Aalto University

OPEN ACCESS

This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Bertinetto, Carlo; Vuorinen, Tapani

How Does Computational Pre-processing Affect Spectral Analysis? An Investigation on Simulated Spectra

Published: 01/01/2013

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Bertinetto, C., & Vuorinen, T. (2013). *How Does Computational Pre-processing Affect Spectral Analysis? An Investigation on Simulated Spectra*. Poster session presented at Scandinavian Symposium on Chemometrics, Stockholm, Sweden.

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

How Does Computational Pre-Processing Affect Spectral Analysis? An Investigation on Simulated Spectra

Aalto University
School of Chemical Technology

Carlo G. Bertinetto, Tapani Vuorinen

Department of Forest Products Technology, School of Chemical Technology, Aalto University, P.O. Box 16300, 00076 Aalto, Finland

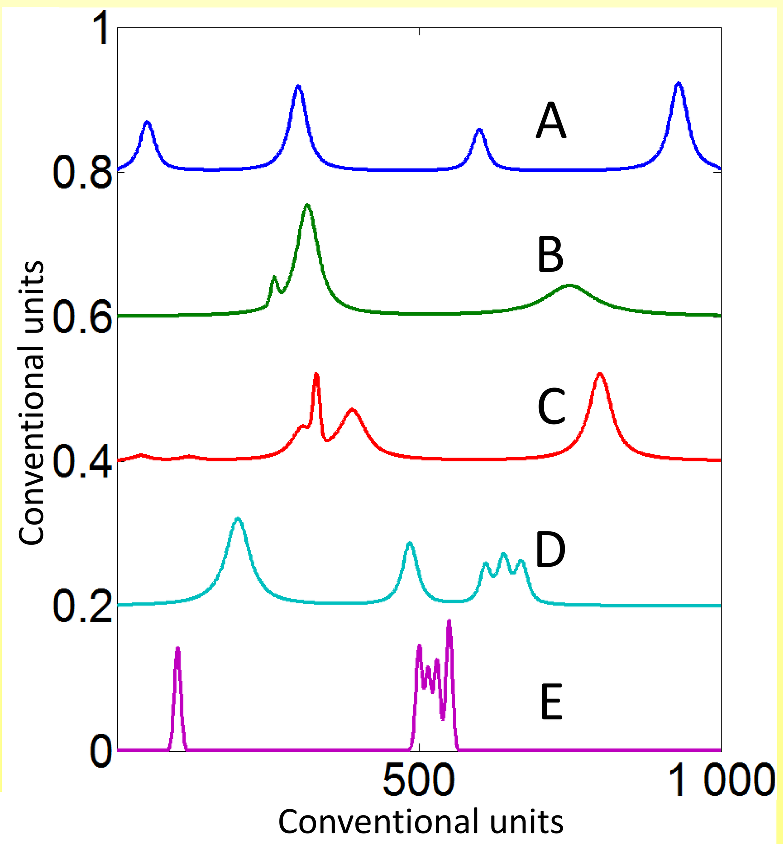
SCOPE OF THE WORK

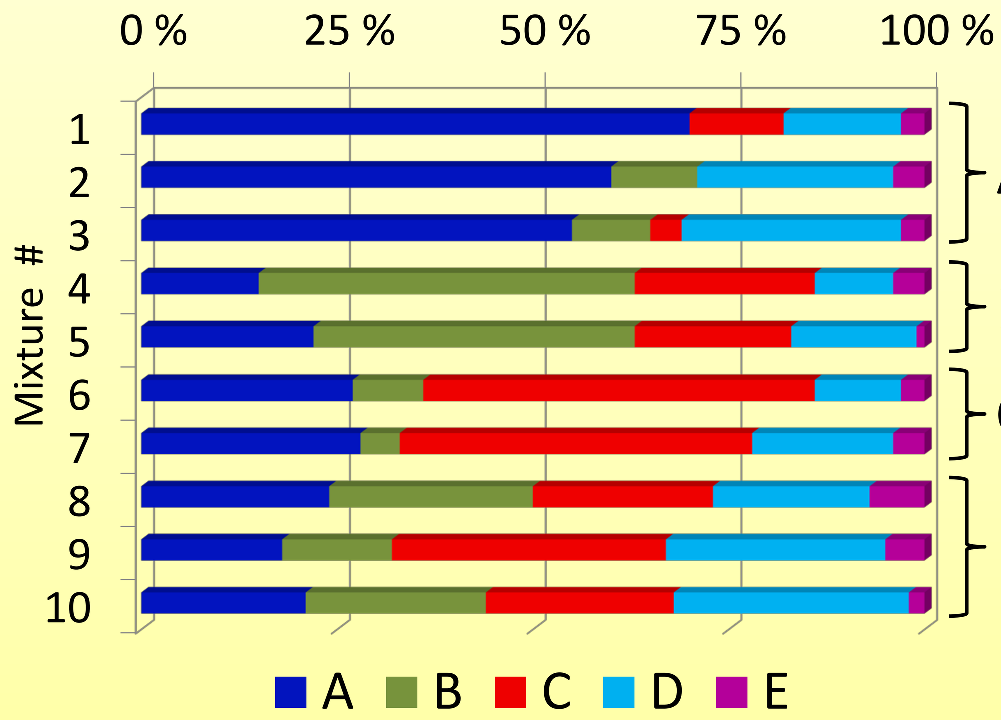
