doi: 10.1255/nirn.1216

The 2010 IDRC software shoot-out at a glance

Benoit Igne,[a] Pierre Dardenne,[b] David Honigs,[c] J. Todd Kuenstner,[d] Karl Norris,[e] Zhenqi Shi[f] and Mark Westerhaus[g]
[a]Duquesne University Center for Pharmaceutical Technology, Pittsburgh, PA, USA
[b]Walloon Agricultural Research Centre, Gembloux, Belgium
[c]Perten Inc, Springfield, IL, USA
[d]Charleston Area Medical Center, Charleston, WV, USA
[e]Consultant, 11204 Montgomery Road, Beltsville, MD, USA
[f]Strategic Process Control, Pittsburgh, PA, USA
[g]Foss, Hillerød, Denmark

As part of the many activities proposed to attendees of the International Diffuse Reflectance Conference (IDRC) taking place biennially in Chambersburg, Pennsylvania, USA, the software shoot-out is always a great occasion to listen to and interact with experienced chemometricians presenting their approach to a common problem. This 2010 IDRC challenge asked participants to develop calibration models for three human blood constituents: haemoglobin, glucose and cholesterol. This biomedical challenge, a cutting edge research area for numerous laboratories throughout the world, was made possible by the help of Karl Norris and Dr J. Todd Kuenstner (Charleston Area Medical Center, West Virginia, USA) who generously donated the data.[1–4]

Blood samples were analysed during the period from 1990 to 1992 with a NIRSystems 6500 spectrometer (Foss NIRSystems, Laurel, MD, USA) equipped with a transmission amplifier mounted in the sample transport accessory. The instrument was configured to a vertical light-path mode using a platform supplied by the manufacturer. This made it possible to use a simple sample presentation for either transmission or reflection measurements. The sample cell was a 2 cm diameter stainless steel cylinder with a quartz window. For transmission measurements, 200 µL of whole blood was transferred from a pipette to the sample cell, providing a sample thickness of 0.6 mm. For reflection measurements, the cell was filled to a sample thickness of at least 2 mm to provide a sample of infinite thickness for the NIR spectral region. The empty cell was used as a reference for transmission measurements and a ceramic reference standard for reflection measurements.

Participants were provided with three data sets for each measurement geometry (calibration, validation and test). Reference measurements were available for calibration and validation sets only. Participants were to develop the best prediction models with the available data and send their predictions of the test set to the shootout chair prior to the beginning of the conference. Criteria for deciding winners included overall prediction statistics of the test sets (both reflection and transmission), novelty and uniqueness of the approach, and clarity of the presentation.

All spectra (reflection and transmission) had 700 variables, from 1100 nm to 2498 nm, with a 2 nm interval. Table 1 describes calibration and validation sample statistics.

The trained chemometrician would notice that calibration and validation sets were suspiciously similar in terms of their distributions. Both sets were actually created from the same data set (fresh blood) while spectra of the test sets were samples frozen for storage and thawed before analysis. This choice was made to test participants' models for robustness. Figure 1 shows spectral data and score plot of all three sets for the transmission data.

Because the data sets were primarily designed for the study of the ability of NIR spectroscopy to predict haemoglobin, only haemoglobin predictions on the test sets were considered when judging calibration model statistics. The approaches taken by the four participants are presented here.

Participant 1

The study of the structure of the data was at the centre of the strategy used by this participant. After recognising that the calibration and validation samples came from the same data set and, more particularly, that some samples were present in duplicate or even triplicate, a reorganisation of the data was performed. Calibration and validation sets were reorganised: for calibration and validation, unique samples and one of the repeated scans for a given sample were used in their respective set while repeated samples were used in a repeatability file (combining calibration and validation).[5] The purpose of this file is to develop an equation that gives the same predicted value across all the conditions represented in the scans.

