Calibration transfer in near infrared spectroscopy

P. Dardenne[a]

[a] Département Qualité des Productions agricoles - Centre de Recherches Agronomiques de Gembloux, 24 Chaussée de Namur, 5030-Gembloux, Belgium. E-mail : dardenne@cragx.fgov.be

Introduction

Calibration transfer in NIR remains a very important topic and many papers have been published on the subject. Recently, Tom Fearn[1] has published an excellent review with 100 references and we suggest the readers to read this paper for more information. Then the present article will not be a review again but a summary of the talk given at the Pittsburgh Conference (New Orleans, 21 March 2002) in the framework of the Tomas Hirschfeld Award and concerning the results of our experience on NIR instrument standardisation. Indeed, we have been running networks of instruments for more than 15 years and calibration transfer was and remains an important subject of our research department.

Why do we need calibration transfer?

The development of robust calibrations is very costly and the main cost comes coming from the reference method analyses which are slow, time consuming and request expensive devices and reagents. Accumulating spectra and reference values for many years leads to very large data bases with sizes very often larger than 1000 samples per commodity. It is obvious that this work can not be reproduced for each instrument. The models must to be used for all the instruments having the same wavelength range and not only for the instruments from the same manufacturer but also for instruments from different companies with the same or shorter wavelength range and for filter instruments.

Despite the effort of the manufacturers to reduce the differences, their instruments do not produce exactly the same response when measuring the same sample. These differences come from many sources. Shenk et al. reported the wavelength accuracy (x axis) and the dynamic range (y axis) as the largest sources[2]. At the last ICNIRS in Kyongju, Korea, we presented a poster[3] on the calibration transfer and we compared the predicted results between one master and 6 "slave" instruments. Even with model coming from 13 000 samples to analyse protein in forages, we observe significant biases and slopes between the instruments. The robustness of a model can help but it is impossible to avoid bias and slope effects when the instruments are unstandardised.

Calibration transfer methods

Slope and bias correction

The oldest and ever used method to transfer calibration is the so-called bias and slope (or skew and bias[1]) correction. This method is very simple and can take in one shot the instrument differences into account but also the sample and/or wet chemistry drifts that can occur. The method consists in measuring a reduced set of samples (10-20) on the target instrument and comparing the predicted results using either the lab values or the predicted values from the master instrument. The samples must be carefully selected to cover the whole range of analytical data and the absorbance range to increase the correlation and decrease the slope confident intervals. The statistical significances of

the bias and the slope are not enough tested and as Tom Fearn said "its simplicity has led to widespread misuse, with perfectly good calibrations being subjected to regular small and unnecessary adjustments using biases estimated from results on one or two samples". The bias significance can be tested by a Ttest:

$$Bias\_Confident\_Limits = \pm (t.SEC)/\sqrt{N} \qquad (1)$$

with *SEC* the Standard Error of Calibration, *N* the number of sample and *t* the value of the t distribution for a given probability α. For instance, for a *SEC*=1, 10 samples and *a*=0.05, the Bias Confident Limits are:

$$BCL = \pm (2,228.1)/\sqrt{10} = \pm 0.70 \qquad (2)$$

This means that with 10 samples the bias must be higher than 70% of the SEC to be significantly different from zero.

The slope can be calculated using the simple linear regression with y_{ref} on the axis and $\hat{y}_{nirs}$ on the X axis:

$$y_{ref} = a + b\hat{y}_{nirs} + e \qquad (3)$$

and the Slope Confident Limits are:

$$Slope\_Confident\_Limits = b \pm t/\sqrt{RSD/SSD_{\hat{y}}} \qquad (4)$$

with RSD the residual standard deviation and $SSD\hat{y}$ the sum of squares of deviations of the "standardisation" samples. If $RSD = 1$, 10 samples, r^2=0.75 and *a*=0.05, $SCL = b \pm 0.38$. This means that with 10 samples and quite low r^2, the correction of the slope must not to be applied because the slope confident interval is too large. An important drawback of the slope and bias correction is the fact the Mahalanobis distances can no longer be used. The new spectra coming from the secondary instrument will always fall outside the calibration data variation indicating "outliers' whereas they are not. The correction must be applied for each model and for each product. Moreover merging spectra in a global data base will lead to use more factors to fit the instrument differences and generally a loss of the performances is observed.

