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:
**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:

- A predominant
- B predominant
- C predominant
- D predominant
- E predominant
- Balanced

...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.

**RESULTS**

Best clusterings, according to the Rand statistic\(^3\), visualized by color maps:

(a perfect clustering would consist of homogeneous horizontal stripes)

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

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

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

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

**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**


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