Spectral analysis is usually preceded by computational Pre-Processing (PP) to improve the signal and highlight the information of interest. However, little systematic study has been carried out on how the final result is affected by different PP methodologies, which are often chosen based on just common sense.

This work addresses this issue through a simulation experiment in which fictitious spectra, corrupted by different types of noise, were processed by various combinations of PP techniques. K-means cluster analysis was then used to recognize the original spectra.

SIMULATED SPECTRA

Spectra were generated by first mixing signals A-E in 10 different proportions…



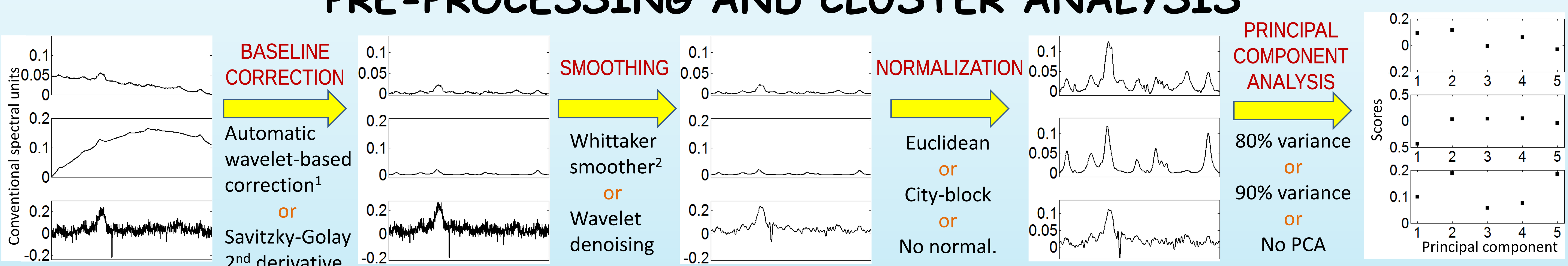


…and subsequently contaminating them with:

- **Baseline distortion**: linear or polynomial, baseline-to-signal ratio (BSR) from 0.5 to 8
- **Random noise**: signal-to-noise ratio (SNR) from 100 to 5
- **Intensity variation**: multiplication by 1, 2, 4 or 8

264 spectra were generated from each mixture, 2640 in total. These spectra simulate common issues encountered in spectral imaging.