Robust calibrations

Some authors recommend to developed robust calibrations to make the transfer easier. This is a good idea, but implies that the models should not be robust when there is no transfer involved. This is not true: all the analysts are trying to make their calibrations as robust as possible even for only one instrument. In Table 1, we sum up results when we tested 7 math pre-treatments to predict protein in forages. The data base contained 1307 samples (several standardised instruments). The test samples are 15 sealed cups measured on 6 instruments.

The best math treatment (smallest bias and RMS) for unstandardised data is SNV-Detrend with a second derivative (SNVD-D277) (RMS=0.75). The best SEPC (standard error of prediction corrected for bias) is obtained with no math (Log(1/R) (SEPC=0.21). Except for Log(1/R) and SNV,

Table 1. Results of the protein prediction of 15 forage sealed cups on 6 instruments using the same PLS models before and after standardisation – Master predictions are used as reference values

		LOG	SNV	DET	SNVD	D277	SNVD D277	MSC	MSC D277
UNSTD	Biases*	0.94	1.18	1.24	0.95	0.63	0.49	1.52	0.56
	RMS**	1.26	1.37	1.35	1.11	0.85	0.75	1.64	0.81
	SEPC**	0.21	0.49	0.22	0.57	0.28	0.54	0.52	0.53
	RSD**	0.16	0.16	0.18	0.23	0.26	0.26	0.25	0.26
STD	RMS**	0.16	0.17	0.18	0.19	0.29	0.31	0.14	0.31
Model	SECV	0.99	0.95	0.97	0.88	0.92	0.84	0.91	0.84

*Mean of absolute biases of 6 unstandardised instruments
** Quadratic mean of 6 instruments.

SEPC's are not acceptable for a transfer using only bias corrections. After standardisation (using the Clone procedure from ISI software)[4] the best RMS is obtained with MSC (RMS=0.14). The conclusions are: i) the transfer between instruments is impossible without at least a bias correction. ii) no math treatment (Log(1/R)) gives the smallest random error. In this trial the model using only Log(1/R) is the more robust model for the unstandardised instruments. iii) all the math treatments lead to good performance after standardisation. Comparing the SECV, the "reproducibility" of the protein determination by NIR is very good among the standardised instruments.

Spectrum correction

Campbell[5] called the transfer "forward" when the old spectra are transformed to the new instrument and "backward" when the new spectra are transformed to be like the old one. This method is much useful when dealing with networks of instruments having the same wavelength range. The "forward" way is used when the new one has a reduced wavelength range as for filter instruments.

The spectral correction procedure whatever the method used is just like a calibration and the same principles must be applied. The selected samples must cover the optical range of the calibration data set. The analytical range is far less important than the optical range. Except for moisture, the spectral changes due to the chemical composition are small comparing to the spectral changes due to particle size for instance. An adequate standardisation set of samples could have the similar composition provided they cover the spectral variation at each wavelength. Changing the particle size and the compaction of the cells can artificially produce an acceptable optical range.

The goal of the standardisation is to reduce the differences between instruments. Then the first step is to evaluate these differences accurately. To be able to evaluate the differences, measurements must be done to avoid any other source of variation like temperature and sampling error. This is the reason why sealed cups are used most of the time. Sealed and waterproof cups are stable over time, but to avoid some tiny changes which can occur during long period of storage, the spectra from instruments are compared only when the cups have been scanned within a period of no longer than one or two weeks. The first way to evaluate instruments is to compute Root Mean Squares (RMS) or Root Mean Squares corrected for bias (RMSC) (or Standard Deviation of the spectral differences) between pairs of spectra. A second way is to use a model and evaluate the standard deviation of the predicted values across instruments. Table 2 reports RMSEP for the protein prediction in 15 forage samples and RMSC of optical differences between a master and 6 instruments before and after standardisation. The standardisation method was the Shenk and Westerhaus algorithm.

