Direct Orthogonal Signal Correction

Chemometrics and Intelligent Laboratory Systems, 56, (2001), 13-25

The data used here to show the effect of DOSC preprocessing is the diesel fuel spectral data obtained from Eigenvector research.

http://software.eigenvector.com/Data/SWRI/index.html

We will use the viscosity data (viscgatest.zip). The data includes 20 high leverage samples (**_hl**) and the remaining samples are split into two random groups (**_ll_a** and **_ll_b**). These spectra can be used to test variable selection and calibration algorithms. For instance, you can use the high leverage samples and one of the other sets to make a calibration model (say the **_hl** and **_ll_a**), then test it on the third set (the **_ll_b**). In all cases the data have been pretty thoroughly weeded: outliers removed, and all samples belong to the same class (all summer fuels, no winter fuels).

The mfile **showresults** works out the following example.

Figure 1 shows the original centered data for the calibration set (left-top) and the test set (left-bottom). On the right side the DOSC corrected spectra are shown.

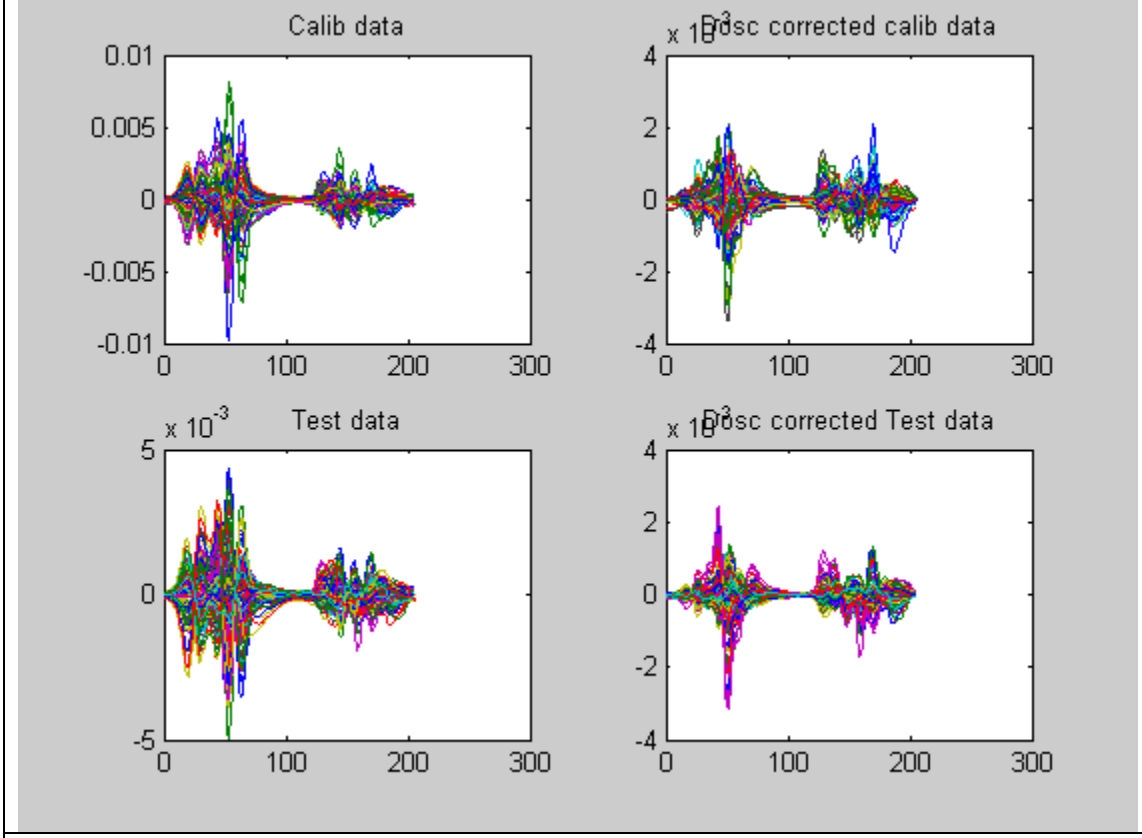

Figure 1: Diesel fuel data before (left) and after (right) DOSC preprocessing.

Looking at the correlation between each wavelength and the viscosity, we can see that this helped quite a bit to get rid of the non-relevant variation in the spectra. Figure 2 shows the correlation between each wavelength of the spectra and the viscosity before (blue, solid) and after (red, dotted) DOSC correction.

