

Multiscale interval PLS for spectral data modeling

Thursday, 5 June 2025 08:20 (20 minutes)

Process analytic technologies (PAT) are routinely used to rapidly assess quality properties in many industrial sectors. The performance of PAT-based models is, however, highly related to their ability to pre-process the spectra and select key wavebands. Amongst the modeling methodologies for PAT, partial least squares (PLS) (Wold, Sjöström and Eriksson, 2001) and interval partial least squares (iPLS) (Nørgaard et al., 2000) models coupled with well-known chemometric pre-processing approaches are the most widespread due to their ease of use and interpretability. As an alternative to classical pre-processing approaches, wavelet transforms (Mallat, 1989) provide a fast framework for feature extraction by convolution of fixed filters with the original signal. The proposed Multiscale interval Partial Least Squares (MS-iPLS) methodology aims to combine the ability of wavelet transforms for feature extraction with those of iPLS for feature selection. To achieve this, MS-iPLS makes use of wavelet transforms to decompose the spectrum into wavelet coefficients at different time-frequency scales, and, afterward, the relevant wavelet coefficients are selected using either Forward addition or Backward elimination algorithms for iPLS. As the wavelet filters are linear, the MS-iPLS model can also be equivalently expressed in the original spectral domain, and thus, the standard PLS approaches can be applied for the sake of interpretability and feature analysis.

In this study, 10 MS-iPLS models variants were constructed using five types of wavelet transforms and two iPLS selection algorithms and compared against 27 PLS benchmarks variants using different chemometric pre-processing and interval selection algorithms. The models were compared in two case studies, addressing a regression problem and a classification problem with real data. The results show that MS-iPLS models can either match or overcome the performance of the PLS benchmark models. For the regression problem, the PLS benchmark models were able to attain the lowest root mean squared error (RMSE), but their performance range was also wider, from an average RMSE of 0.11 (best model) to 2.46 (worst model), with most models being on the lower end. In contrast, the MS-iPLS models were consistently on the upper end, with an average RMSE ranging from 0.13 (best model) to 0.50 (worst model).

In the classification problem, MS-iPLS attained the best performance with an average accuracy of 92.7%, while the best PLS benchmark model had an average accuracy of 89.0%.

Similarly to the PLS benchmarks models, MS-iPLS still requires an exhaustive search for the optimal wavelet transform for each case study. However, with MS-iPLS the number of models to explore was significantly reduced (by a factor of 3, i.e., 1/3) without compromising on performance or interpretability.

Type of presentation

Contributed Talk

Primary author: KUHLMANN DE CANAVIRI, Lara (Fachhochschule Dortmund)

Presenter: KUHLMANN DE CANAVIRI, Lara (Fachhochschule Dortmund)

Session Classification: Fake Session

Track Classification: Spring Meeting