Table 2. RMSEP of protein determinations and RMSC of spectral differences between a master and 6 slave instruments before and after standardisation.

Instruments	UNSTD		STD	
	CP RMSEP	Spectral RMSC*	CP RMSEP	Spectral RMSC*
Inst 1	1.04	6981	0.15	576
Inst 2	0.16	3990	0.09	317
Inst 3	1.24	9238	0.17	431
Inst 4	1.38	7396	0.21	626
Inst 5	0.67	11704	0.10	750
Inst 6	0.48	12189	0.16	487

*RMSC units : microlog on the Log(1/R) spectra

The main goal of an instrument standardisation is to obtain after the process differences between instruments which are less than the sampling errors[6]. Table 3 reports some average values that are good to have in mind when dealing with calibration transfer. A RMSC between 2 standardised instruments measuring the same sealed cup (RMSC≈1000 µlog) must be lower than the RMS of the sampling error (obtained by comparing 2 or several refillings of the same powder – RMSC≈2000 µlog). For very homogenous or liquid samples, the use of sealed cups can be avoided providing the sampling error is at the same magnitude than the instrument error. The most difficult transfer cases occur with fresh and inhomogeneous samples like fresh silage when the samples cannot be preserved. In this case, the best solution is to have the instruments to be adjusted side by side at the same location to scan new samples using the same cups without any refilling.

Table 3. Typical RMSC between spectra according the average optical densities and taken on one or two cloned instruments

Material	RMSC (Log10^-6)
Ceramic – one instrument	30
Soya meal (0.4 AU) (same cup – powder) – One instrument	200
Rapeseed (0.8 AU) (same cup – whole grain) – One instrument	500
Soya meal (0.4 AU) (same cup – powder) 2 Standardised instruments	1000
Soya meal (0.4 AU) (same sample – refilling) – One instrument	2000
Rapeseed (0.8 AU) (whole grain – refilling) – One instrument	6000

The question about the choice of the cloning samples is also very often raised. Why not using reference standards (ei grey standards, NIST standards, etc)? Having tried several times this kind of inert standards, we observed that these standards give differences between instruments which are different from the differences obtained from natural products. There are some kind of interactions between instrument & product which lead to wrong corrections. Figure 1 reports the spectra from 2 instruments of a natural product (corn silage) and a reference tile (0.3AU-LabSphere®). The differences between instruments are different according to the samples even at the wavelengths where the curves intersect around 1410 nm. This is the reason why the standardisation between instruments has to be performed by means of natural product samples.

Moreover, as it was said, the standardisation is very similar to a calibration process and any extrapolation is strictly forbidden. A correction made between 0.2-0.6 optical densities cannot be applied on spectra outside this range. In our agricultural applications, we currently run several types of corrections: one for all the products which are dried and ground, one for all the unground grains

(cereals) and unground feed pellets, one for fresh forages and specific corrections for the other products like meat, cheese, fruit juices, wines, etc. or particular seeds like rapeseed or sunflower.

