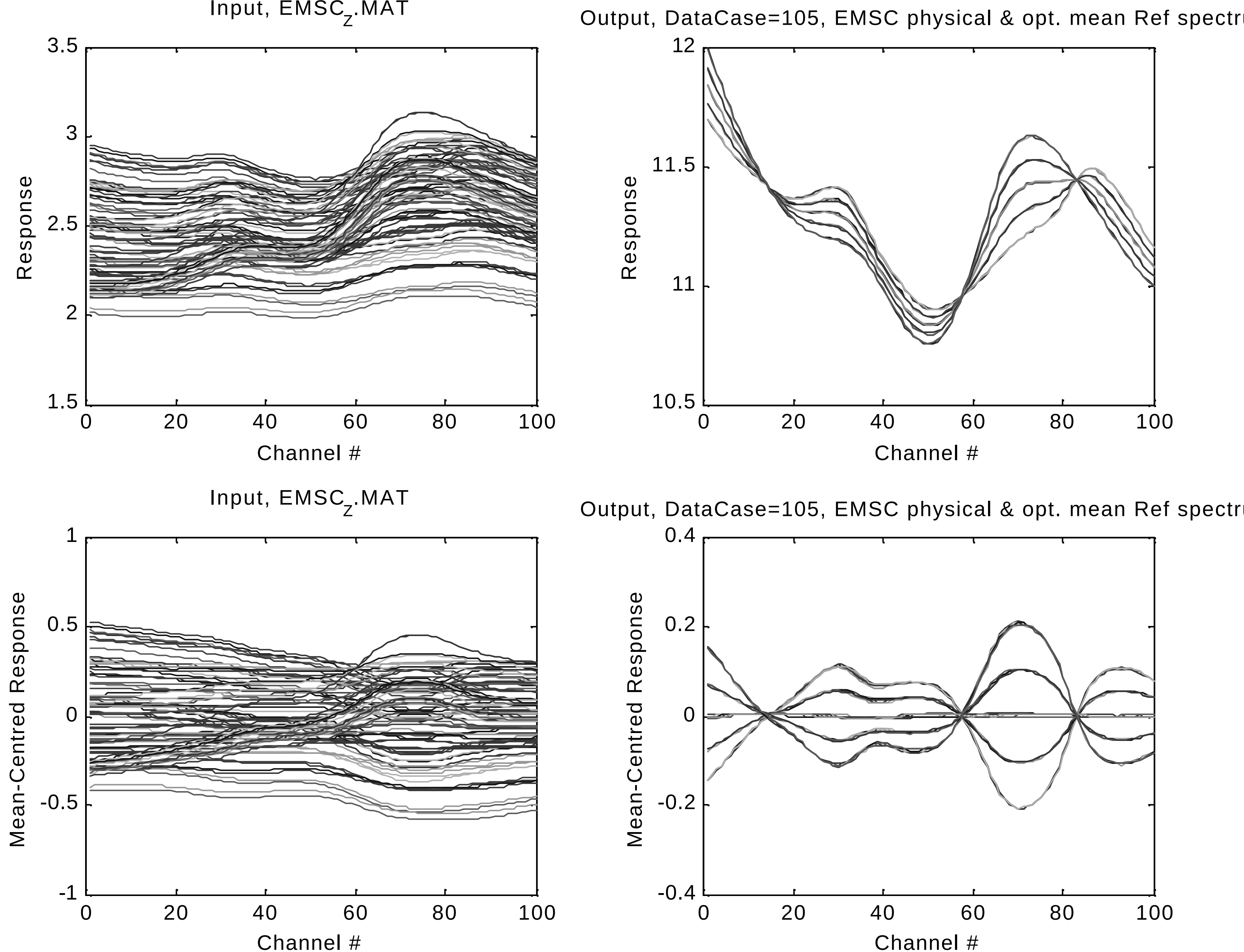

Input, $EMSC_Z$.MAT
Response
Channel #
3.5
3
2.5
2
1.5
0
20
40
60
80
100


Output, DataCase=153, EMSC, opt. a Good component
Response
Channel #
3
2.8
2.6
2.4
2.2
2
0
20
40
60
80
100

ent

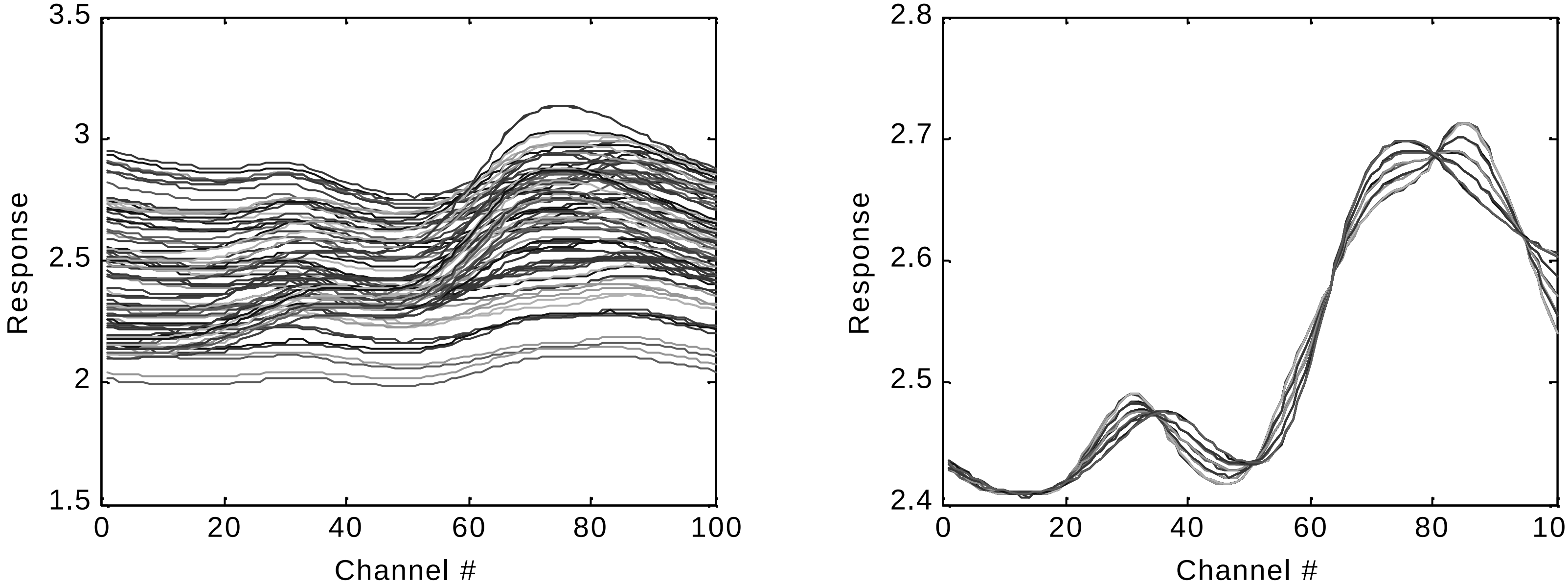

Input, EMSC_Z.MAT
Response
Channel #
3.5
3
2.5
2
1.5
0
20
40
60
80
100


Output, DataCase=105, EMSC physical & opt. mean Ref spectr
Response
Channel #
12
11.5
11
10.5
0
20
40
60
80
100