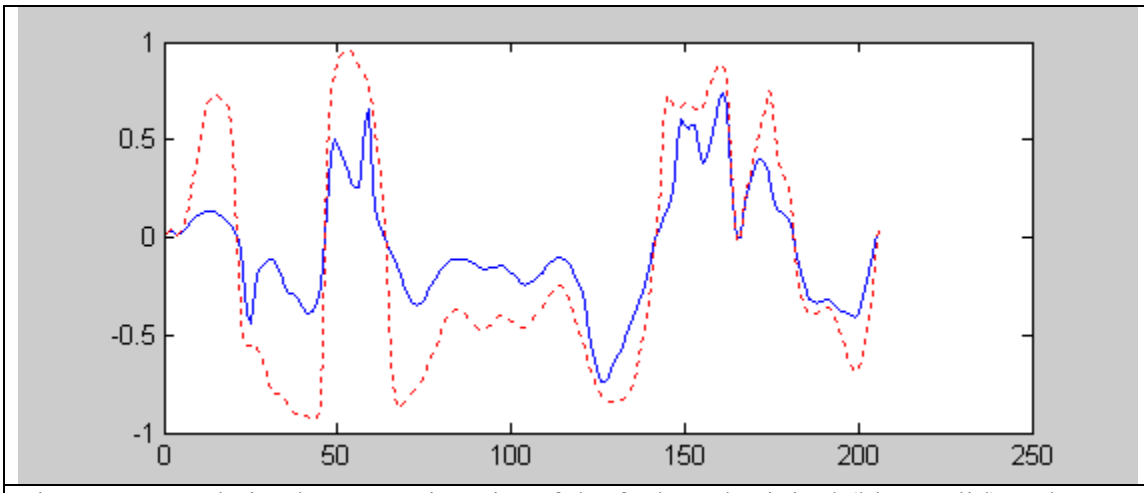

Figure 2: Correlation between viscosity of the fuels and original (blue, solid) and DOSC corrected (red, dotted) spectra.

Finally, using the DOSC scores we can monitor whether the DOSC correction for new samples was 'in control' (see Figure 3). One can draw control limits around the DOSC scores of the calibration samples and check whether the scores of the test samples are inside the limits.

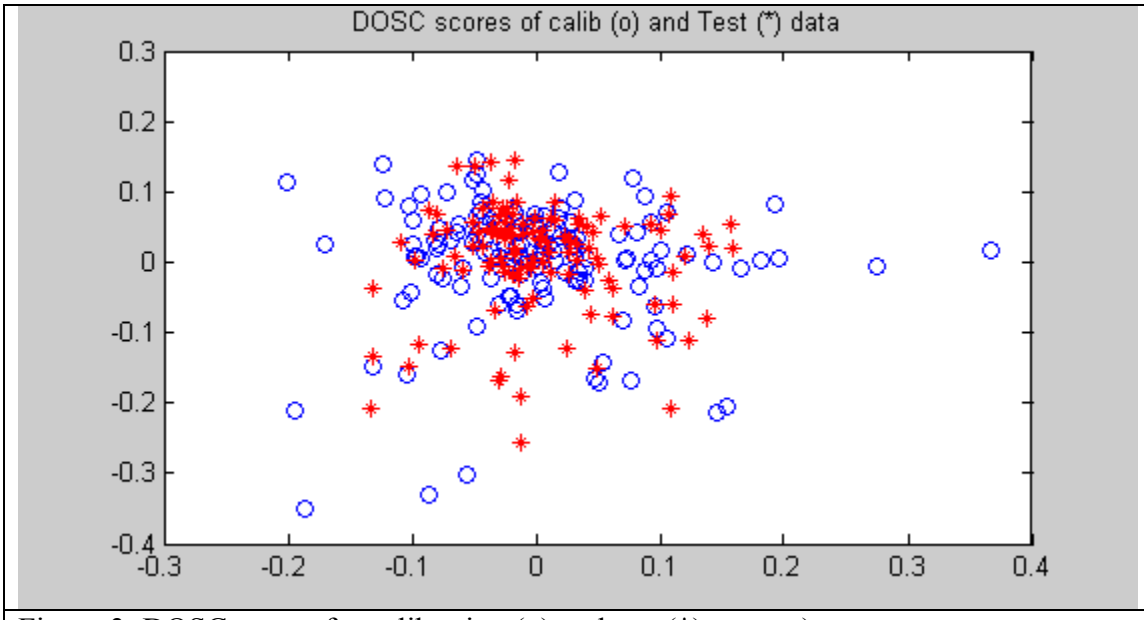

Figure 3: DOSC scores for calibration (o) and test (*) spectra).