doi: 10.1255/nirn.1331

Summary of the 2012 IDRC software shoot-out

Benoît Igne,[a] Paolo Berzaghi,[b] Dongsheng Bu,[c] Pierre Dardenne,[d] Peter Tillmann[e] and Mark Westerhaus[f]
[a]Duquesne University Center for Pharmaceutical Technology, Pittsburgh, PA, USA
[b]Università di Padova, Legnaro PD, Italy
[c]Bristol-Myers Squibb, New Brunswick, NJ, USA
[d]Walloon Agricultural Research Centre, Gembloux, Belgium
[e]VDLUFA Qualitätssicherung NIRS GmbH, Kassel, Germany
[f]Foss, Hillerød, Denmark

As one of the traditional events of the International Diffuse Reflectance Conference (IDRC), a biennial meeting which takes place in Chambersburg, Pennsylvania, USA, the software shoot-out is a great occasion to learn from and interact with experienced chemometricians presenting their approach to a common multivariate analysis problem. This 2012 edition challenged participants to develop a calibration model for an active pharmaceutical ingredient (Escitalopram) using samples generated on a laboratory scale and predicting test and validation tablets created at intermediate- and large-scale manufacturing. This shoot-out was highly representative of the work performed by pharmaceutical scientists and was made possible by the help of Professor Engelsen who generously donated the data.[1]

Near infrared (NIR) transmittance spectra were recorded from 4000 cm^{-1} to 14,000 cm^{-1}, with a resolution of 16 cm^{-1} and 128 co-added scans per sample. The spectrometer was an FT-NIR model MB-160 (ABB Bomem, Mannheim, Germany) equipped with a Tablet Samplir and an InGaAs detector. Background transmittance spectra were recorded using a Spectralon 99% standard reflectance tile and were used to convert the tablet spectra to absorbance units. Tablets contained the active ingredient, microcrystalline cellulose (about 80%), and minor components such as talc and magnesium stearate. Samples were evaluated for their content of the active ingredient.

Participants were provided with three data sets (calibration, test and validation) with reference measurements being available for calibration and test sets only. Contestants were to develop the best prediction model with the available data and send their predictions of the validation set to the shoot-out chair prior to the beginning of the conference. Criteria for deciding winners included prediction statistics of the validation set, novelty and uniqueness of the approach, and clarity of the presentation. In addition, the audience was asked to vote for their favourite approach.

Because of the availability of the data on the internet (www.models.kvl.dk/Tablets) some modifications were performed to the original data so that participants would not be in a position to easily find the solution to the problem. All spectral data were slightly trimmed (7645–10,500 cm^{-1}) and samples for each set were randomly selected amongst the available data for each manufacturing scale. Note that the origin of the data was not communicated until the day of the shoot-out session. The reference error was 3.5% of the active value; calibration and test data summary characteristics are presented in Table 1.

While the statistics appear to show that the data are quite similar, Figure 1 provides a whole new insight into the structure present in the various sets. Although the spectra show mainly two clusters, the principal component analysis of the calibration data and the projection of the test and validation sets onto that model show three to four clusters. These clusters corresponded to four different dosage values for the pharmaceutical drug. More precisely, each set was created with tablets of different sizes (90, 125, 188 and 250 mg) but presenting the same relative drug load. This variability in tablet shapes impacted on the sample pathlength and thus the spectral data. In addition, the test data were film-coated. Participants were not given this additional information about the data as the goal of the shoot-out was to challenge contestants to create precise and accurate models that would be robust to scale and manufacturing conditions. The approaches taken by five of the participants are presented here.

Participant 1

While a large variability existed in the data, this participant, along with all other participants, noted that most of the variance in the calibration data was between 8700 cm^{-1} and 8950 cm^{-1} as indicated on Figure 2. Figure 2 presents the data corrected for multiplicative scatter by MSC. Realising that pharmaceutical active ingredients often exhibit sharp peaks compared to the excipients and that the change in composition was mainly reflected in one spectral zone, Participant 1 decided to focus on this spectral range to develop his model. However, to confirm this hypothesis, a global PLS model was developed which revealed that the regression coefficients had large values in the previously-stated spectral range and coefficients close to zero elsewhere. Employing cross-validation during the PLS model development, the participant noticed a high degree of redundancy in the data which was evidenced by an error of cross-validation twice as small as the reference error.

To cope with this problem, various models were developed and optimised based on the prediction statistics of the test set. The final model was a multiple linear regression using only three variables selected from the range related to the active ingredient, after MSC correction and first derivative with gap and smoothing (four-point window).