An automatic screening of the most efficient pre-treatment was carried out by searching six wavelengths for the development of stepwise multiple linear regres-

Table 1. Calibration and test sets statistics.

		n		Min		Max		Mean		Std	
		Refl.	Trans.	Refl.	Trans.	Refl.	Trans.	Refl.	Trans.	Refl.	Trans.
Haemo. ($g\,dL^{-1}$)	Cal.	173	143	10.3	10.6	17.3	17.3	13.8	13.6	1.7	1.6
	Val.	58	47	10.6	10.3	17.6	17.3	13.6	13.9	1.4	1.6
Gluc. ($mg\,dL^{-1}$)	Cal.	173	143	46.0	46.0	457.0	457.0	92.0	90.9	56.6	54.2
	Val.	58	47	46.0	46.0	159.0	303.0	82.0	96.4	21.2	44.0
Cholest. ($mg\,dL^{-1}$)	Cal.	173	143	99.0	99.0	358.0	358.0	209.3	210.8	48.5	45.7
	Val.	58	47	99.0	99.0	358.0	358.0	216.3	223.0	46.2	49.9

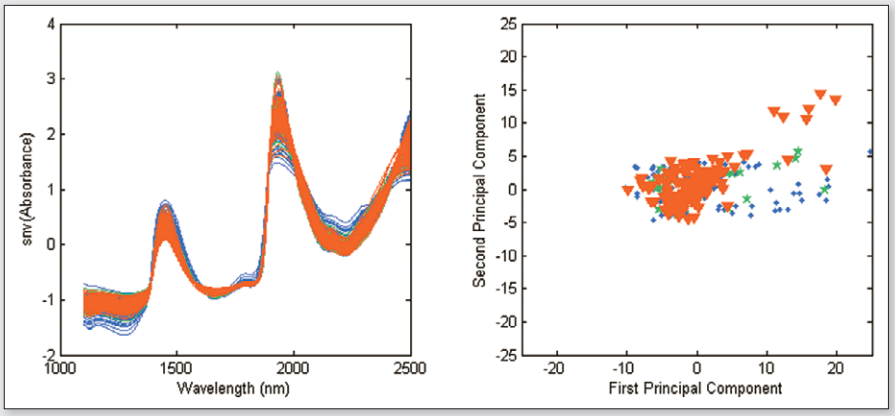

Figure 1. Visual depictions of the spectral data in transmission. Orange: the test set; green: validation set; blue: calibration set.

sion (MLR) models using 60 different pre-processing techniques. This regression method was chosen to further enhance model robustness and extrapolation abilities. The combination standard normal variate[6] + detrend[6] + Norris derivative[7] was found as the best pre-treatment on the validation set. The selected wavelengths were tuned by "manual" step-up MLR using permutations of the terms and by looking not only at the minimum *SEP* but at the highest F-test of the B coefficients: those selected were 1286 nm, 1626 nm, 2042 nm, 2058 nm, 2182 nm and 2222 nm. None of the selected wavelengths belonged to water bands. This strategy was applied for both reflectance and transmittance geometries.

Participant 2

After pre-processing the spectral data using harmonisation (Perten Instruments proprietary method), it was noted that for reflection data, an instrumental effect was most likely responsible for distorting the water peak at 1930 nm, compared to the 1450 nm water band. This squashing was not as severe for transmission data. To avoid these distortions, spectral data were limited to the range 1100–1828 nm.

The calibration data were then regressed using the Honigs Regression technique as provided by Perten Instruments. The Honigs Regression is a learning technique that adapts to trends or changes in sample spectra. In this particular case, there was not much difference between standard PLS and the Honigs regression technique, both in the specific equation generated and the prediction results. The dominant error source in the data seemed to be the lab value accuracy or the preparation of the samples. When this is the case, every regression technique tends to give similar results. The other reason the HR performs similarly to PLS is that the spectra are relatively simple and behave normally. Finally, the calibration set was static and used at one time. Under these conditions the learning characteristics of the HR simply do not come into play.

Participant 3

Recognising the need to mitigate interferences from water on the spectra, participant 3 decided to first investigate the use of Extended Multiplicative Scattering correction[8] and several orthogonalisation techniques [net analyte signal,[9] orthogonal signal correction[10] and region orthogonal signal correction (ROSC)[11]]. Second, an original calibration procedure was used to obtain the final model; (1) a moving window-based PLS (MWPLS)[12] was applied to select the optimal wavelength regions for subsequent PLS modelling; (2) separate PLS models were built for individually-selected wavelength regions; (3) a stepwise linear regression was performed on the predicted concentrations from separate PLS models on individual wavelength regions to select the predicted concentrations showing the significant correlation with the haemoglobin concentration; and (4) a final MLR model was built based on previously selected predicted concentrations.