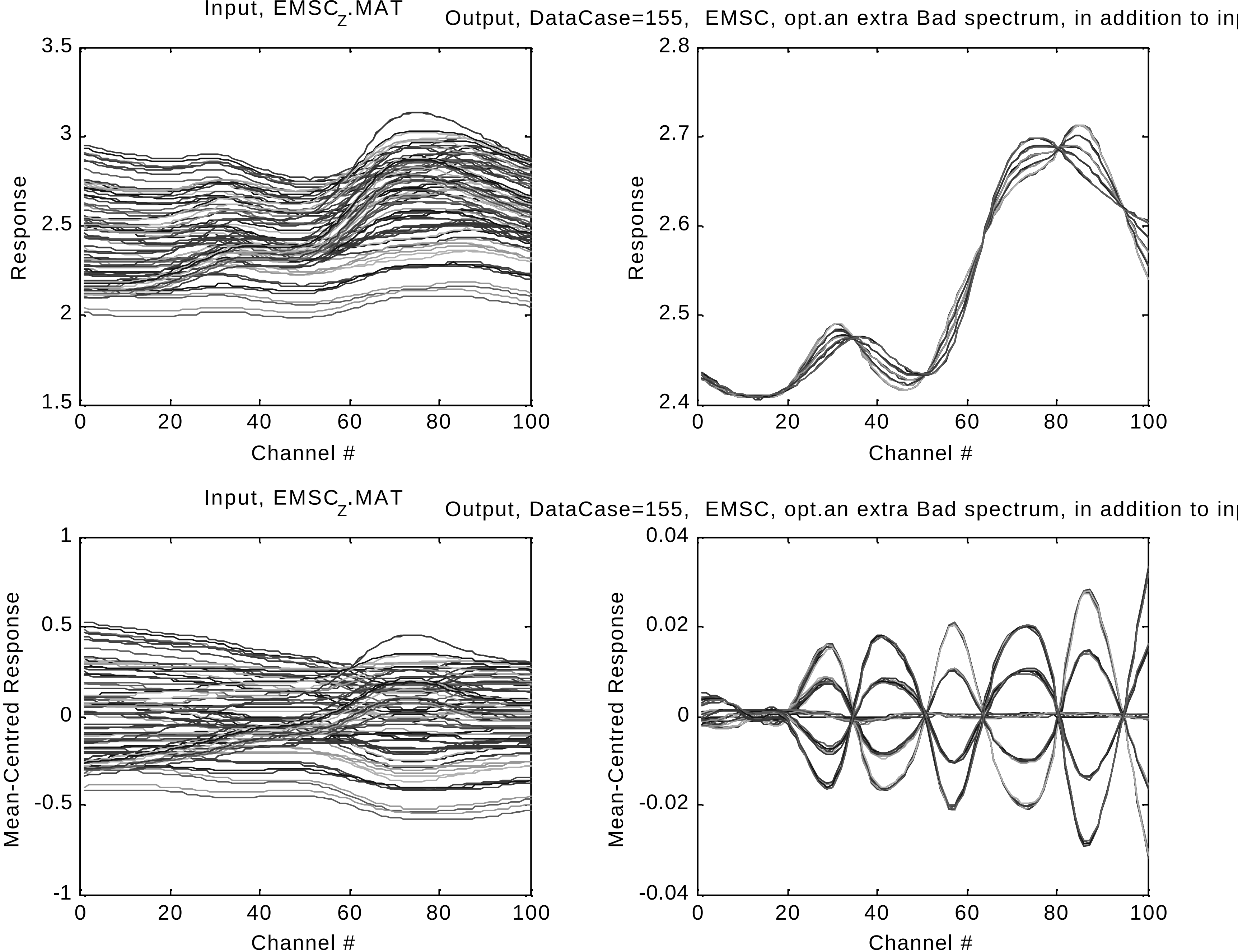


Input, $EMSC_Z$.MAT
Output, DataCase=105, EMSC physical & opt. mean Ref spectru
Response
Channel #
3.5
3
2.5
2
1.5
12
11.5
11
10.5
0
20
40
60
80
100
Input, $EMSC_Z$.MAT
Output, DataCase=105, EMSC physical & opt. mean Ref spectru
Mean-Centred Response
1
0.5
0
-0.5
-1
0.4
0.2
-0.2
-0.4

Input, $EMSC_Z$.MAT Output, DataCase=155, EMSC, opt.an extra Bad spectrum, in addition to inp

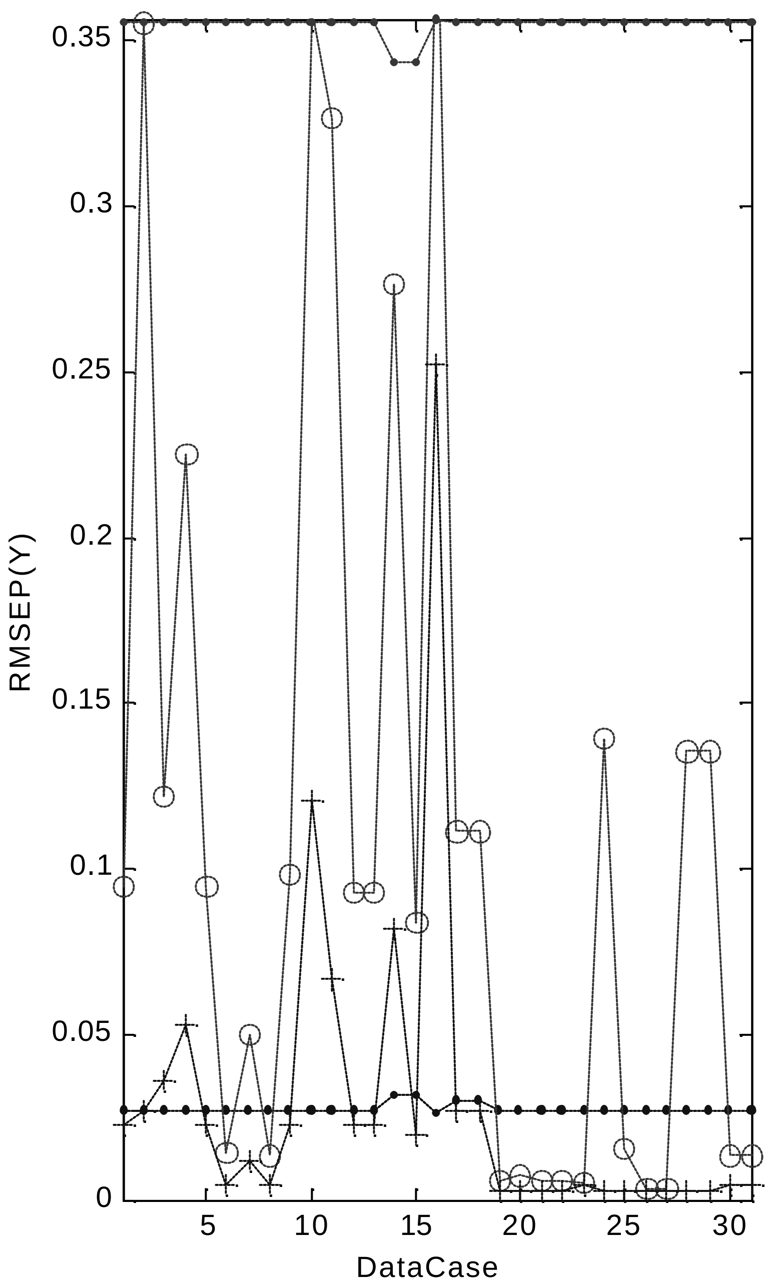


to inp



Input, EMSC_Z.MAT
Output, DataCase=155, EMSC, opt.an extra Bad spectrum, in addition to inp
Response
Channel #
Input, EMSC_Z.MAT
Output, DataCase=155, EMSC, opt.an extra Bad spectrum, in addition to inp
Mean-Centred Response
Channel #