PRE-PROCESSING AND CLUSTER ANALYSIS



All 216 combinations of the following techniques were tested (only one option per PP step is shown in this example figure)

K-MEANS CLUSTER ANALYSIS: Euclidean or City-block or Cosine

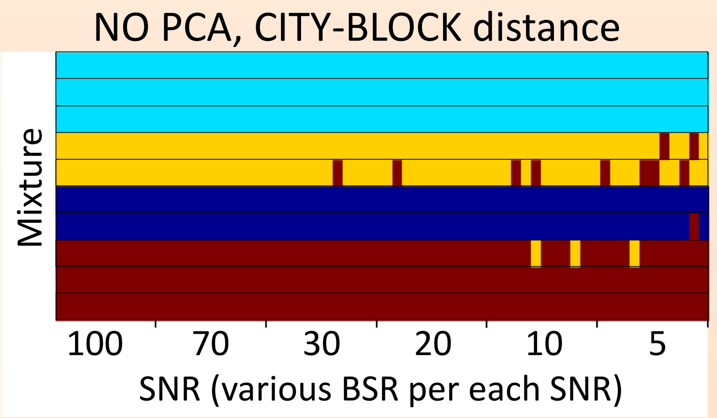
4 clusters or 10 clusters

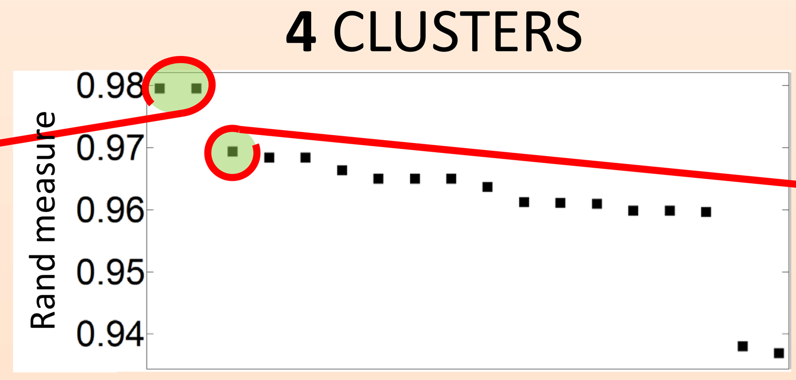
RESULTS

Best clusterings, according to the Rand statistic[3], visualized by color maps:
(a perfect clustering would consist of homogeneous horizontal stripes)