For reflection spectra, EMSC was chosen to mitigate water interferences. After combining EMSC with first derivative, the MWPLS indicated three wavelength regions for ideal model performance on haemoglobin—these were 1120–1316 nm, 1520–1920 nm and 2000–2390 nm. Based on stepwise linear regression, the predicted concentration from the wavelength region 1520–1920 nm was rejected because of its insignificance ($\alpha = 0.05$). The predicted concentrations from the remaining two wavelength regions were retained in the final MLR model.

For transmission spectra, ROSC with 10 components removed was used as the optimal pre-processing. After combining ROSC with SNV as the spectral pre-treatment, MWPLS indicated three optimal wavelength regions for modelling haemoglobin, similar to those found in reflection. After building separate PLS models for individual wavelength regions, the predicted concentrations from all three wavelength regions were proved significant by stepwise linear regression. A final MLR model using the predicted concentrations from all three regions was established.

Participant 4

A variable selection approach was also chosen by participant 4. The goal was to develop models that would be simple enough to avoid overfitting the calibration data, extrapolate well to the test set and yet be accurate enough to be useful. First, through a careful evaluation of the spectral data, participant 4 noticed that spectral offset and scale effects were responsible for much of the variability. Also, the use of third derivative showed non-random patterns above 2430 nm. Finally, an instrumental effect resulting in squashing the water bands was observed. This squashing was not as severe for transmission data, but additional noise was observed between 1870 nm and 2050 nm. Water bands and noisy regions were then removed from the spectral analysis.

A division approach was chosen to remove the scale effect. Since there was also an offset effect, a scale correction that was independent of the offset was also needed. The difference between two wavelengths was chosen as the divisor. For the numerator, three individual wavelengths were chosen. The numerator and denominator wavelengths were chosen to be close to each other because, through experience, such data are better suited for linear models. To select the wavelengths, a program to do an all possible combination of wavelengths search within 100 wavelengths windows was created. The program targeted denominator differences sufficiently far from zero and ensured that the numerator wavelengths were separated by at least three

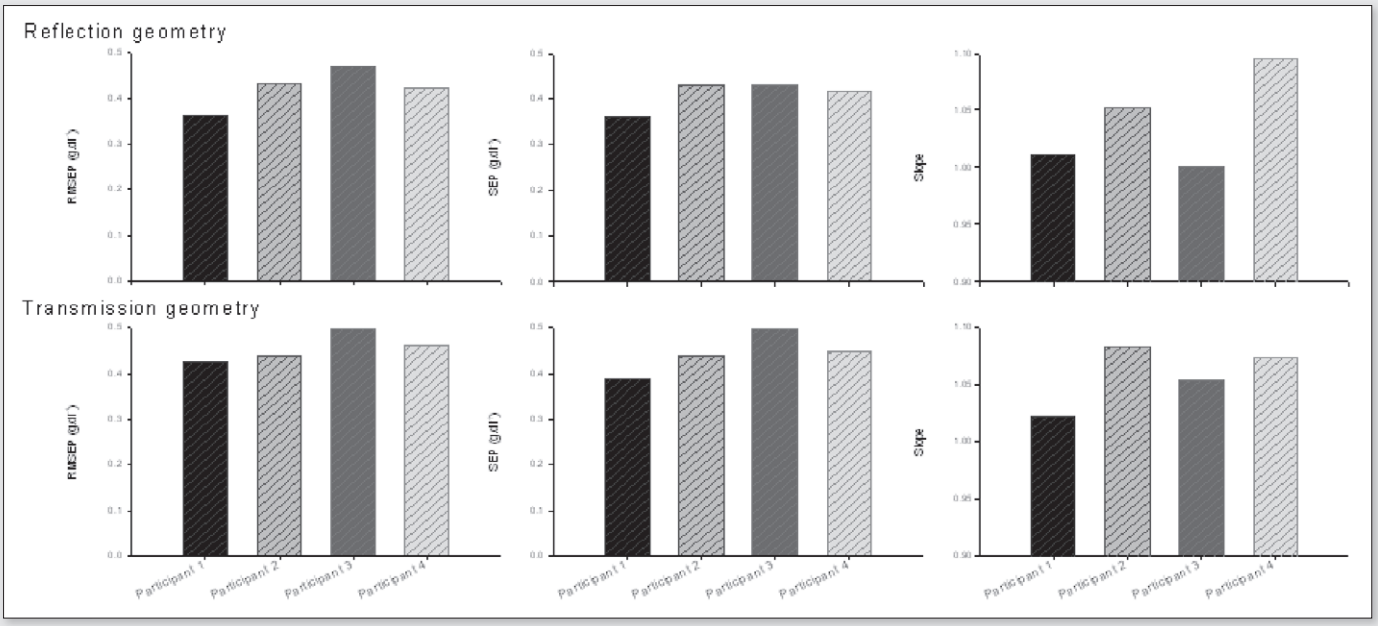

Figure 2. Test statistics for both reflection and transmission geometries.