Comparison of DataCases, r=1 PC, b= A_{Opt} PCs; ...=input data

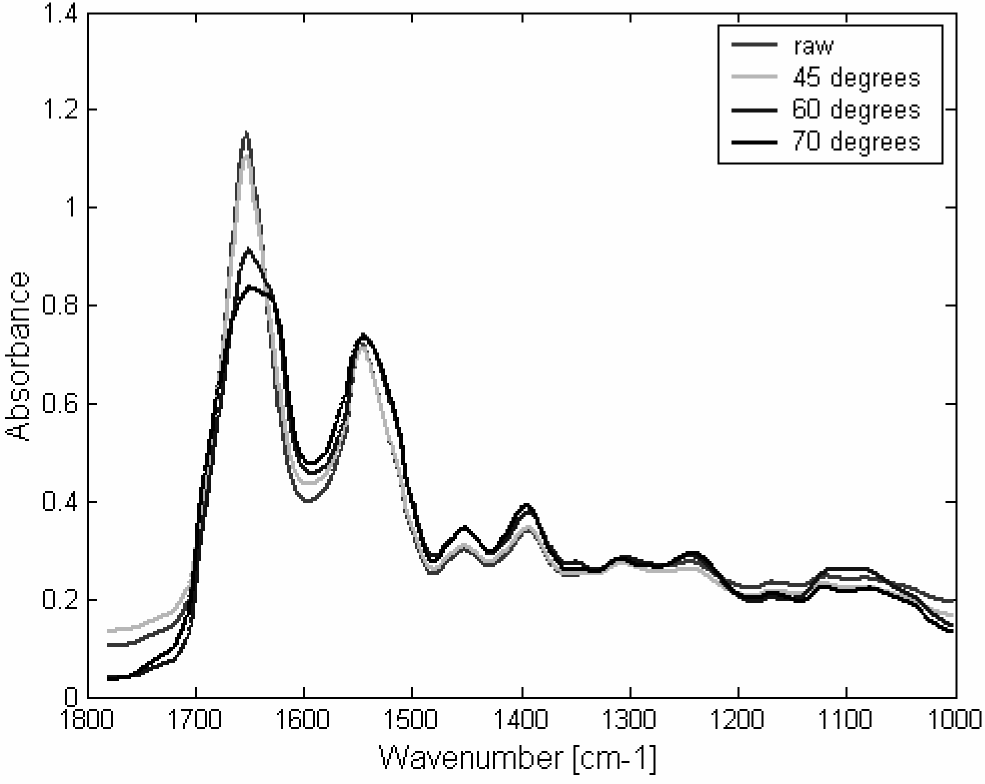


Automatic comparison of many different pre-processing alternatives for a given data set

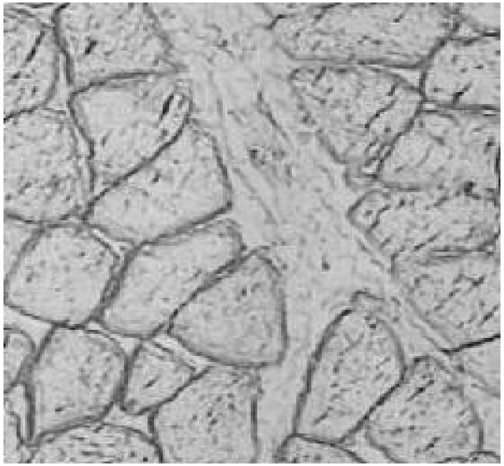
Comparative study of chemometric and conventional pre-processing methods for FT-IR spectra of biological material

A. Kohler, C. Kirschner, A. Oust and H. Martens

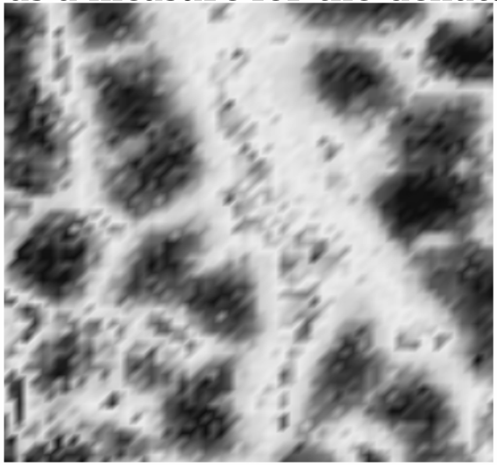
FT-IR of heated bovine muscle:



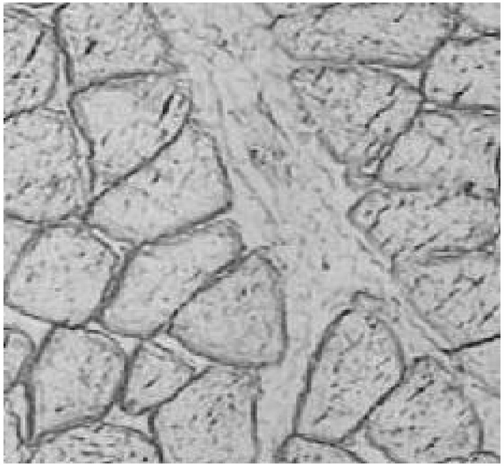
Vis. image of muscle



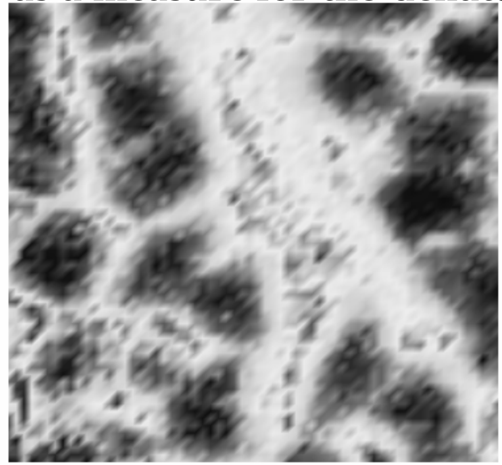
Chemical map from I_{1630}/I_{1654} as a measure for the denaturation level



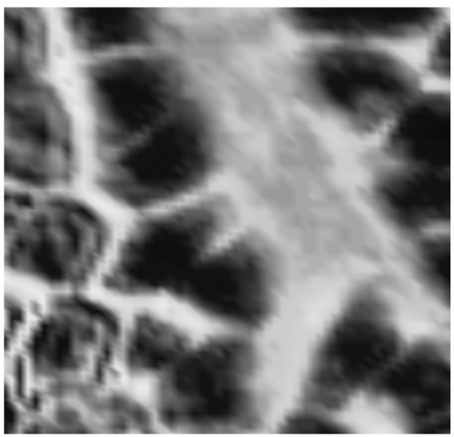
Vis. image of muscle



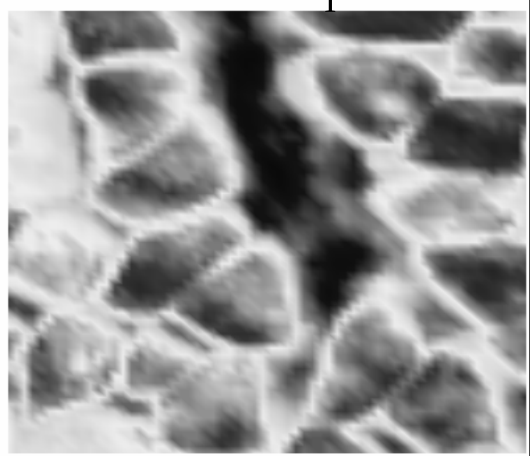
Chemical map from I_{1630}/I_{1654} as a measure for the denaturation level