A network of instruments is based on one instrument considered as the 'master' instrument. What happens if the master itself has technical failures and even worse if it disappears? One can consider a secondary instrument as the new master provided this latter has been properly standardised and gives the same response than the original master. We did instrument corrections in a cascade scheme till 4 levels without significant loss of information[7]. Another solution is to measure the sealed cups every 2-3 weeks. If we assume that the sealed cups do not change during such a period, the previous measurements constitute the "master' and the newest ones the slave. A given instrument can be standardised on itself. Table 4 reports RMSC between spectra of the same 30 sealed cups measured over time on the same instrument. Between periods of 9 to 79 days the differences are far lower than the sampling errors and the spectra would be used to standardise the instrument on itself if it would have been needed.
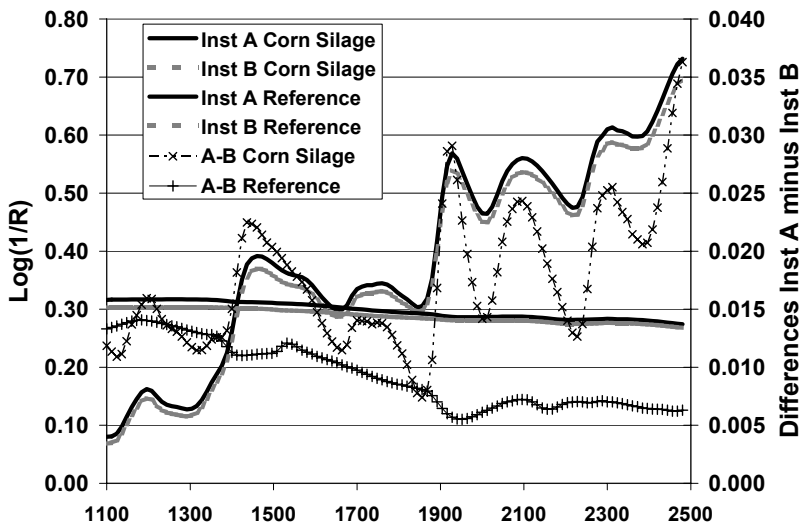

Figure 1. Spectra of a natural product (corn silage) and a reference standard (0.3AU) on 2 instruments and their differences.

Table 4. RMSC between spectra for 30 sealed cups measured on the same instrument at different dates – Comparison between two consecutive dates.

mm/dd/yy	Delay(days)	RMSC(µLog)	mm/dd/yy	Delay(days)	RMSC(µLog)
11/09/00			08/16/01	42	537
12/11/00	32	502	09/17/01	32	788
01/20/01	40	170	10/04/01	17	649
02/01/01	12	224	11/30/01	57	480
03/23/01	50	483	01/22/02	53	369
04/25/01	33	591	02/13/02	22	275
06/13/01	49	441	05/03/02	79	450
07/05/01	22	530	05/13/02	10	504

The Shenk and Westerhaus patented method[4] is a 2 step correction. After the correction of the wavelength shift based on first derivatives and second degree polynomial regression, the absorbencies are corrected data point by data point using single linear regressions[8]. All the other regression techniques can be applied like DS, PDS, FIR, Wavelet Tranform, and ANN. The reader is invited to go to the Fearn's article[1] to get the corresponding references. At the Pittcon conference, a successful example of the PDS application was presented to transfer a data set from a NIRSystems 5000 to a Bran&Luebbe 500 for the prediction of a wheat flour quality index predicted on whole grains. The SMATCH routine included in the SESAME package from Bran&Luebbe Gmbh had been used. In this case the transfer had been done 'backward' and a new PLS model had been computed on the NIRSystems spectra corrected to look like the B&L ones.