Table 1. Calibration and test sets statistics.

		n	Min	Max	Mean	Std
Active (%, w/w)	Cal.	89	4.74	9.79	7.55	1.48
	Test	72	5.12	8.48	7.39	1.16

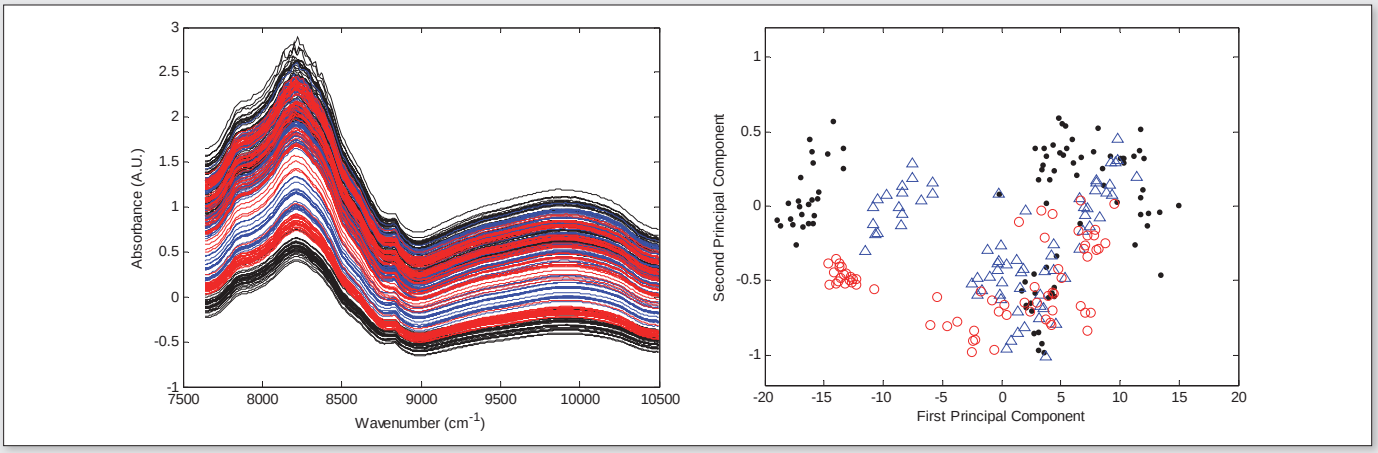

Figure 1. Visual depictions of the spectral data. Calibration samples in black, test samples in blue and validation samples in red. The PCA analysis was performed on mean-centred data.

Participant 2

After a visual evaluation of the data, MSC and first Savitzky–Golay derivative were chosen to reduce spectral variations arising from tablet thickness and hardness changes. A second step consisted in identifying regions of interest in the spectra. Spectral ranges 7752–8620 cm^{-1} and 7692–8000 cm^{-1} were pre-selected to cover the chemical absorption ranges of the active, magnesium stearate and microcrystalline cellulose. Finally, the model complexity was evaluated and four partial least squares (PLS) components were selected based on the predicted ability of the model in calibration and test along with an evaluation of the complexity of the formulation.

An optimisation routine was then employed to fine-tune the derivative settings and the variable ranges; this optimisation was based on the precision and accuracy of the test set only. Following the optimisation, a PLS model using four latent variables and spectral ranges from 8460 cm^{-1} to 8606 cm^{-1} and from 7722 cm^{-1} to 7937 cm^{-1} pretreated by MSC and first derivative (23-point window and third order polynomial) was used. To ensure adequate model robustness and limit over-fitting, the Hotelling's T^2, Q residual and uncertainty values of the validation data were monitored.

Participant 3

Using a principal component analysis (PCA) developed on the calibration set correlation matrix, the participant noticed that a significant number of samples from the validation set did not overlap with the calibration data, thus indicating a need to build a model with good extrapolation abilities.

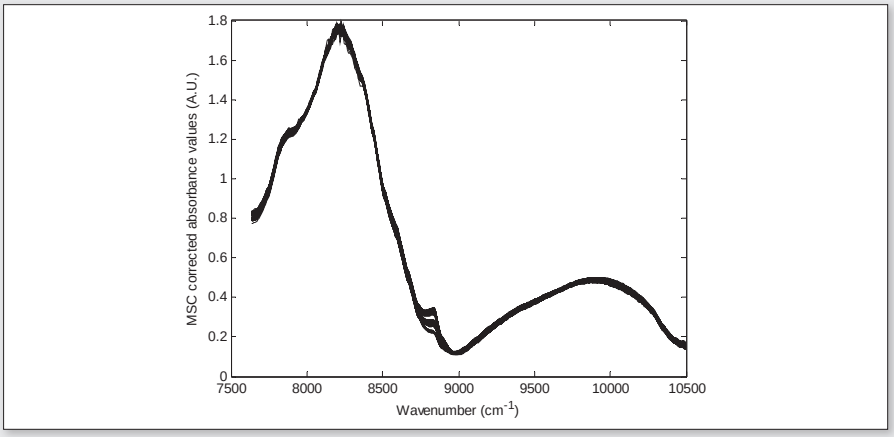

Figure 2. MSC-corrected calibration spectra.

In addition, after pre-treating the spectral data with MSC, the participant noticed the 8696–9091 cm^{-1} region as being the most informative. Finally, a study of the correlation between each variable and the active content showed that the absorption value at 8842 cm^{-1} had the highest correlation (0.987). That same variable had the largest spectral standard deviation, thus ensuring that it had the most information with minimum relative noise.