EMSC a_i



EMSC b_i



The correlation coefficients for

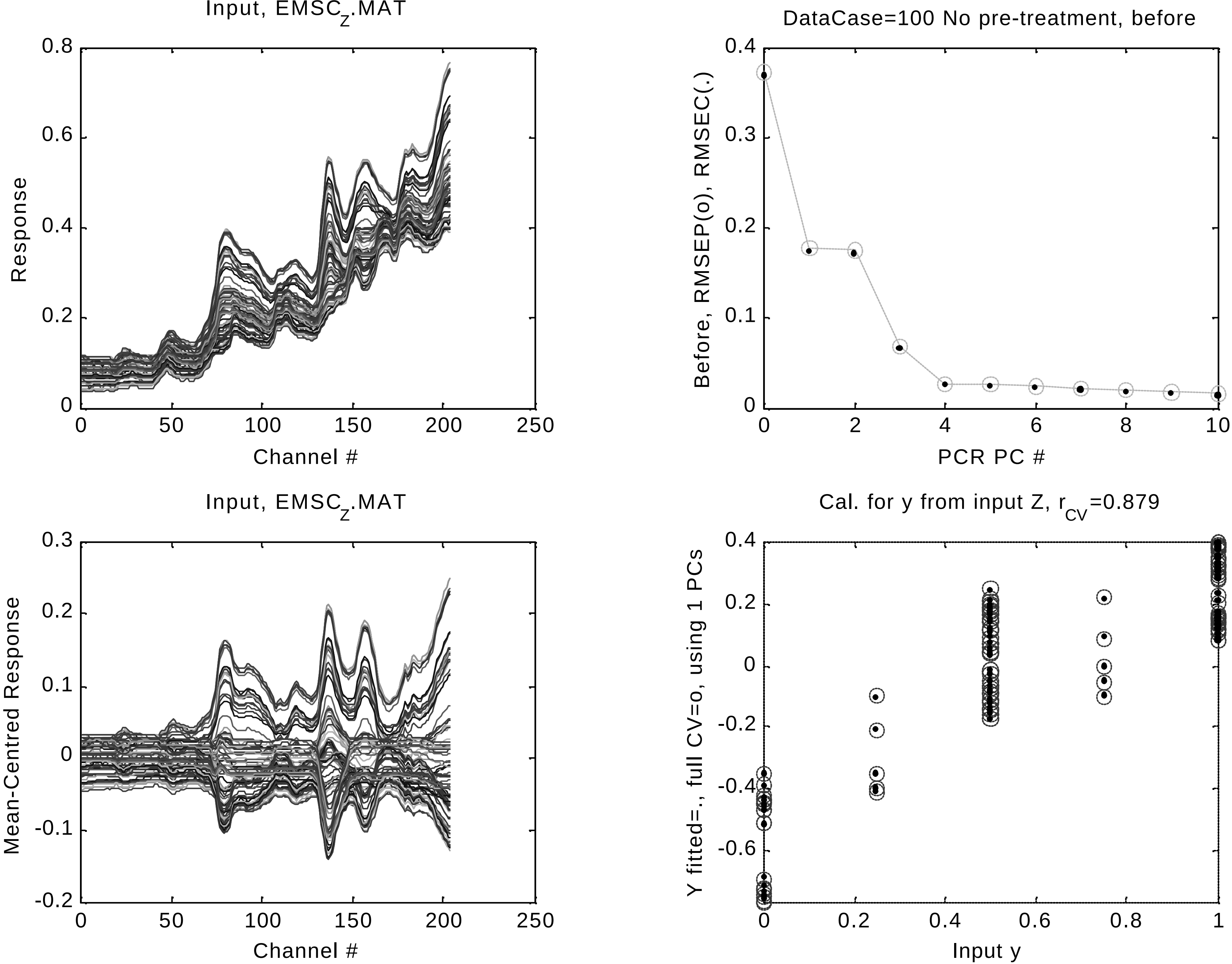
X= pre-processed FT-IR spectra of myofibres and

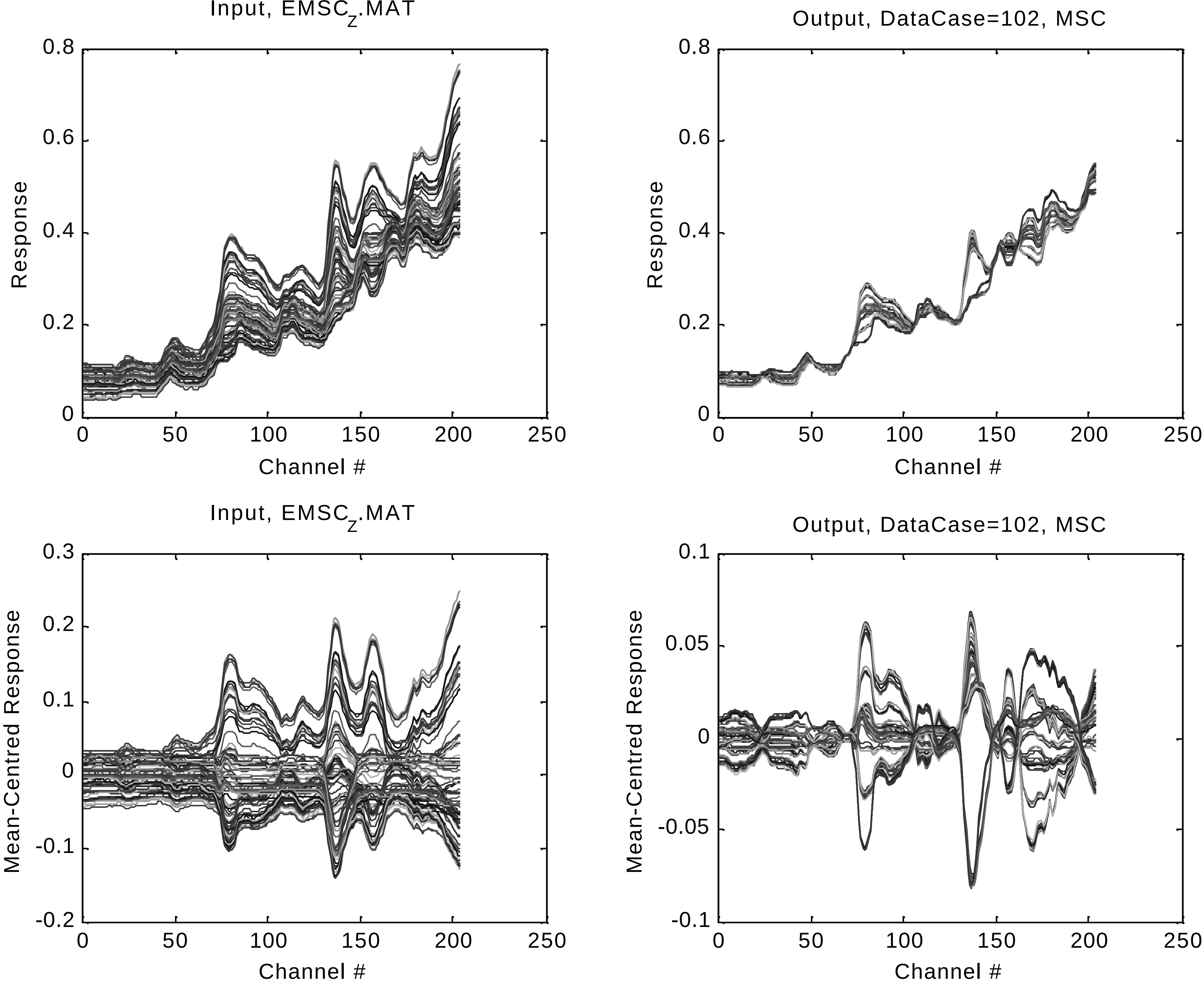
y = the different temperatures used for the heat treatment