Different instrument types

A new challenge appears with the coming on the market of the diode array instruments. The measurements with the Zeiss are carried out using a rotating cut. Each sample is scanned 4 times during 3 sec and the averages are used for the calculation. Figure 2 represents the available spectral information respectively from a NIRSystems 5000 and a Zeiss Corona NIR45 (DA). The goal is to use the NIRSystems information to build models for the DA. For the next experiment 97 samples of dried and ground maize silage were measured on both instruments. The Zeiss spectra were interpolated to get the same increment as the monochromator (2 nm). Among the 97 samples, 10 were selected to cover the absorbance range. A PDS (Matlab routine STDGEN.M from Eigenvectors® - windows of 3 data points and 1 PLS factor) transfer method was used with these 10 samples to modify the monochromator file and create a 'virtual' DA file from which new models are computed. The 97 'actual' samples are then predicted with the 'virtual' models.
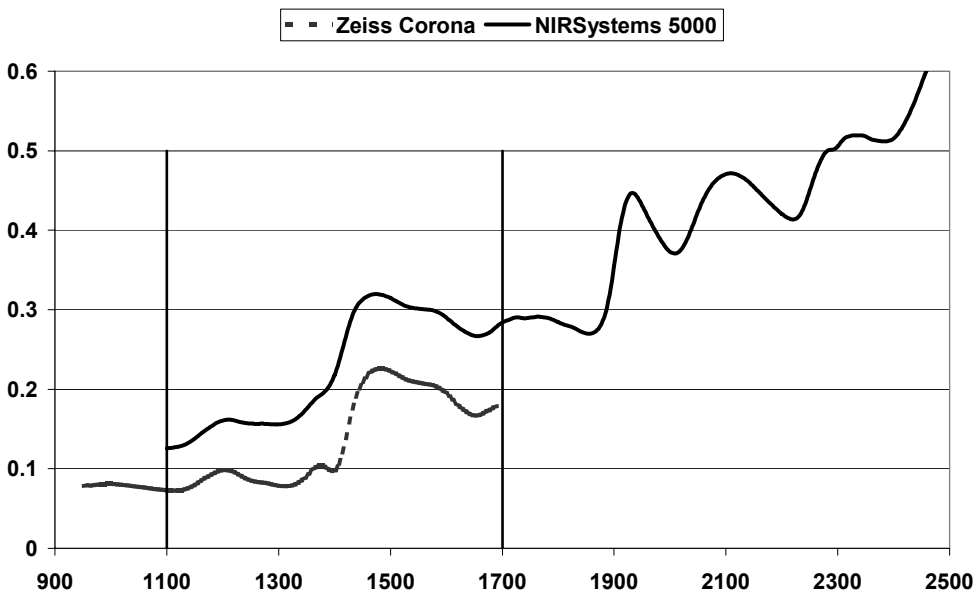

Figure 2. Spectral information of a monochromator NIRSystems 5000 (PbS detectors) and a diode array (InGaAs) Zeiss Corona NIR45

Table 5 reports the PLS calibration and the transfer results. The first column contains the SECV's for the monochromator models using the full wavelength range. The second contains the SECV's for the monochromator using the same range as the DA instrument. The third one is the SECV's for the models built from the actual DA spectra. The fourth reports the SECV's from the 'virtual' models and the last one the SEP when the 'actual' DA spectra are predicted with the 'virtual' models. The comments are: i) The accuracies obtained on the monochromator within the reduced range (1100-1680nm) instead of the full range (1100-2500nm) are not significantly modified. The SECV's remain very similar. ii) The actual models from the DA give SECV's just a little bit higher than the values obtained with the monochromator but they remain definitely acceptable. iii) The transformed 'virtual' spectra can reproduce the same accuracies than the original ones. iv) The prediction of the actual spectra with the 'virtual' models gives acceptable results except for protein. This first experiment is promising and more tests must be done. Anyway, the use of new DA instruments is not limited by the need of new calibration development from scratch. For instance, Greensill[9] has tested different methods and mentioned the Wavelet Transform technique as the best one to standardise two DA instruments. Further developments and experiments must be carried out to select the most efficient methods or to set up new ones.

Table 5. Calibration and transfer results between a monochromator NIRSystems 5000 and a diode array Zeiss Corona NIR45.

	NIRS5000	NIRS5000	Zeiss	NIRS5000	Predict
	1100-2500	1100-1680	1100-1680	STDGEN.M	Zeiss
	n=97	n=97	n=97	n=97	n=97
	SECV	SECV	SECV	SECV	RMSEP
Ash	0.31	0.32	0.37	0.35	0.46
Protein	0.25	0.28	0.32	0.31	0.76
CF	1.05	1.04	1.07	1.12	1.12
NDF	1.90	1.94	2.06	1.88	2.03
ADF	1.06	1.06	1.13	1.13	1.18
ADL	0.23	0.23	0.26	0.28	0.32
Starch	2.18	2.26	2.41	2.30	2.24
OMD	1.57	1.55	1.83	1.72	2.01