wavelengths. Even though many potential models were evaluated (which is one way to overfit data), they were all simple models. The lack of overfitting was also supported by monitoring the wavelength search and seeing that similar models had similar accuracy. The model was finally enhanced by switching from three-wavelength MLR to three-factor PLS, and including neighbouring wavelengths that improved the fit. The idea here was that PLS would replace individual wavelengths with small wavelength regions, providing the benefits of averaging or smoothing the data while maintaining the model accuracy.

Results

Figure 2 presents test results for each participant. Root mean squared error of prediction (*RMSEP*) and standard error of prediction (*SEP*) values are presented along with prediction slopes as these three statistics indicated conflicting results for some of the approaches (good accuracy but large slope).

The four participants chose very different approaches to get overall results that were very similar. The more parsimonious models (MLR based models) seemed to be outperforming full spectrum methods in terms of accuracy and precision. However, the effect of the different approaches on the slopes varied greatly and was not always consistent through sampling geometries. With overall best statistics and the public votes, participant 1 won the 2010 IDRC Shoot-Out, followed by participant 4 and 2.

The data are available on the IDRC website (http://www.idrc-chambersburg.org). The authors would like to thank the 2010 IDRC chair Dr Charles Miller and the Council for Near-Infrared Spectroscopy for providing funding and support for the conference.

References

1. J.T. Kuenstner and K.H. Norris, "Spectrophotometry of human haemoglobin in the near infrared region from 1000 to 2500 nm", *J. Near Infrared Spectrosc.* **2,** 59–65 (1994).
2. J.T. Kuenstner, K.H. Norris and W.F. McCarthy, "Measurement of hemoglobin in unlysed blood by near-infrared spectroscopy", *Appl. Spectrosc.* **48(4),** 421 (1994).
3. J.T. Kuenstner and K.H. Norris, "Near infrared haemoglobinometry", *J. Near Infrared Spectrosc.***3,** 11–18 (1995).
4. K.H. Norris and J.T. Kuenstner, "Rapid measurement of analytes in whole blood with NIR transmittance", in *Leaping Ahead with Near Infrared Spectroscopy*, Ed by G.D. Batten, P.C. Flinn, L.A. Welsh and A.B. Blakeney. NIR Spectroscopy Group, RACI, Australia, pp. 431–436 (1995).
5. M.O. Westerhaus, "Improving repeatability of NIRS calibrations across instruments", *Proceedings of the Third International Conference on Near Infrared Spectroscopy*, Ed by R. Biston and N. Bartiaux-Thill. Agricultural Research Centre, Gembloux, Belgium, pp. 671–674 (1991).
6. R.J. Barnes, M.S. Dhanoa and S.J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra", *Appl. Spectrosc.* **43(5),** 772–776 (1989).
7. K.H. Norris, "Applying Norris derivatives. Understanding and correcting the factors which affect diffuse transmittance spectra", *NIR news* **12(3),** 6–9 (2001).
8. H. Martens and E. Stark, "Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy", *J. Pharm. Biomed. Anal.* **9(8),** 625–635 (1991).
9. R.P. Cogdill, T. Herkert, C.A. Anderson and J.K. Drennen, "Synthetic calibration for efficient method development: analysis of tablet API concentration by near-infrared spectroscopy", *J. Pharm. Innov.* **2,** 93–105 (2007).
10. T. Fearn, "On orthogonal signal correction", *Chemometr. Intell. Lab. Syst.* **50,** 47–52 (2000).
11. P. Du Yi, Y.Z. Liang, S. Kasemsumran, K. Maruo and Y. Ozaki, "Removal of interference signals due to water from *in vivo* near-infrared (NIR) spectra of blood glucose by region orthogonal signal correction (ROSC)", *Anal. Sci*. **20(9),** 1339–1345 (2004).
12. J.H. Jiang, R.J. Berry, H.W. Siesler and Y. Ozaki, "Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data", *Anal. Chem*. **74(14),** 3555–3565 (2002).