

# International Diffuse Reflectance Conference 2010

## Chemometrics ShootOut Rules of Procedure

May 2010

The dataset for the IDRC ShootOut 2010 was graciously provided by Karl Norris. It is a biomedical application where blood samples were analyzed during the time period from 1990 to 1992 with an NIRSystems 6500 spectrometer with a transmission amplifier mounted in the sample transport. 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  $\mu$ l 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 was used for reflection measurements. Blood samples were then evaluated for their content in Hemoglobin, Glucose, and Cholesterol.

This year's challenge will not only test participant's ability to develop the best possible models, in reflectance and transmittance for the three blood constituents, but also require them to guide the judges and audience through their reasoning during the calibration development effort. Participants are to:

- 1) Develop best possible models in transmittance for Hemoglobin, Glucose, and Cholesterol on XcalTrans
  - a. Validate on XvalTrans (we provide reference values)
  - b. Test on XtestTrans (we do NOT provide reference values)
- 2) Develop best possible models in reflectance for Hemoglobin, Glucose, and Cholesterol on XcalReflect
  - a. Validate on XvalReflect (we provide reference values)
  - b. Test on XtestReflect (we do NOT provide reference values)
- 3) Transfer calibration model for Hemoglobin only from transmittance to reflectance and from reflectance to transmission. **This is an optional exercise and will not be considered to determine winners.**
  - a. Test on dataset XtestReflect for transmittance model
  - b. Test on dataset XtestTrans for reflectance model
- 4) Detail their reasoning when selecting pre-treatment methods.

- 5) Reflect on which approach should be used to develop calibration models for blood constituents with respect to calibration statistics, their knowledge of the two different measurement geometries, and patient consideration.
- 6) Reflect on why hemoglobin content is so much easier to predict than other two constituents, based on provided data (pure components included).

Participants who wish to compete for prizes **must submit** their predictions of XtestTrans and XtestReflect (**by July 30, 2010**) in an EXCEL (\*.xls) file to

Fred McClure  
Email: [fmac@nc.rr.com](mailto:fmac@nc.rr.com)

**AND**

Benoit Igne  
Email: [igneb@duq.edu](mailto:igneb@duq.edu)

Criteria for deciding winners include: (1) Prediction statistics of the test sets, (2) novelty, uniqueness and clarity of the presentation, (3) timing (staying within time assigned), (4) software evaluation (why you chose the software over other options), and (5) quality of answers to questions 3,4, and 5. An audience vote will be taken and the results of this vote will be considered by the judges for determining the winners. Winners will be announced during the banquet on Thursday night. Prizes this year will be as follows: 1<sup>st</sup> Prize: \$200, 2nd Prize: \$100, 3<sup>rd</sup> Prize: \$50.

***Submissions received after July 31, 2010 will not be eligible for prizes, although they may be presented at the Conference, at the discretion of the organizers.*** Decisions of the judges are final.

To ensure consistency among participants, participants are asked to report the following calibration and validation statistics:

1. Coefficient of determination
2. Root mean square error of calibration/cross-validation/prediction
3. Standard error of calibration/ cross-validation/ prediction
4. Bias of calibration/ cross-validation/ prediction

To determine test set statistics, judges will use the following definitions of the above terms:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y} - y)^2}{n}} \quad SE = \sqrt{\frac{\sum_{i=1}^n (\hat{y} - y)^2 - \frac{\left(\sum_{i=1}^n \hat{y} - y\right)^2}{n}}{n-1}} \quad Bias = \frac{\sum_{i=1}^n (\hat{y} - y)}{n}$$

There have been four publications in which some of the spectra were used:

J.T. Kuenstner, and K.H. Norris, 1994. Spectrophotometry of human hemoglobin in the near infrared region from 1000-2500nm, *J. Near Infrared Spectroscopy*, 2,59-65.

J.T. Kuenstner, K.H. Norris, and W.F. McCarthy, 1994. Measurement of Hemoglobin in Unlysed Blood by Near-Infrared Spectroscopy, *Applied Spectroscopy*, 48, 4.

J.T. Kuenstner and K.H. Norris, 1995. Near infrared hemoglobinometry. *J. Near Infrared Spectroscopy*, 3, 11-18.