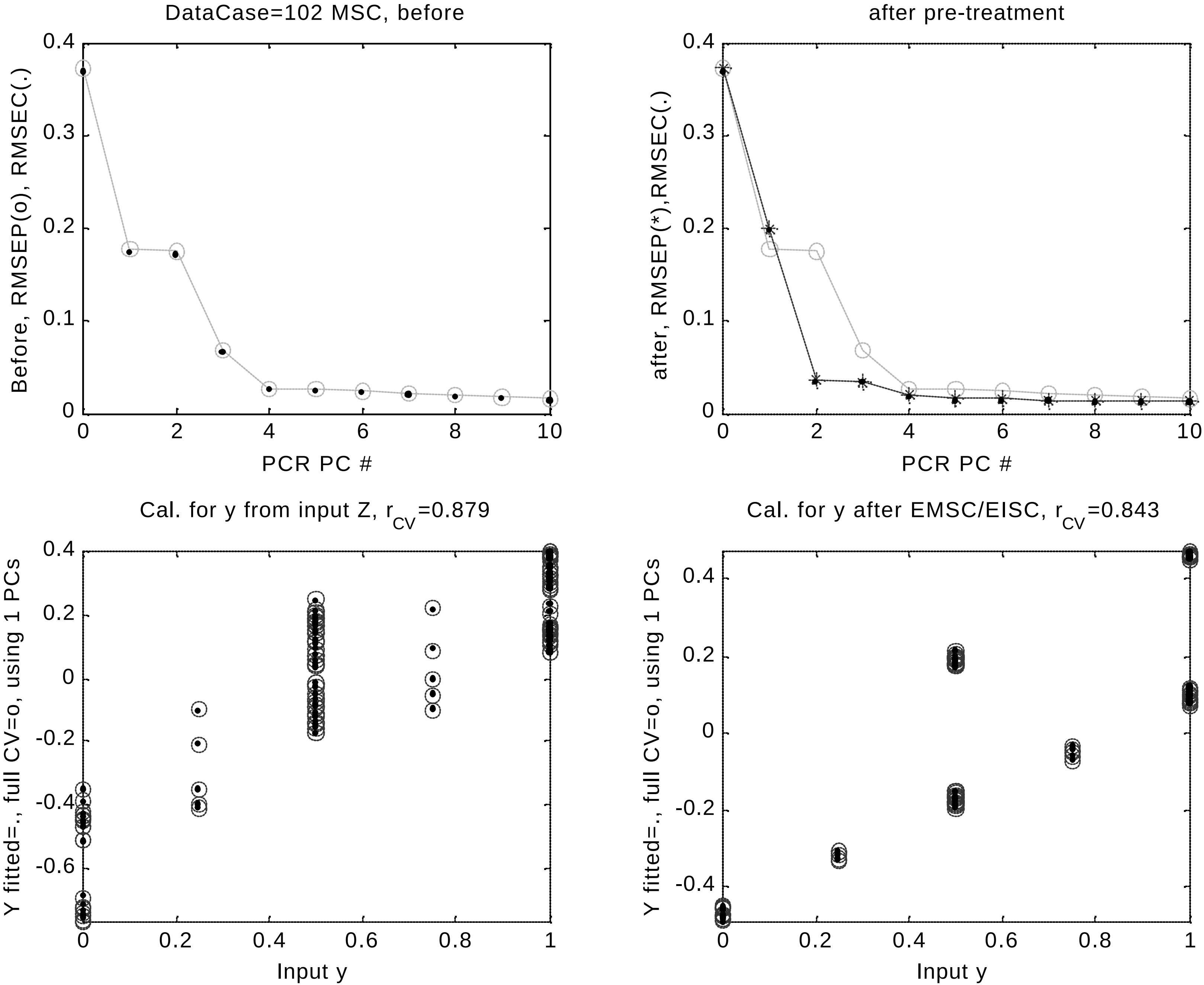
$1000\text{-}1780\ cm^{-1}$	Original spectra	Derivative + vector normalisation	MSC	EMSC
Correlation coefficient All variables	0.92 (10 PCs)	0.91 (5 PCs)	0.92 (6 PCs)	0.94 (7 PCs)
Correlation coefficient Selected variables	0.93 (8 PCs, 120 variables)	0.91 (1 PC, 103 variables)	0.92 (2 PCs, 215 variables)	0.95 (3 PCs, 182 variables)

NIR reflectance of
protein and starch mixtures (5 mixtures)
in different bottles and
at different water contents.