Using a forward multiple linear regression including the 8842 cm^{-1} variable, another variable was included in the model. However, this caused the error in the prediction of the test set to become larger. Thus, a manual search of the optimal pre-treatment was performed until the bias of test prediction was minimised. It was found that a one-variable model using MSC and a first derivative (Norris derivative) with 25 data points for the gap and 15 for the smoothing gave the best calibration and test statistics. Finally, a simple regression model with one term was used to predict the validation set.

Participant 4

After a visual inspection of the data, and recognising that the transmission measurement was responsible for pathlength differences, the standard normal variate (SNV) procedure was used to correct the data, being aware that, because of the large scatter of data, the mean used by MSC would not be meaningful. A small spectral region was then selected (8888–8810 cm^{-1}).

A preliminary PLS calibration limited to this region gave a root mean square error (*RMSE*) on the test set close to double the standard error of the reference method. It could also be seen from the PLS output that a model with one factor was almost as good as a model with four factors (chosen

by the chemometric software). This was a reasonable observation given that the product analysed was primarily a mixture of cellulose and the active ingredient. Therefore a simple model with one or two factors should be sufficient to describe the system. A model using MLR confirmed this decision; one wavelength was enough to achieve precision similar to that of any PLS model. The selected wavelength was similar to that chosen by participant 3 (8842 cm^{-1}).

However, to ensure model transferability and its ability to provide diagnostics, PLS was preferred to MLR as it allowed comparison of the Mahalanobis distance between sample sets to ensure that the model was not overfitting. Keeping in mind that the calibration set, test and validation sets came from different tablet presses (laboratory size, intermediate and industrial size) this proved to be a good indicator of the model transferability.

Participant 5

In addition to noticing the API peak, the major offset within the datasets and the smaller offsets due to scattering, this participant compared the distribution of reference values with respect to spectral grouping. Histograms of the lab values exhibited clusters which were not related to the spectral groups, thus indicating that pathlength differences had to be corrected. Although the low and high absorption regions of the spectra were noisy, the region of the absorption peak had low noise. Thus, a full spectrum scatter correction would not provide the best spectral data. A weighted MSC pretreatment, which excluded spectral regions with high absorption peak variation, was used.

While considering using a multiple linear regression (MLR) model, the participant decided to build PLS models based on the variables selected for MLR plus a few neighbours; in the participant's stated experience, such methods usually perform better than MLR. An exhaustive search of simple one- and two-factor PLS models using two short spectral regions was performed and it was found that a one-factor model using absorption values at 9118 cm^{-1}, 9110 cm^{-1}, 9103 cm^{-1} and 8879 cm^{-1} performed well on the calibration and test sets.

A second method was investigated which used division by the difference between the transmittance values at two variables to remove the scale variation among the spectra. A search was made to find the best two data points for the divisor while fitting a two data-point MLR model to the resulting spectra. The best performing model on the calibration and test sets used MLR with variables 9018 cm^{-1} and 8918 cm^{-1} after dividing by the difference between the transmittance values at 8810 cm^{-1} and 8856 cm^{-1}. This model performed better than the weighted MSC and four data-point PLS model. Predictions on the validation set were the average of the predicted values from the two models.

In order to determine if the prediction errors were due to the incorrect model or to reference method errors, the calibration set was condensed. Using the spectral and reference values grouping information, 11 groups with seven or eight samples per group were identified. Spectral data and reference values within each group were averaged to get 11 condensed spectra and reference values. Using the first modeling technique, the root mean square error of prediction (*RMSEP*) was reduced by a factor of 3.2 with slightly better performance on the test set, indicating that most of the observed error was due to laboratory method errors.

Table 2. Validation statistics.

Participant	1	2	3	4	5
RMSEP	0.41	0.44	0.44	0.48	0.43
SEP	0.41	0.43	0.42	0.42	0.40
Bias	–0.04	0.07	0.14	–0.23	0.16
r^2	0.86	0.84	0.85	0.86	0.86

Results

Table 2 presents the validation results for each participant. *RMSEP*, standard error of prediction (*SEP*) and bias are presented along with coefficient of determination.

The five participants chose quite similar approaches to get results that were very close. With overall best statistics and the public votes, participant 1 won the 2012 IDRC Shoot-Out, followed by participant 5 and 2. However, the two other participants had very similar statistics.

The data are available on the IDRC website (www.idrc-chambersburg.org/index.html). The authors would like to thank the 2012 IDRC chair Dr David Burns and the Council for Near-Infrared Spectroscopy for providing funding and support for the conference. The next conference will take place in Chambersburg from 2 to 8 August 2014.

Reference

1. M. Dyrby, S.B. Engelsen, L. Nørgaard, M. Bruhn and L. Lundsberg Nielsen, "Chemometric quantitation of the active substance in a pharmaceutical tablet using near infrared (NIR) transmittance and NIR FT Raman spectra", *Appl. Spectrosc.* **56(5),** 579–585 (2002). doi: 10.1366/0003702021955358