The ultimate network

The latest development in the field of NIR networking is certainly the concept proposed by Dr J. Shenk (Foss-Infrasoft International with the package called RINA (Remote Internet Nir Analysis)). The idea is to have the spectra and models in a database available on an Interrnet server. Using a browser the site is accessed and after introducing a user name and a password any spectrum can be sent to be predicted. The server returns the results under different formats within a few seconds. The requirements are the Foss-ISI file structures and standardised instruments. Combined with the 'Local' concept (for each spectrum and each parameter to be predicted a PLS temporary model is built with the N neighbours selected in the data base)[10], RINA offers the next main advantages: i) The spectral data are protected and the models can not be copied. ii) The network operator can supervise all the operations performed on the different sites and can be more efficient in support and operation. iii) The routine analyses are easier for the user. He does not need to know the type of product he measures: global and neighbourhood distances guarantee the results validity.

Conclusion

Transfer of calibration would be a question of the past if the manufacturers succeeded to remove all the differences between instruments. This can be improved for dispersive and FT instruments. The coming of lower cost diode array instruments emphasizes the need of transfer methods because the diode arrays will never give the same response.

The main advantages of the instrument standardisation and networking are:

- The models from large and expensive data sets can be shared among networks of instruments. The databases can be transformed to other reduced spectral ranges. The standardisation allows keeping the outlier detection in routine mode.

- Databases from many standardised instruments can be merged to save cost of reference chemistry and make models more robust.

- The standardisation allows the maintenance of instruments constant overtime even after severe repairs.

Acknowledgments

We were very grateful and felt very honoured for receiving the 2002 Tomas Hirschfeld Award. We would like to close this brief summary by thanking:
- the members of the International Committee for Near Infrared Spectroscopy,
- Bran&Luebbe Gmbh for sponsoring the award,
- Pittsburgh Conference 2002 for the invitation,
- Prof. John S. Shenk, who shared his experience and knowledge,
- all my co-workers at CRAGx. A particular acknowledgment to my former director, Robert Biston, who during his career played an essential role to promote NIR and our research works.

References

1. T. Fearn, *J. Near Infrared Spectrosc.* **9**, 229(2001).
2. S.S. Shenk, J.J. Workman and M.O. Westerhaus, in *Handbook of Near Infrared Analysis*, Ed. Burns et Ciurczak, p383–431,(1992).
3. P. Dardenne, I.A. Cowe, P. Berzaghi, P.C. Flinn, M. Lagerholm, J.S. Shenk and M.O. Westerhaus. In *Near Infrared Spectroscopy; Proceedings of the 10th International Conference.* Ed by A.M.C. Davies and R.K. Cho. NIR Publications, Chichester, UK pp 23-28 (2002),.
4. J.S. Shenk and M.O. Westerhaus, US Patent No. 4,866,644 (12 September 1989).
5. B.H. Campbell. In *Near Infrared Spectroscopy: The Future Wave*, Ed by A.M.C. Davies and P. Williams. NIR Publications, Chichester, UK, pp218-220, (1996).
6. J.S. Shenk, M.O. Westerhaus and W.C. Templeton, Jr. *Crop Science.* **25**, 159-161 (1985).
7. P. Dardenne, R. Biston and G. Sinnaeve. In *Near Infrared Spectroscopy: Bridging the gap between Data Analysis an NIR applications*. Ed. K.I.Hildrum, T.Isaksson, T.Naes nad T.Tandgberg. Ellis Horwood, Chivhester, UK, pp-453-458 (1992).
8. E. Bouveresse, D.L. Massart and P. Dardenne. *Anal. Chem.* **67**, 1381-1389 (1995).
9. C.V. Greensill and K.B. Walsh. *J. Near Infrared Spectrosc.* **10**, 27–35 (2002).
10. G. Sinnaeve, P. Dardenne and R. Agneessens. *J. Near Infrared Spectrosc.* **2**, 163-175(1994).