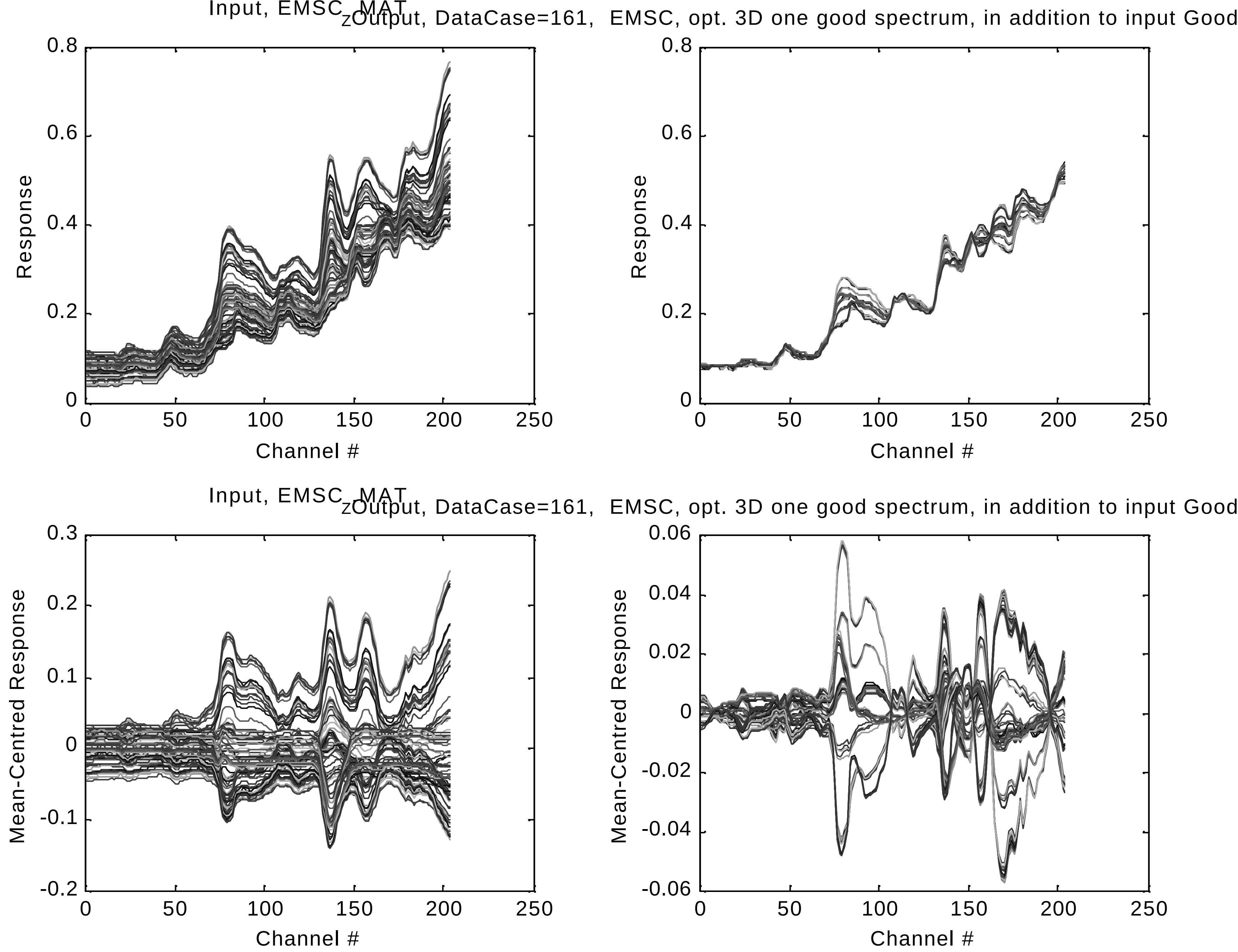
Data from Xuxin Lai, Biocentrum DTU

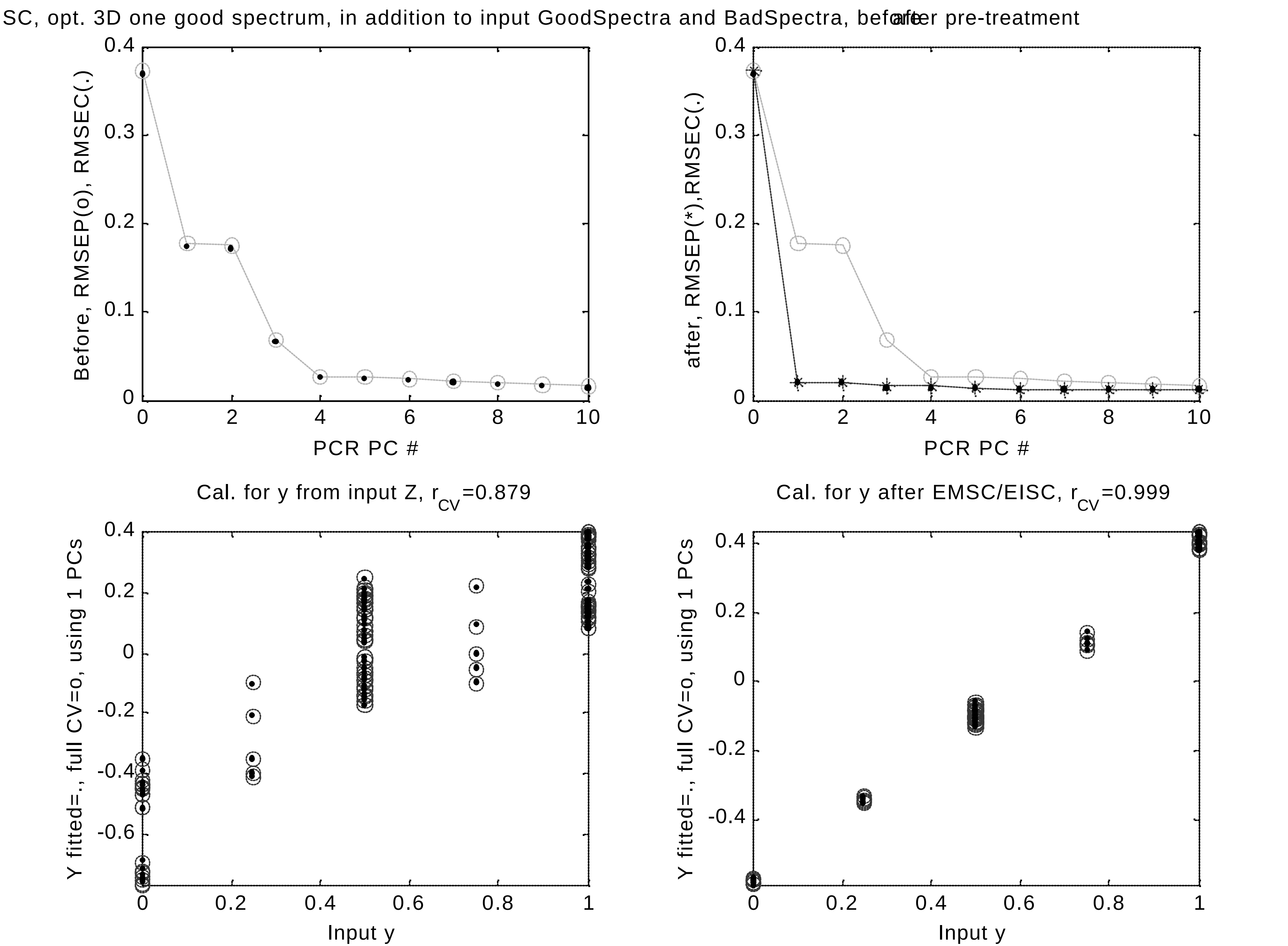

Input, $EMSC_Z$.MAT
Response
Channel #
DataCase=100 No pre-treatment, before
Before, RMSEP(o), RMSEC(.)
PCR PC #
Input, $EMSC_Z$.MAT
Mean-Centred Response
Channel #
Cal. for y from input Z, r_{CV}=0.879
Y fitted=., full CV=o, using 1 PCs
Input y


Input, EMSC_Z.MAT
Output, DataCase=102, MSC
Input, EMSC_Z.MAT
Output, DataCase=102, MSC
Response
Mean-Centred Response
Channel #
0
50
100
150
200
250
0.8
0.6
0.4
0.2
0.3
0.1
-0.1
-0.2
0.05
-0.05



DataCase=102 MSC, before
after pre-treatment
Before, RMSEP(o), RMSEC(.)
after, RMSEP(*),RMSEC(.)
PCR PC #
PCR PC #
Cal. for y from input Z, r_{CV}=0.879
Cal. for y after EMSC/EISC, r_{CV}=0.843
Y fitted=., full CV=o, using 1 PCs
Y fitted=., full CV=o, using 1 PCs
Input y
Input y


Input, EMSC_MAT
zOutput, DataCase=161, EMSC, opt. 3D one good spectrum, in addition to input Good
Response
Channel #
0
0.2
0.4
0.6
0.8
0
50
100
150
200
250
Input, EMSC_MAT
zOutput, DataCase=161, EMSC, opt. 3D one good spectrum, in addition to input Good
Mean-Centred Response
-0.2
-0.1
0
0.1
0.2
0.3
-0.06
-0.04
-0.02
0
0.02
0.04
0.06


SC, opt. 3D one good spectrum, in addition to input GoodSpectra and BadSpectra, before
after pre-treatment
Before, RMSEP(o), RMSEC(.)
PCR PC #
after, RMSEP(*),RMSEC(.)
PCR PC #
Cal. for y from input Z, r_{CV}=0.879
Y fitted=., full CV=o, using 1 PCs
Input y
Cal. for y after EMSC/EISC, r_{CV}=0.999
Y fitted=., full CV=o, using 1 PCs
Input y

NIT whole wheat grains

Pedersen, D.K., Martens, H., Pram Nielsen, J. and Balling Engelsen, S. Light absorbance and light scattering separated by Extended Inverted Multiplicative Signal Correction (EIMSC). Analysis of NIT spectra of single wheat seeds. Applied Spectroscopy 2002, **56**(9) 1206-1214.

EISC vs EMSC

MSC: $\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + e_i$

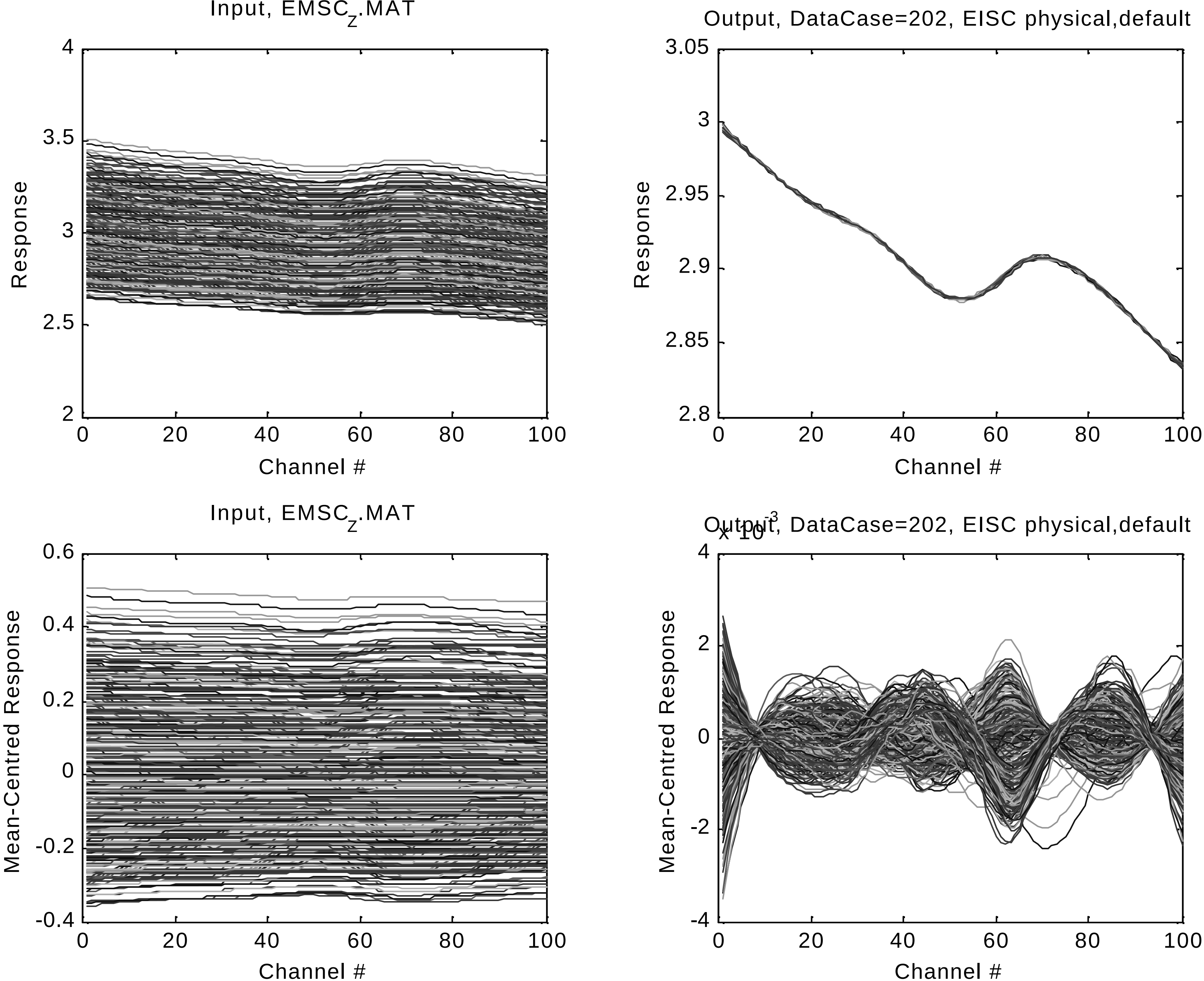
EMSC: $\mathbf{z}_i = a_i\mathbf{1}' + b_i\mathbf{m}' + h_i\mathbf{k}' + d_i\lambda + e_i\lambda^2 + e_i$