K.H. Norris, and J.T. Kuenstner, 1995. Rapid measurement of analytes in whole blood with NIR transmittance. In: *Leaping Ahead with Near Infrared Spectroscopy*. (ed. G.D. Batten, P.C. Flinn, L.A. Welsh, and A.B. Blakeney), Published by the NIR Spectroscopy Group, 1/21 Vale Street, Melbourne, Victoria, Australia. 431-436.1995. (Proceedings of the 6th International Conference on Near Infrared Spectroscopy)

Information about the data:

All spectra have 700 variables, from 1100 to 2498 nm, with a 2 nm interval.

**Whole Blood- In reflectance:**

		<b>n</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Std</b>
<b>Hemo.</b>	Cal	173	10.30	17.30	13.81	1.71
	Val	58	10.60	17.00	13.67	1.49
<b>Gluc.</b>	Cal	173	46.00	457.00	92.04	56.65
	Val	58	46.00	159.00	82.00	21.26
<b>Cholest.</b>	Cal	173	99.00	358.00	209.28	48.50
	Val	58	99.00	358.00	216.26	46.27

Additionally, 194 samples are available for the blind test and 10 for standardization.

XcalReflect: spectral data – calibration  
XvalReflect: spectral data – validation  
XtestReflect: spectral data – test

YcalReflect: reference data – calibration  
YvalReflect: reference data – validation

XstdReflect: spectral data – standardization  
YstdReflect: reference data – standardization

### Whole Blood- In transmittance:

		<b>n</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Std</b>
<b>Hemo.</b>	Cal	143	10.60	17.30	13.68	1.64
	Val	47	10.30	17.30	13.94	1.65
<b>Gluc.</b>	Cal	143	46.00	457.00	90.95	54.29
	Val	47	46.00	303.00	96.45	44.09
<b>Cholest.</b>	Cal	143	99.00	358.00	210.78	45.75
	Val	47	99.00	358.00	222.98	49.99

Additionally, 138 samples are available for the blind test and 10 for standardization.

XcalTrans: spectral data – calibration  
XvalTrans: spectral data – validation  
XtestTrans: spectral data – test

YcalTrans: reference data – calibration  
YvalTrans: reference data – validation

XstdTrans: spectral data – standardization  
YstdTrans: reference data – standardization

Transmittance spectra of several pure components known to be present in whole blood are also provided, to aid in the development and evaluation of models:

Water  
Urea  
Globulin  
Albumin  
Cholesterol  
Hemoglobin

When no sample name is available in files, pure spectra are stored in the above order.

Additional Information:

1. Transfer samples (XstdTrans and XstdReflect), scanned on both reflectance and transmittance, are provided to allow participants to perform the transfer of calibration models from transmittance to reflectance and from reflectance to transmittance.
2. Pure spectra of the main component of blood are provided. They were scanned in transmittance only.
3. Datasets were provided in different format to hopefully accommodate most of the participants' software.

3.1. MATLAB format

The folder named IDRC2010MATLAB contains datasets saved in MATLAB format. Files names are as follows:

IDRCShootOut2010Reflect.mat  
IDRCShootOut2010Transmit.mat  
IDRCShootOut2010PureCompTransmit.mat

3.2. The Unscrambler format

The folder named UNS contains datasets saved in The Unscrambler format. Files names are as follows:

IDRC2010ReflectUNC.00D  
IDRC2010TransmitUNC.00D  
IDRC2010TransPureCompUNC.00D

3.3. JCAMP-DX format

The folder named JCAMP-DX contains datasets saved in JCAMP-DX format. Files names are as follows:

IDRC2010calReflect.JDX  
IDRC2010valReflect.JDX  
IDRC2010testReflect.JDX  
IDRC2010stdReflect.JDX  
  
IDRC2010calTransmit.JDX  
IDRC2010valTransmit.JDX  
IDRC2010testTransmit.JDX  
IDRC2010stdTransmit.JDX  
  
IDRC2010TransPureComp.JDX

### 3.4. NSAS

The folder named NSAS contains datasets saved in NSAS format. Files names are as follows:

ReflectCal.da  
ReflectVal.da  
ReflectTest.da  
ReflectStd.da

TransCal.da  
TransVal.da  
TransTest.da  
TransStd.da

TransPure.da

For additional information or comments, please contact Fred McClure ([fmac@nc.rr.com](mailto:fmac@nc.rr.com))  
**and** Benoit Igne ([igne@duq.edu](mailto:igne@duq.edu)).