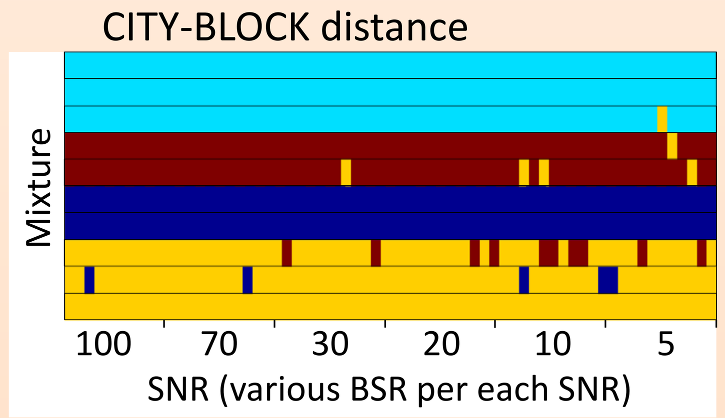
4 CLUSTERS

WAVELET baseline correction
CITY-BLOCK normalization
NO PCA, CITY-BLOCK distance



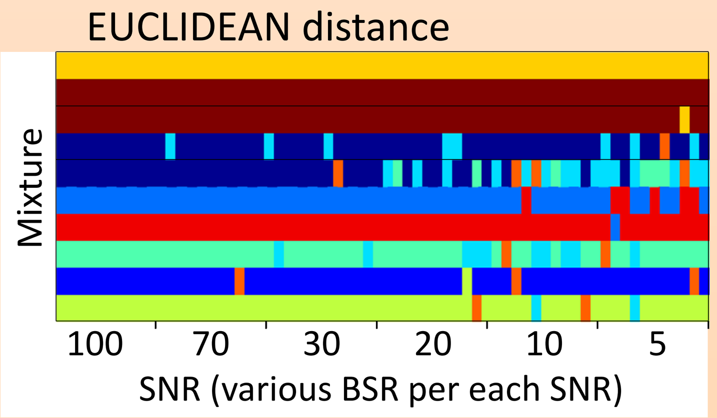


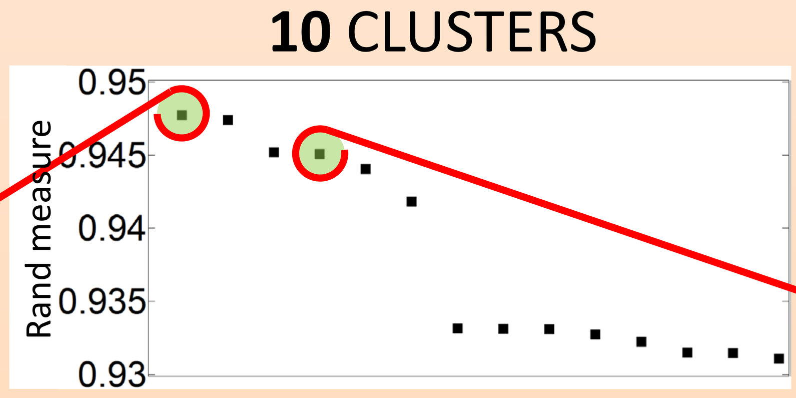
WAVELET baseline correction
EUCLIDEAN normalization
16 PCs (90% explained variance)
CITY-BLOCK distance



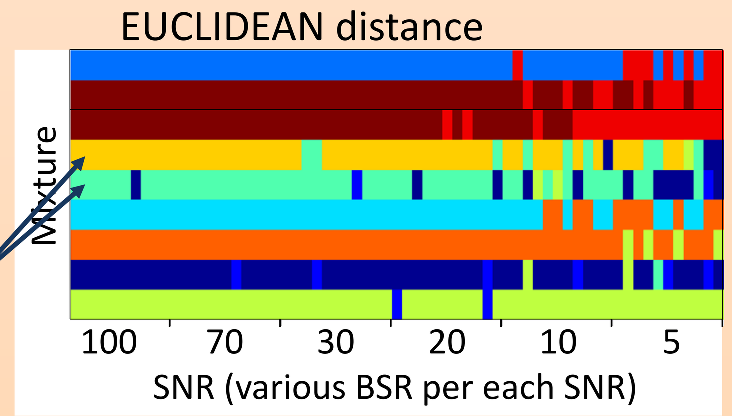
10 CLUSTERS

WAVELET baseline correction
EUCLIDEAN normalization
4 PCs (80% explained variance)
EUCLIDEAN distance





WAVELET baseline correction
CITY-BLOCK normalization
3 PCs (80% explained. variance)
EUCLIDEAN distance



Worse clustering of noisy spectra, but better distinction between mixtures 6-7 and higher silhouette coefficient

Further observations:

- No significant influence of the smoothing method was observed
- Good clustering can be achieved even when n. PCs < 5 (n. of components)
- Some PPs found in the literature[4] appear far from optimal in this study

CONCLUSIONS

For the proposed spectra, the best combinations of PP techniques to distinguish the relevant mixtures were found. The effect of each PP method was investigated.

These results can provide a guideline for a more effective spectral analysis and multivariate curve resolution.

REFERENCES

[1] C. G. Bertinetto, T. Vuorinen; *Appl. Spectr.* Submitted.
[2] P. H. C. Eilers; *Anal. Chem.* **2003**, 75(14): 3631-3636.
[3] W. M. Rand; J. Am. Stat. Ass. **1971**, 66(336): 846–850.
[4] P. N. Perera, M. Schmidt, P. J. Schuck, P. D. Adams; *Anal. Chim. Acta* **2011**, 702: 172−177.

Contact:
Carlo G. Bertinetto
Post-doc researcher
carlo.bertinetto@aalto.fi

13th Scandinavian Symposium of Chemometrics, June 17-20 2013, Stockholm archipelago, Sweden