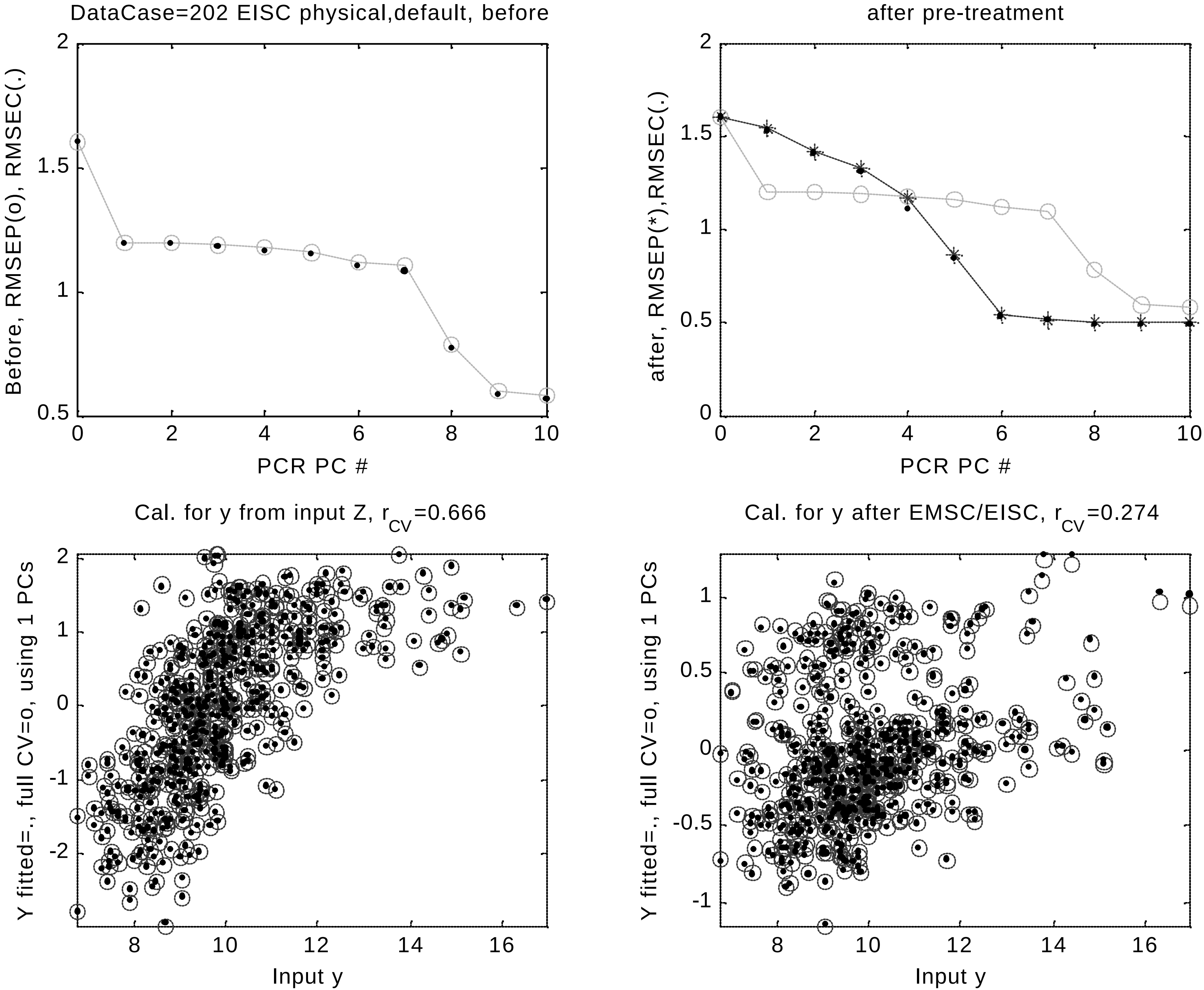
$\mathbf{z}_{i,\mathrm{Corrected}} = (\mathbf{z}_i - a_i\mathbf{1}' - d_i\lambda - e_i\lambda^2)/b_i$

ISC: $\mathbf{m}' = a_i\mathbf{1}' + b_i\mathbf{z}_i + g_i$

EISC: $\mathbf{m}' = a_i\mathbf{1}' + b_i\mathbf{z}_i + h_i\mathbf{k} + d_i\lambda + e_i\lambda^2 + g_i$

$\mathbf{z}_{i,\mathrm{Corrected}} = a_i\mathbf{1}' + b_i\mathbf{z}_i + h_i\mathbf{k} + d_i\lambda + e_i\lambda^2$


Input, EMSC$_Z$.MAT
Response
Channel #
Output, DataCase=202, EISC physical,default
Response
Channel #
Input, EMSC$_Z$.MAT
Mean-Centred Response
Channel #
Output, DataCase=202, EISC physical,default
x 10^{-3}
Mean-Centred Response
Channel #


DataCase=202 EISC physical,default, before
Before, RMSEP(o), RMSEC(.)
PCR PC #
after pre-treatment
after, RMSEP(*),RMSEC(.)
PCR PC #
Cal. for y from input Z, r_{CV}=0.666
Y fitted=., full CV=o, using 1 PCs
Input y
Cal. for y after EMSC/EISC, r_{CV}=0.274
Y fitted=., full CV=o, using 1 PCs
Input y

New method ≈ Direct Orthogonalization:

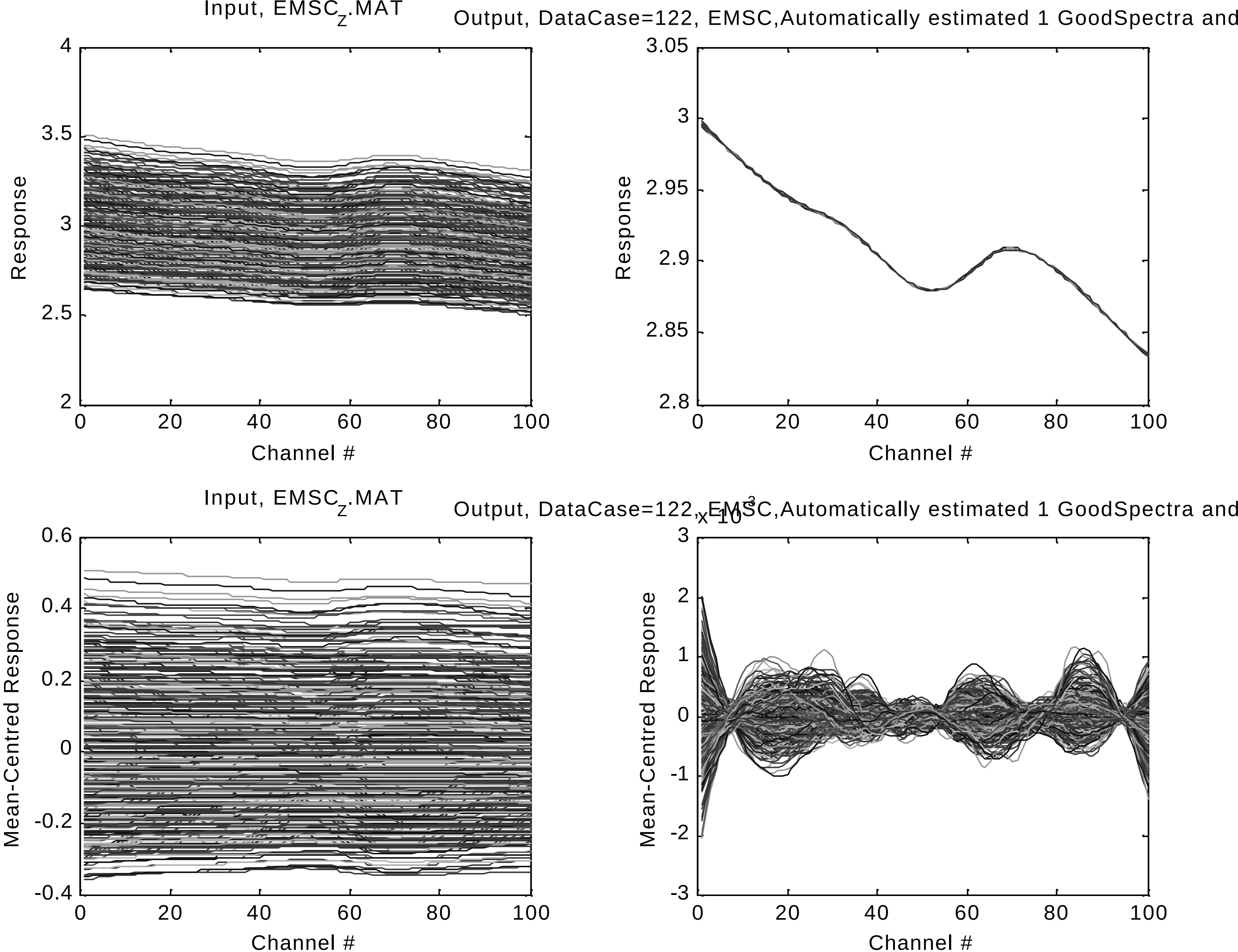
Estimate unknown “good” and “bad” spectra after projection of spectra **Z** on **y**:

$$\mathbf{Z} = \mathbf{y}\mathbf{k}_{\mathrm{Good}}{}' + \mathbf{E}$$

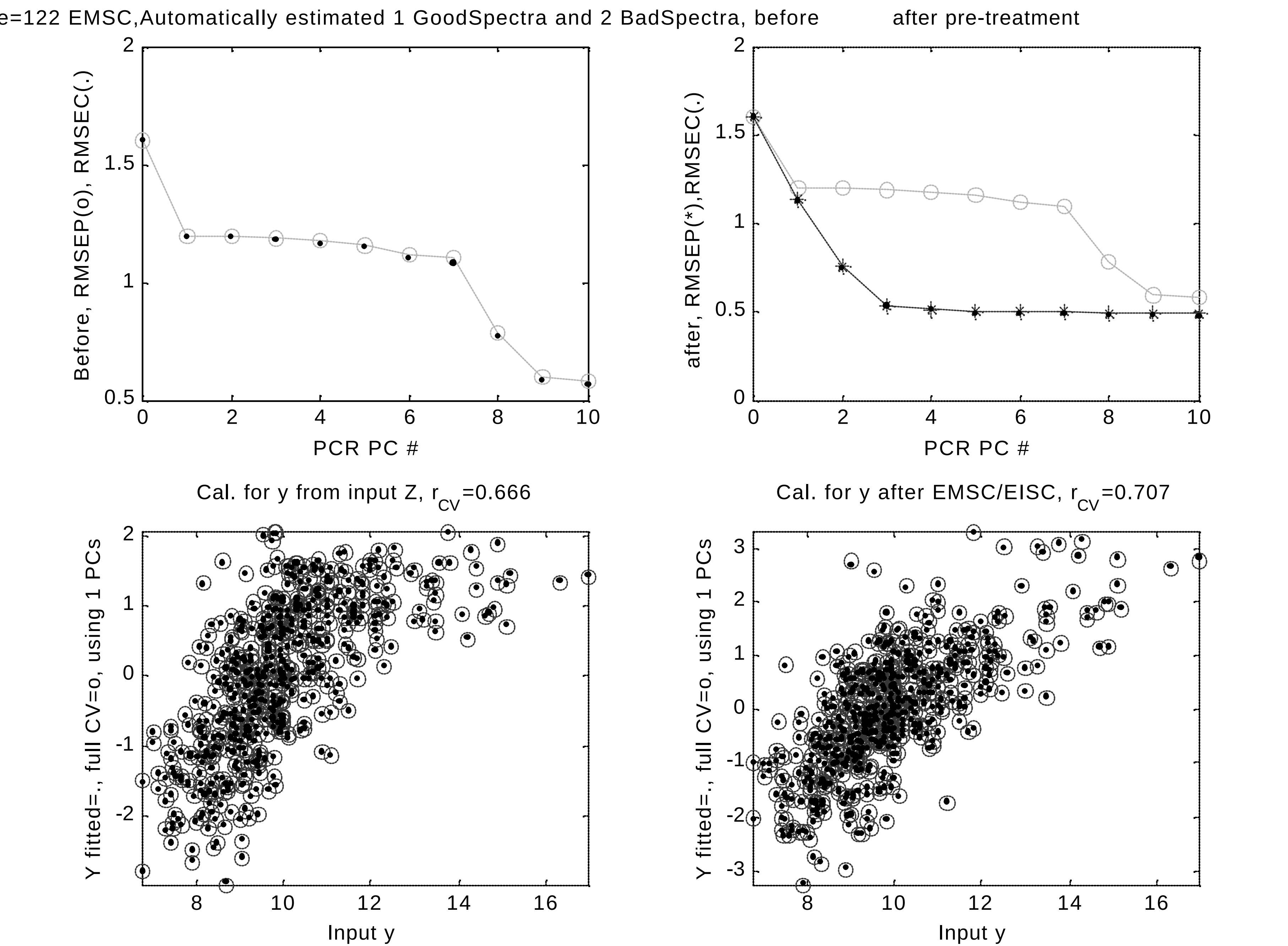
Estimate $\mathbf{k}_{\mathrm{Good}}$ and **E** by regression

$$\mathbf{k}_{\mathrm{Bad},1,2} = \mathrm{svd}(\mathbf{E})$$

In EMSC: Estimate their concentrations, and subtract the effects of the “bad” spectra.



Input, $EMSC_Z$.MAT
Output, DataCase=122, EMSC,Automatically estimated 1 GoodSpectra and
Response
Channel #
Input, $EMSC_Z$.MAT
Output, DataCase=122, EMSC,Automatically estimated 1 GoodSpectra and
$x\ 10^{-3}$
Mean-Centred Response
Channel #



e=122 EMSC,Automatically estimated 1 GoodSpectra and 2 BadSpectra, before
after pre-treatment
Before, RMSEP(o), RMSEC(.)
PCR PC #
after, RMSEP(*),RMSEC(.)
PCR PC #
Cal. for y from input Z, r_CV=0.666
Y fitted=., full CV=o, using 1 PCs
Input y
Cal. for y after EMSC/EISC, r_CV=0.707
Y fitted=., full CV=o, using 1 PCs
Input y

Summary

Have shown:

MSC

EMSC (physical effects)

Summary

Have shown:

MSC

EMSC (physical effects)

SIS

EMSC (phys. effects + known chem. spectra, “good”/”bad”)

Summary

Have shown:

MSC

EMSC (physical effects)

SIS

EMSC (phys. effects + known chem. spectra, “good”/”bad”)

* EISC

Summary

Have shown:

MSC

EMSC (physical effects)

SIS

EMSC (phys. effects + known chem. spectra, “good”/”bad”)

* EISC

* Automatically *estimated* chem. spectra, (“good”/”bad”)

* Automatically *optimized* model spectra (ref, “good” or ”bad”) spectra

* *Re-estimated weights* for the LS parameter estimation in EMSC/EISC

Conclusions:

* Separated chemical and physical information
  by model-based pre-processing.
* Reduced required # of PCs in calibration model,
  from 5 to 1, which is expected chemically.
* Many different appearances of the same input data set,
  depending on the preprocessing.
  Rotations in a subspace!

Conclusions:

* Separated chemical and physical information
by model-based pre-processing.
* Reduced required # of PCs in calibration model,
from 5 to 1, which is expected chemically.
* Many different appearances of the same input data set,
depending on the preprocessing.
Rotations in a subspace!

http://www.models.kvl.dk/source/EMSCtoolbox/index.asp